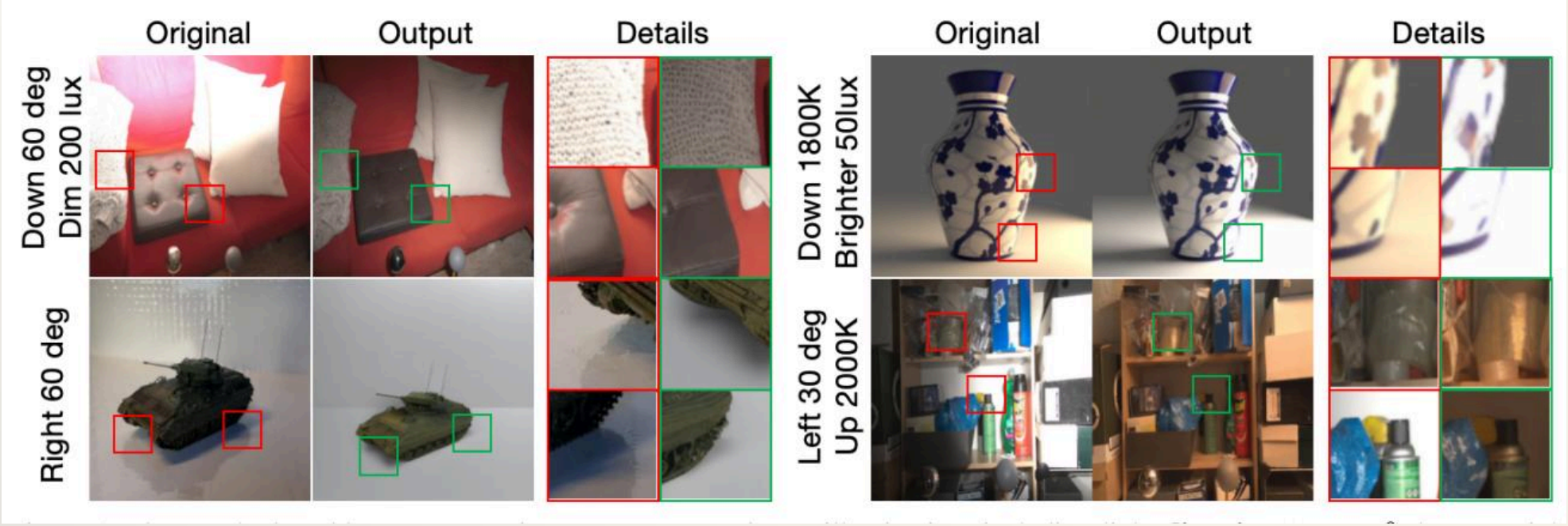
Learning Latent Proxies for Controllable Single-Image Relighting

Haoze Zheng, Zihao Wang, Xianfeng Wu, Yajing Bai, Yexin Liu, Yun Li, Xiaogang Xu, Harry Yang

HKUST, HKPOLYU, CUHK

The Challenge of Single-Image Relighting

- Relighting from a single image is fundamentally ill-posed
- Small lighting changes cause nonlinear variations in shading, shadows, and specularities
- Geometry and materials remain completely unobserved



Limitations of Existing Approaches

Intrinsic methods rely on dense, fragile PBR supervision like G-buffers

Latent diffusion methods lack physical grounding and precise control

Heavy supervision limits scalability and generalizability

Key Insight: Sparse Physical Cues

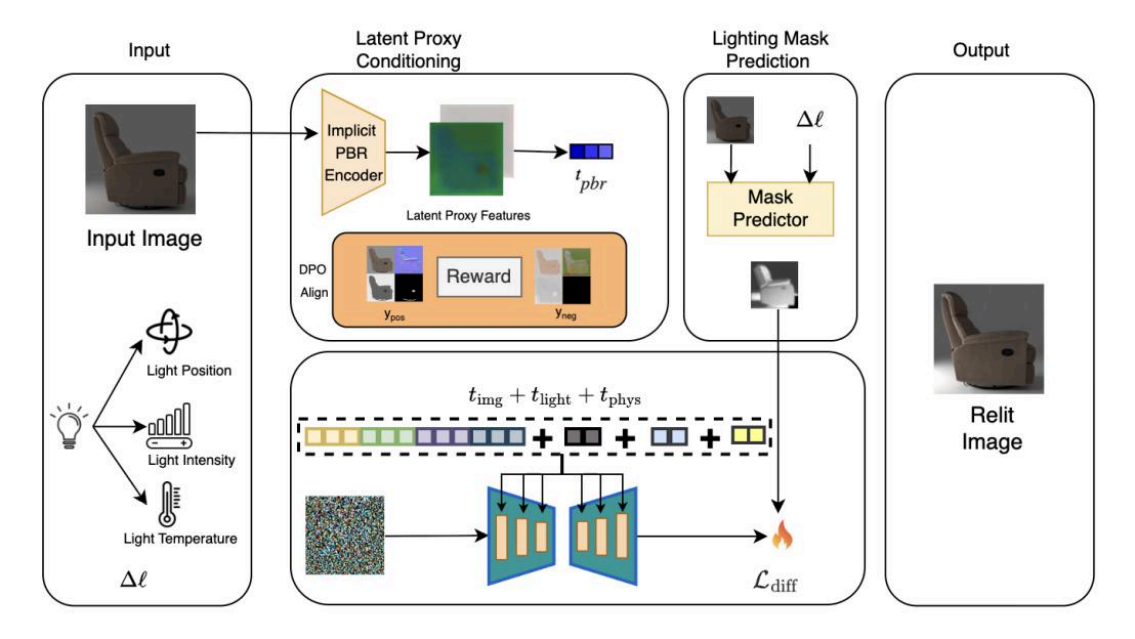
A full intrinsic decomposition is unnecessary for accurate relighting

Sparse, physically meaningful cues are sufficient to guide a diffusion model

Introduces LightCtrl for fine-grained, physically consistent control

LightCtrl Framework Overview

- Unifies scalable generative modeling with lightweight physical guidance
- Injects minimal physical cues via a few-shot latent proxy
- Employs a lighting-aware mask for spatial selectivity



Formulating the Relighting Problem

STEP 1 **Input**

Provided with a source image and target illumination parameters

STEP 2 **Relative Lighting**

Compute the geometric and photometric differences between light sources

STEP 3 **Synthesis**

Condition the diffusion model to produce the target image

$$\hat{x}_s^{\ell_t} = f_\theta(x_s^{\ell_s}, \Delta\ell)$$

Few-Shot Latent Proxy Conditioning

Compact Representation

Predicts compact material and geometry cues instead of full intrinsics

Physical Properties

Extracts albedo, normals, roughness, and metallicity into a latent map

Sparse Supervision

Learned in a few-shot manner with limited PBR labels

Integration

Projected into a single token to constrain the denoising trajectory

Latent Proxy Optimization

—Captures natural smoothness of albedo and roughness

—Enforces unit-normal consistency and quasi-binary metallic behavior

$$\mathcal{L}_{\text{proxy}} = \lambda_a \|a - \hat{a}\|_1 + \lambda_n (1 - \langle n, \hat{n} \rangle) + \lambda_r \|r - \hat{r}\|_1 + \lambda_m \, \text{BCE}(m, \hat{m})$$

Lighting-Aware Spatial Masking

STEP 1 **Motivation**

Illumination changes only affect subsets of pixels like shadows and specularities

STEP 2 **Ground Truth**

Derived from radiometric differences in linear luminance

STEP 3 **Prediction**

Inferred from source appearance and relative lighting encoding

STEP 4 **Application**

Modulates the noise reconstruction loss to emphasize lighting-sensitive areas

$$M_\theta = f_\theta(x_s^{\ell_s}, \Delta\ell)$$

DPO-Based Post-Training

STEP 1 **Issue**

Sparse supervision can lead to inconsistent relighting or over-smoothed outputs

STEP 2 **Reward Formulation**

Positive samples from ground-truth PBR; negative from model predictions

STEP 3 **Update Rule**

Increases likelihood of higher-reward predictions without modifying the base model

$$\Delta r = r(y_{\text{pos}}) - r(y_{\text{neg}})$$

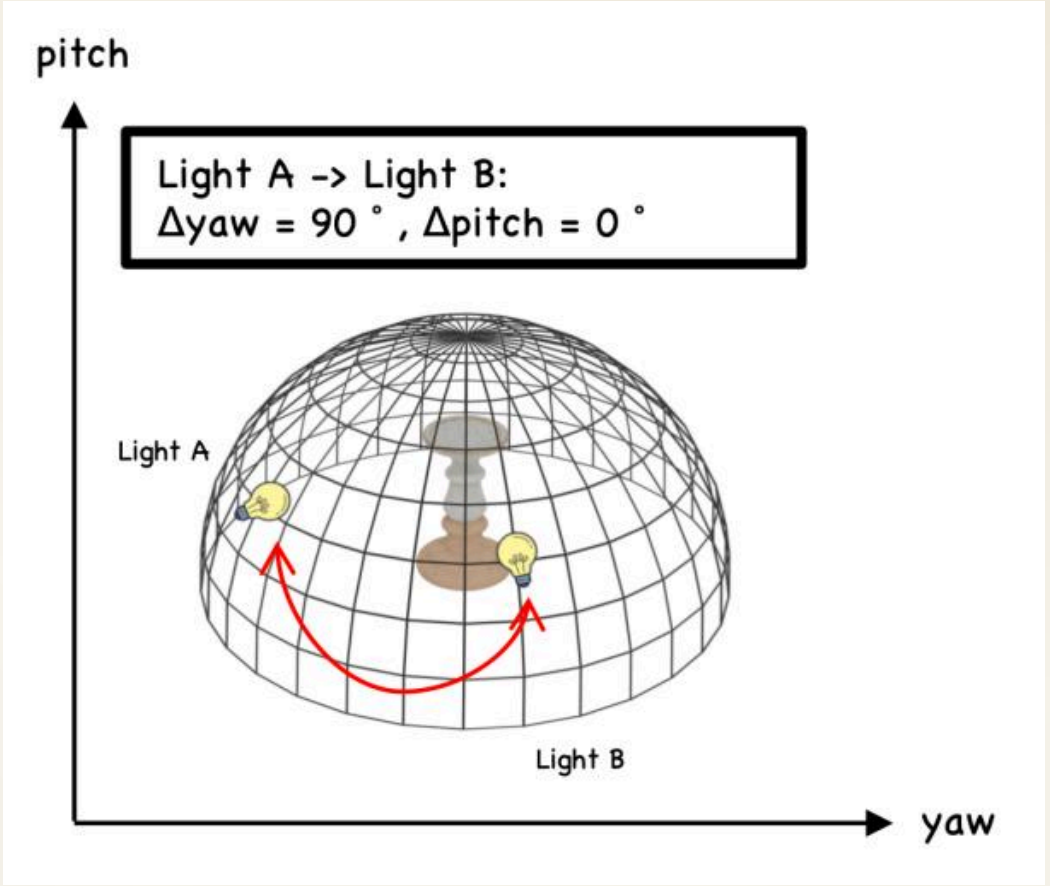
ScaLight Dataset Comparison

- Large-scale synthetic dataset containing 300K+ objects
- Systematically varied illumination with complete camera-light metadata

DATASET	OBJECTS	LIGHTING CONTROL	POSE / LIGHT
OpenIllumi	64	Limited	None
OWL	8	Limited	Limited
Rene	20	Limited	None
LightProp	80K+	Full	None
ScaLight (Ours)	300K+	Full	Full

Relative Illumination Encoding

- User edits are specified in intuitive 2D yaw-pitch space
- Mapped to unit direction vectors and projected to Spherical Harmonics
- Ensures smooth, rotation-aware continuous embeddings



Scene-Level Relighting Results (MIIW)

- Model generalizes from object-centric training to full complex scenes
- Consistently outperforms diffusion baselines on indoor benchmarks

METHOD	RMSE (LOWER)	SSIM (HIGHER)	PSNR (HIGHER)
RGB↔X	0.389	0.425	8.25
IC-Light	0.413	0.337	7.94
LumiNet	0.139	0.904	17.20
Ours	0.167	0.655	18.30

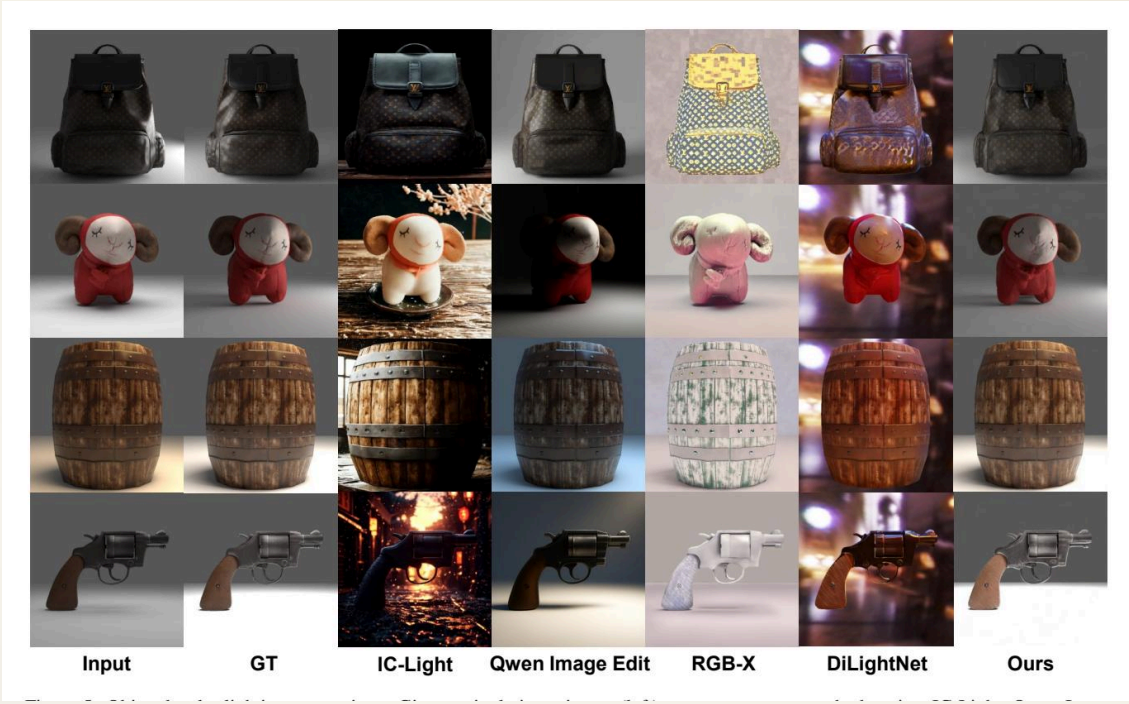
Object-Level Quantitative Evaluation

- Evaluated across color temperature, position, and energy variations
- Achieves state-of-the-art PSNR and robust RMSE scores

METHOD	TEMP RMSE	TEMP PSNR	POS RMSE	POS PSNR	ENERGY RMSE	ENERGY PSNR
IC-Light	0.397	8.21	0.375	8.65	0.380	8.63
DreamLight	0.349	9.40	0.221	13.3	0.221	13.5
x2rgb	0.472	7.81	0.487	8.32	0.484	8.89
LBM	0.064	27.8	0.084	23.1	0.073	25.3
Luminet	0.172	15.8	0.146	17.8	0.164	16.2
Ours (full)	0.053	30.2	0.074	25.6	0.083	27.1

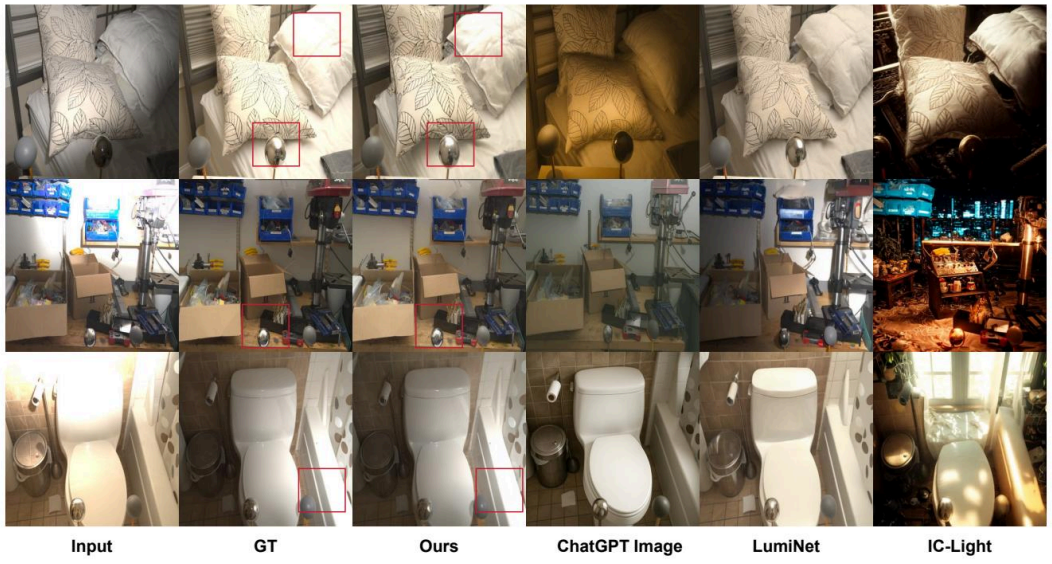
Qualitative Object-Level Relighting

- Preserves shading discontinuities and specular structure
- Maintains consistent albedo without unwanted color shifts
- Outperforms intrinsic and latent-only diffusion models



Generalizing to Real Indoor Scenes

- Produces consistent adjustments to global scene structure
- Shadows and highlights adapt coherently across surfaces
- Avoids shadow discontinuities common in other baselines



User Preference Study

- Evaluated by 35 participants on perceptual realism
- Achieved highest preference in both controlled object and complex scene settings

METHOD	SCENE-LEVEL PREFERENCE	OBJECT-LEVEL PREFERENCE
LumiNet	12.83%	4.3%
IC-Light	9.37%	11.45%
NanoBanana	22.07%	2.8%
Ours	55.73%	81.45%

Ablation Study: Component Impact

Without Proxy

Degrades performance notably under color and intensity shifts

Without Mask

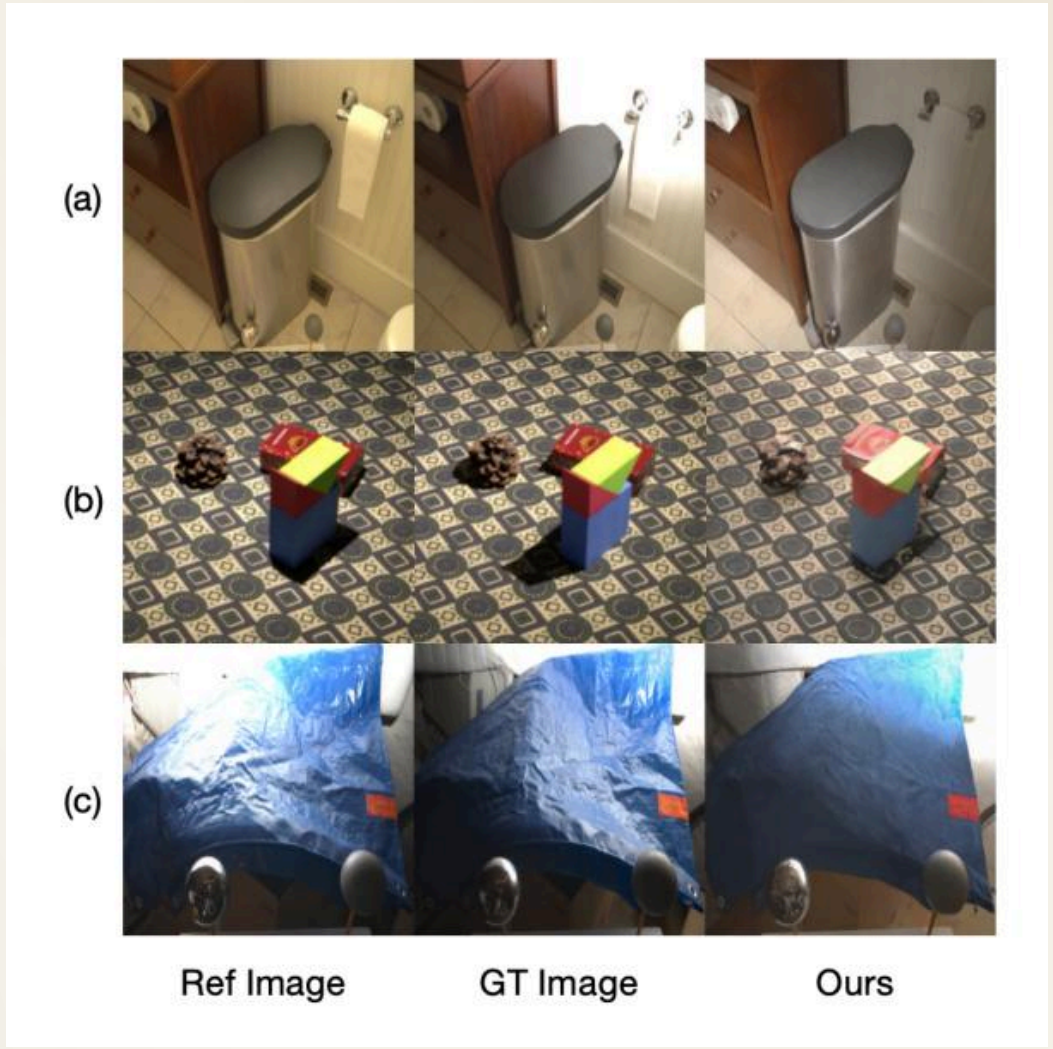
Over-edits appearance-invariant regions, blurring shadows

Without DPO

Causes largest performance drop, producing noisy and distorted albedo

Limitations and Future Work

- Struggles with sharp long-range global cast shadows in complex scenes
- May over-smooth fine structural details in highlight-heavy specular regions
- Future work will focus on richer geometry reasoning and shadow diagnostics



Conclusion

Thank You