StyLitGAN: Image-based Relighting via Latent Control

Anand Bhattad James Soole D.A. Forsyth
University of Illinois Urbana-Champaign
https://anandbhattad.github.io/stylitgan/

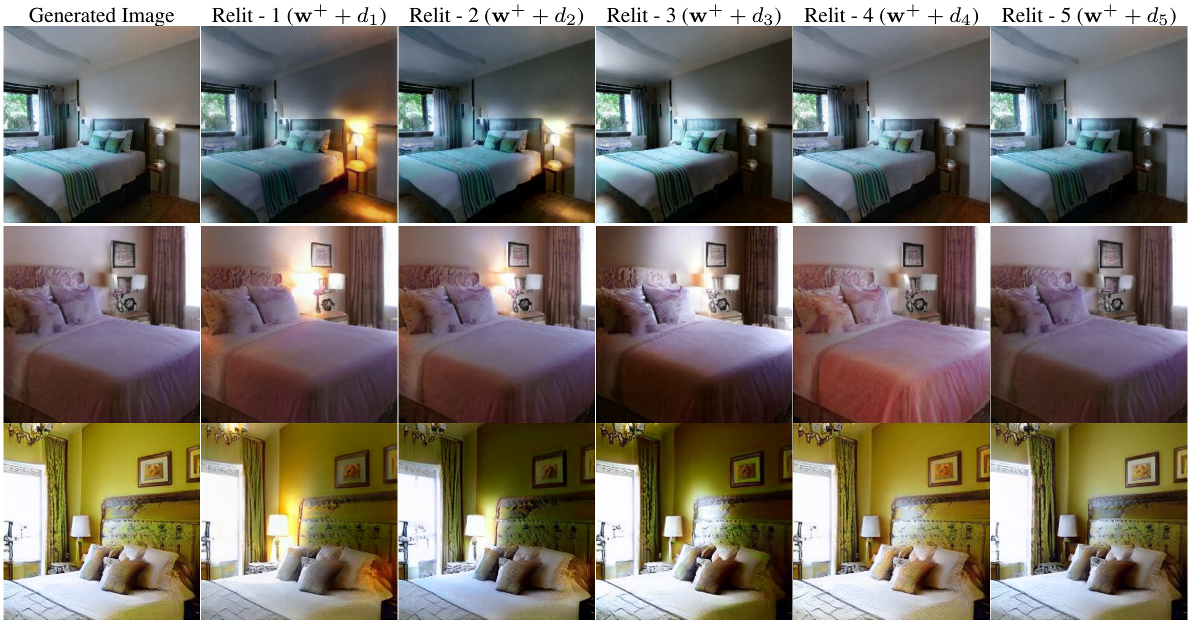

Generated Image
Relit - 1 ($\mathbf{w}^+ + d_1$)
Relit - 2 ($\mathbf{w}^+ + d_2$)
Relit - 3 ($\mathbf{w}^+ + d_3$)
Relit - 4 ($\mathbf{w}^+ + d_4$)
Relit - 5 ($\mathbf{w}^+ + d_5$)

Figure 1. StyLitGAN identifies directional vectors (d_i) within StyleGAN's style space ($\mathcal{W}^+$) which, when added to the $\mathbf{w}^+$ style code, effectively modify the lighting of generated images while preserving their geometry and albedo. This process eliminates the need for per-image search or model fine-tuning. The first column displays images generated from StyleGAN2, with subsequent columns illustrating the same scenes, each relit using a specific direction. These relighting directions (d_i) are derived through a forward selection method, ensuring diversity and avoiding cherry-picking. The directional effects are consistent across different scenes: for instance, d_1 activates an orange-tinged bedside lamp, d_2 a less intense white-tinged lamp, d_3 introduces strong directional light from the window, and so on, demonstrating diverse relighting capabilities of StyLitGAN.

Abstract

We describe a novel method, StyLitGAN, for relighting and resurfacing images in the absence of labeled data. StyLitGAN generates images with realistic lighting effects, including cast shadows, soft shadows, inter-reflections, and glossy effects, without the need for paired or CGI data. StyLitGAN uses an intrinsic image method to decompose an image, followed by a search of the latent space of a pretrained StyleGAN to identify a set of directions. By prompting the model to fix one component (e.g., albedo) and vary another (e.g., shading), we generate relighted images by adding the identified directions to the latent style codes. Quantitative metrics of change in albedo and lighting diversity allow us to choose effective directions using a forward selection process. Qualitative evaluation confirms the effectiveness of our method.

1. Introduction

Scene perception dramatically shifts with varying lighting conditions - a sunlit room takes on a different character as daylight fades, and interior spaces transform with the flick of a switch. Similarly, surface changes, like a wall's paint color, change not only the wall's appearance but also the overall image due to light reflection. Despite the impressive realism achieved by current generative models like StyleGAN [21–23], they fall short in dynamically controlling scene lighting, a key aspect of realistic image generation.

In this work, we present StyLitGAN, a novel approach that extends the editing capabilities of StyleGAN [37, 44, 45, 52]. StyLitGAN uniquely manipulates style codes to selectively change lighting while preserving other image attributes like albedo and geometry. This selective editing addresses a critical gap in current generative methods, which typically lack the precision to independently control individual scene components.

Our method, StyLitGAN, first uses StyleGAN to produce a set of images and then decomposes these generated images into albedo, diffuse shading, and glossy effects using an off-the-shelf, self-supervised network [13]. We then search for style code edits by prompting StyleGAN to produce images that (a) are diverse, but (b) have the same albedo (and so geometry and material) as the original generated images. Our search selects the most effective relighting directions in a data-driven manner.

Our approach generates images with realistic lighting effects, including cast shadows, soft shadows, inter-reflections, and glossy effects. Importantly, we observe that style code edits produce consistent effects across images. For instance, as seen in Fig. 1, adding the first and second directions tends to switch on bedside lamps (columns Relit-1 and Relit-2), while adding the fourth direction increases the light intensity from outside the window (column Relit-4). Since StyLitGAN can generate any image that a vanilla StyleGAN can, but also generate images that are out of distribution, one would expect FID scores to increase over StyleGAN; this happens.

The recent method GAN-control [38] controls lighting on face images using an attribute procedure. We show, in contrast to StyLitGAN, GAN-control fails on indoor scenes, likely because the attribute vocabulary is too easily subverted by the complex lighting effects in indoor scenes.

We demonstrate applications of StyLitGAN to standard computer vision problems. Using the Multilum dataset of [28], we show that predictions from a SOTA surface normal predictor [20] vary significantly when lighting is changed. Finetuning this normal predictor using StyLitGAN images suppresses this effect. The improvement is comparable with that obtained by finetuning with true multiilluminant images (which are very difficult to obtain in quantity).

2. Related Work

Image Manipulation: A significant literature deals with manipulating and editing images [3, 9, 10, 14, 16, 26, 32, 34, 43, 53]. Editing procedures for generative image models [15] are important, because they demand compact image representations with useful, disentangled interpretations. StyleGAN [21–23] is currently de facto state-of-the-art for editing generated images, likely because its mapping of initial noise vectors to style codes which control entire feature layers produces latent spaces that are heavily disentangled and so easy to manipulate. Recent editing methods include [7, 35, 37, 44, 45, 52], with a survey in [46]. The architecture can be adapted to incorporate spatial priors for authoring novel and edited images [12, 27, 42]. In contrast to this literature, we show how to fix one physically meaningful image factor while changing another. Doing so is difficult because the latent spaces are not perfectly disentangled, and we must produce a diverse set of changes in the other factor.

Relighting using StyleGAN: Relighting faces using StyleGAN can be achieved with Stylerig [42], but this method requires a 3D morphable face model. In contrast, StyLitGAN does not require a 3D model and can be extended to complex indoor scenes, which is not possible with Stylerig. Yang et al. [47] uses semantic label attributes to train a binary classifier to find latent space directions that represent indoor and natural lighting, but this method cannot produce diverse relighting effects. We also find Yang et al's relighting to change color or albedo. In contrast, StyLitGAN generates diverse realistic relighting effects without changing albedo and without requiring any labeled attributes.

StyleFlow [1] and GAN-control [38] require a parametric model to express lighting, such as spherical harmonics. These methods are limited to relighting faces and do not result in realistic relighting of rooms. Our experiments using GAN-control for rooms result in large geometry and albedo change. In contrast, StyLitGAN can produce relighted images without changing geometry or albedo. We also note that rooms are more challenging to relight than faces due to significant long-scale inter-reflection effects, diverse shadow patterns, stylized luminaires, stylized surface albedos, and surface brightnesses that are not a function of surface normal alone. These factors make it difficult to apply GAN control directly to rooms. Additionally, none of these methods have the ability to resurface or recolor rooms, though StyLitGAN can also edit color or materials while preserving the lighting of the scene.

Other Face Relighting methods use carefully collected supervisory data from light-stages or parametric spherical harmonics [29, 31, 36, 40, 51]. ShadeGAN [30], Rendering with Style [5], and Volux-GAN [41] use a volumetric rendering approach to learn the 3D structure of the face and the illumination encoding. Volux-GAN [41] also requires image decomposition from [31] that is trained using carefully curated light-stage data. In comparison, we neither require any explicit 3D modeling of the scene nor labeled and curated data for training the image decomposition model.

3. Approach

We follow convention and manipulate StyleGAN [23] by adjusting the $\mathbf{w}^+$ latent variables. We do not modify StyleGAN, but instead, seek a set of lighting directions $\mathbf{d}_i$ which are independent of $\mathbf{w}^+$ and which have desired effects on the generated image. We obtain these directions by constructing losses that capture the desired outcomes, then search for directions that minimize these losses. We find all directions at once and use 2000 randomly generated images for this search. Once found, these lighting directions are applicable to all other generated images. Our search procedure only

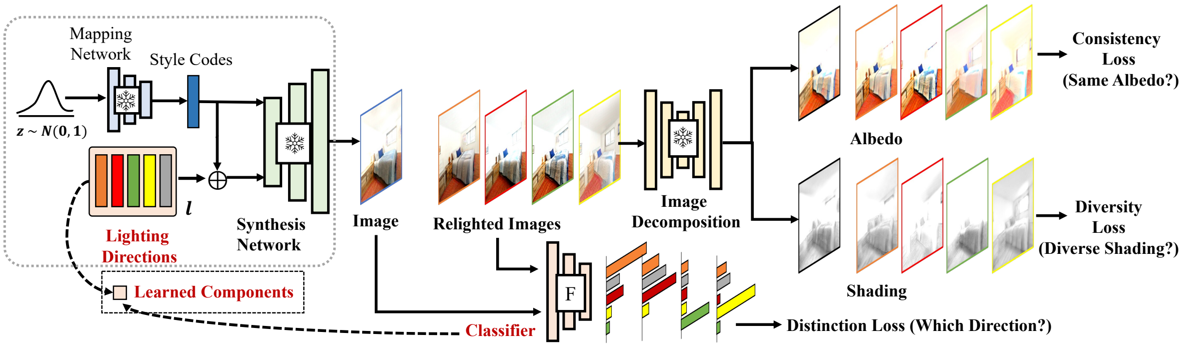


Figure 2. How StyLitGAN works: We generate an image from random Gaussian noise using a pretrained StyleGAN. We also generate novel relighted versions (16 in our case) of the same image using randomly initialized latent directions ($\mathbf{d}$) that are added to $\mathbf{w}^+$ latent style codes. We train a classifier (F) that takes in all the pairs of relighted and original images and predicts the relighting direction applied to them. We apply a distinction loss and jointly update the latent directions and the classifier. Next, we generate the decomposition of these images from a pretrained decomposition model (D). We then apply losses that force StyLitGAN to find latent directions such that the albedo does not change (consistency loss), but the image does (diversity loss).

sees each image once.

Fig. 2 summarizes our procedure and we call our model StyLitGAN. Our method consists of two stages. The first stage involves decomposing images using a pretrained model. The second stage *jointly* searches for directions and trains a classifier F. The classifier F predicts the latent direction applied to image pairs. It's a classification task with a fixed number of jointly learned latent directions. While classifier F doesn't directly know whether latent directions relate to lighting, our use of image decomposition losses ensures these directions are lighting-related. The consistency loss maintains albedo, and the diversity loss ensures diverse shading changes, both properties expected when changing lighting conditions. Thus, while jointly updating the classifier F and latent directions, these losses ensure the discovered directions pertain to relighting. We now elaborate on our search for directions and losses in detail.

Base StyleGAN Models: We use baseline pretrained models from [48] that use a dual-contrastive loss to train StyleGAN for bedrooms, faces and churches. We also use baseline pretrained StyleGAN2 models from [12] for conference rooms, kitchens, dining, and living rooms.

Decomposition: We decompose images into albedo, shading, and gloss maps (gloss; only when available) as $A \times S + G$, where A models albedo effects and S and G model shading and gloss effects respectively. We use the method of Forsyth and Rock [13], which is easily adapted because it is self-supervised and uses only samples from statistical models derived from Land's Retinex theory [24]. We can construct many decompositions using their approach by changing the statistical spatial model parameters. We evaluate several such decomposition models under several hyperparameter settings and create a large pool of relighting directions. We finalize our directions using a forward selection process that provides minimal albedo and geometry shift with a large relighting diversity (Section 4).

Relighting a scene should produce a new, realistic image where the shading has changed but the albedo has not. Write $I(\mathbf{w}^+)$ for the image produced by StyleGAN given style codes $\mathbf{w}^+$, and $A(I)$, $S(I)$, and $G(I)$ for the albedo, shading, and gloss respectively recovered from image I. We search for multiple directions $\mathbf{d}_i$ such that: (a) $A(I(\mathbf{w}^+ + \mathbf{d}_i))$ is very close to $A(I(\mathbf{w}^+))$ – so the image is a relighted version of $I(\mathbf{w}^+)$, a property we call **persistent consistency**; (b) the images produced by the different directions are linearly independent – **relighting diversity**; (c) which direction was used can be determined from the image, so that different directions have visibly distinct effects – **distinctive relighting**; and (d) the new shading field is not strongly correlated to the albedo – **independent relighting**. Not every shading field can be paired with a given albedo, otherwise there would be nothing to do. We operate on the assumption that edited $\mathbf{w}^+$ will result in realistic images [6].

Recoloring: Alternatively, we may wish to edit scenes where the colors or materials of objects have changed, but the lighting hasn't. Because shading conveys a great deal of information about shape, we can find these edits using modified losses by seeking consistency in the shading field.

Persistent Consistency: The albedo decomposition of both the relighted scene: $A_R = A(I(\mathbf{w}^+ + \mathbf{d}_i))$ and the original: $A_O = A(I(\mathbf{w}^+))$ must be the same; where R refers to relighted images and O refers to StyleGAN generated images. We use a Huber loss and a perceptual feature loss [19, 50] from a VGG feature extractor (Φ) [39] at various feature layers (j) to preserve persistent effects (geometry, appearance and texture) in the scene.

$$\mathcal{L}_{const}(A_O, A_R) = \begin{cases} \frac{1}{2}\left[A_O - A_R\right]^2 & \text{for } |A_O - A_R| \leq \delta, \\ \delta\left(|A_O - A_R| - \delta/2\right) & \text{otherwise.} \end{cases} \tag{1}$$

$$\mathcal{L}_{per}(A_O, A_R) = ||\Phi_j(A_O) - \Phi_j(A_R)||_2. \tag{2}$$

Relighting Diversity: We want the set of relighted images produced by the directions to be diverse on a long scale so that regions that were in shadow in one image might be bright in another. For each $S(\mathbf{w}^{+} + \mathbf{d}_i)$, we stack the two shading and gloss: S and G and compute a smoothed and downsampled vector $\mathbf{t}_i$ from these maps. We then compute $\mathcal{L}_{div}(S, G)$ (diversity loss) which compels these $\mathbf{t}_i$ to be linearly independent and encourages diversity in relighting.

$$\mathcal{L}_{div}(S, G) = -\log\det N \tag{3}$$

where i^{th} & j^{th} component of N is $t_i^{\intercal} t_j$

Distinctive Relighting: A network might try to cheat by making minimal changes to the image. Directions $\mathbf{d}_i$ should have the property that $\mathbf{d}_i$ is easy to impute from $I(\mathbf{w}^{+} + \mathbf{d}_i)$. We train a classifier joint with the search for directions. This classifier accepts $I(\mathbf{w}^{+})$ and $I(\mathbf{w}^{+} + \mathbf{d}_i)$ and must predict i. The cross-entropy of this classifier supplies our loss:

$$\begin{aligned} &\min_{l,F} \mathcal{L}_{dist}(I(\mathbf{w}^{+}),\ I(\mathbf{w}^{+} + d_i)) \\ &= -\sum_{i=1}^{M} d_i \log F(I(\mathbf{w}^{+}),\ I(\mathbf{w}^{+} + d_i)) \end{aligned} \tag{4}$$

Saturation Penalty: Our diversity loss might cheat and obtain high diversity by generating blocks of over-saturated or under-saturated pixels. To discourage these effects, we apply a saturation penalty over number of pixels within a certain threshold.

$$\begin{aligned} \mathcal{L}_{sat} = \lambda_{oversat}\Big[&\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} max(0, I_{i,j} - s)^2\Big] \\ +\lambda_{undersat}\Big[&\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} max(0, s - I_{i,j})^2\Big] \end{aligned} \tag{5}$$

where $\lambda_{oversat}$ and $\lambda_{undersat}$ are the penalty weights for over-saturation and under-saturation respectively, H and W are the height and width of the images, $I_{i,j}$ is the pixel intensity at pixel location (i, j), and s is the saturation threshold (i.e., the maximum allowed pixel intensity). The penalty is computed as the mean squared difference between the pixel intensity and the saturation threshold.

Recoloring requires swapping albedo and shading components in all losses, except we do not use decorrelation loss while recoloring. Obtaining good results requires quite a careful choice of loss weights (λ coefficients). We experiment with several λ coefficients for both these edits (Section 4 and supplementary).

4. Model and Directions Selection

We prompt StyleGAN to find style code directions that: (a) do not change albedo, and (b) strongly change the image. We use a variety of different image decomposition models to obtain directions across multiple different hyperparameter settings. We have no particular reason to believe that a single model will give only good directions or all good directions. We then find a subset of admissible models. We must choose admissible models using a plot of albedo change versus diversity because there is no way to weigh these effects against one another. However, relatively few methods are admissible – see Figure 3. We then pool all directions from all of the admissible models and use forward selection to find a small set of polished directions in this pool.

Scoring Albedo Change: We use SuperPoints [18] to find 100 interest points in the original StyleGAN-generated image. Around each interest point, we form a 8×8 patch. We then compare these patches with patches in the same locations for multiple different relightings of that image. If the albedo in the image does not change, then each patch will have the same albedo but different lighting.

Given two color image patches $\mathbf{p}$ and $\mathbf{q}$, viewed under different lights, we must measure the difference between their albedos $d_a(\mathbf{p}, \mathbf{q})$. Write p_{ij} for the RGB vector at the i, j'th location ($1 \leq i \leq M$, $1 \leq j \leq N$) and write $p_{ij,k}$ for the k'th RGB component at that location. The intensity of the light may change without the albedo changing, so this problem is homogeneous (i.e. for $\lambda, \mu > 0$, $d_a(\mathbf{p}, \mathbf{q}) = d_a(\lambda\mathbf{p}, \mu\mathbf{q})$). Assume that the illumination intensity changes, but not color. The patches are small, so the illumination field on a patch can be modeled as a linear function, so there are albedos $\mathbf{a}$, $\mathbf{b}$ such that $p_{ij} = (p_x i + p_y j + p_c)a_{ij}$ and $q_{ij} = (q_x i + q_y j + q_c)b_{ij}$. If the two patches have similar albedo, there will be p_x etc. such that $p'_{ij} = (q_x i + q_y j + q_c)p_{ij}$ is the same as $q'_{ij} = (p_x i + p_y j + p_c)q_{ij}$. We measure the cosine distance

$$d_a(\mathbf{p}, \mathbf{q}) = 1 - \max_{p_x, \ldots q_c} \frac{\sum_{ijk} p'_{ijk} q'_{ijk}}{\sqrt{\sum_{ijk}(p'_{ijk})^2}\sqrt{\sum_{ijk}(p'_{ijk})^2}} \tag{6}$$

The relevant maximum can be calculated by analogy with canonical correlation analysis (Supplementary).

Scoring Lighting Diversity: Illumination cone theory [2] yields that any non-negative linear combination of k shadings is a physically plausible shading. To determine if an image is new, we relax the non-negativity constraint and so must ensure that it cannot be expressed as a linear combination of existing images. In turn, we seek a measure of the linear independence of a set of images. This measure should: be large when there is a strong linear dependency; and not grow too fast when the images are scaled. Write $\mathbf{x}_i$ for the i'th image, and $\mathcal{X}$ for the matrix whose i, j'th component is $\mathbf{x}_i\mathbf{x}_j$. Then $-\log\det\mathcal{X}$ is very large when the $\mathbf{x}_i$ is close to linearly dependent, but does not scale too fast when the images are scaled.

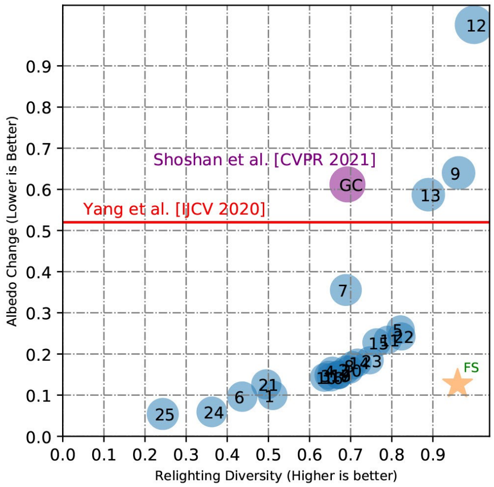


Figure 3. Each model from StyLitGAN produces 16 directions. Models differ by choice of hyperparameters and intrinsic image decomposition. We evaluate models by albedo change and by diversity, averaged across a small fixed validation set of test scenes. As the figure shows, there is typically a payoff, but some models are not admissible. Figure 4 shows examples from some of the models considered here. We exclude inadmissible models, then pool all directions from all other models, and apply a forward selection procedure (section 4). This yields 16 strong relighting directions (the star). In our comparison with Yang et al. [47], we focus on the changes in albedo since their method identifies only a single relighting direction. This limits the scope for evaluating relighting diversity. Yang et al. [47] aggressively change scene albedo, while our StyLitGAN ensures only lighting changes. Additionally, we compare with GAN-Control [38], which, while attempting relighting, often changes the scene layout, leading to large albedo change and increased diversity score due to layout variations.

Decomposition Models Investigated: We searched 25 instances in total obtained with different hyperparameter settings from three families of decomposition. The first family is the SOTA unsupervised model of [13], which decomposes images into albedo and shading using example images drawn from statistical models. The second is a variant of that family that decomposes into albedo, shading, and gloss decomposition. The third is an albedo, shading, and gloss decomposition that models fine edges in the albedo rather than the shading field. These models were chosen to represent a range of possible decompositions, but others could yield better results. The key point is that we can choose a model from a collection by a rational process.

Selecting Directions: Our approach for selecting directions involves creating a scatter plot of 25 instances with various hyperparameters and image decomposition models. We find 16 directions for each instance in our final experiments, and the search for 16 directions takes about 14 minutes on an A40 GPU. We experimented with different numbers of directions of order 2^n for n=2, 3, 4, 5, 6, 7 and found that 16 directions (n=4) strike a better balance between relighting diversity and albedo change. However, finding multiple directions is challenging because the search space is complex and high-dimensional, and we lack ground truth to supervise the search. Therefore, we apply a two-step process to find effective directions and filter out any bad directions.

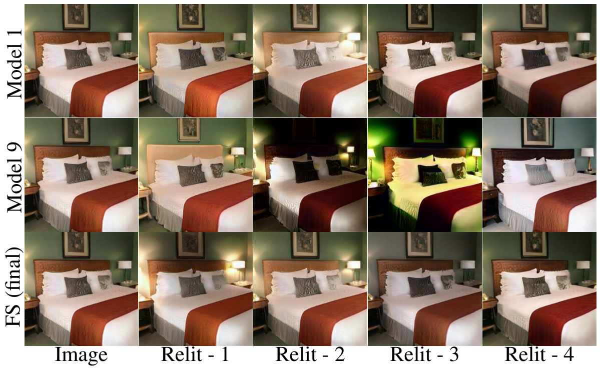


Figure 4. The **bottom** row shows scene relightings obtained using our final, forward selected, set of directions (star of Figure 3). For comparison, we also show scene relightings from different models obtained from StyLitGAN shown in Figure 3 (model numbers correspond to numbers on that figure). Note how most models are capable of producing some good directions, but not all directions from a given model may be good.

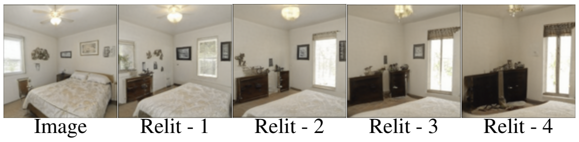


Figure 5. GAN-control (GC) [38] cannot relight complex scenes like bedrooms; its interpretable relighting completely changes the layout of the scene (subsequent columns correspond to stronger light from the right).

We first identify and discard inadmissible models that are located behind the Pareto frontier. We then select the top 10 admissible models based on their average albedo change when applied over a large set of fixed validation images. Our goal is to select the best relighting directions from these admissible models. To achieve this, a forward selection process is employed, which involves selecting a subset of directions from the set of admissible directions.

Forward Selection Process: To select the best directions, we begin with all directions from the admissible models, resulting in 160 directions from 10 models. These directions are then filtered to remove "bad" directions that produce relighting similar to the original image or shading that does not vary across pixels, resulting in 108 directions.

Next, we use a greedy process to select the best 16 directions from the remaining 108 directions. We evaluate each direction one at a time and add it to the pool if it provides a large diversity score while incurring a small penalty for large albedo change. This process continues until the desired number of directions are selected. The forward selection

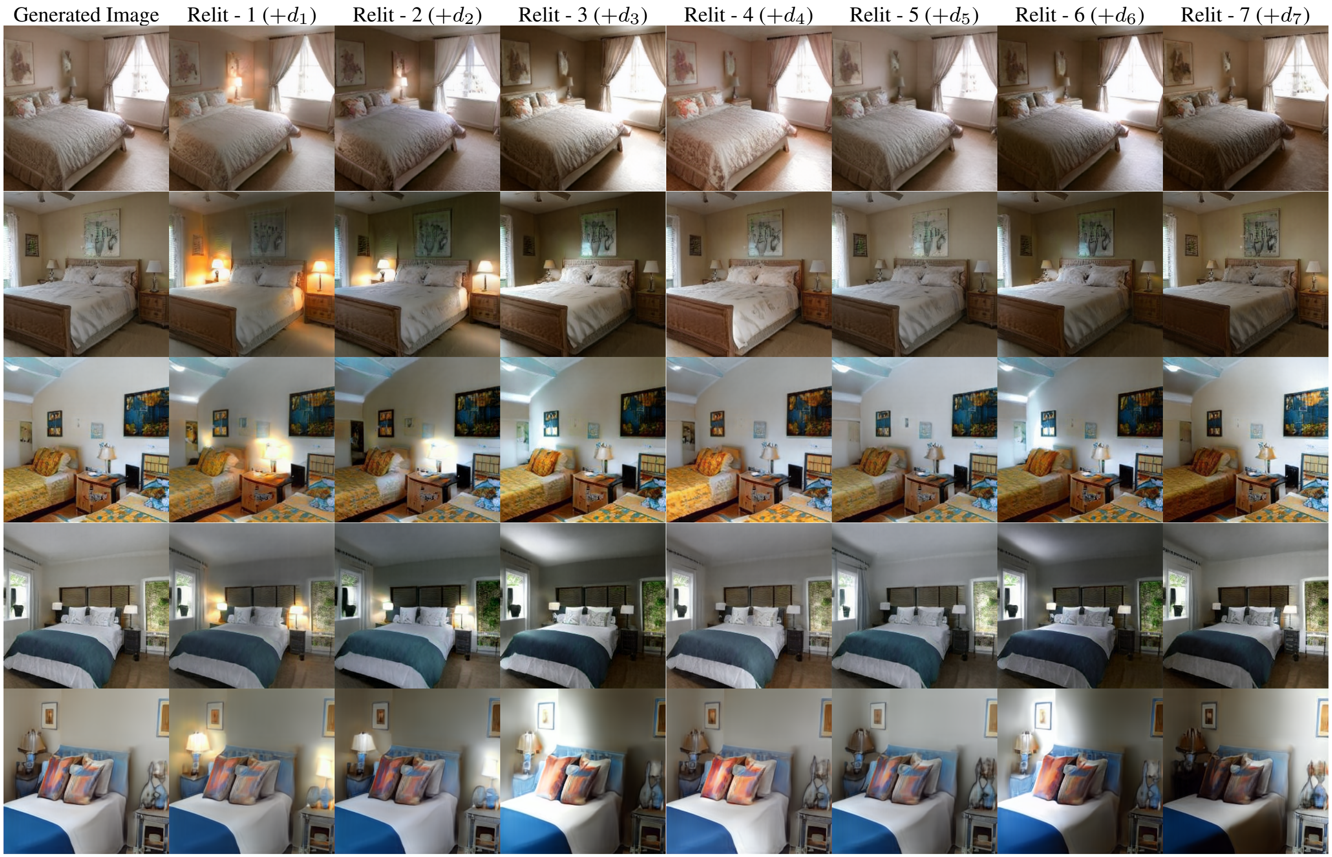


Figure 6. **First column** images generated by the original StyleGAN. **Other columns** show images obtained from $\mathbf{w}^+ + \mathbf{d}_i$, our relighting directions added to the style codes ($\mathbf{w}^+$) of the images in the first column. These directions have been chosen to fix albedo, but change shading. Note: each column appears to show the same scenes (by row) but with different illumination. Lighting varies aggressively, and the individual latent direction $\mathbf{d}_i$ has persistent semantics – each column corresponds to a type of illumination. Notice relighting changes around ceilings (when visible), walls, and bedside lamps. Another observation that can be made is that all relightings are with respect to world coordinates and not camera coordinates.

process is fast and efficient, taking less than a minute.

The resulting scores from the forward select 16 directions are marked with a *star* in golden color in Fig. 3. The directions obtained with this process are significantly better than individual models alone. A qualitative ablation is in Fig. 4.

5. Experiments

Qualitative Evaluation: StyLitGAN produces realistic images that are out of distribution but known to exist for straightforward physical reasons. Because they're out of distribution, current quantitative evaluation tools do not apply. We evaluate realism qualitatively. Further, there is no direct comparable method. However, we show relighting comparisons to a recent SOTA method that is physically motivated and trained with CGI data [25] (Figure in Supplementary). For relighting, our method should generate images that: are clearly relightings of a scene; fix geometry and albedo but visibly change shading; and display complicated illumination effects, including soft shadows, cast shadows and gloss. For resurfacing, our method should generate images that: are clearly images of the original layout, but with different materials or colors or changes in furniture; and display illumination effects that are consistent with these color changes. As Figure 7 shows, our method meets these goals. Figure 8 and Figure 9 show interpolation sequences for a relighting between two directions and scaling only one direction. Note that the lighting changes smoothly, as one would expect. A figure in supplementary shows that our relighting and recoloring directions are largely disentangled.

Quantitative Evaluation: Figure 3 shows how we can evaluate albedo shift and lighting shift. Further, in Table 1 we show we can generate image datasets with *increased* FID [17, 33] compared to the base comparison set (we use clean-FID [33]). This is strong evidence our method can produce a set of images that is a strict superset of those that the vanilla StyleGAN can produce.

Generality: We have applied our method to StyleGAN instances trained on different datasets – Conference Room, Kitchen, Living Room, Dining Room, Church, and Faces

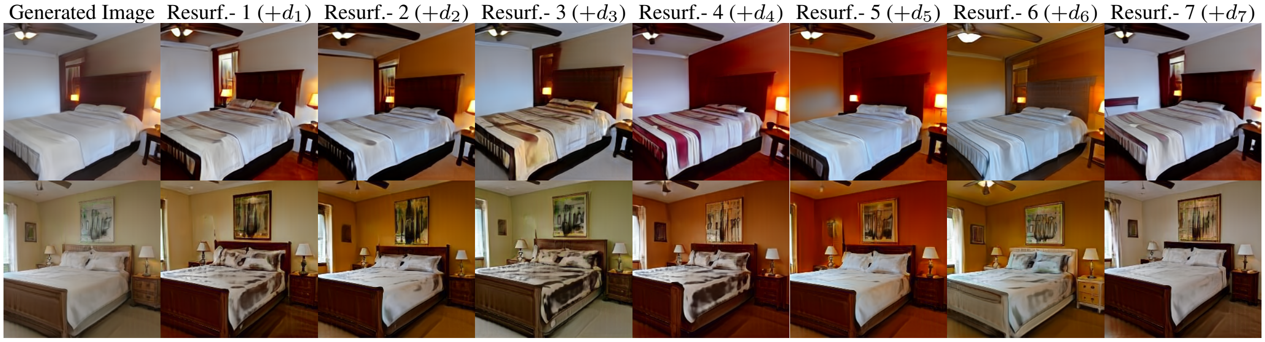


Figure 7. Resurfacing Generated Images. Instead of relighting images, we can generate resurfaced images by swapping our consistency and diversity loss. We apply diversity loss to change the albedo and consistency loss to maintain the shading and overall illumination. The **first column** shows images generated by the original StyleGAN, and the **other columns** show images obtained from $\mathbf{w}^+ + \mathbf{d}_i$ for our resurfacing or recoloring directions. Each column shows the same scene as in the first column, but with varying surface colors and materials, while the individual latent direction $\mathbf{d}_i$ retains its semantics.

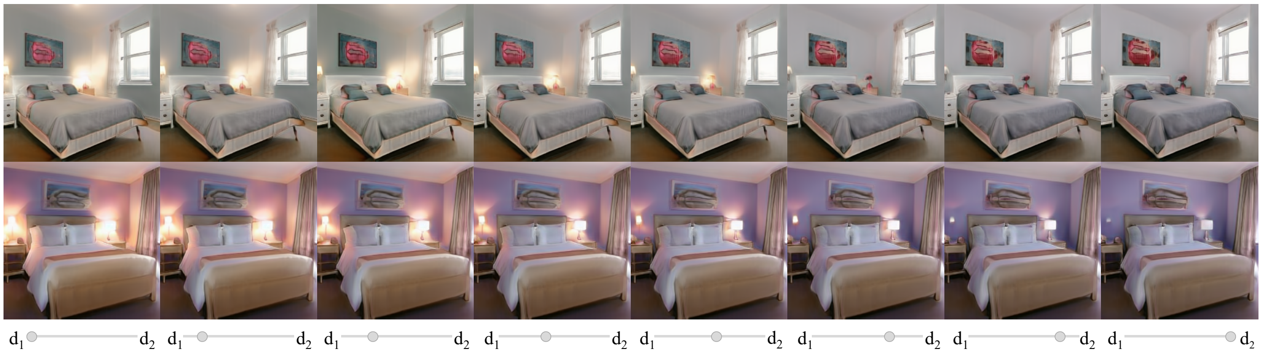


Figure 8. Controllability: The first and last columns of the figure show relighted images generated using our Relit-1 and Relit-5 directions. The bottom section of the figure features a user-controllable slider that enables adjusting the weight of the relighting effects produced by these two directions. Moving the slider from left to right results in a seamless interpolation between the two lighting directions, providing users with precise control over the relighting of the generated images.

Table 1. FID measures distribution shift and not realism. Our generated images are realistic and are out-of-distribution because of large illumination and color changes in the images. This results in large FID scores. KDL in the table is for kitchen, dining and living room which are jointly trained [12].

Type	Bedroom	KDL	Conference	Church	Faces
StyleGAN (SG)	5.01	5.86	9.35	3.80	5.02
SG + Relighting (RL)	14.23	6.87	10.48	12.12	37.87
SG + Resurfacing (RS	17.03	9.41	10.63	18.60	34.06
SG + RL + RS	21.39	11.68	12.71	21.08	37.40

(results in Figure 10).

Comparison to controlontrol: GAN-control (GC) [38] represents lighting with a spherical harmonic predictor pretrained on a parameterized 3D face reconstruction model [8]. This model does not apply to indoor lighting (among other problems, it predicts all points with the same normal have the same shading). We trained a GC model on the LSUN Bedroom dataset with 2 subspaces z^k – *illumination* corresponding to spherical harmonic coefficients, and *other* to represent all other structural information. The model was trained for 800 epochs. GC produces images whose structure varies wildly with any lighting changes, resulting in large albedo changes (*GC* in Figures 3, 5). The difficulty appears to be that the attribute predictor is easily subverted; if the lighting representation cannot produce an image that is (say) dark on the left side, the albedo is adjusted instead.

6. Downstream Applications

Lighting Variance in Surface Normal Prediction: The Multilum dataset [28] provides images of 1000 various indoor scenes, each under 25 lighting conditions, physically relit and captured. A SOTA normal predictor (Omnidata [11, 20]) applied to the test set produces surface normal predictions that vary significantly with changes of light in a fixed scene (Figure 11, purple bar). Finetuning Omnidata with the Multilum training dataset significantly reduces this variance (Figure 11, orange bar); but multiple lightings of a fixed scene are very hard to find. Finetuning with StyLitGAN relights produces improvements that are comparable; using seven distinct relights (Figure 11, pale orange bar) is slightly worse than using 25 distinct relights (Figure 11, yellow bar). The resulting improvement is *not* at the cost

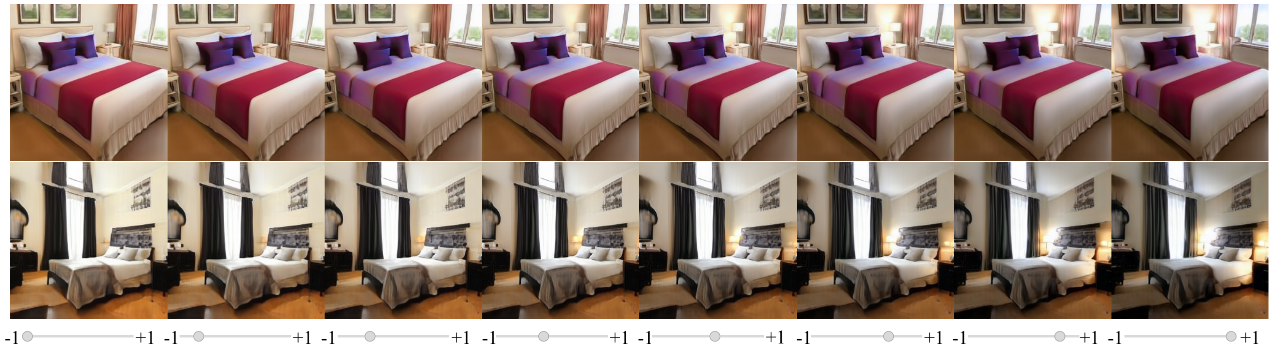
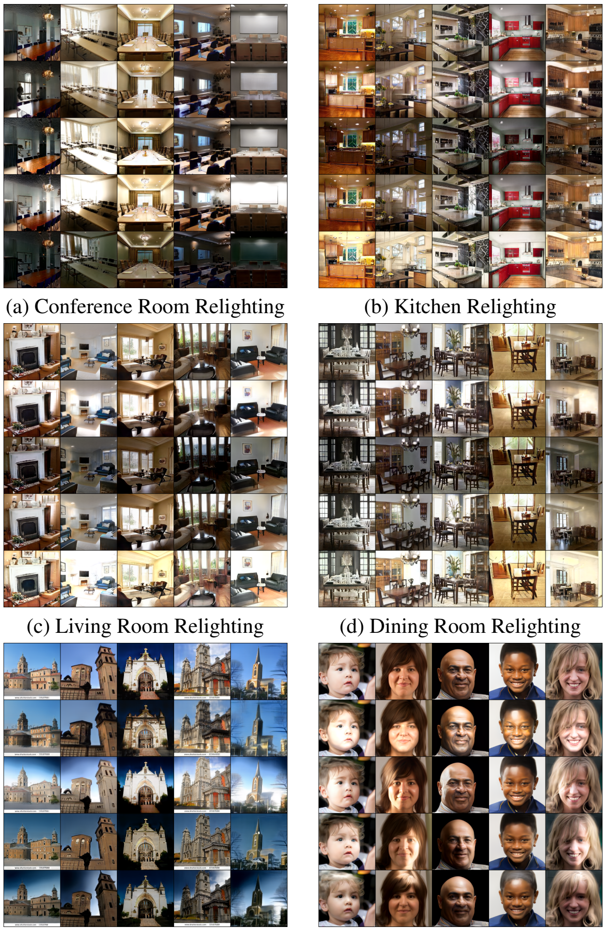


Figure 9. Scaling Directions: The figure depicts the persistent and smooth effects of applying the direction at different scalar coefficients. We use our Relit-2 direction. A slider at the bottom allows the user to adjust the weight of the relighting direction, producing a seamless interpolation when increasing or decreasing the intensity of the chosen direction. The relighting effects range from a well-lit room with the bedside lamp off to weak external lighting with a bedside lamp on.



(a) Conference Room Relighting

(b) Kitchen Relighting

(c) Living Room Relighting

(d) Dining Room Relighting

(e) Relighting outdoor Churches

(f) Portrait Relighting

Figure 10. StyLitGAN extends to finding relighting directions for StyleGANs trained on other datasets.

of base accuracy. Figure 12 compares the accuracy of various finetuned models on the Taskonomy test set [49] (recall this involves thousands of frames each in 10 blocks; we show results by block). Note that finetuned methods mostly show slight accuracy improvements over the base model, but losses in some blocks result in means that match.

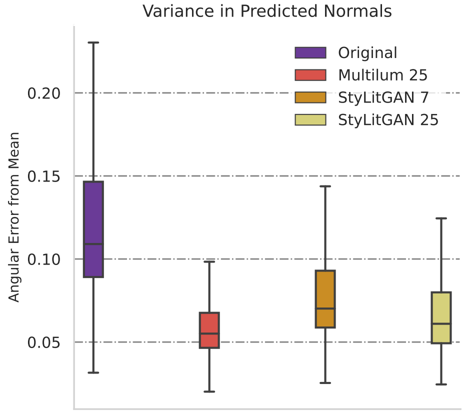


Figure 11. Normal Variance to Lighting reduces when fine-tuning on our relighting dataset. Purple boxplot shows normal variance under relighting for real test scenes from Multilum using the Omnidata normal predictor; orange shows the result of finetuning using Multilum training data; light orange and yellow show the result of finetuning using StyLitGAN images (7 and 25 per scene respectively). The measure is angular error in radians from the mean prediction of a scene for each relit image in the Multilum test set (30 scenes, 25 lightings each). See Figure 16 in Supplementary.

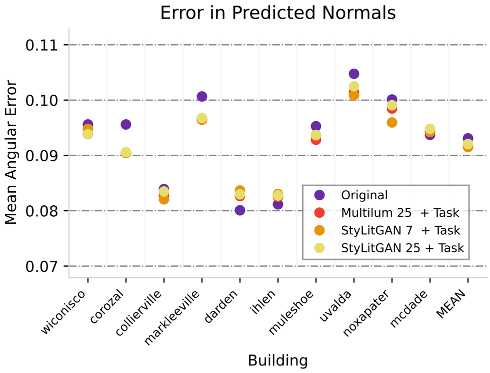


Figure 12. A surface normal predictor is finetuned to increase prediction consistency for relit images of the same scene. General prediction performance on Taskonomy test images parallels that of the original model. Mean across buildings given in last column.

Acknowledgment

We thank Aniruddha Kembhavi, Derek Hoiem, Min Jin Chong, and Shenlong Wang for their feedback and suggestions. This material is based upon work supported by the National Science Foundation under Grant No. 2106825 and by a gift from the Boeing Corporation.

References

[1] Rameen Abdal, Peihao Zhu, Niloy J Mitra, and Peter Wonka. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. *ACM Transactions on Graphics (ToG)*, 40(3):1–21, 2021. 2

[2] Peter N Belhumeur and David J Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 1998. 4

[3] Anand Bhattad and D.A. Forsyth. Cut-and-paste object insertion by enabling deep image prior for reshading. In *2022 International Conference on 3D Vision (3DV)*. IEEE, 2022. 2

[4] Anand Bhattad, Viraj Shah, Derek Hoiem, and D. A. Forsyth. Make it so: Steering stylegan for any image inversion and editing, 2023. 1

[5] Prashanth Chandran, Sebastian Winberg, Gaspard Zoss, Jérémy Riviere, Markus Gross, Paulo Gotardo, and Derek Bradley. Rendering with style: combining traditional and neural approaches for high-quality face rendering. *ACM Transactions on Graphics (ToG)*, 40(6):1–14, 2021. 2

[6] Min Jin Chong and David Forsyth. Jojogan: One shot face stylization. *arXiv preprint arXiv:2112.11641*, 2021. 3

[7] Min Jin Chong, Hsin-Ying Lee, and David Forsyth. Stylegan of all trades: Image manipulation with only pretrained stylegan. *arXiv preprint arXiv:2111.01619*, 2021. 2

[8] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *IEEE Computer Vision and Pattern Recognition Workshops*, 2019. 7

[9] Aditya Deshpande, Jiajun Lu, Mao-Chuang Yeh, Min Jin Chong, and David Forsyth. Learning diverse image colorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6837–6845, 2017. 2

[10] Alexei A Efros and William T Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346, 2001. 2

[11] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10786–10796, 2021. 7

[12] Dave Epstein, Taesung Park, Richard Zhang, Eli Shechtman, and Alexei A Efros. Blobgan: Spatially disentangled scene representations. *arXiv preprint arXiv:2205.02837*, 2022. 2, 3, 7

[13] D.A. Forsyth and Jason J Rock. Intrinsic image decomposition using paradigms. *TPAMI*, 2022 in press. 2, 3, 5, 1

[14] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016. 2

[15] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014. 2

[16] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340, 2001. 2

[17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *arXiv preprint arXiv:1706.08500*, 2017. 6

[18] Le Hui, Jia Yuan, Mingmei Cheng, Jin Xie, Xiaoya Zhang, and Jian Yang. Superpoint network for point cloud oversegmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5510–5519, 2021. 4

[19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 2016. 3

[20] Oğuzhan Fatih Kar, Teresa Yeo, Andrei Atanov, and Amir Zamir. 3d common corruptions and data augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18963–18974, 2022. 2, 7

[21] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34, 2021. 1, 2

[22] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[23] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. CVPR*, 2020. 1, 2

[24] Edwin H Land. The retinex theory of color vision. *Scientific american*, 1977. 3

[25] Zhengqin Li, Jia Shi, Sai Bi, Rui Zhu, Kalyan Sunkavalli, Miloš Hašan, Zexiang Xu, Ravi Ramamoorthi, and Manmohan Chandraker. Physically-based editing of indoor scene lighting from a single image. *arXiv preprint arXiv:2205.09343*, 2022. 6, 1

[26] Zicheng Liao, Hugues Hoppe, David Forsyth, and Yizhou Yu. A subdivision-based representation for vector image editing. *IEEE transactions on visualization and computer graphics*, 2012. 2

[27] Huan Ling, Karsten Kreis, Daiqing Li, Seung Wook Kim, Antonio Torralba, and Sanja Fidler. Editgan: High-precision semantic image editing. *Advances in Neural Information Processing Systems*, 34, 2021. 2

[28] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A multi-illumination dataset of indoor object appearance. In *2019 IEEE International Conference on Computer Vision (ICCV)*, Oct 2019. 2, 7

[29] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews,

Epic Games, Andreas Lehrmann, and AI Borealis. Learning physics-guided face relighting under directional light. 2020. 2

[30] Xingang Pan, Xudong Xu, Chen Change Loy, Christian Theobalt, and Bo Dai. A shading-guided generative implicit model for shape-accurate 3d-aware image synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 2

[31] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Transactions on Graphics (TOG)*, 40(4):1–21, 2021. 2

[32] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2

[33] Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. On aliased resizing and surprising subtleties in gan evaluation. In *CVPR*, 2022. 6

[34] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001. 2

[35] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2287–2296, 2021. 2

[36] Soumyadip Sengupta, Brian Curless, Ira Kemelmacher-Shlizerman, and Steven M Seitz. A light stage on every desk. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 2

[37] Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 1, 2

[38] Alon Shoshan, Nadav Bhonker, Igor Kviatkovsky, and Gerard Medioni. Gan-control: Explicitly controllable gans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14083–14093, 2021. 2, 5, 7

[39] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2015. 3

[40] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Transactions on Graphics*, 2019. 2

[41] Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escolano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. Volux-gan: A generative model for 3d face synthesis with hdri relighting. *arXiv preprint arXiv:2201.04873*, 2022. 2

[42] Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhofer, and Christian Theobalt. Stylerig: Rigging stylegan for 3d control over portrait images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6142–6151, 2020. 2

[43] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2

[44] Andrey Voynov and Artem Babenko. Unsupervised discovery of interpretable directions in the gan latent space. In *International conference on machine learning*, pages 9786–9796. PMLR, 2020. 1, 2

[45] Zongze Wu, Dani Lischinski, and Eli Shechtman. Stylespace analysis: Disentangled controls for stylegan image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12863–12872, 2021. 1, 2

[46] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. Gan inversion: A survey. *arXiv preprint arXiv: 2101.05278*, 2021. 2

[47] Ceyuan Yang, Yujun Shen, and Bolei Zhou. Semantic hierarchy emerges in deep generative representations for scene synthesis. *International Journal of Computer Vision*, 2020. 2, 5

[48] Ning Yu, Guilin Liu, Aysegul Dundar, Andrew Tao, Bryan Catanzaro, Larry S Davis, and Mario Fritz. Dual contrastive loss and attention for gans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6731–6742, 2021. 3

[49] Amir R. Zamir, Alexander Sax, William B. Shen, Leonidas J. Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018. 8

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 3

[51] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019. 2

[52] Jiapeng Zhu, Yujun Shen, Deli Zhao, and Bolei Zhou. In-domain gan inversion for real image editing. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2020. 1, 2

[53] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2017. 2

StyLitGAN: Image-based Relighting via Latent Control

Supplementary Material

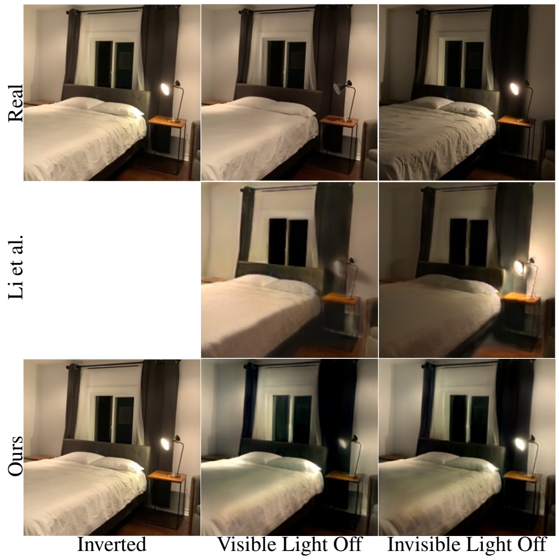


Figure 13. With an appropriate inversion method [4], we can relight real scenes. **Top row:** shows three images of the same scene in different lightings, obtained from [25]. **Middle row:** shows relightings of the first image, obtained by Li *et al.* by inverse rendering, changing luminaire parameters, then forward rendering [25]. **Bottom left:** image generated by passing latent variables from our inversion of the top left 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.

7. Choice of Decomposition

The choice of decomposition matters for relighting without change in geometry and albedo. The best-performing decomposition that was admissible 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. 15. 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.

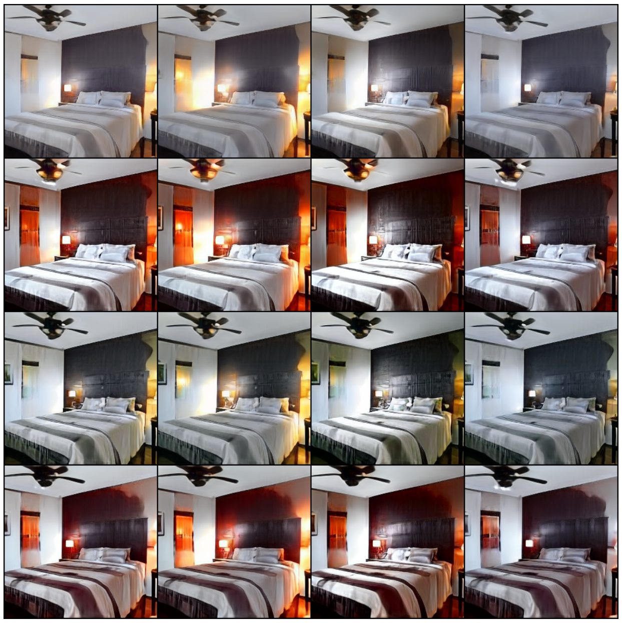


Figure 14. Simultaneous Joint Relighting and Resurfacing. **Columns** show different relights for a fixed resurfacing, and **rows** show different resurfacing for a fixed relight. The interactions between relighting and resurfacing are largely disentangled.

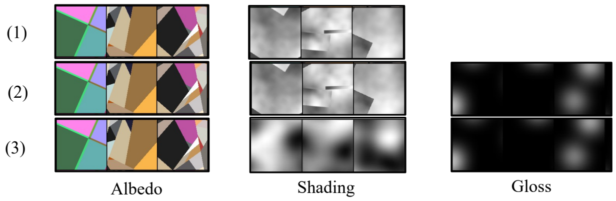


Figure 15. We compare three different decomposition models derived from [13]. We find decomposition (3) to provide significantly better results with small geometry shifts after relighting. In general, the directions obtained with this decomposition are admissible. The other two decompositions resulted 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).

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

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

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

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

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

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

9. 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.

10. 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.

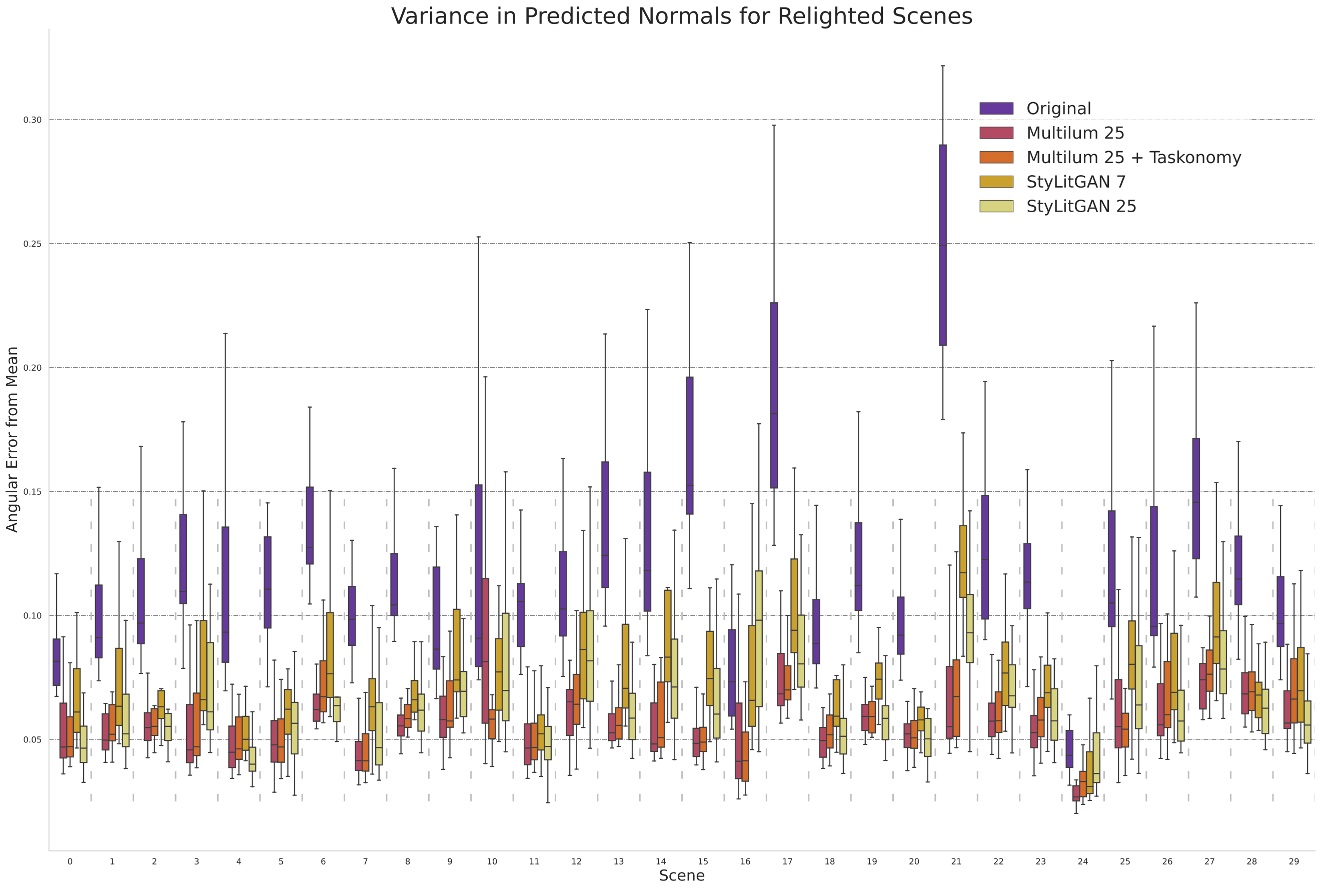


Figure 16. Scene-specific performance of a finetuned normal predictor on full Multilum test set, a comprehensive version of Figure 11.