

000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 046 047 048 049 050 051 052 053 TRANSLIGHT: IMAGE-GUIDED CUSTOMIZED LIGHTING CONTROL WITH GENERATIVE DECOUPLING

Anonymous authors

Paper under double-blind review

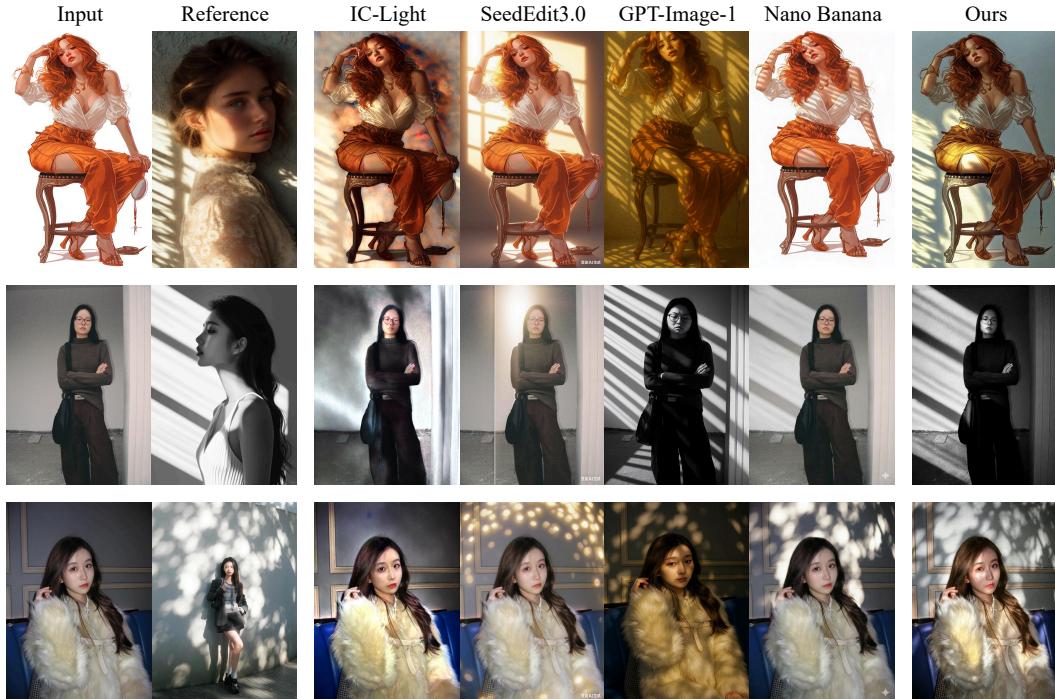


Figure 1: **Transfer light effects from reference to input image.** Our TransLight is a novel method that high-fidelity transfers light effects with obvious geometric structure from reference images to input images. The background conditioned model of IC-Light (Zhang et al., 2025) exhibits an overall darker tone and may introduce content from the reference image into the resulting images. SeedEdit3.0 (Wang et al., 2025) can not perceive the type, direction and structure of light effects in the reference image. GPT-Image-1 (OpenAI, 2025) partially understands the light effects in the reference image, but alters the human ID. The interaction between the character and the light effect in the transferred result of Nano Banana (Google DeepMind, 2025) is not natural enough.

ABSTRACT

Most existing illumination-editing approaches fail to simultaneously provide customized control of light effects and preserve content integrity. This makes them less effective for practical lighting stylization requirements, especially in the challenging task of transferring complex light effects from a reference image to a user-specified target image in [portrait photography scenes](#). To address this problem, we propose **TransLight**, a novel framework that enables high-fidelity and high-freedom transfer of light effects. Extracting the light effect from the reference image is the most critical and challenging step in our method. The difficulty lies in the complex geometric structure features embedded in light effects that are highly coupled with content in real-world scenarios. To achieve this, we first present Generative Decoupling, where two fine-tuned diffusion models are used to accurately separate image content and light effects, generating a newly

054 curated, million-scale dataset of image–content–light triplets. Then, we employ
 055 IC-Light as the generative model and train our model with our triplets, inject-
 056 ing the reference lighting image as an additional conditioning signal. The re-
 057 sulting TransLight model enables customized and natural transfer of diverse light
 058 effects. Notably, by thoroughly disentangling light effects from reference images,
 059 our generative decoupling strategy endows TransLight with highly flexible illumi-
 060 nation control. Experimental results establish TransLight to successfully transfer
 061 light effects with geometric structure across disparate images, **especially in por-**
 062 **trait photography**, delivering more customized illumination control than existing
 063 techniques and charting new directions in illumination harmonization and editing.

1 INTRODUCTION

067 Recent rapid iteration of image illumination editing methods (Zhang et al., 2021; Dastjerdi et al.,
 068 2023; Choi et al., 2024; Fortier-Chouinard et al., 2024; Kim et al., 2024; Ren et al., 2024; Phongth-
 069 awei et al., 2024; Zeng et al., 2024a; Jin et al., 2024; Zhang et al., 2025; Magar et al., 2025;
 070 Chaturvedi et al., 2025) has led to significant advancements in manipulating lighting attributes
 071 within the field of image editing. Relighting of human portraits, as a specific branch in this area,
 072 holds broad application prospects and high practical value. Existing human portraits relighting ap-
 073 proaches can generally be divided into two categories: visual-guided and text-guided illumination
 074 adjustment. Visual-guided methods (Kim et al., 2024; Ren et al., 2024; Chaturvedi et al., 2025) of-
 075 ten employ high dynamic range (HDR) maps or additional background images to composite human
 076 portraits into new lighting environments. However, this typically leads to a loss of content integrity
 077 in the original image. Text-guided methods (Zhang et al., 2025; Cha et al., 2025) accept textual
 078 descriptions of the desired light effects as input to perform image relighting. It is evident that tex-
 079 tual descriptions alone are insufficient to achieve customized control over lighting properties such
 080 as direction, position and geometric structure. Although these methods can achieve striking visual
 081 effects, the application scenarios for relighting human portraits through background replacement or
 082 text prompts remain limited. In real-world image editing scenarios, users may prefer personalized
 083 lighting modifications (e.g., adding or removing specific light effects) while preserving the original
 084 content of the image. For example, a user might request: *“Add a rainbow lens flare similar to the*
 085 *reference image to my outdoor travel photo.”* Existing relighting approaches often struggle to fulfill
 such customized lighting stylization requirements.

086 In this work, we focus on the highly challenging task of transferring light effects from a reference
 087 image to a target image **in portrait photography scenes**. To achieve this, we first need to address
 088 two main challenges: (1) *decoupling light effects from content to prevent content leakage* and (2)
 089 *constructing a suitable data structure for light effect transfer training*. Although this task is broadly
 090 similar to well-established style transfer, both aiming to replicate specific attributes from one refer-
 091 ence image onto another image, it requires additional consideration of geometric information such
 092 as the direction, structure, and position of the light effects. This leads to a stronger coupling between
 093 the light effects and the content, making many existing strategies (Wang et al., 2024a;b; Xing et al.,
 094 2024; Jeong et al., 2024) designed to mitigate content leakage in style transfer task less applicable
 095 to our case. Aiming at the challenges mentioned above, we propose **Generative Decoupling** as the
 096 solution. Specifically, we fine-tune two diffusion models using natural images without light effects
 097 and light material images, enabling them to remove and extract light effects from input images. After
 098 sufficient training, our light extraction model is capable of effectively isolating the most prominent
 099 light effects without incorporating any other content. This largely avoids the issue of content leakage
 100 in most cases. Moreover, we construct over one million image-content-light triplets for light effect
 transfer task training using our generative decoupling strategy.

101 After that, we propose TransLight, a **novel** method capable of transferring light effects with obvious
 102 geometric structure from a reference image to another image. **It is worth noting that we focus on**
 103 **visually perceptible light effects with distinct geometric structures, such as light beam in Tyndall**
 104 **Effect, iridescent halo, and lens flare, among others.** We employ IC-Light (Zhang et al., 2025) as
 105 the fixed generative network and train our model using the image-content-light triplets. During
 106 inference, the target light effect is decoupled from the reference image, effectively mitigating content
 107 leakage and enabling greater transfer flexibility. Users can freely adjust the transferred result by
 modifying the position and orientation of the light effect image. Experimental results demonstrate

UJpe: W1, Q1

108 that our TransLight possesses a powerful capability for customized light effects transfer. It can retain
 109 the background information of the source images while naturally incorporating target light effects
 110 from reference images as shown in Figure 1. Moreover, our method achieves a Light-FID score of
 111 6.02, outperforming the SOTA illumination editing method IC-Light, which obtains 10.05.
 112

113 The main contribution can be summarized as follows:

- 114 • We propose **Generative Decoupling** to decouple content and light in natural images using
 115 two diffusion models. Among them, light removal model eliminates prominent light effects
 116 from natural images, *e.g.*, lens flares, Tyndall effects, *etc*. Light extraction model is the first
 117 capable of separating light effects from images without retaining any original content.
- 118 • We construct an image-content-light triplet generation pipeline based on our generative
 119 decoupling strategy. We first select images with prominent light effects from a massive
 120 database, and then use our light removal and light extraction models to generate the corre-
 121 sponding content and light effects images, respectively. After filtering step, we obtain over
 122 one million training samples.
- 123 • We introduce **TransLight**, a *novel* image-guided method that allows accurate transfer of
 124 light effects with obvious geometric structure from a reference image to a target image
 125 *in portrait photography scenes*. Our method supports translation, flipping, and numeri-
 126 cal scaling of the extracted light effects, enabling adjustments to the position, orientation,
 127 and intensity of the light effects in transferred result. Experimental results show that our
 128 TransLight achieves visually striking light effects across a wide range of scenarios.

130 2 RELATED WORK

132 2.1 IMAGE STYLE TRANSFER

134 Image style transfer aims to transfer the style of a reference image to the target content image (Chung
 135 et al., 2024; Gao et al., 2024; Wang et al., 2023b; Zhang et al., 2023b; Frenkel et al., 2024; Wang
 136 et al., 2024b; Qi et al., 2024; Lei et al., 2025; Xing et al., 2024). This process is fundamentally
 137 similar to light effects transfer, which aims to replicate the intrinsic attributes of a reference image
 138 onto another image, while preserving the content consistency of the target image. In this task, a
 139 well-studied problem is how to disentangle the content and style of the reference image to avoid
 140 content leakage. StyelAligned (Hertz et al., 2024) proposes replacing full attention with shared
 141 attention during style injection to reduce content leakage. FreeTuner (Xu et al., 2024) employs a
 142 disentanglement strategy that separates the generation process into two stages to effectively mitigate
 143 concept entanglement. InstantStyle (Wang et al., 2024a) and Swapping self-attention (Jeong et al.,
 144 2024) minimize the introduction of content information by injecting style features into specific at-
 145 tention layers of the generation network. The aforementioned methods are specifically designed to
 146 disentangle abstract stylistic attributes from images and have proven effective in mitigating content
 147 leakage. However, they lack the capability to precisely perceive the geometric and structural char-
 148 acteristics of light effects, making it difficult to extract light. As a result, these approaches are not
 149 well-suited for the task of light effects transfer.

150 2.2 LIGHT REMOVAL

152 Light removal aims to eliminate unwanted illumination from images while meticulously preserving
 153 the underlying scene content. The most representative application is the removal of flares from im-
 154 ages (Dai et al., 2023; Wang et al., 2023a; Jin et al., 2022; Qiao et al., 2021). (Wu et al., 2021)
 155 collected 2K real reflective flare data using a camera to train the first general lens flare removal
 156 network. After that, (Dai et al., 2022) further constructed 7K synthetic flare data to reduce the domain
 157 gap with real-world scenarios. (Qiao et al., 2021) propose the first unpaired flare removal dataset
 158 and present a deep framework with light source aware guidance for single-image flare removal. Re-
 159 cently, Illuminet (Xu et al., 2025) proposed a two-stage pixel-to-pixel generative model based on the
 160 U-Net (Ronneberger et al., 2015a) model, that achieved both image flare removal and image light
 161 enhancement. In addition, LuminaBrush (Team, 2024) convert images to “uniformly-lit” appear-
 162 ances by fine-tuning the DiT (Peebles & Xie, 2023) model through LoRA (Hu et al., 2022). While
 163 these methods work well in specific situations, their limited training dataset makes it hard for them

162 to perform satisfactorily in complex scenes, which means they can not adequately support the our
 163 decoupling requirements.
 164

165 **2.3 LIGHTING CONTROL WITH DIFFUSION MODELS**
 166

167 Diffusion models have recently found widespread application in highly challenging tasks that de-
 168 mand precise control over image illumination, including portrait relighting and indoor lighting
 169 adjustment. For portrait relighting, methods such as DiFaReli (Ponglertnapakorn et al., 2023),
 170 Lite2Relight (Rao et al., 2024), and Holo-Relighting (Mei et al., 2024) specifically target the re-
 171 lighting of human faces. Relightful Harmonization (Ren et al., 2024) and Total Relighting (Pandey
 172 et al., 2021) manipulate the illumination of the foreground portrait using background conditions.
 173 SwitchLight (Kim et al., 2024) controls portrait illumination by using high dynamic range (HDR)
 174 images as conditions. Text2Relight (Cha et al., 2025) generates relighted portrait using text prompt
 175 while keeping the original contents. Furthermore, IC-Light (Zhang et al., 2025), trained extensively
 176 on large-scale datasets using the principle of light transport consistency, achieves impressive ca-
 177 pabilities in text prompt and background controllable portrait illumination generation. For indoor
 178 lighting adjustment, common operations include introducing new light sources or regulating the
 179 illumination intensity of those already present in a scene (Zhang et al., 2024; Wang et al., 2022;
 180 Magar et al., 2025). SpotLight (Fortier-Chouinard et al., 2024) control the local lighting of an object
 181 through a coarse shadow. By training a ControlNet (Zhang et al., 2023a), ScribbleLight (Choi et al.,
 182 2024) controls image light effects based on user scribbles. Latent Intrinsics (Zhang et al., 2024) and
 183 LumiNet (Xing et al., 2025) are capable of modifying the illumination states of input images based
 184 on the light effects from a reference image. DiffusionRenderer (Liang et al., 2025) use specified
 185 environment map to accurately estimates geometry and material buffers and generates photorealistic
 186 images. In addition, in the field of video generation, Light-A-Video (Zhou et al., 2025) and Re-
 187 lightVid (Fang et al., 2025) use IC-Light to add artistic lighting and shadow effects to the video
 188 generation process, and VidCRAFT3 (Zheng et al., 2025) control the light direction using per-frame
 189 3D vector in Cartesian coordinates. While the aforementioned light control methods achieve good
 190 results in specific scenarios, their customized control over light effects is highly limited. They can-
 191 not precisely and comprehensively adjust details such as light position, and shape while preserving
 192 image content. Our TransLight aims to provide a solution.
 193

UIpe: W1, W2

194 **3 METHOD**
 195

196 In this section, we first provide a clear task definition for the light effect transfer task: *Given a user*
 197 *input image I , a reference image R and user control parameters p (e.g., translation, flip, intensity),*
 198 *finally obtain the edited result image I_L .* The most ideal generation result is that I_L has the same light
 199 effects as the reference image R , mainly manifested in the consistency of the geometric structure
 200 and the type of the light effect in the two images. Moreover, I_L should maintain complete content
 201 consistency with the input image I . The light effect transfer task we define is formally similar to the
 202 illumination condition transfer task in LumiNet (Xing et al., 2025), yet exhibits clear distinctions in
 203 many aspects. We provide a detailed discussion in the supplementary material. We systematically
 204 illustrate the overall framework for achieving the task of light effect transfer in Figure 2. This
 205 framework comprises three key modules: the model training of generative decoupling, the image-
 206 content-light triplet construction pipeline, and the training of our TransLight.
 207

UIpe: W1, W2

208 **3.1 GENERATIVE DECOUPLING**
 209

210 We use two diffusion models to decouple the image content and light effects, which we refer to as
 211 generative decoupling. We design a simple yet effective strategy to train our light removal and light
 212 extraction models. Figure 2 (a) shows our training process.

213 **Source data collection.** We first employ InternVL2.5 (Chen et al., 2024) to select 1M images with-
 214 out light effects from a massive database. Then, we collect 100K light material images, all featuring
 215 pure black backgrounds. Specifically, we generate approximately 90K light material images using
 216 FLUX.1-schnell (Labs, 2024) and collect an additional 10K real images from public datasets (Dai
 217 et al., 2022; Wu et al., 2021). The details of the prompts used in the image collection process are
 218 given in the appendix.



Figure 2: **Overall framework.** (a) We fine-tune two diffusion models initialized with the weights of IC-Light. During training, the input synthesized image I_S is the direct addition result of no light image I and light material image L . (b) This pipeline involves selecting relevant data using a vision language model, decoupling image content and lighting via light removal and extraction models, and filtering out poor generation results. (c) Our TransLight training consists of two stages: LoRA fine-tuning for content-preserving lighting editing and ControlNet training to inject light effects.

Training input condition image construction. During training, the input condition images provided to the model are those containing prominent light effects. However, directly using real-world images would make it difficult to obtain effective supervision signals. Therefore, we simulate such lighting-affected images by applying preprocessing to the images from the source data. We apply two distinct preprocessing methods to obtain I_S and I_R . The synthesis of no light image I and light material L into I_S involves no complex operations and can be expressed by $I_S = aI + bL$, where $a, b \sim U(0, 1)$. This simple and direct operation is designed to simulate real-world scenarios in which lighting effects exhibit prominent geometric structural characteristics. In addition, to simulate strong lighting contrasts such as pronounced shadows and highlights commonly observed in real-world scenes, we relight the image I to generate the I_R . We use relighting model IC-Light and image editing methods FLUX.1-Kontext-dev (Labs et al., 2025) and Qwen-Image-Edit (Wu et al., 2025) as the relighting methods, and use diverse prompts to obtain I_R .

Training strategy. We fine-tune two diffusion models to obtain the light removal and light extraction models. We initialize the U-net (Ronneberger et al., 2015b) architecture using the weights from IC-Light (Zhang et al., 2025). The training input of the U-net (Ronneberger et al., 2015b) architecture is the 8-channels concatenation representation of noisy latent and input condition image latent. For light removal model training, the input condition image consists of two parts: synthesized image I_S and relighted image I_R . The prediction target is image I . For light extraction model training, the input image is I_S and the prediction target is light effect image L . Experimental results demonstrate that even with simple synthesis operations, our fully trained light extraction model achieves good performance in real-world scenarios.

270
271

3.2 IMAGE-CONTENT-LIGHT TRIPLET CONSTRUCTION PIPELINE

272
273

To make light effects transfer possible, we construct a large volume of image-content-light triplets. As shown in Figure 2 (b), our triplet construction pipeline contains three steps.

274
275
276

Selection. Initially, we select target data from a massive proprietary database. We employ InternVL2.5 for this large-scale batch selection. We use highly stringent prompts to instruct the vision language model to determine if images contain strong light effects.

277

Generation. We use light removal model and light extraction model to decompose an image with strong light effects into an image containing only content and an image containing only light effects. For the light removal process, we use InternVL2.5 to evaluate the generated results. If the results do not meet requirements, we perform a second round of light removal. In the light extraction process, we also use InternVL2.5 to generate a description of the image’s light effects. This description, when combined with the image as input to the light extraction model, notably enhances the model’s ability to separate light.

284

Filtering. We perform similarity-based filtering on the generated results to improve training data quality, discarding samples that exhibit notably poor generation outcomes. Specifically, after decoupling the input image I_L into a no light image I and a light effect image L , we obtain I_S using the simple synthesis strategy. Then, we extract features from I_S , I_L , and I separately using the DINOv2 (Oquab et al., 2023), and compute the cosine similarity of these features. The specific calculation method is as follows:

290

$$\cos(\alpha) = \text{Sim}_{\text{cos}}(\mathcal{D}(I_L), \mathcal{D}(I)), \quad (1)$$

291

$$\cos(\beta) = \text{Sim}_{\text{cos}}(\mathcal{D}(I_L), \mathcal{D}(I_S)), \quad (2)$$

293

where \mathcal{D} stands for DINOv2 and Sim_{cos} stands for cosine similarity. During our triplet construction, we aim for a significant degree of light effect removal, which necessitates that the similarity between I and I_L falls below a threshold γ . Concurrently, we desire a salient and clean separated light effect, thus expecting an increased similarity between I_S and I_L . This means that the feature similarity among these images should satisfy the following expression:

298

$$(\cos(\beta) > \cos(\alpha)) \wedge (\cos(\alpha) < \gamma), \quad (3)$$

299

where γ is set to 0.98. We set $\gamma = 0.98$ for the trade off between data quality and data quantity. The larger the threshold γ is, the more data remains after filtering, but the data quality varies. Conversely, the smaller the threshold γ is, the less data remains after filtering, but the data quality is higher.

67f2: W2

300

3.3 TRANSLIGHT

301

302

Figure 2 (c) shows the training process of our TransLight. We adopt a two-stage training strategy because the training objectives of the two modules differ. The lora fine-tuning in the first stage is to enable the model to keep the background unchanged. The second stage of training is aimed at providing complete light effect images and enhancing the degree of conditional injection.

67f2: W1;
Y35F: Q1

303

Training stage 1. First, we use the SD1.5 (Rombach et al., 2022) version of the IC-Light (Zhang et al., 2025) as our generation network. This is done to fully leverage its powerful illumination editing capabilities, which were acquired through training on large-scale datasets. We freeze the parameters of the generation network during training. We first perform LoRA (Hu et al., 2022) fine-tuning on the generation network, with the goal of enabling the model to preserve the background unchanged during the generation process and enhance its content consistency. The input during this stage is a simple addition of the content image and the background light effects image. We concatenate the latent-space features of this single image with the latent noise along the channel dimension and feed the result into the first layer of the U-Net, as in IC-Light. This stage aims to enable the model to generate realistic images from composite inputs based on content and light effects image. Additionally, we use only the background portion of the light effect to avoid affecting the foreground person. Specifically, we use background removal model BiRefNet (Zheng et al., 2024) to obtain the foreground mask M for the content image. The background light effects image L_M is equivalent to $L \times (1 - M)$.

SUH4: Q1

323

Training stage 2. In the stage 2, we inherit and fix the LoRA weights trained in the stage 1, and train an additional ControlNet (Zhang et al., 2023a) to inject the light effects image into the generation

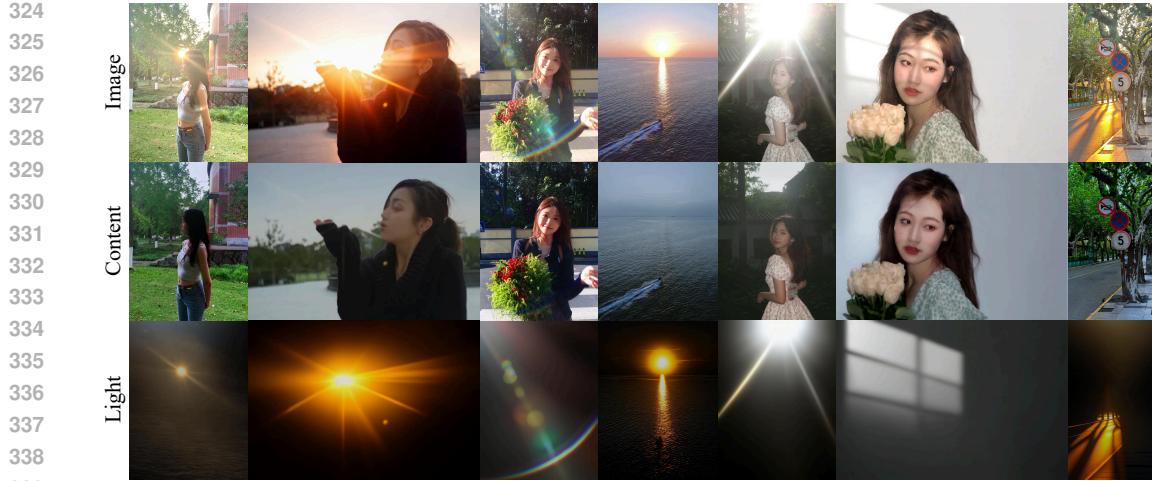


Figure 3: **Visualizations of our generative decoupling.** Our light removal model can eliminate light effects from images while preserving the underlying content unchanged as shown in the second line. Additionally, our light extraction model is capable of isolating the light effects from the image without introducing other extraneous objects.

process. To ensure the effectiveness of the ControlNet training, we only use the content image with the light effect removed as the additional channel input for the U-net structure during this stage. The input image for ControlNet is the extracted light effect image.

67f2: Q3

Inference stage. [The complete inference pipeline requires three components: the light extraction model, LoRA weights, and ControlNet weights.](#) The light extraction model first extracts light effects from a reference image, which are then fed into the ControlNet. After that, the background light effect is directly added to the content image, providing a more direct and effective condition control. After encoding into latent space via the VAE encoder (Kingma et al., 2013), the result is concatenated channel-wise with Gaussian noise and fed into the generator, yielding the final output through multi-step denoising. Notably, our TransLight allows for flexible adjustment of the following light effect attributes:

- **Position.** Apply horizontal or vertical translation to the light effect image.
- **Direction.** Perform flipping or rotating operations on the light effect image.
- **Intensity.** Modify the multiplicative coefficient of the light effect image overlaid on the content image.

4 EXPERIMENTS

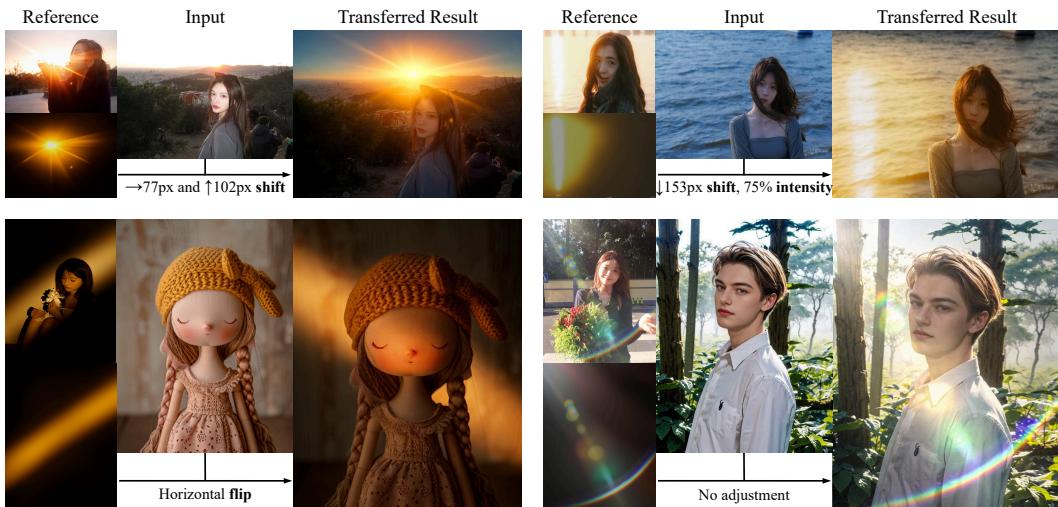
4.1 TRAINING DETAILS

Generative Decoupling. We fine-tune the SD1.5 (Rombach et al., 2022) version of IC-Light (Zhang et al., 2025) model. For both light removal and light extraction model training, we use the AdamW optimizer (Loshchilov & Hutter, 2017) with the learning rate of 1e-5. We train two models on 8 \times L40s GPUs with a batch size of 128. We only use MSE loss to train our models as the normal SD1.5 training. For light removal model training, the training input image consists of two parts: the synthesized image I_S and the relighting image I_R . During training, the sampling ratio of these two parts of image is 1 : 1, and the ratio of relighting images generated by IC-Light, FLUX.1-Kontext-dev and Qwen-Image-Edit is 2 : 1 : 1. We use a fixed prompt and train the light removal model for 2K iterations. For light extraction model training, the input image is the synthesized image I_S and the optimization target is to generate the light effects contained in the normal image. [To enhance the robustness of the light extraction model, when synthesizing \$I_S\$, we perform appropriate data augmentation operations on \$L\$, such as flipping, cropping, and random masking, etc.](#) We use the caption corresponding to light effect image as prompts and train the light extraction model for 26K iterations. The caption is obtained using InternVL2.5.

378
 379 **Table 1: Quantitative results.** IC-Light
 380 uses its background conditioned model, with
 381 ground truth serving as the background input.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
IC-Light	15.34	0.6784	0.2756
Ours	19.58	0.7931	0.1982

TransLight. We update the parameters of LoRA and ControlNet using the AdamW optimizer with a learning rate of 1e-5 for the two stages. Considering that our training data scale is on the order of millions, we set the rank of LoRA to 128 and train it for 10K iterations in stage 1. The training process for stage 2 requires 40K iterations. We train our TransLight on $16 \times$ L40s GPUs with a batch size of 128.



403
 404 **Figure 4: Visualizations of the specific details about the light transfer process.** Our TransLight
 405 uses the generative decoupling to break down such a difficult task. We first extract the light effect
 406 using our thoroughly fine-tuned light extraction model, as shown in the reference image below, and
 407 then composite it onto the user-specified image. Our TransLight allows flexible control over the
 408 position and direction of the light effects, thereby yielding visually authentic results with a high
 409 degree of freedom.

4.2 PERFORMANCE OF GENERATIVE DECOUPLING

410
 411 During the process of constructing triplets using generative decoupling, all the source images with
 412 obvious light effects are selected from the massive database using InternVL2.5. When quantitatively
 413 evaluating the light effect removal capability of the light removal model, we also use InternVL2.5 to
 414 judge the generated images, and use this as the basis for determining whether the removal is qual-
 415 ified. We evaluate the light effect removal performance on 500 real-world images, achieving a suc-
 416 cess rate of 85.8%. To further intuitively demonstrate the performance of our generative decoupling,
 417 we present visual results of decoupled generations in Figure 3. Our method achieves high-quality
 418 decoupled generation on images with various types of light effects across different scenes.

4.3 PERFORMANCE OF TRANSLIGHT

421
 422 In this section, we present the generated results of our method and comparisons with other ap-
 423 proaches using both subjective and objective evaluation methods. For objective comparisons, we
 424 first use metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM),
 425 and Learned Perceptual Image Patch Similarity (LPIPS) to evaluate the similarity between a single
 426 generated image and ground truth. Considering that the ground truth of the transfer result can not be
 427 obtained during real scene reasoning, we use the real image itself with as both the ground truth and
 428 the reference image when calculating the quantitative indicators. Therefore, using the background
 429 replacement relighting model in the visual-guided methods can, to a certain extent, achieve the
 430 transfer of light effects with obvious geometric structures to the content image while retaining the
 431 background content of the original image. It should be noted that, since representative background

432
 433 **Table 2: Light-FID.** “Content” in the first row
 434 refers to result of lighting-removed content image.
 435 “Content+Light” refers to the result formed by
 436 directly adding content and light effect maps. IC-
 437 Light uses its background conditioned model, with
 438 ground truth serving as the background input.

Method	Light-FID
Content	10.61
Content+Light	8.40
IC-Light	10.05
Relighted data	7.72
Unfiltered data	6.49
Only LoRA	7.94
Only ControlNet	8.44
LoRA+ControlNet (joint)	7.42
Ours	6.02

439
 440 **Table 3: User study.** For “IC-Light (bg)”, we
 441 use the reference image as the background input
 442 to the background-conditioned IC-Light. For
 443 “IC-Light (prompt)”, we use the description de-
 444 rived from the reference image. For Nano Ba-
 445 nana and Seedream 4.0, we provide both the in-
 446 put image and the reference image to the image
 447 editing model simultaneously and use appropri-
 448 ate prompt to complete the light effect transfer
 449 task.

Method	Consistency Pass Rate	
	Light	Content
IC-Light (prompt)	4.0%	12.0%
IC-Light (bg)	8.4%	8.8%
Nano Banana	39.5%	61.3%
Seedream 4.0	32.4%	31.1%
Ours	84%	73.8%

450
 451
 452 replacement relighting methods such as Total Relighting (Pandey et al., 2021) and Relightful Harmonization (Ren et al., 2024) are not publicly available, we adopt the background conditioned model of IC-Light as the comparison baseline. We select 1000 triplet samples from the large-scale dataset constructed using the generative decoupling strategy, which are not involved in the training process. Our TransLight uses the light effect images from triplets for condition generation. Regarding IC-Light, we apply its background conditioned model, using the ground truth as the background input. As shown in Table 1, our method clearly outperforms IC-Light.

453
 454
 455
 456
 457
 458
 459
 460 In addition to the above-mentioned similarity metrics computed on individual images, we believe
 461 that calculating the distributional distance between the generated images and the real images pro-
 462 vides a more objective evaluation of the overall model performance. Therefore, we sample 12,000
 463 triplets that are not used during training. We compute the Fréchet Inception Distance (FID) (Heusel
 464 et al., 2017) to evaluate the light transfer capability of the model, which we refer to as **Light-FID**.
 465 The Light-FID reflects the distance of the data distribution between the transfer result images and
 466 the real images with obvious light effects. We report the result in Table 2. When performing infer-
 467 ence using the background conditioned model of IC-Light, the ground truth is also directly used as
 468 the background input. Our method demonstrates significant advantages over IC-Light. Moreover,
 469 using only the light effects from the light extraction model and adding them to the content image
 470 yields better results than IC-Light.

471 For subjective comparison in real-world application scenarios, we show some visualizations in
 472 Figure 1. To fully highlight the advantages of our TransLight, we also show the result of the
 473 SOTA relighting method IC-Light, alongside the closed-source SOTA image editing methods
 474 SeedEdit3.0 (Wang et al., 2025), GPT-Image-1 (OpenAI, 2025) and Nano Banana (Google Deep-
 475 Mind, 2025). The reference image is used as background input for IC-Light’s background con-
 476 ditioned model. For SeedEdit3.0, GPT-Image-1 and Nano Banana, we directly provide both input
 477 image and reference image to the model simultaneously. We experiment with different prompts to
 478 instruct the model and select the best output from the results. As illustrated in Figure 1, current
 479 relighting and image editing methods are unable to fulfill our expected light effect transfer re-
 480 quirements. IC-Light incorporates the content of the reference image into the generated results, yet the
 481 overall style tends to be darker. Although SeedEdit3.0 is capable of modifying the light effects
 482 in the target image, it fails to transfer light effects according to user preferences from a reference
 483 image. The generated results of the GPT-Image-1 show noticeable changes in the character ap-
 484 pearance. The light effect transfer capability of Nano Banana most closely resembles that of our
 485 proposed approach, as it can understand the type and structure of light effects from the reference im-
 486 age. However, the visual results indicate weak interaction between the light effects and the human
 487 subjects, resulting in a perceived lack of realism. Moreover, free adjustment of light effects position
 488 and direction enables more natural integration with the target image, resulting in visually superior

486 outcomes as shown in Figure 4. Our TransLight successfully implements image-guided customized
 487 lighting control via transferring light effects from a reference image to a target image.
 488

489 **4.4 USER STUDY**
 490

491 We conduct a user study to evaluate the performance of our model on the light effect transfer
 492 task. In total, we collect voting responses from 15 participants. We asked the participants to
 493 vote respectively from the dimension of light effect consistency and content consistency, and
 494 select the results that meet the requirements in each samples. Good light effect consistency
 495 means that the light effect in the generated image closely matches that in the reference
 496 image. Good content consistency indicates that neither the subject nor the background in the
 497 generated image undergoes noticeable changes. For each sample, we present the generation
 498 results of five methods: prompt-conditioned IC-Light (Zhang et al., 2025), background-conditioned
 499 IC-Light, Nano Banana (Google DeepMind, 2025), Seedream 4.0 (Seedream et al., 2025), and our
 500 TransLight, displayed in random order. We show the result in Table. 3. Our TransLight significantly
 501 outperforms all competing methods in both light effect consistency and content consistency.
 502

UJpe: W5

502 **4.5 ABLATION STUDY**
 503

504 **Training strategy.** In this part, we conduct ablation studies on the two-stage training strategy.
 505 Specifically, we compute the Light-FID scores under three settings: (1) only LoRA fine-tuning,
 506 (2) only ControlNet training, and (3) joint two-stage training. The results are reported in
 507 Table. 2. Experimental results demonstrate that both stages contribute to improving TransLight’s
 508 performance. Moreover, due to the discrepancy in their respective training objectives, training the
 509 LoRA and ControlNet modules separately yields better results than joint optimization.
 510

67f2: W3, Q3;
Y35F: Q1

511 **Training data.** We conduct an experiment in which TransLight is trained on relighted data instead.
 512 Specifically, we employ the text-conditioned model of IC-Light to relight real images that originally
 513 contain no light effects, and then extract the corresponding light effects images from the generated
 514 results to form the triplet data. Using this approach, we also generate one million training samples.
 515 The Light-FID is reported in Table 2. This result further justify the rationality of our motivation to
 516 construct triplets for training by leveraging real-world data through generative decoupling.
 517

518 **Table 4: Ablation of filtering threshold γ .**
 519 We report the number of training samples and
 520 the Light-FID scores corresponding to different
 521 values of the threshold γ .

γ	Training samples	Light-FID
1	~1.5M	6.41
0.99	~1.3M	6.22
0.98	~1.2M	6.02
0.97	~0.9M	6.54
0.96	~0.6M	6.83
0.95	~0.5M	7.00

522 **Data filtering.** To validate the effectiveness of
 523 the third step, filtering in the receipt construc-
 524 tion pipeline, we conduct experimental verifica-
 525 tion. Specifically, we train our model using one
 526 million unfiltered triplet samples and compute the
 527 Light-FID score as shown in Table 2. Training with
 528 filtered data enables our model to perform better.
 529 Moreover, the threshold γ significantly affects both
 530 the quantity and quality of the data. We perform
 531 filtering under different values of γ and train the
 532 model using the resulting eligible images, and the
 533 experimental results are presented in Table. 4. As
 534 shown in the table, we set $\gamma = 0.98$ for the trade
 535 off between data quality and data quantity.
 536

67f2: W2, Q2;
SUh4: W2

537 **5 CONCLUSION**
 538

539 In this paper, we propose TransLight, a novel lighting-control generation method that enables the
 540 transfer of light effects from reference images to target images in portrait photography scenes. To
 541 achieve this, we introduce generative decoupling strategy, which decouples the content and light ef-
 542 ffects in real-world images by fine-tuning two diffusion models. Based on this strategy, we construct
 543 over one million high-quality triplet samples, which are then used to train our TransLight. Experi-
 544 mental results demonstrate that our method achieves impressive light effect transfer capabilities,
 545 enabling flexible customization of lighting-controlled image generation. Our TransLight provides a
 546 solid and flexible foundation for future developments in image-guided light effect transfer, and we
 547 are highly optimistic about its promising application potential in practical lighting editing tasks.
 548

540 REFERENCES
541

542 Junuk Cha, Mengwei Ren, Krishna Kumar Singh, He Zhang, Yannick Hold-Geoffroy, Seunghyun
543 Yoon, HyunJoon Jung, Jae Shin Yoon, and Seungryul Baek. Text2relight: Creative portrait re-
544 lighting with text guidance. In Proceedings of the AAAI Conference on Artificial Intelligence,
545 volume 39, pp. 1980–1988, 2025.

546 Sumit Chaturvedi, Mengwei Ren, Yannick Hold-Geoffroy, Jingyuan Liu, Julie Dorsey, and Zhixin
547 Shu. Synthlight: Portrait relighting with diffusion model by learning to re-render synthetic faces.
548 In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 369–379, 2025.

549 Zhe Chen, Jiannan Wu, Wenhui Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong
550 Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning
551 for generic visual-linguistic tasks. In Proceedings of the IEEE/CVF conference on computer
552 vision and pattern recognition, pp. 24185–24198, 2024.

553 Jun Myeong Choi, Annie Wang, Pieter Peers, Anand Bhattad, and Roni Sengupta. Scribblelight:
554 Single image indoor relighting with scribbles. arXiv preprint arXiv:2411.17696, 2024.

555 Jiwoo Chung, Sangeek Hyun, and Jae-Pil Heo. Style injection in diffusion: A training-free approach
556 for adapting large-scale diffusion models for style transfer. In Proceedings of the IEEE/CVF
557 conference on computer vision and pattern recognition, pp. 8795–8805, 2024.

558 Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A
559 phenomenological nighttime flare removal dataset. Advances in Neural Information Processing
560 Systems, 35:3926–3937, 2022.

561 Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime
562 smartphone reflective flare removal using optical center symmetry prior. In Proceedings of the
563 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20783–20791, 2023.

564 Mohammad Reza Karimi Dastjerdi, Jonathan Eisenmann, Yannick Hold-Geoffroy, and Jean-
565 François Lalonde. Everlight: Indoor-outdoor editable hdr lighting estimation. In Proceedings
566 of the IEEE/CVF international conference on computer vision, pp. 7420–7429, 2023.

567 Ye Fang, Zeyi Sun, Shangzhan Zhang, Tong Wu, Yinghao Xu, Pan Zhang, Jiaqi Wang, Gordon
568 Wetzstein, and Dahua Lin. Relightvid: Temporal-consistent diffusion model for video relighting.
569 arXiv preprint arXiv:2501.16330, 2025.

570 Frédéric Fortier-Chouinard, Zitian Zhang, Louis-Etienne Messier, Mathieu Garon, Anand Bhattad,
571 and Jean-François Lalonde. Spotlight: Shadow-guided object relighting via diffusion. arXiv
572 preprint arXiv:2411.18665, 2024.

573 Yarden Frenkel, Yael Vinker, Ariel Shamir, and Daniel Cohen-Or. Implicit style-content separation
574 using b-lora. In European Conference on Computer Vision, pp. 181–198. Springer, 2024.

575 Junyao Gao, Yanchen Liu, Yanan Sun, Yinhao Tang, Yanhong Zeng, Kai Chen, and Cairong Zhao.
576 Styleshot: A snapshot on any style. arXiv preprint arXiv:2407.01414, 2024.

577 Google DeepMind. Gemini 2.5. <https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/>, 2025.

578 Amir Hertz, Andrey Voynov, Shlomi Fruchter, and Daniel Cohen-Or. Style aligned image generation
579 via shared attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
580 Recognition, pp. 4775–4785, 2024.

581 Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter.
582 Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in
583 neural information processing systems, 30, 2017.

584 Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang,
585 Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. ICLR, 1(2):3, 2022.

594 Zhongyun Hu, Xin Huang, Yaning Li, and Qing Wang. Sa-ae for any-to-any relighting. In European
 595 Conference on Computer Vision, pp. 535–549. Springer, 2020.

596

597 Jaeseok Jeong, Junho Kim, Yunjey Choi, Gayoung Lee, and Youngjung Uh. Visual style prompting
 598 with swapping self-attention. arXiv preprint arXiv:2402.12974, 2024.

599

600 Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and
 601 Noah Snavely. Neural gaffer: Relighting any object via diffusion. In The Thirty-eighth Annual
602 Conference on Neural Information Processing Systems, 2024.

603

604 Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer
 605 decomposition meets light-effects suppression. In European Conference on Computer Vision, pp.
 606 404–421. Springer, 2022.

607

608 Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, and Sanghyun Woo. Switchlight:
 609 Co-design of physics-driven architecture and pre-training framework for human portrait relight-
 610 ing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,
 611 pp. 25096–25106, 2024.

612

613 Diederik P Kingma, Max Welling, et al. Auto-encoding variational bayes, 2013.

614

615 Black Forest Labs. Flux. <https://github.com/black-forest-labs/flux>, 2024.

616

617 Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril
 618 Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, Kyle Lacey,
 619 Yam Levi, Cheng Li, Dominik Lorenz, Jonas Müller, Dustin Podell, Robin Rombach, Harry Saini,
 620 Axel Sauer, and Luke Smith. Flux.1 kontext: Flow matching for in-context image generation and
 621 editing in latent space, 2025. URL <https://arxiv.org/abs/2506.15742>.

622

623 Mingkun Lei, Xue Song, Beier Zhu, Hao Wang, and Chi Zhang. Stylestudio: Text-driven style
 624 transfer with selective control of style elements. In Proceedings of the Computer Vision and
625 Pattern Recognition Conference, pp. 23443–23452, 2025.

626

627 Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Chih-Hao Lin, Jun Gao,
 628 Alexander Keller, Nandita Vijaykumar, Sanja Fidler, et al. Diffusion renderer: Neural inverse
 629 and forward rendering with video diffusion models. In Proceedings of the Computer Vision and
630 Pattern Recognition Conference, pp. 26069–26080, 2025.

631

632 Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. arXiv preprint
633 arXiv:1711.05101, 2017.

634

635 Nadav Magar, Amir Hertz, Eric Tabellion, Yael Pritch, Alex Rav-Acha, Ariel Shamir, and Yedid
 636 Hoshen. Lightlab: Controlling light sources in images with diffusion models. arXiv preprint
637 arXiv:2505.09608, 2025.

638

639 Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon
 640 Jung, and Vishal M Patel. Holo-relighting: Controllable volumetric portrait relighting from a
 641 single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
642 Recognition, pp. 4263–4273, 2024.

643

644 Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A multi-illumination dataset
 645 of indoor object appearance. In 2019 IEEE international conference on computer vision (ICCV),
 volume 2, 2019.

646

647 OpenAI. GPT-Image-1. <https://openai.com/index/introducing-4o-image-generation/>, 2025.

648

649 Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov,
 650 Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning
 651 robust visual features without supervision. arXiv preprint arXiv:2304.07193, 2023.

652

653 Rohit Pandey, Sergio Orts-Escalano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph
 654 Rhemann, Paul E Debevec, and Sean Ryan Fanello. Total relighting: learning to relight portraits
 655 for background replacement. ACM Trans. Graph., 40(4):43–1, 2021.

648 William Peebles and Saining Xie. Scalable diffusion models with transformers. In Proceedings of
 649 the IEEE/CVF international conference on computer vision, pp. 4195–4205, 2023.
 650

651 Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Varun Jampani, Amit
 652 Raj, Pramook Khungurn, and Supasorn Suwajanakorn. Diffusionlight: Light probes for free by
 653 painting a chrome ball. In Proceedings of the IEEE/CVF conference on computer vision and
 654 pattern recognition, pp. 98–108, 2024.

655 Puntawat Ponglernapakorn, Nontawat Tritrong, and Supasorn Suwajanakorn. Difareli: Diffusion
 656 face relighting. In Proceedings of the IEEE/CVF international conference on computer vision,
 657 pp. 22646–22657, 2023.
 658

659 Tianhao Qi, Shancheng Fang, Yanze Wu, Hongtao Xie, Jiawei Liu, Lang Chen, Qian He, and Yong-
 660 dong Zhang. Deadiff: An efficient stylization diffusion model with disentangled representations.
 661 In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.
 662 8693–8702, 2024.

663 Xiaotian Qiao, Gerhard P Hancke, and Rynson WH Lau. Light source guided single-image flare
 664 removal from unpaired data. In Proceedings of the IEEE/CVF International Conference on
 665 Computer Vision, pp. 4177–4185, 2021.
 666

667 Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun BR, Fangneng Zhan, Tim Weyrich, Bernd
 668 Bickel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, et al. Lite2relight: 3d-aware
 669 single image portrait relighting. In ACM SIGGRAPH 2024 Conference Papers, pp. 1–12, 2024.

670 Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, HyunJoon Jung, Guido
 671 Gerig, and He Zhang. Relightful harmonization: Lighting-aware portrait background replace-
 672 ment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,
 673 pp. 6452–6462, 2024.
 674

675 Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-
 676 resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF
 677 conference on computer vision and pattern recognition, pp. 10684–10695, 2022.

678 Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks
 679 for biomedical image segmentation. In Medical image computing and computer-assisted
 680 intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9,
 681 2015, proceedings, part III 18, pp. 234–241. Springer, 2015a.
 682

683 Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for
 684 biomedical image segmentation. In International Conference on Medical image computing and
 685 computer-assisted intervention, pp. 234–241. Springer, 2015b.

686 Team Seedream, Yunpeng Chen, Yu Gao, Lixue Gong, Meng Guo, Qiushan Guo, Zhiyao Guo,
 687 Xiaoxia Hou, Weilin Huang, Yixuan Huang, et al. Seedream 4.0: Toward next-generation multi-
 688 modal image generation. arXiv preprint arXiv:2509.20427, 2025.
 689

690 LuminaBrush Team. Luminabrush: Illumination drawing tools for text-to-image diffusion models,
 691 2024.

692 Guangcong Wang, Yinuo Yang, Chen Change Loy, and Ziwei Liu. Stylelight: Hdr panorama genera-
 693 tion for lighting estimation and editing. In European conference on computer vision, pp. 477–492.
 694 Springer, 2022.
 695

696 Haofan Wang, Matteo Spinelli, Qixun Wang, Xu Bai, Zekui Qin, and Anthony Chen. In-
 697 stantstyle: Free lunch towards style-preserving in text-to-image generation. arXiv preprint
 698 arXiv:2404.02733, 2024a.
 699

700 Haofan Wang, Peng Xing, Renyuan Huang, Hao Ai, Qixun Wang, and Xu Bai. Instantstyle-
 701 plus: Style transfer with content-preserving in text-to-image generation. arXiv preprint
arXiv:2407.00788, 2024b.

702 Peng Wang, Yichun Shi, Xiaochen Lian, Zhonghua Zhai, Xin Xia, Xuefeng Xiao, Weilin Huang,
 703 and Jianchao Yang. Seededit 3.0: Fast and high-quality generative image editing. [arXiv preprint](https://arxiv.org/abs/2506.05083)
 704 [arXiv:2506.05083](https://arxiv.org/abs/2506.05083), 2025.

705 Yufei Wang, Renjie Wan, Wenhan Yang, Bihan Wen, Lap-Pui Chau, and Alex C Kot. Removing
 706 image artifacts from scratched lens protectors. In [2023 IEEE International Symposium on Circuits](#)
 707 and [Systems \(ISCAS\)](#), pp. 1–5. IEEE, 2023a.

708 Zhizhong Wang, Lei Zhao, and Wei Xing. Stylediffusion: Controllable disentangled style transfer
 709 via diffusion models. In [Proceedings of the IEEE/CVF international conference on computer](#)
 710 [vision](#), pp. 7677–7689, 2023b.

711 Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng ming Yin, Shuai
 712 Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang,
 713 Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, Hang Zhang, Hao Meng, Hu Wei, Jingyuan
 714 Ni, Kai Chen, Kuan Cao, Liang Peng, Lin Qu, Minggang Wu, Peng Wang, Shuteng Yu, Tingkun
 715 Wen, Wensen Feng, Xiaoxiao Xu, Yi Wang, Yichang Zhang, Yongqiang Zhu, Yujia Wu, Yuxuan
 716 Cai, and Zenan Liu. Qwen-image technical report, 2025. URL <https://arxiv.org/abs/2508.02324>.

717 Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and
 718 Jonathan T Barron. How to train neural networks for flare removal. In [Proceedings of the](#)
 719 [IEEE/CVF International Conference on Computer Vision](#), pp. 2239–2247, 2021.

720 Peng Xing, Haofan Wang, Yanpeng Sun, Qixun Wang, Xu Bai, Hao Ai, Renyuan Huang, and
 721 Zechao Li. Csgo: Content-style composition in text-to-image generation. [arXiv preprint](https://arxiv.org/abs/2408.16766)
 722 [arXiv:2408.16766](https://arxiv.org/abs/2408.16766), 2024.

723 Xiaoyan Xing, Konrad Groh, Sezer Karaoglu, Theo Gevers, and Anand Bhattacharjee. Luminet: Latent
 724 intrinsics meets diffusion models for indoor scene relighting. In [Proceedings of the Computer](#)
 725 [Vision and Pattern Recognition Conference](#), pp. 442–452, 2025.

726 Lizhi Xu, Liqiang Zhu, Yaodong Wang, and Yao Wang. Illuminet: A two-stage model for effective
 727 flare removal and light enhancement under complex lighting conditions. [Expert Systems with](#)
 728 [Applications](#), pp. 127638, 2025.

729 Youcan Xu, Zhen Wang, Jun Xiao, Wei Liu, and Long Chen. Freetuner: Any subject in any style
 730 with training-free diffusion. [arXiv preprint arXiv:2405.14201](https://arxiv.org/abs/2405.14201), 2024.

731 An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang
 732 Gao, Chengan Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu,
 733 Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin
 734 Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang,
 735 Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui
 736 Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang
 737 Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger
 738 Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan
 739 Qiu. Qwen3 technical report. [arXiv preprint arXiv:2505.09388](https://arxiv.org/abs/2505.09388), 2025.

740 Hao-Hsiang Yang, Wei-Ting Chen, and Sy-Yen Kuo. S3net: A single stream structure for depth
 741 guided image relighting. In [Proceedings of the IEEE/CVF Conference on Computer Vision and](#)
 742 [Pattern Recognition](#), pp. 276–283, 2021.

743 Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. Dilightnet:
 744 Fine-grained lighting control for diffusion-based image generation. In [ACM SIGGRAPH 2024](#)
 745 [Conference Papers](#), pp. 1–12, 2024a.

746 Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan,
 747 Ling-Qi Yan, and Miloš Hašan. Rgb-x: Image decomposition and synthesis using material- and
 748 lighting-aware diffusion models. In [ACM SIGGRAPH 2024 Conference Papers](#), SIGGRAPH '24,
 749 New York, NY, USA, 2024b. Association for Computing Machinery. ISBN 9798400705250. doi:
 750 10.1145/3641519.3657445. URL <https://doi.org/10.1145/3641519.3657445>.

756 Lvmi Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image
 757 diffusion models. In Proceedings of the IEEE/CVF international conference on computer vision,
 758 pp. 3836–3847, 2023a.

759 Lvmi Zhang, Anyi Rao, and Maneesh Agrawala. Scaling in-the-wild training for diffusion-based
 760 illumination harmonization and editing by imposing consistent light transport. In The Thirteenth
 761 International Conference on Learning Representations, 2025.

762 Xiao Zhang, William Gao, Seemandhar Jain, Michael Maire, David Forsyth, and Anand Bhattad.
 763 Latent intrinsics emerge from training to relight. Advances in Neural Information Processing
 764 Systems, 37:96775–96796, 2024.

765 Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio
 766 Orts-Escalano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport
 767 for relighting and view synthesis. ACM Transactions on Graphics (TOG), 40(1):1–17, 2021.

768 Yuxin Zhang, Nisha Huang, Fan Tang, Haibin Huang, Chongyang Ma, Weiming Dong, and Chang-
 769 sheng Xu. Inversion-based style transfer with diffusion models. In Proceedings of the IEEE/CVF
 770 conference on computer vision and pattern recognition, pp. 10146–10156, 2023b.

771 Peng Zheng, Dehong Gao, Deng-Ping Fan, Li Liu, Jorma Laaksonen, Wanli Ouyang, and Nicu
 772 Sebe. Bilateral reference for high-resolution dichotomous image segmentation. CAAI Artificial
 773 Intelligence Research, 2024.

774 Sixiao Zheng, Zimian Peng, Yanpeng Zhou, Yi Zhu, Hang Xu, Xiangru Huang, and Yanwei Fu.
 775 Vidcraft3: Camera, object, and lighting control for image-to-video generation. arXiv preprint
 776 arXiv:2502.07531, 2025.

777 Yujie Zhou, Jiazi Bu, Pengyang Ling, Pan Zhang, Tong Wu, Qidong Huang, Jinsong Li, Xiaoyi
 778 Dong, Yuhang Zang, Yuhang Cao, et al. Light-a-video: Training-free video relighting via pro-
 779 gressive light fusion. arXiv preprint arXiv:2502.08590, 2025.

780

781 APPENDIX

782 A DATASET DETAILS

783 A.1 LIGHT MATERIAL

784 During the training of both the light removal and light extraction models, we use approximately
 785 100K light material images, of which 10K are sourced from public datasets (Dai et al., 2022; Wu
 786 et al., 2021), and the remaining 90K are generated using the FLUX.1-schnell (Labs, 2024). We em-
 787 ploy the following prompt for generation: *Generate a light material image: <random-example>,*
*788 the background color is pure black, true and natural, do not show anything other than light ef-
 789 fects.* The “random-example” refers to a randomly selected prompt from our designed set of diverse
 790 descriptions of light effects. We provide some examples below:

- 791 • Mystical light beams traversing a clear yet particulate-rich environment, showcasing the
 792 diffusion and beauty of light.
- 793 • Organic patterns of random color and random shape light forming soft, diffused spots with
 794 varied shapes and sizes, as if passing through a complex, uneven medium.
- 795 • A single, intense light source creating a bright spot. The light should have a soft glow
 796 around it, with subtle gradients transitioning from bright white to deep black. Include lens
 797 flares and light streaks for added realism.
- 798 • Abstract a small light with curved rainbow effect, no objects, vibrant light streaks, dynamic
 799 glow, detailed texture, photorealistic.
- 800 • An abstract underwater scene featuring dynamic light rays piercing through a rippling water
 801 surface. The surface is textured with subtle waves and ripples. The overall atmosphere is
 802 ethereal, focusing solely on the interplay of light and water.

810

811 • Abstract warm oval patches of light beams casting soft shadows on a dark background wall,
812 mimicking the dramatic lighting often used in portrait photography. Focus on the interplay
813 of light and shadow without any objects or figures.

814 • The glow reflected on the lake surface, with its floating light leaping like gold.

815 • The golden side light projects regular light and dark patterns on the deep black background,
816 suggesting the warmth and depth of the hidden light source.

817 We provide some examples of light material generated by Flux in Figure 5.

818



819 Figure 5: **Visualizations of the light material generated by Flux.**

820

821 A.2 IMAGE-CONTENT-LIGHT TRIPLETS

822

823 We first employ the vision-language model InternVL2.5 (Chen et al., 2024) to filter approximately
824 2 million images with prominent light effects from a collection of over 20 million natural images.
825 Then, we apply our generative decoupling strategy to these datasets for decoupled generation. After
826 the third filtering operation, approximately 1.2 million high-quality triplets that meet expectations
827 are obtained. To accommodate the training requirements of the diffusion model, we use InternVL2.5
828 to generate brief captions for the images. [The triplet data construction process took approximately
829 12 days, and during this period, we used 16 L40s GPUs for all steps. It takes approximately one day
830 to select images with strong light effects from massive data using VLM. It took 5 days and 6 days
831 respectively to extract the light effects from these images and remove the light effects.](#)

SUH4: W2

832 We categorize all triplets in our dataset into nine light-effect types: 1.Light Beam in Tyndall
833 Effect; 2.Dappled Light; 3.Backlighting Effect; 4.Iridescent Halo; 5.Lens Flare; 6.Bright Light
834 Source; 7.Gleaming Water Reflection; 8.Projecting Light Patch; 9.Volumetric Light Rays. Using
835 InternVL2.5-8B, we count the number of images associated with each category. The distribution of
836 sample counts across each category is shown in Figure 6. To provide a more intuitive understanding
837 of the light effects associated with each category, we present representative examples in Figure 7.
838 Unlike traditional indoor scene data, our data mainly consists of portrait photography and natural
839 landscapes, among which portrait data accounts for about 80%

SUH4: W2;
UIpe: Q1

859 A.3 THE USAGE OF VISION LANGUAGE MODEL

860

861 As shown in Figure 2, in the overall framework for implementing the light effect transfer task, the
862 vision language model InternVL2.5 (Chen et al., 2024) plays a very important role. We leverage the
863 InternVL2.5 model to pre-filter a large-scale database, ensuring high-quality inputs for subsequent
864 generative decoupling training and triplets construction.

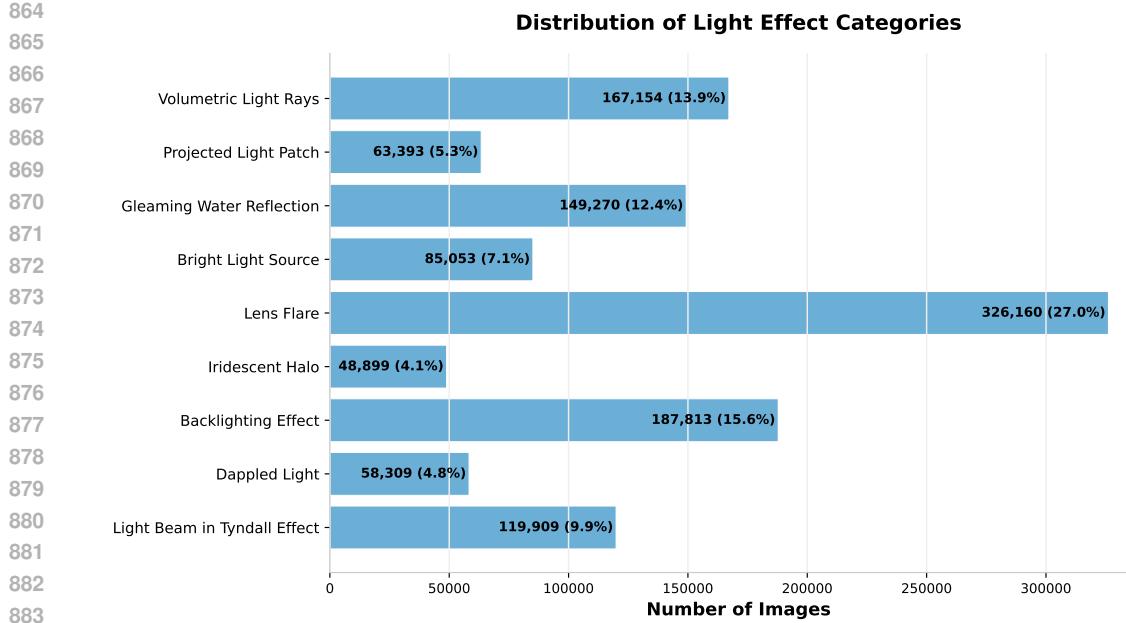


Figure 6: **Light effect distribution.** We categorize light effects into nine classes and perform a statistical analysis on our million-scale triplet dataset.

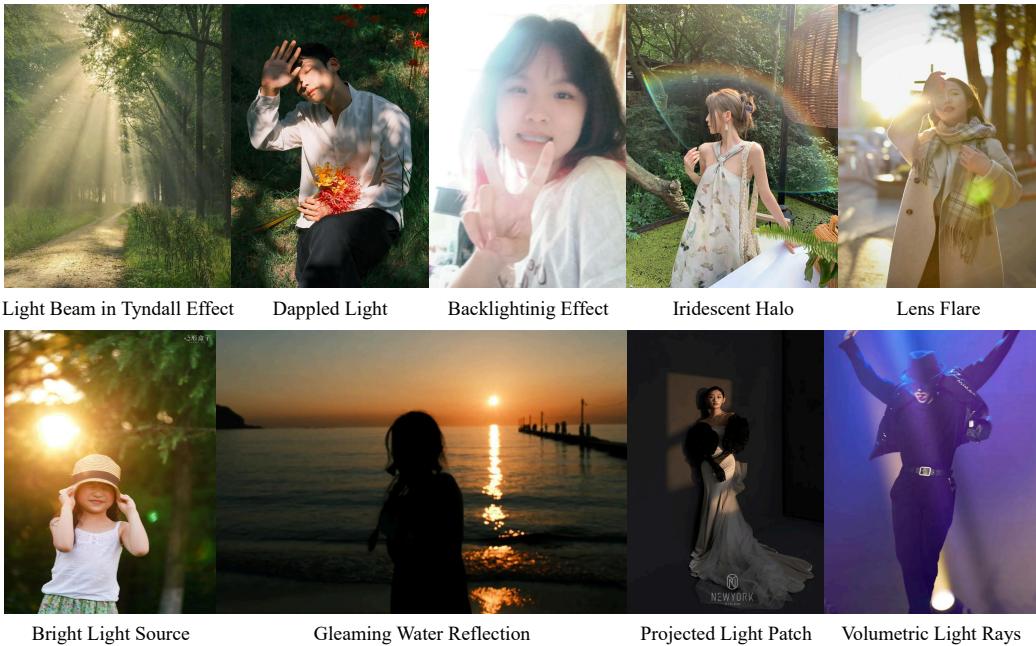


Figure 7: **Examples of different types of light effects.**

We employ the prompt: “*Is there no obvious lighting effect in the image and is the person under uniform lighting?*” to determine whether an image does not contain light effects. We first use this prompt to obtain approximately 1 million images. These images, which exhibit highly similar data distributions in terms of lighting attributes, are used for generative decoupling training process. And this prompt is also applied to judge the validity of the light removal results.

Furthermore, we employ the prompt: “*Does this image exhibit extremely strong lighting effects with dramatic light/shadow contrast, visible light beams/rays, prominent lens flares/light spots, or intense dominant light sources?*” to determine whether an image contains obvious light effects. This prompt is employed to select the source images used for triplet dataset construction.

918
 919 **Table 5: Quantitative results on the MIIW dataset.** Although our method is not specifically
 920 designed for indoor scenes, it remains highly competitive.

Methods	Labels	RMSE \downarrow	SSIM \uparrow
Input Img	-	0.384	0.438
SA-AE (Hu et al., 2020)	Light	0.288	0.484
SA-AE (Hu et al., 2020)	-	0.443	0.300
S3Net (Yang et al., 2021)	Depth	0.512	0.331
S3Net (Yang et al., 2021)	-	0.499	0.336
Latent-Intrinsic (Zhang et al., 2024)($\sigma=4$)	-	0.326	0.232
Latent-Intrinsic (Zhang et al., 2024)	-	0.297	0.473
RGB-X (Zeng et al., 2024b)	-	0.256	0.476
LumiNet (Xing et al., 2025)	-	0.180	0.647
Ours	-	<u>0.212</u>	<u>0.627</u>



933 Figure 8: **Visual Comparison on MIIW (Murmann et al., 2019) dataset.**

945 B COMPARISON WITH INDOOR SCENE RELIGHTING METHODS

946
 947 In the Related Work section of the main text, we briefly review several physically-based relighting
 948 methods (Zhang et al., 2024; Xing et al., 2025; Liang et al., 2025) for indoor scenes. Although the
 949 light effect transfer task we define bears some formal similarities to the tasks addressed by these
 950 methods, our TransLight differs significantly from them in terms of design motivation, application
 951 scenario, and prediction targets.

- 952 • **Design motivation:** Our TransLight aims to enable the flexible transfer of geometrically
 953 structured light effects that are difficult to describe with simple text to other images,
 954 whereas physically-based indoor relighting methods primarily seek to achieve improved
 955 illumination consistency or coherence within a scene.
- 956 • **Application scenario:** Our TransLight focuses on portrait photography scenarios with
 957 broad practical applicability, aiming to enhance the aesthetic quality of images, whereas
 958 relighting methods such as LumiNet are primarily targeted at indoor scenes, such as
 959 bedrooms and living rooms.
- 960 • **Prediction targets:** Our TransLight can add customized light effects with distinct
 961 geometric structures to the input image. Moreover, our TransLight can also adjust the
 962 illumination state of the image by extracting light effects from the reference image to make
 963 the transfer result more realistic. For example, it can modify the shadow of the person's
 964 face according to the direction and position of the light effect, as shown in the example
 965 in the second sample of Figure 1. The target of previous physically-based methods is to
 966 render the input image to a new illumination status, such as converting an “off” desk lamp
 967 to an “on” state based on the state of the reference image, or modifying the direction of the
 968 shadow from left to right according to a specific Env.Map.

Y35F: W1;
 UJpe: W1, W2

Y35F: W1;
 UJpe: W1, W2

Y35F: W1;
 UJpe: W1, W2

970 We follow the evaluation method of Latent Intrinsic (Zhang et al., 2024) and conduct zero-shot
 971 evaluation of our TransLight on the MIIW dataset (Murmann et al., 2019) to enable a thorough
 972 comparison. We randomly select an image and its 12 reference image for light effect transfer



(a) State transfer of lamps in indoor scenes



(b) Light transfer in portrait photography scenes

Figure 9: **Visual Comparison between LumiNet and TransLight under different scenes.** LumiNet excels at transferring illumination status, whereas TransLight is better suited for transferring light effects themselves.

inference. We evaluate all 30 scenes in the test set by computing the RMSE and SSIM between the generated images and the reference images, and report the average scores. The results are shown in the Table 5. We also provide visualization comparison on MIW (Murmann et al., 2019) dataset in Figure. 8. Although our method is not specifically designed for indoor scenes, it nevertheless demonstrates strong generalization capability.

UJpe: W3, W5

Moreover, we also provide a qualitative comparison with LumiNet in Figure. 9. We conduct comparisons on two distinct tasks: state transfer of lamps in indoor scenes and light effect transfer in



Figure 10: **Visual Comparison of Light Removal Results.** The first row shows the input images, the second row presents the results generated by the model trained without using I_R , and the third row displays the results generated by the model trained with both I_S and I_R .

portrait photography scenes. As evidenced by the presented results, it is evident that our TransLight and LumiNet are fundamentally designed for different tasks. Our TransLight can not change the lamp in the input image from the off state to the on state based on a reference image. Similarly, LumiNet cannot add an iridescent halo surrounding the person to a portrait photo either.

UIpe: W4, Q4

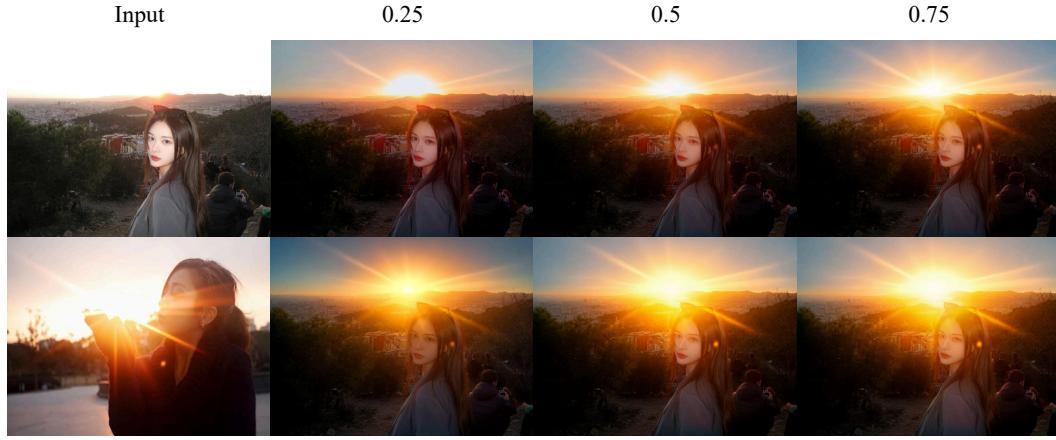


Figure 11: **Transferred results under different intensity coefficients**

C MORE VISUALIZATIONS

During the training of the light removal model, in addition to the synthesized image I_S , the relighted image I_R is also provided as a conditional input. We employ relighting methods to simulate strong illumination contrasts present in real-world scenes, such as shadows and highlights, with the aim of enhancing the generalization capability of the light removal model in eliminating complex light effects. To intuitively illustrate the role of I_R , we present in Figure .10 the generated results of the

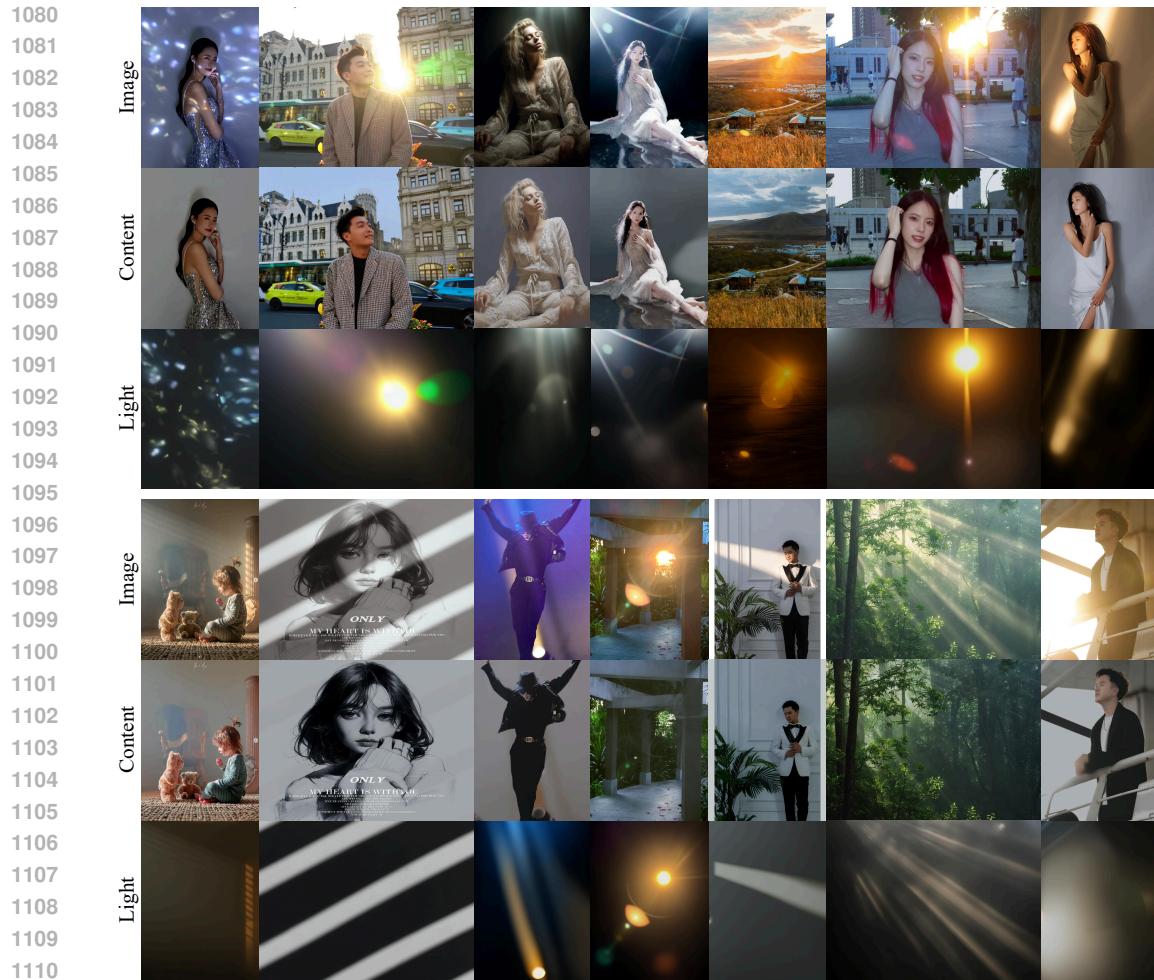


Figure 12: **Visualizations of our generative decoupling.** Our light removal model can eliminate light effects from images while preserving the underlying content unchanged as shown in the second line. Additionally, our light extraction model is capable of isolating the light effects from the image without introducing other extraneous objects.

light removal model trained with and without the inclusion of I_R . When I_R is included as training data, the light removal model is able to simultaneously remove light effects as well as shadows and highlights on the subject’s face.

We present more visualizations of our generative decoupling in Figure 12. [In Figure 13, we show more qualitative comparison results of our TransLight on the task of light effect transfer.](#) Our TransLight accurately captures the structural details of light effects and effectively transfers them onto the input image. Moreover, TransLight demonstrates superior performance in preserving content consistency and significantly outperforms existing methods such as IC-Light and Seedream 4.0 in mitigating content leakage. In addition to controlling the position and direction of light effects, our TransLight further enables adjustment of their intensity in the transferred results. We show an example in Figure 11.

UIpe: Q6;
SUh4: W4

D LIMITATION AND FAILURE CASE

The primary limitation of our TransLight framework lies in its dependence on the performance of the light extraction model. The quality of the final transferred output is heavily influenced by how effectively the light effect is extracted from the reference image. This implies that when the

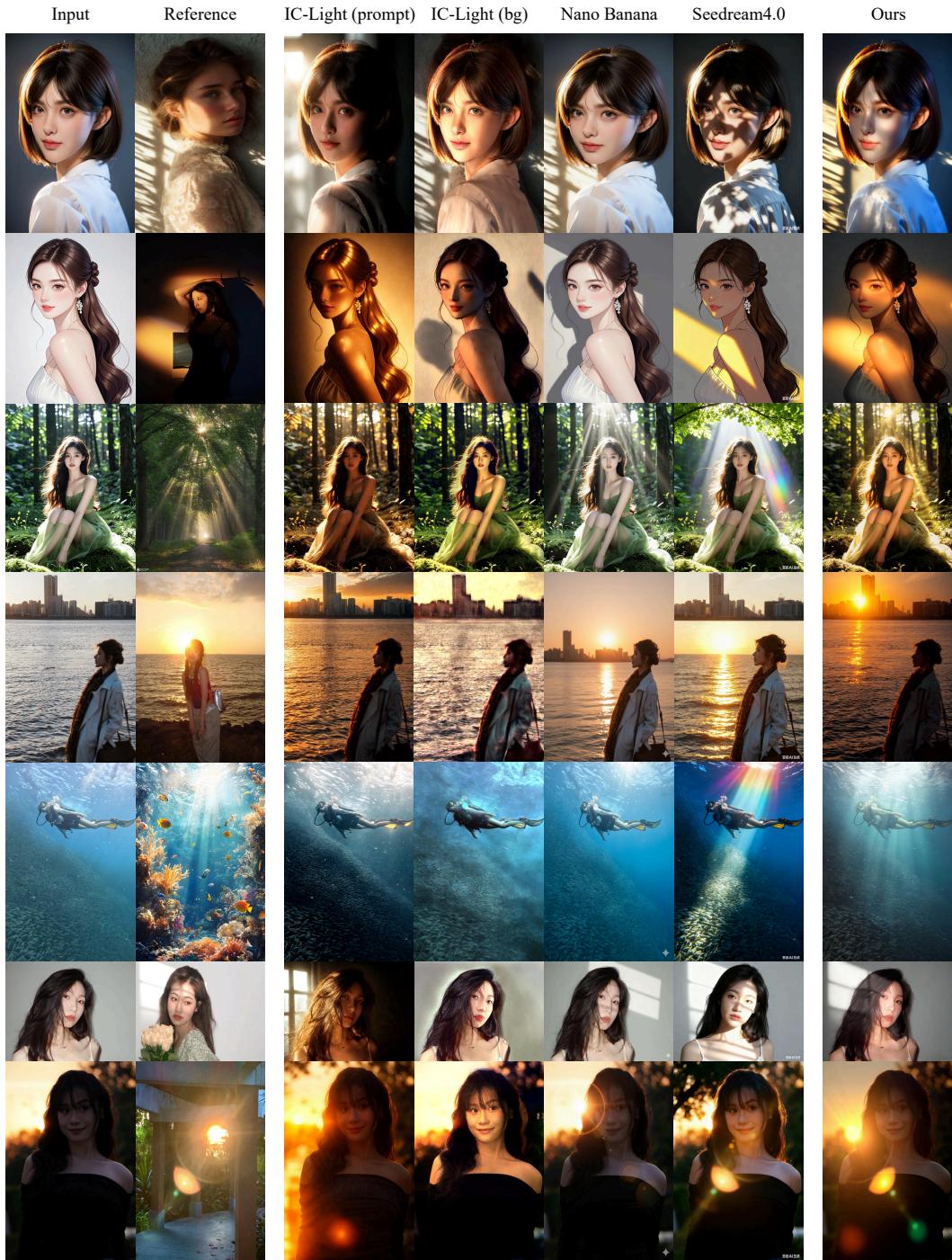
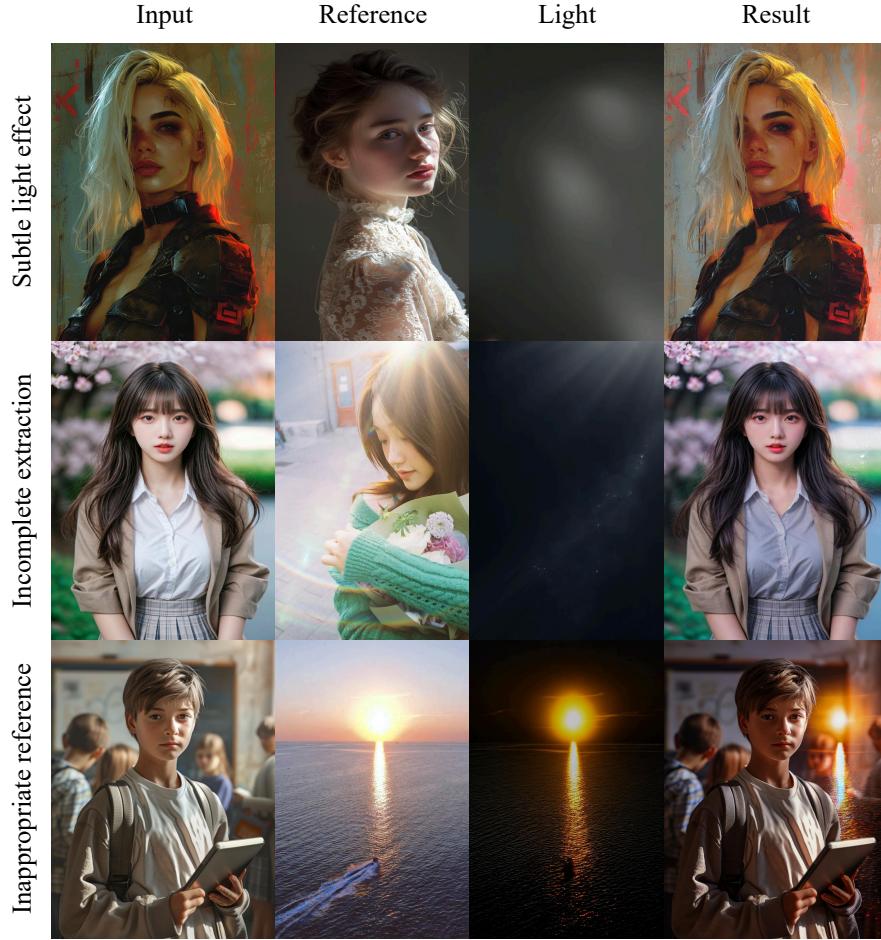


Figure 13: **Visualization comparison.** IC-Light (prompt) uses its prompt-conditioned model, and the prompt is derived from the reference image. IC-Light (bg) uses its background-conditioned model. For Nano Banana and Seedream 4.0, we provide both the input image and the reference image to the image editing model.

light effect in the user-provided reference is subtle or indistinct, our method may fail to produce satisfactory results. We show some failure cases in Figure. 14. When the light effect in the reference image is subtle or when the light extraction model fails to adequately capture the lighting effect, this may result in less pronounced transfer outcomes. In addition, when the reference image itself

1188
 1189
 1190
 1191
 1192
 1193
 1194
 1195
 1196
 1197
 1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218



1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241

Figure 14: **Failure cases.** For subtle light effect case, the transfer degree is not obvious. For incomplete extraction case, since it is impossible to ensure 100% perfect disentanglement of light effects, the generated results may fall short of expectations. For inappropriate reference case, large scene mismatch may be meaningless for the light effect transfer task we defined, and it is unnecessary to pay too much attention to such case.

deviates too much from the input image scene, forced transfer may also lead to inappropriate results. This is because the light effect transfer task in such cases is inherently unreasonable. Given that the primary limitation of our method lies in the light extraction model's capacity to accurately capture light effects, a promising direction for future work in the context of light effect transfer is to further enhance the model's ability to disentangle and represent light effects.

67f2: Q4;
 UJpe: W6, Q5;
 SUh4: W3

E DISCUSSION

In this paper, we define a novel task that has not yet received significant attention: light effect transfer. This task aims to transfer the light effects from a reference image onto a target image. We think such a capability has substantial potential in real-world image-editing scenarios. For instance, after capturing a portrait, users may wish to enhance the visual appeal through appropriate illumination editing. Notably, users typically desire to preserve the original image content while incorporating artistic light effects, which may be sourced from high-quality image templates found on the internet or social media platforms. Motivated by this practical need, we propose TransLight.

Our approach presents a feasible solution to the light effect transfer task by first decoupling the light effect from the reference image into a standalone light-only image, and then compositing it onto the

target image. This strategy significantly mitigates the risk of content leakage, and by adjusting the position and orientation of the extracted light image, enables highly flexible and controllable light transfer with great ease. Although the transfer performance of our TransLight is less pronounced when the lighting effect in the reference image is subtle, this does not diminish the method’s practical applicability or potential for real-world deployment. We have extensively validated our method across diverse reference image scenarios, demonstrating that it achieves lighting editing outcomes that are beyond the reach of existing illumination or image editing techniques. We believe that light effect transfer possesses a broad spectrum of application scenarios and represents a task eminently worthy of in-depth exploration and research. Furthermore, we are confident that our TransLight framework offers a plausible and valuable direction for advancing this field of study.

E.1 STATEMENT ON THE USAGE OF LARGE LANGUAGE MODELS (LLMs)

In this paper, we employ Qwen 3 (Yang et al., 2025) and Gemini 2.5 flash (Google DeepMind, 2025) to perform grammatical revision and linguistic refinement of the manuscript. We ensure that the accuracy and validity of the theoretical explanations and conclusions presented in this paper are not influenced by LLMs.