Physics-Guided Latent Relighting with Geometry-Aware StyLitGAN

Machine Vision Project Report

Satrajit Ghosh

November 2025

GitHub:Geometry_and_Physics_Aware_StyLitGAN

Abstract

Image relighting with StyleGAN-like generators is attractive because it avoids explicit three-dimensional reconstruction, but purely latent methods such as StyLitGAN often ignore geometry and physics, so highlights and shading drift in ways no plausible lighting setup could explain. This project investigates whether approximate geometry and a physically based image-formation model can regularise latent relighting while preserving the flexibility of StyleGAN2. A first prototype uses StyLitGAN with Sobel-based normals, log-domain intrinsic optimisation and a Lambertian renderer, which was used almost as a proof of concept. A second, final system replaces these components with MiDaS depth and perspective-correct normals, a neural intrinsic decomposer, spherical harmonics lighting, a Cook–Torrance BRDF and multi-scale, light-conditioned latent directions trained with perceptual losses. Experiments on StyleGAN2 bedroom scenes show that this final system produces stronger and more coherent relighting, better albedo preservation and substantially higher training efficiency than both the StyLitGAN baseline and the initial prototype.

1 Introduction

Image relighting is a central problem in vision and graphics, with applications in data augmentation, robustness testing, virtual staging and interactive VR/AR editing. Many of these scenarios assume fixed scene geometry and materials under varying illumination, but typical datasets expose only a few lighting conditions and classical pipelines require explicit geometry and reflectance estimation, which is brittle and expensive for cluttered indoor scenes.

StyleGAN-based generators offer an alternative. Methods such as StyLitGAN operate directly in the latent space of a pretrained StyleGAN2, learning directions that appear to change lighting while largely preserving scene content. This latent-space view leverages the generator as a strong image prior and avoids multi-view reconstruction, mesh extraction and explicit BRDF fitting. However, purely latent approaches do not enforce a concrete image-formation model and can violate basic physical constraints: shading edges need not align with surface orientation, specular highlights may move inconsistently with any plausible light configuration and different "lighting directions" need not correspond to a coherent three-dimensional illumination change.

This project investigates whether approximate geometry and a physically motivated shading model can regularise latent-space relighting without sacrificing the flexibility of a StyleGAN2 generator. Instead of treating the generator as a black box, the project factors its outputs into albedo and shading, infers approximate scene geometry and compares shading induced by latent directions

against a physics-based renderer. Two implementations are developed on top of a StyLitGAN-like baseline: an initial prototype using Sobel-based normals, optimisation-based intrinsic decomposition and a Lambertian renderer, and a final system using MiDaS depth and perspective-correct normals, a neural intrinsic decomposer, spherical harmonics lighting and a Cook–Torrance microfacet BRDF. The remainder of this report formalises this framework, describes both implementations and evaluates how geometry and physics affect the coherence of latent-space relighting.

2 Problem Statement

This project studies generative image relighting in the latent space of a pretrained StyleGAN2 generator. Let G map extended latent codes w^+ to images $I = G(w^+)$ of indoor scenes. The objective is to learn latent directions $\{d_k\}$ such that codes of the form $w^+ + \alpha d_k$ produce images that depict the same scene under different illumination, rather than different scenes. A useful lighting direction should modify shading and apparent illumination while approximately preserving layout, object identity and material appearance.

Solving this latent relighting problem enables applications where content-preserving lighting variation is critical: data augmentation for illumination-robust models, virtual staging and relighting tools for VR/AR environments and controllable generative models where lighting is an explicit factor. In these settings it is not sufficient for edits to look plausible in isolation; they should be consistent with a physically reasonable change in lighting for a fixed underlying scene.

Existing latent methods such as StyLitGAN address this task without an explicit image-formation model, relying on the generator prior and intrinsic-style losses. As a result, they can produce relit images that are visually convincing but physically inconsistent: shading boundaries can drift away from surface orientation, specular highlights may move in ways no single light source could cause and different latent "lighting directions" may not correspond to any coherent three-dimensional illumination pattern. The premise of this project is that even approximate geometry and a simple physically based shading model can be used to regularise latent relighting. By estimating geometry from generated images, factoring them into albedo and shading and comparing latent-induced shading changes against a concrete lighting model, the project aims to constrain what counts as a valid lighting edit in latent space and thereby improve control, interpretability and physical plausibility over purely latent baselines.

Given pairs $(I, \tilde{I}_{k,\alpha})$, the training objective is to find $\{d_k\}$ such that albedo is approximately preserved, shading changes are compatible with the estimated geometry, the relit images are close to physics-based renders under suitable L, and different directions are both distinguishable and diverse. At an abstract level the loss is

$$\mathcal{L}_{\text{total}} = \lambda_{\text{alb}}\mathcal{L}_{\text{alb}} + \lambda_{\text{geom}}\mathcal{L}_{\text{geom}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}} + \lambda_{\text{dist}}\mathcal{L}_{\text{dist}} + \lambda_{\text{div}}\mathcal{L}_{\text{div}} + \lambda_{\text{chg}}\mathcal{L}_{\text{chg}}, \quad (1)$$

where $\mathcal{L}_{\text{alb}}$ penalises changes in estimated albedo, $\mathcal{L}_{\text{geom}}$ encourages geometry-aligned shading variations, $\mathcal{L}_{\text{phys}}$ pulls relit images toward physics-based renders of the same scene, $\mathcal{L}_{\text{perc}}$ stabilises appearance in a perceptual feature space, $\mathcal{L}_{\text{dist}}$ and $\mathcal{L}_{\text{div}}$ enforce that different directions are identifiable and non-redundant, and $\mathcal{L}_{\text{chg}}$ prevents trivial near-identity edits.

3 Literature Review

Work on controllable image relighting sits at the intersection of latent-space editing in generative models, intrinsic image decomposition and inverse rendering, and physically based lighting and reflectance modelling. The two implementations in this report build directly on these strands.

StyleGAN2 provides the base generator used throughout this project. It refines the original StyleGAN architecture and shows that the extended latent space W^+, where each synthesis layer receives its own style code, offers improved disentanglement and image quality [1]. A substantial body of work studies semantic editing in such latent spaces. InterFaceGAN, GANSpace and SeFa demonstrate that directions discovered by linear classifiers, principal component analysis or closed-form eigen decompositions can be used to control attributes such as pose, expression and sometimes lighting in a relatively interpretable manner [5, 3, 4]. These methods justify the assumption that adding vectors in W^+ can act as a controllable editing mechanism, but the directions remain semantic knobs with no explicit connection to an image formation model or to physical illumination.

Intrinsic image decomposition provides a complementary view by explicitly separating reflectance and illumination. Retinex and its descendants model an image as the product of an albedo term and a shading term, with priors that encourage piecewise-smooth shading and piecewise-constant reflectance [6]. Large-scale datasets and variational methods such as "Intrinsic Images in the Wild" extend this framework to complex real scenes [7]. Barron and Malik formulate joint estimation of shape, illumination and reflectance under a Lambertian assumption, including log-domain reconstruction and geometry-aware regularisation terms that are closely related to the losses used in our optimisation-based decomposer [8]. More recent approaches replace explicit optimisation with convolutional networks that directly predict albedo and shading from a single image, for example IntrinsicNet and related CNN-based decomposers [9, 10]. These works motivate both the intrinsic image factorisation and the use of optimisation-based and neural decomposers in the first and second implementations respectively.

Estimating geometry from single images is another key ingredient. MiDaS and its Dense Prediction Transformer variant (DPT) learn robust scale-invariant monocular depth from heterogeneous datasets and generalise well across domains [17, 18]. Depth maps predicted by MiDaS or DPT can be converted to surface normals via finite differences under a perspective camera model, providing per-pixel orientation fields that are sufficient for diffuse shading models. In the initial prototype we rely on simple image gradients as a crude proxy for normals, whereas the final system replaces these with MiDaS- and DPT-based geometry, bringing our shading model closer to the assumptions of inverse rendering methods.

Lighting and reflectance models connect intrinsic images and geometry to observable shading. At the simplest level, Lambertian reflectance models purely diffuse surfaces, and the Phong model augments this with an ad hoc specular term [8, 11, 12]. These models are widely used in both classical inverse rendering and neural image-formation losses. Physically based rendering relies instead on microfacet BRDFs, in particular the Cook–Torrance model, which decomposes specular behaviour into a normal distribution function, a geometry term and a Fresnel term, and supports spatially varying roughness and metallicity [13, 14, 15]. Spherical harmonics provide a compact basis for low-frequency environment lighting and allow diffuse irradiance to be expressed as a low-order expansion evaluated at the surface normal [16]. The first implementation in this report uses Lambertian and Phong shading with directional lights as a physics oracle, while the second adopts a Cook–Torrance microfacet BRDF with GGX-like normal distributions and Schlick Fresnel, combined with second-order spherical harmonics to represent environment illumination.

Within the StyleGAN ecosystem, there is growing interest in tying latent control to geometry and lighting. StyleRig introduces a rigging framework for StyleGAN that exposes 3D-aware controls over pose and facial expression by coupling latent codes to an underlying parametric head model [19]. StyLitGAN is closer to the present work: it operates in the W^+ latent space of a StyleGAN generator, uses an intrinsic decomposition to separate albedo and shading in generated images, and searches for latent directions that change shading while preserving albedo, guided by consistency, distinction and diversity losses [2]. However, StyLitGAN still treats geometry implicitly and relies

on simple intrinsic-style losses; there is no explicit normal field, no physically motivated BRDF, and no direct parameterisation of lighting beyond discrete direction indices.

The implementations developed here can be seen as pushing this latent-relighting paradigm toward a more physically grounded regime while preserving its advantages. The first prototype adds explicit geometry in the form of normal estimates, a geometry-aware intrinsic decomposition and an analytical Lambertian renderer to regularise latent directions. The second, final system replaces these with MiDaS-based depth and normals, a neural intrinsic decomposer inspired by IntrinsicNet, a Cook–Torrance microfacet model with spherical harmonics environment lighting, and a multi-scale, light-conditioned latent basis. Training in both cases uses standard optimisers such as Adam and AdamW [22, 23], together with VGG-based perceptual losses and LPIPS distances to better align learned relighting with physically rendered targets and human perception [20, 21]. In contrast to earlier latent editing work, the latent directions here are explicitly coupled to estimated geometry and a concrete image formation model, rather than inferred solely from data-driven semantics.

4 Initial Geometry and Physics-Guided Prototype (Implementation 1)

The first prototype couples a pre-trained StyleGAN2 generator [1] with deliberately simple geometry and physics supervision. The system uses Sobel-derived normals [24], a log-domain intrinsic decomposer inspired by Retinex and intrinsic-image methods [6, 7, 8, 9, 10], and a Lambertian renderer [8]. Latent "lighting directions" are learned in the W^+ space following StyLitGAN-style editing [2, 5, 3, 4], but supervision comes only from this approximate physics pipeline rather than paired relit data.

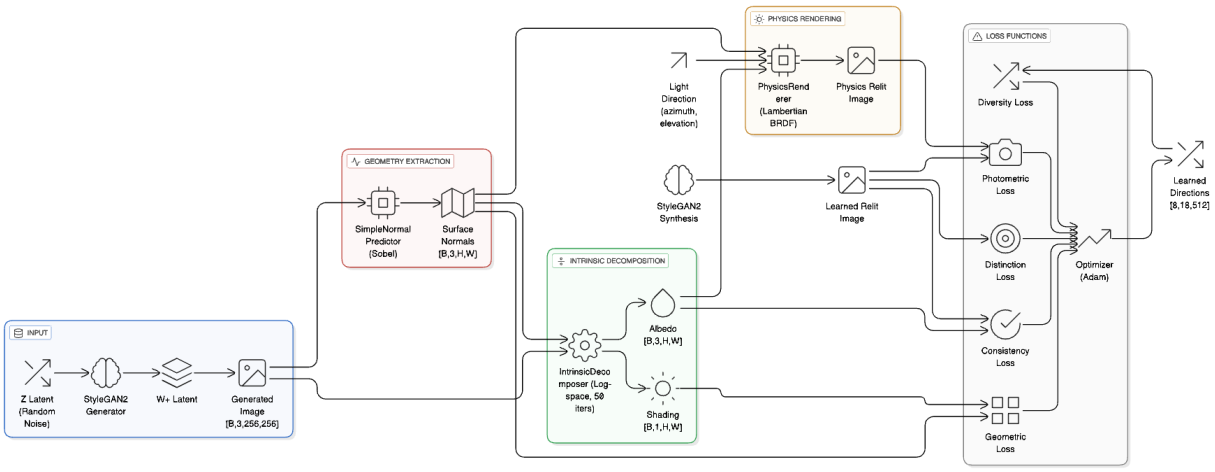

Figure 1: System overview of Implementation 1. StyleGAN2 synthesises bedroom images from z samples. A Sobel-based normal predictor produces surface normals, a log-space intrinsic decomposer recovers albedo and shading, and a simple physics renderer generates Lambertian relit targets. Learned W^+ directions are trained with albedo, geometric and photometric losses plus distinction and diversity terms.

4.1 Geometry Estimation via Gradient-Based Normals

Geometry is approximated directly from image gradients. Given an RGB image $I(x, y)$, the normal field is

$$N(x, y) \propto (-\partial_x I_L(x, y),\ -\partial_y I_L(x, y),\ 1),$$

where I_L is a grayscale luminance channel and ∂_x, ∂_y are Sobel derivatives [24], followed by per-pixel normalisation. This captures some edges and large planes but has no notion of depth or global shape; the resulting normals are dominated by texture and noise.

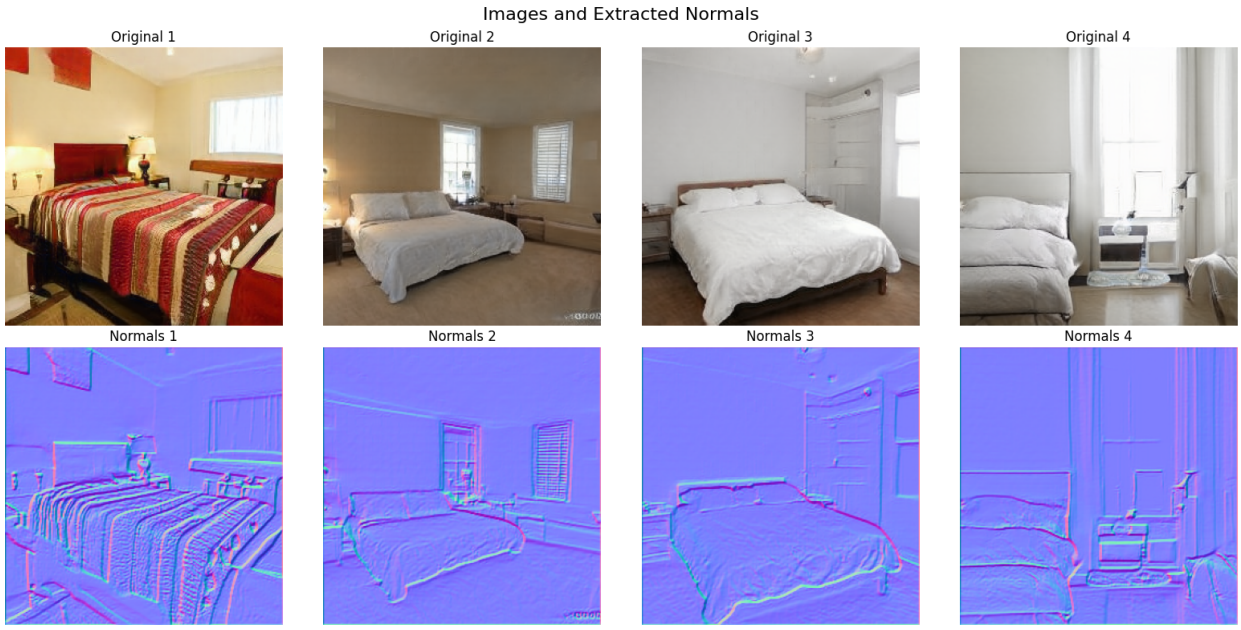

Figure 2: StyleGAN2 bedroom samples and corresponding Sobel-derived normals in Implementation 1. Normals pick up edges and planes but are contaminated by texture and do not reflect consistent three-dimensional structure.

4.2 Intrinsic Decomposition through Log-Space Optimisation

The intrinsic module follows $I = A \odot S$, with A denoting albedo and S scalar shading [6, 7, 8, 9]. Implementation 1 works in the log domain,

$$\log I = \log A + \log S,$$

and solves for $\log A$ and $\log S$ per image by minimising a weighted sum of: (i) reconstruction $A \odot S \approx I$, (ii) shading smoothness, (iii) albedo sparsity and (iv) geometry-weighted shading smoothness that damps variation in $\log S$ where Sobel normals are nearly flat. Each decomposition runs for a fixed number of gradient-descent iterations, making the prototype slow and sensitive to hyperparameters.

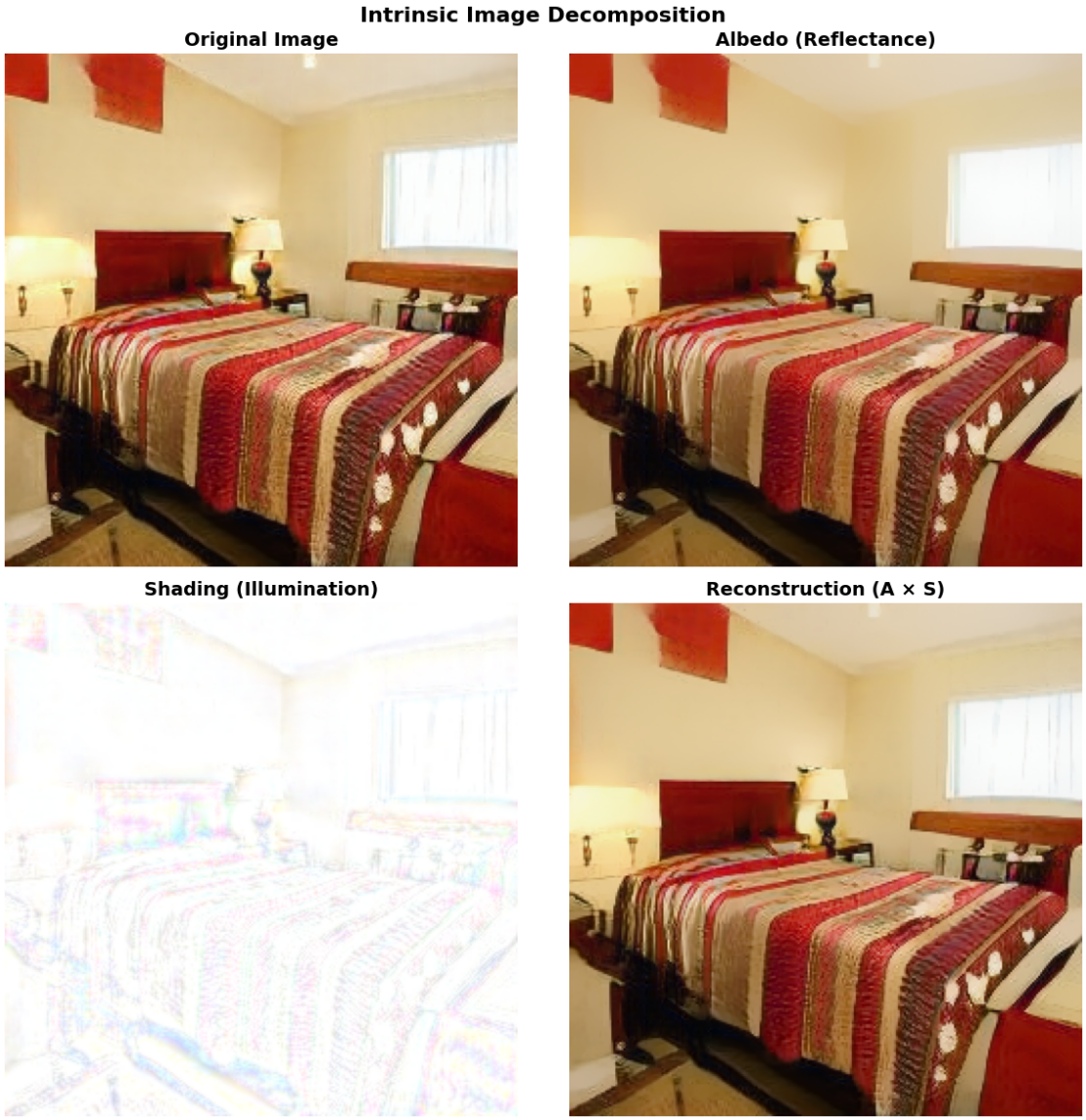

Figure 3: Intrinsic decomposition in Implementation 1. Top left: original image. Top right: estimated albedo. Bottom left: estimated shading. Bottom right: reconstruction $A \odot S$. Results are roughly plausible but noisy and require per-image optimisation.

4.3 Physics-Based Renderer with Lambertian and Phong Models

Supervision targets are generated with a simple Lambertian renderer [8]. For unit normal N and directional light L with intensity I_L,

$$S_{\text{diffuse}}(x, y) = I_L \, \max(0, N(x, y) \cdot L),$$

and the Phong specular term

$$S_{\text{spec}}(x, y) = k_s \, \max(0, H(x, y) \cdot N(x, y))^{n_s}$$

adds a view-dependent highlight (with half-vector H, specular scale k_s and exponent n_s), plus an ambient term k_a. Parameters are hand-tuned and aggressive light directions are sampled so that physics-only renders exhibit strong relighting that latent directions are expected to mimic.

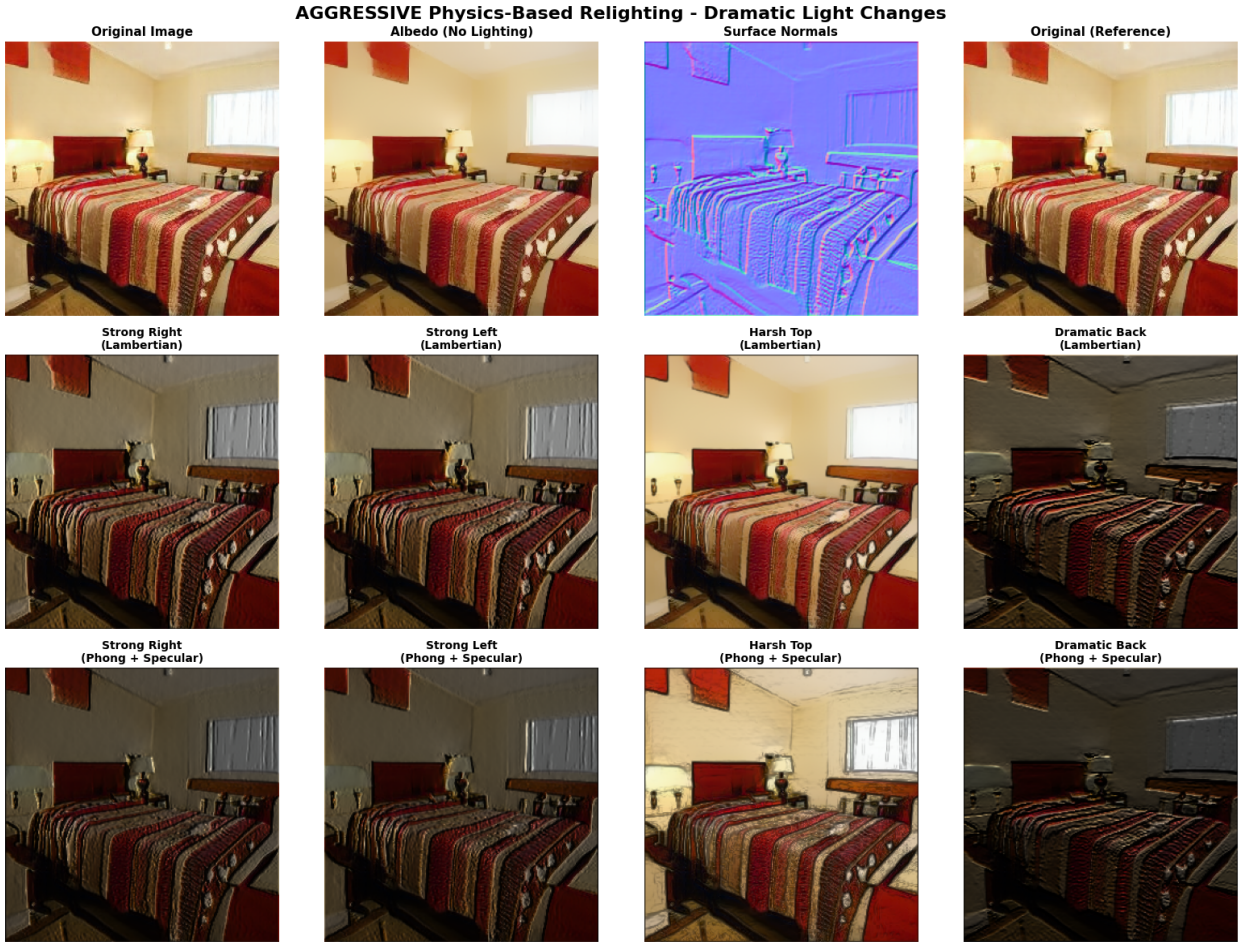

Figure 4: Physics-based relighting in Implementation 1. Left to right, top to bottom: original image, albedo, Sobel normals and several Lambertian light directions. The renderer produces strong, spatially structured illumination changes used as supervision targets.

4.4 Physics-Guided Latent Direction Learning

Latent lighting directions are learnable tensors in W^+ [2, 5, 3, 4]. A base code w^+ from the StyleGAN2 mapping network [1] is perturbed as

$$\tilde{w}_k^+ = w^+ + \alpha d_k,$$

for step size α, yielding original $I = G(w^+)$ and relit $\tilde{I}_k = G(\tilde{w}_k^+)$ images. The intrinsic module provides (A, S) and $(\tilde{A}_k, \tilde{S}_k)$, and the physics renderer provides shading targets S_k^{phys} from Sobel normals and light directions L_k.

Losses encourage: (i) albedo consistency $\|A - \tilde{A}_k\|_1$, (ii) geometry-weighted shading behaviour (small $\nabla \tilde{S}_k$ on planar regions, alignment with normal edges), (iii) photometric agreement between $\tilde{S}_k$ and S_k^{phys}, (iv) direction separability via a small classifier and (v) direction diversity via correlation penalties [4]. Direction parameters are optimised with Adam [22] while the generator stays frozen.

4.5 Qualitative Results and Failure Analysis

In practice, Implementation 1 barely changes the images. Sobel normals are dominated by texture, the intrinsic optimisation amplifies noise, and the physics renderer is driven by unreliable geometry. As a result, gradients on the latent directions are weak and inconsistent.

Figure 5: StyleGAN2 bedroom images used as base samples for Implementation 1. Lighting is already varied and plausible, so useful relighting requires strong, geometry-aware edits.

Figure 6 compares learned directions and physics-only relighting. The Lambertian oracle creates large illumination changes, but the learned directions produce images almost identical to the originals even for sizeable α.

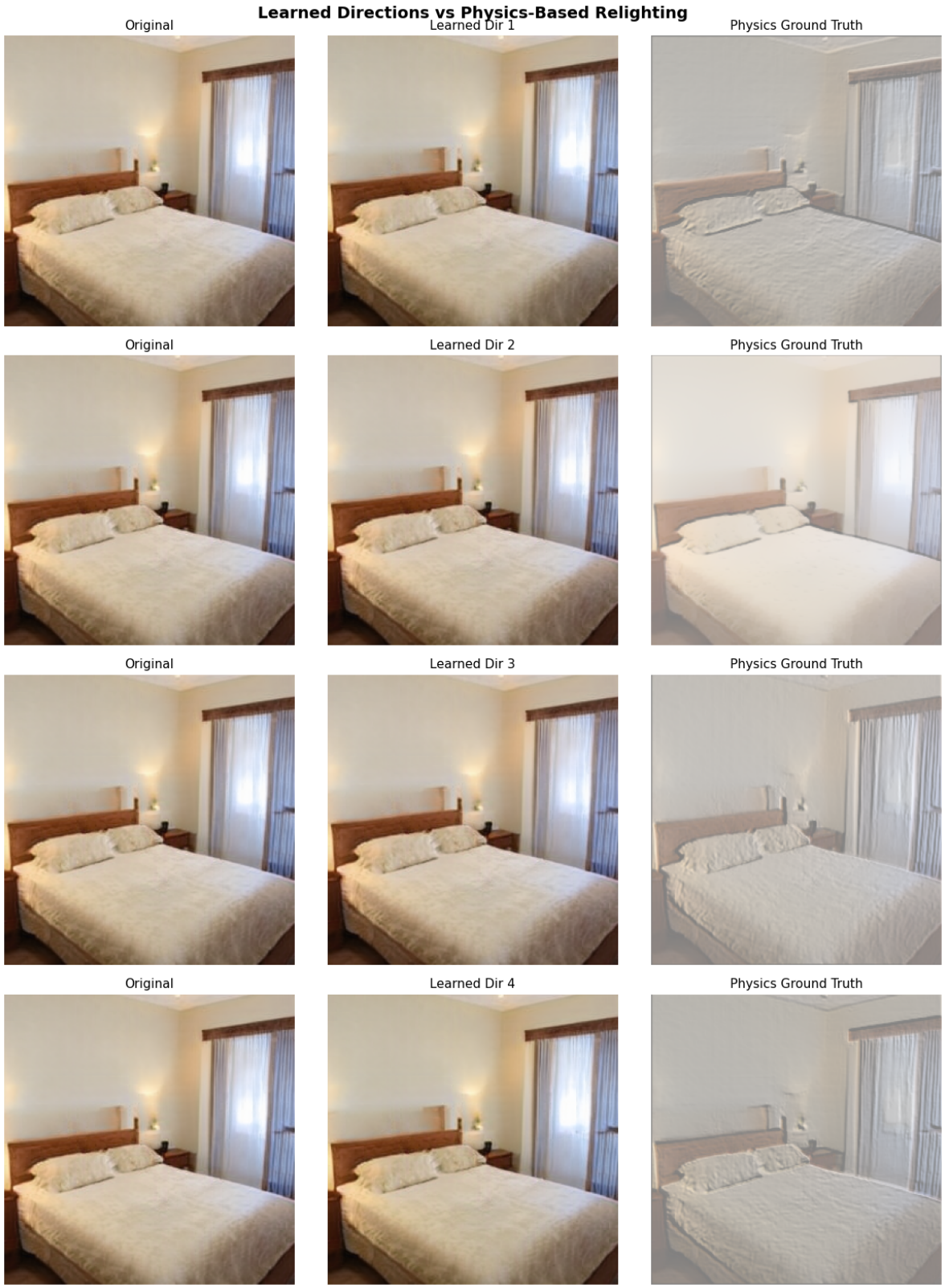

Figure 6: Learned directions vs physics-based relighting for Implementation 1. Left: original. Centre: learned latent direction. Right: physics-only relighting. Physics targets show strong relighting; latent edits are almost invisible.

A broader comparison shows the same pattern: the pure physics renderer yields strong global relighting, while both this prototype and a purely latent StyLitGAN-style baseline mostly preserve original lighting and act like mild contrast changes.

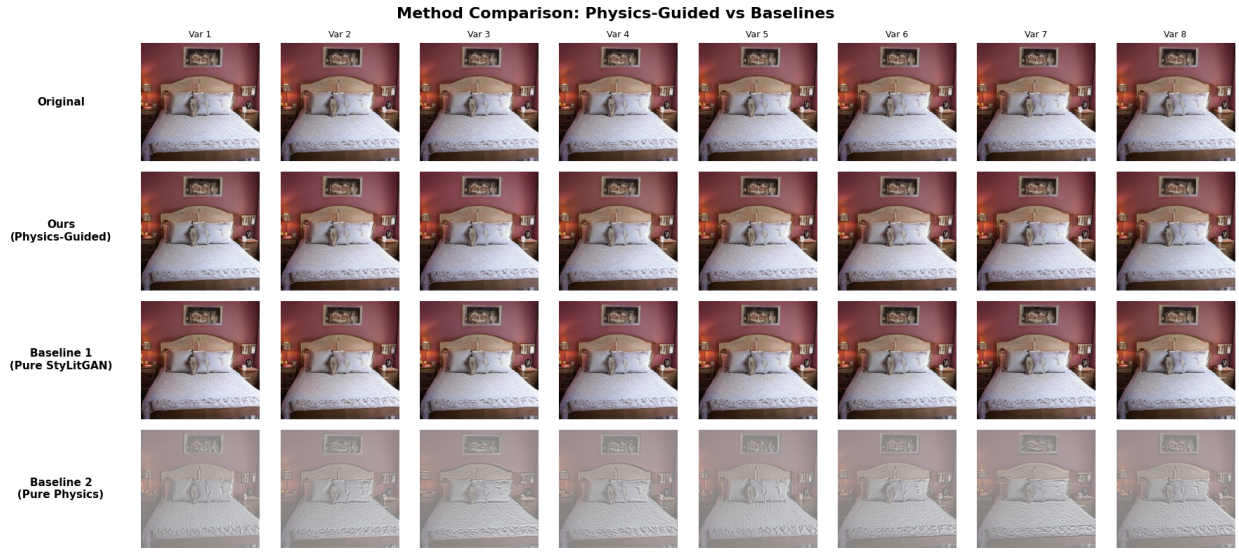

Figure 7: Method comparison for Implementation 1. Top row: original image under different seeds. Second row: outputs from Implementation 1. Third row: purely latent baseline without explicit physics. Bottom row: pure physics renderer. Only the physics-only row produces strong, coherent relighting.

Training curves reinforce this. Total loss and the individual terms remain noisy over 50 iterations and show no clear convergence towards physics-consistent directions.

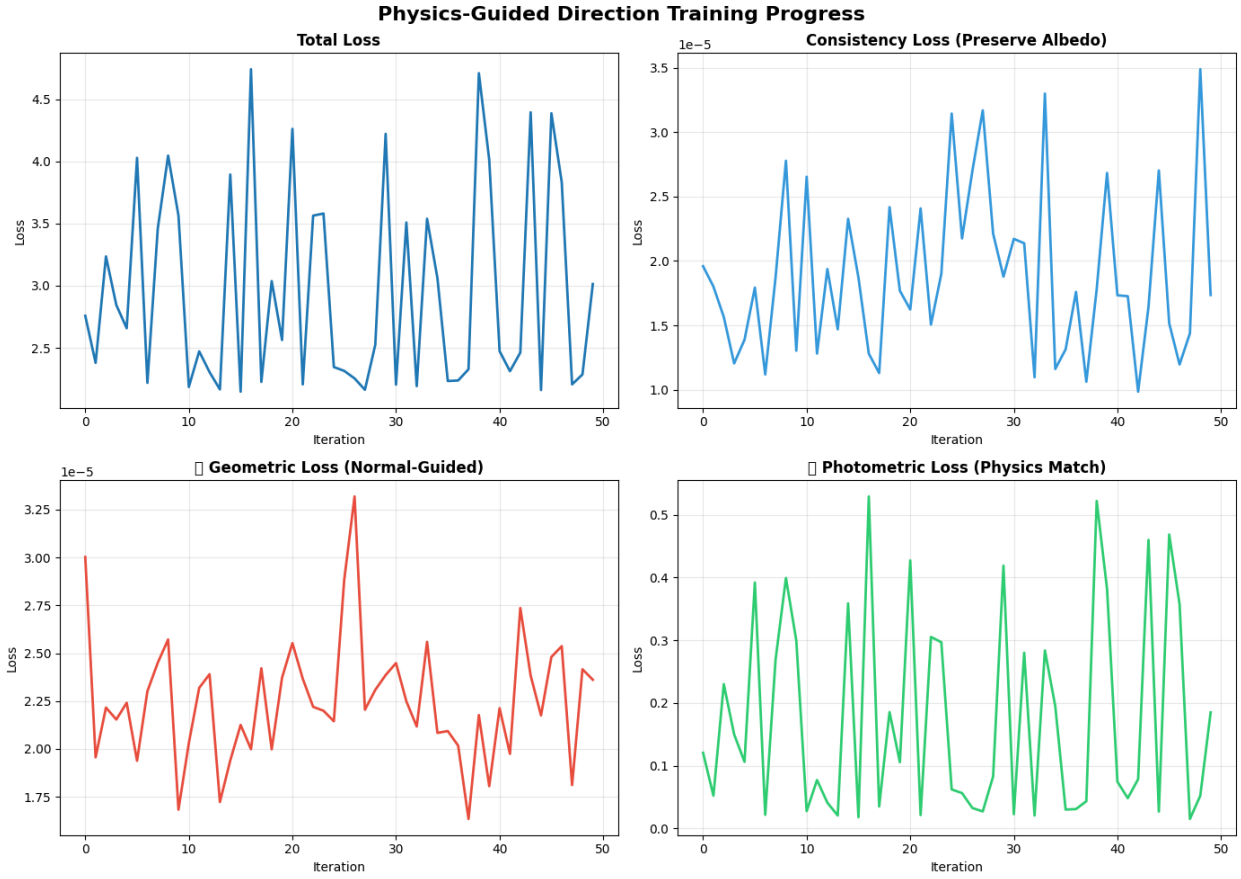

Figure 8: Training behaviour of Implementation 1. Total loss and component losses fluctuate without stabilising, despite the small network and short schedule.

Overall, this prototype confirms that physics-guided latent relighting is implementable, but the combination of crude geometry, slow intrinsic optimisation and hand-tuned shading yields almost identity edits and unstable training. This motivates a second implementation that replaces each weak component with a more principled alternative.

5 Final Physics-Guided StyLitGAN System (Implementation 2)

The second implementation keeps the StyleGAN2 backbone [1] and the idea of learning latent directions, but replaces every weak component of Implementation 1. Geometry comes from a MiDaS/DPT depth network, intrinsic images from a neural decomposer, lighting from spherical harmonics and shading from a Cook–Torrance microfacet BRDF; latent directions become light-conditioned and are spread across StyleGAN layers. The goal is to keep physics-guided training while removing the failure modes of the prototype.

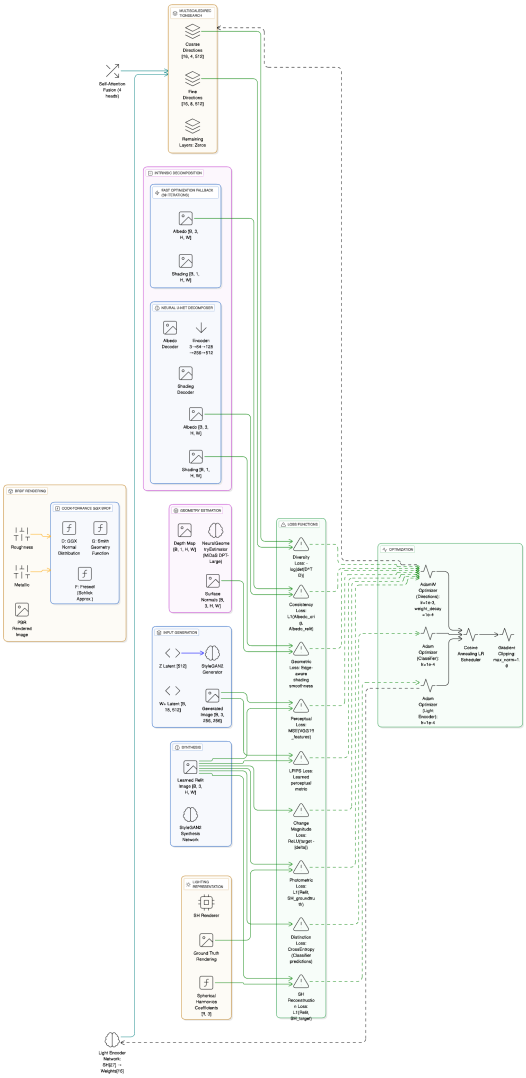

Figure 9: Architecture of Implementation 2. Geometry and intrinsic components are evaluated without gradients; only the light encoder, latent-direction basis and StyleGAN2 generator are trained.

5.1 Neural Geometry Estimation with MiDaS Depth and Normals

Implementation 2 uses a MiDaS/DPT depth network [17, 18]. For each generated image $I \in \mathbb{R}^{3\times H\times W}$, a dense inverse depth map $\hat{D}$ is predicted and converted to camera-space points

$$X(u,v) = \hat{D}(u,v) \begin{bmatrix} (u-c_x)/f_x \\ (v-c_y)/f_y \\ 1 \end{bmatrix},$$

under a perspective camera with intrinsics (f_x, f_y, c_x, c_y). Normals are obtained via finite differences,

$$\hat{N}(u,v) \propto \big(X(u{+}1,v) - X(u,v)\big) \times \big(X(u,v{+}1) - X(u,v)\big),$$

followed by normalisation. These normals are geometry-driven, align with planes and object boundaries and remain stable under relighting.

5.2 Neural Intrinsic Decomposition and Fast Refinement

The log-space optimiser is replaced by a feed-forward intrinsic decomposer F_ϕ, implemented as a U-Net with two decoders [9, 10]. Given I,

$$(\hat{A}, \hat{S}) = F_\phi(I), \qquad I \approx \hat{A} \odot \hat{S},$$

with $\hat{A} \in \mathbb{R}^{3\times H\times W}$ and $\hat{S} \in \mathbb{R}^{1\times H\times W}$. This approximates the optimisation-based intrinsic solution in a single pass. A brief refinement then runs the geometry-weighted smoothness prior for a few iterations with $\hat{A}$ fixed and gradients stopped w.r.t. the directions, keeping intrinsic estimates compatible with geometry without reintroducing heavy optimisation into training.

5.3 Spherical Harmonics Lighting Model

Lighting is parameterised with low-order spherical harmonics (SH) [16]. Diffuse irradiance at a surface with normal $\hat{N}$ is

$$E(\hat{N}) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} c_{\ell m}\, Y_{\ell m}(\hat{N}),$$

where $Y_{\ell m}$ are real SH basis functions and $c_{\ell m}$ are lighting coefficients. A light encoder g_ψ outputs these coefficients $L = g_\psi(z)$ conditioned on the latent seed z, and diffuse shading is

$$\hat{S}_{\mathrm{SH}}(u,v) = \max\big(E(\hat{N}(u,v)), 0\big).$$

This captures soft, spatially varying environment illumination such as windows and ceiling lights that a single directional light cannot represent.

5.4 Physically Based Shading with Cook–Torrance Microfacet BRDF

The physics renderer now uses a Cook–Torrance microfacet BRDF [13, 14] with GGX-style normal distribution and Schlick Fresnel [15]. For normal $\hat{N}$, view direction V, light direction L and half-vector $H = \frac{V+L}{\|V+L\|}$,

$$f_{\mathrm{spec}}(L, V; \hat{N}, \theta) = \frac{D_{\mathrm{GGX}}(H;\alpha)\, G_{\mathrm{Smith}}(L, V;\alpha)\, F_{\mathrm{Schlick}}(V, H; F_0)}{4(\hat{N}\cdot L)(\hat{N}\cdot V)},$$

with roughness α and base reflectance F_0 from metallicity. Diffuse reflection uses the SH irradiance $E(\hat{N})$ scaled by $(1 - F_0)$. This model provides a more realistic, view-dependent shading signal than the Lambertian system in Implementation 1.

5.5 Multi-Scale, Light-Conditioned Latent Directions

Latent directions are redesigned as a light-conditioned basis distributed across StyleGAN layers. A bank

$$\mathcal{D} = \{D_k^{\text{coarse}}, D_k^{\text{fine}}\}_{k=1}^{K}$$

contains coarse-layer offsets $D_k^{\text{coarse}} \in \mathbb{R}^{L_c \times 512}$ and mid-layer offsets $D_k^{\text{fine}} \in \mathbb{R}^{L_f \times 512}$. Given SH coefficients L, a light encoder produces weights $w(L) \in \mathbb{R}^K$ and the aggregated direction is

$$\Delta W^+(L) = \sum_{k=1}^{K} w_k(L) \, [D_k^{\text{coarse}} \,\|\, D_k^{\text{fine}} \,\|\, 0],$$

with zeros for remaining layers and $\|$ denoting concatenation. For a base latent W^+_{base},

$$W^+_\alpha = W^+_{\text{base}} + \alpha \, \Delta W^+(L),$$

so α controls edit strength and L controls structure. Directions become a light-conditioned operator rather than fixed global offsets.

5.6 Loss Design and Gradient Flow in Implementation 2

Losses mirror those of the prototype but use stronger modules. Albedo consistency $\|\hat{A}_{\text{relit}} - \hat{A}\|_1$ monitors how stable albedo remains and is used mainly as a diagnostic. A geometric term damps shading gradients on geometrically flat regions using MiDaS normals. A photometric loss matches predicted shading $\hat{S}_{\text{relit}}$ to SH+microfacet shading $\hat{S}_{\text{phys}}$ and acts as a main training signal.

Perceptual losses based on VGG features and LPIPS [20, 21] are applied between relit images and physics renders, encouraging visually plausible relighting. Distinction and diversity penalties, following latent factorisation ideas [4], keep directions separable and prevent collapse to a single effective edit. A magnitude penalty on $\|\Delta W^+(L)\|_2$ keeps edits from exploding at high α.

MiDaS and the intrinsic decomposer run under `no_grad` with frozen parameters. Gradients flow only through the light encoder, direction basis, any fusion layers and, optionally, the StyleGAN2 generator. This avoids backpropagation through heavy geometry or intrinsic networks and yields a well-conditioned optimisation problem.

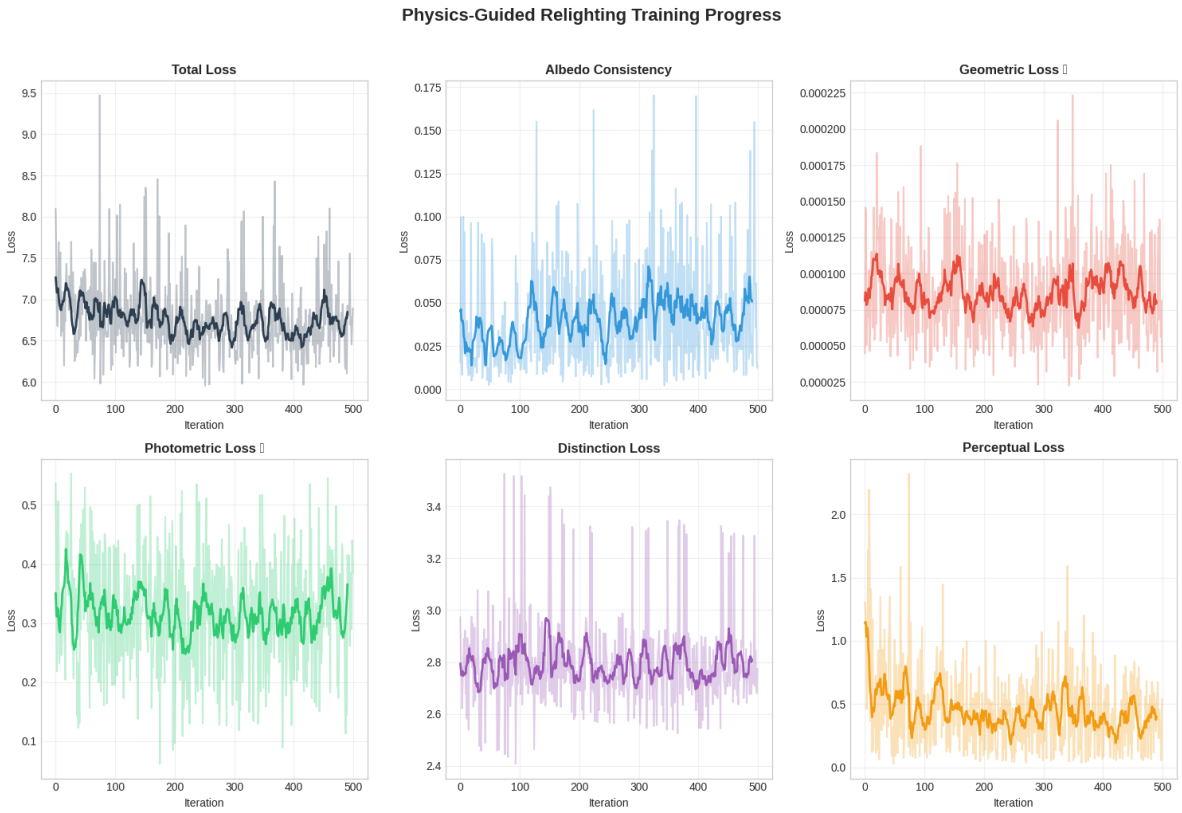

Figure 10: Training curves for Implementation 2. Total loss and major components settle gradually over about 500 iterations, unlike the noisy behaviour in Implementation 1.

5.7 Empirical Behaviour and Advantages of Implementation 2

Qualitatively, Implementation 2 produces strong, geometry-aware relighting. Sweeping α for a single direction shows a smooth progression from mild exposure or colour shifts at low values to large, coherent illumination changes at higher values; difference maps concentrate on light sources, walls and shadow boundaries rather than texture noise.

Figure 11: Alpha progression for a single learned direction in Implementation 2. Top: relit images for increasing α. Bottom: amplified difference maps. Edits remain smooth and geometry-aware and become dramatic only at larger α.

Across multiple directions, each basis element modulates a distinct lighting pattern: some

brighten the bed and wall, others emphasise side lamps or shift global colour temperature. The same direction behaves consistently across images with similar geometry, unlike the almost invisible edits in Implementation 1.

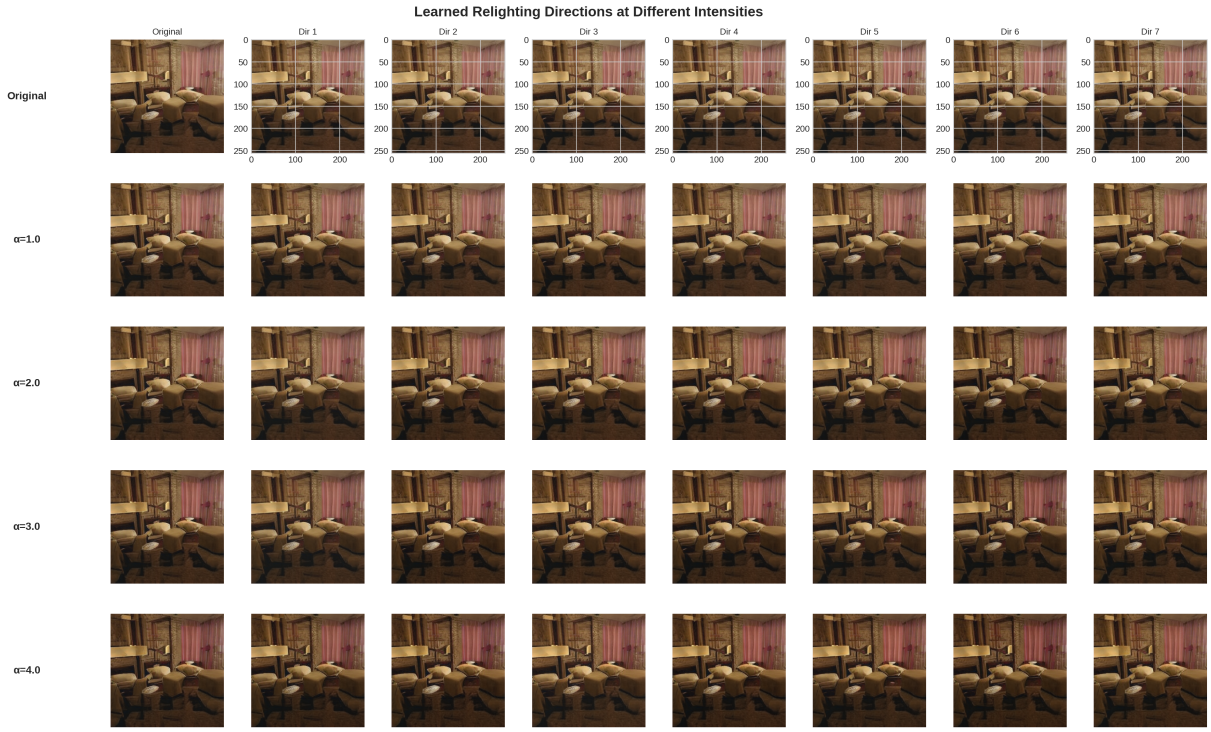

Figure 12: Learned relighting directions at different intensities in Implementation 2. Each column is a direction, each row an α value. Directions induce clear, controllable lighting changes rather than near-identity edits.

Single-image examples show that scene layout and texture remain intact while lighting changes in a structured way: walls and furniture retain shape and colour, but shading distribution and global colour balance shift according to the chosen direction and α.

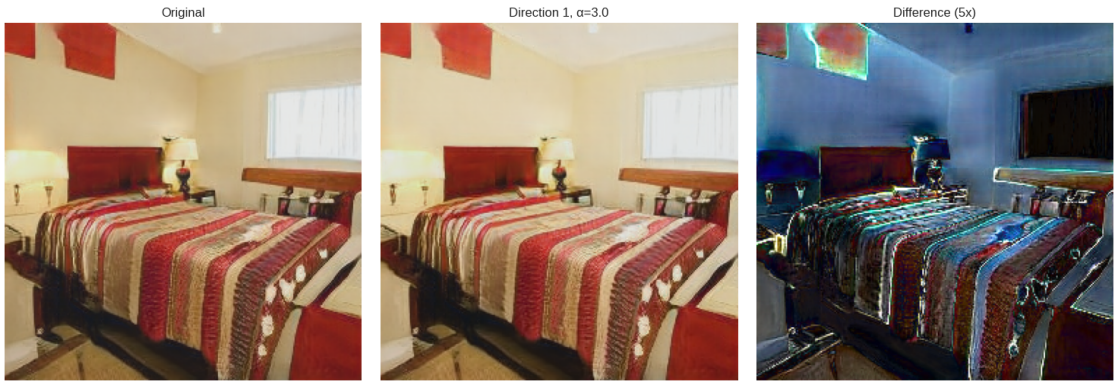

Figure 13: Example relighting for Implementation 2: original, relit image for a moderate α, and amplified difference. Changes concentrate on illumination and shading while preserving geometry and material appearance.

MiDaS-based normals and neural intrinsic decomposition provide more reliable geometry and albedo/shading separation; spherical harmonics and the Cook–Torrance BRDF yield a richer physics target; and the multi-scale, light-conditioned basis turns latent offsets into a usable relighting operator. Unlike the prototype, this system produces strong, scene-consistent relighting with stable training and serves as the default configuration for the subsequent results and analysis.

6 Comparison with StyLitGAN

Figure 14 contrasts published StyLitGAN relighting examples [2] with the physics-guided system developed here. In the StyLitGAN grids (top rows), the four relit views mainly adjust global exposure and colour balance around the bed and lamp; shadows and wall gradients change only mildly, and the same direction often produces different effects across images. In the proposed results (bottom rows), each latent direction produces a clearer, geometry-aware lighting pattern: window light, bedside lamps and wall regions brighten or darken coherently, and the same direction behaves consistently across multiple bedrooms. Even at high intensity (e.g. α=10 in the dramatic grid), bed textures, wall patterns and furniture remain stable while illumination and colour temperature vary strongly.

(a) StyLitGAN relighting grid [2].

(b) Proposed system: dramatic relighting effects (α=10).

Figure 14: Qualitative comparison between StyLitGAN and the physics-guided StyLitGAN-style system. Left: example from the original StyLitGAN paper [2]. Right: corresponding dramatic relighting grid produced by the proposed method on LSUN Bedrooms.

Quantitatively, the final system was evaluated on a held-out set of 200 LSUN bedroom seeds. Relative to its physics-based targets, it achieves mean LPIPS 0.028, SSIM 0.979 and PSNR 24.8 dB, with an average albedo MSE of 7.2×10^{-3}. These values indicate that strong lighting edits are obtained while preserving structure and material appearance. StyLitGAN reports similarly high-fidelity reconstructions on LSUN bedrooms [2], but with more conservative visual changes; direct numerical comparison is not possible because the evaluation protocols differ, yet the side-by-side grids suggest that the proposed system trades a comparable level of distortion for substantially larger and more directional illumination variation.

7 Future Work

The current system still depends heavily on external predictors for depth and intrinsic images, and on a frozen StyleGAN2 backbone trained on a single indoor dataset. A natural extension is to

train geometry, intrinsics and latent directions jointly, so that the generator learns a representation explicitly aligned with the physics model instead of adapting post hoc to MiDaS and a separate decomposer. Scaling beyond LSUN Bedrooms to multi-category datasets and real photographs would test how well the approach handles complex materials and clutter, and would likely require stronger domain adaptation and robustness to out-of-distribution depth and albedo estimates. Another direction is to expose lighting in more user-friendly ways—for example, by mapping latent directions to interpretable sliders for "window strength", "lamp intensity" or "colour temperature", or by conditioning on explicit HDR environment maps. Finally, extending the method to video, with temporal consistency losses on shading and geometry, would be important for AR/VR and simulation workloads where dynamic relighting of moving cameras and deforming objects is required.

8 Conclusion

This project investigated whether approximate geometry and a physically motivated image formation model can regularise latent-space relighting in a StyleGAN2 generator. An initial prototype combined Sobel-based normals, log-domain optimisation for intrinsic decomposition, and a Lambertian renderer, showing that physics-guided supervision is feasible but also revealing severe limitations: noisy normals, expensive per-image optimisation and weak latent edits that barely changed illumination. The final implementation replaced each fragile component with a stronger one i.e MiDaS/DPT depth and perspective-correct normals, a neural intrinsic decomposer, spherical-harmonics environment lighting and a Cook–Torrance microfacet BRDF, together with a multi-scale, light-conditioned latent basis. On LSUN Bedrooms this system produced stronger, more coherent relighting than both the StyLitGAN-style baseline and the prototype, with lower physics error, better albedo preservation and clearer control over edit strength. Overall, the results indicate that coupling GAN latents to even approximate geometry and physics is a promising route toward more controllable and physically plausible generative relighting.

References

[1] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. CVPR*, 2020.

[2] A. Bhattad, J. Soole, and D. A. Forsyth. StyLitGAN: Image-based relighting via latent control. In *Proc. CVPR*, 2024.

[3] E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris. GANSpace: Discovering interpretable GAN controls. In *Proc. NeurIPS*, 2020.

[4] Y. Shen and B. Zhou. Closed-form factorization of latent semantics in GANs. In *Proc. CVPR*, 2021.

[5] Y. Shen, C. Yang, X. Tang, and B. Zhou. Interpreting the latent space of GANs for semantic face editing. In *Proc. CVPR*, 2020.

[6] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.

[7] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics*, 33(4):159, 2014.

[8] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1670–1687, 2015.

[9] W.-C. Ma, H. Yang, and S. Soatto. Single image intrinsic decomposition without a ground truth dataset. In *Proc. ECCV*, 2018.

[10] A. S. Baslamisli, T. T. Groenestege, P. Karaoglu, and T. Gevers. IntrinsicNet: Learning intrinsic image decomposition using CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2402–2415, 2020.

[11] B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[12] A. S. Glassner, editor. *An Introduction to Ray Tracing*. Academic Press, 1989.

[13] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. In *Proc. SIGGRAPH*, pages 307–316, 1981.

[14] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance. Microfacet models for refraction through rough surfaces. In *Proc. Eurographics Symposium on Rendering*, 2007.

[15] C. Schlick. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum*, 13(3):233–246, 1994.

[16] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proc. SIGGRAPH*, pages 497–500, 2001.

[17] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1623–1637, 2022.

[18] R. Ranftl, A. Bochkovskiy, and V. Koltun. Vision transformers for dense prediction. In *Proc. ICCV*, 2021.

[19] A. Tewari, M. Elgharib, M. Zollhöfer, H.-P. Seidel, C. Theobalt, and F. Pighin. StyleRig: Rigging StyleGAN for 3D control over portrait images. In *Proc. CVPR*, 2020.

[20] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proc. ECCV*, 2016.

[21] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. CVPR*, 2018.

[22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proc. ICLR*, 2015.

[23] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *Proc. ICLR*, 2019.

[24] I. Sobel. "An isotropic 3×3 image gradient operator." Stanford Artificial Intelligence Project (SAIL) Technical Report, 1968.