TransLight: Image-Guided Customized Lighting Control with Generative Decoupling

Zongming Li[* 1 2] **Lianghui Zhu**[1] **Haocheng Shen**[2] **Longjin Ran**[2] **Wenyu Liu**[1] **Xinggang Wang**[1]

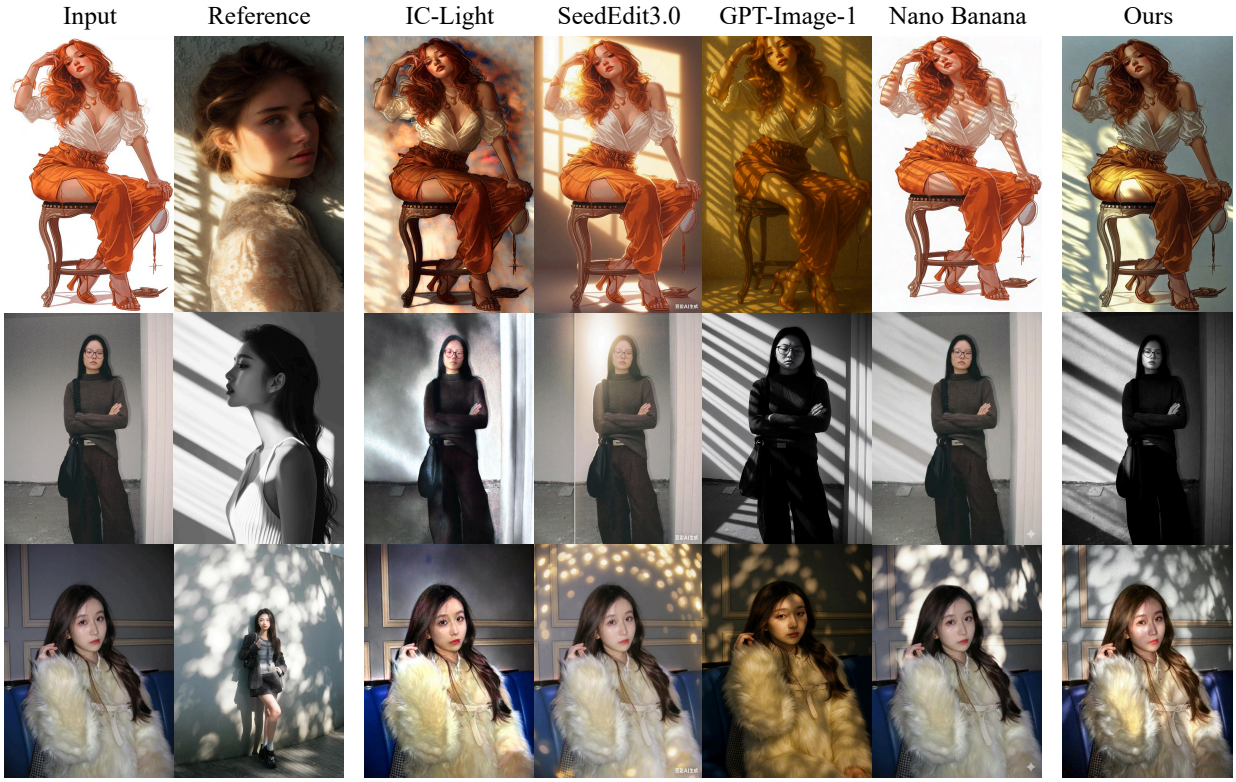

Figure 1. **Transfer light effects from reference to input image.** Our TransLight is a novel method that high-fidelity transfers light effects with obvious geometric structure from reference images to input images. In comparison, IC-Light's (Zhang et al., 2025) background-conditioned model tends to produce darker outputs and may leak reference content into results; SeedEdit3.0 (Wang et al., 2025) fails to capture the type, direction, or structure of lighting; GPT-Image-1 (OpenAI, 2025) partially understands lighting but alters facial identity; and Nano Banana (Google DeepMind, 2025) yields unnatural interactions between subjects and transferred light effects.

Abstract

Most existing illumination-editing methods struggle to jointly offer customized lighting control and preserve content integrity, limiting their effectiveness especially in transferring complex light effects from a reference to a target image in portrait photography. To address this problem, we propose **TransLight**, a novel framework that enables high-fidelity and high-freedom transfer of geometric-structured light effects such as Tyndall beams and specular highlights. Extracting light effects from the reference image is the most critical and challenging step, as real-world lighting contains complex geometric structures tightly coupled with image content. To achieve this, we propose Generative Decoupling, using two

* Work was done during Zongming Li's internship at vivo Mobile Communication Co., Ltd.

[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, WuHan, China
[2]vivo Mobile Communication Co., Ltd, China. Correspondence to: Xinggang Wang <xgwang@hust.edu.cn>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

fine-tuned diffusion models to accurately separate image content and lighting, and create a new million-scale dataset of image–content–light triplets. We then adopt IC-Light as the generative model, training it on these triplets with the reference lighting image as an additional conditioning signal. The resulting model enables customized and natural transfer of diverse light effects. Notably, by fully disentangling light effects from reference images, our generative decoupling strategy gives TransLight highly flexible illumination control. Experiments show that TransLight successfully transfers geometric-structured lighting effects across diverse images in portrait photography, offering more customized control than existing methods and charting new directions in illumination harmonization and editing. Our project page is *hustvl.github.io/TransLight*.

1. Introduction

Recent rapid iteration of image illumination editing methods (Zhang et al., 2021; Dastjerdi et al., 2023; Choi et al., 2025; Fortier-Chouinard et al.; Kim et al., 2024; Ren et al., 2024; Phongthawee et al., 2024; Zeng et al., 2024a; Jin et al., 2024; Magar et al., 2025; Chaturvedi et al., 2025) has led to significant advancements in manipulating lighting attributes within the field of image editing. Relighting of human portraits, as a specialized task in this field, holds broad application prospects and high practical value. Existing approaches generally fall into two categories: visual-guided and text-guided illumination adjustment. Visual-guided methods (Kim et al., 2024; Ren et al., 2024; Chaturvedi et al., 2025) often employ high dynamic range (HDR) maps or additional background images to composite human portraits into new lighting environments, but this typically compromises the content integrity of the original image. Text-guided methods (Zhang et al., 2025; Cha et al., 2025) accept textual descriptions of the desired light effects as input to perform image relighting. Textual descriptions alone are clearly insufficient for customized control over lighting properties such as direction, position, and geometric structure. While these methods can produce striking visual effects, their applicability to portrait relighting via background replacement or text prompts remains limited. In real-world image editing scenarios, users may prefer personalized lighting modifications (e.g., adding or removing specific light effects) while preserving the original content of the image. For example, a user might request: *"Add a rainbow lens flare similar to the reference image to my outdoor travel photo."* Existing relighting approaches often fail to meet such customized lighting stylization requirements.

In this work, we focus on the highly challenging task of transferring light effects from a reference image to a target image in portrait photography. To achieve this, we must address two main challenges: (1) *decoupling light effects from content to prevent content leakage* and (2) *constructing a suitable data structure for light effect transfer training*. Although this task is broadly similar to well-established style transfer, both aiming to replicate specific attributes from one reference image onto another image, it requires additional consideration of geometric information such as the direction and structure of the light effects. This leads to a stronger coupling between the light effects and the content, making many existing strategies (Wang et al., 2024a;b; Xing et al., 2024; Jeong et al., 2024) designed to mitigate content leakage in style transfer task less applicable to our case. To address the challenges above, we propose **Generative Decoupling**. Specifically, we fine-tune two diffusion models using natural images without light effects and light material images, enabling them to remove and extract light effects from input images. After sufficient training, our light extraction model effectively isolates dominant light effects without retaining extraneous content, largely avoiding content leakage. Moreover, we construct over one million image-content-light triplets for light effect transfer task training using our generative decoupling strategy.

After that, we propose TransLight, a novel method capable of transferring light effects with obvious geometric structure from a reference image to another image. It is worth noting that we focus on visually perceptible light effects with distinct geometric structures, such as light beam in Tyndall Effect, iridescent halo, and lens flare, among others.

We employ IC-Light (Zhang et al., 2025) as the fixed generative network and train our model using the image-content-light triplets. During inference, the target light effect is decoupled from the reference image, effectively reducing content leakage and enhancing transfer flexibility. Users can freely adjust the result by modifying the position and orientation of the light effect. Experimental results show that our TransLight achieves strong customized light effect transfer, preserving the background of the source image while naturally integrating target lighting from the reference image, as shown in Figure 1. Moreover, our method achieves a Light-FID score of 6.02, outperforming the SOTA illumination editing method IC-Light, which obtains 10.05.

The main contribution can be summarized as follows:

- We propose **Generative Decoupling** to decouple content and light in natural images using two diffusion models. Among them, light removal model eliminates prominent light effects from natural images, *e.g.*, lens flares, Tyndall effects, *etc*. Light extraction model is the first capable of separating light effects from images without retaining any original content.

- We construct an image-content-light triplet generation pipeline based on our generative decoupling. We first select images with prominent light effects from a massive database, and then use our light removal and light extraction models to generate the corresponding content and light effects images, respectively. After filtering step, we obtain over one million training samples.

- We introduce **TransLight**, a novel image-guided method that allows accurate transfer of light effects with obvious geometric structure from a reference image to a target image in portrait photography scenes. Our method supports translation, flipping, and numerical scaling of the extracted light effects, enabling adjustments to the position, orientation, and intensity of the light effects in transferred result. Experimental results show that our TransLight achieves visually striking light effects across a wide range of scenarios.

2. Related Work

2.1. Image Style Transfer

Image style transfer aims to transfer the style of a reference image to the target content image (Chung et al., 2024; Gao et al., 2025; Wang et al., 2023; Zhang et al., 2023b; Frenkel et al., 2024; Wang et al., 2024b; Qi et al., 2024; Lei et al., 2025; Xing et al., 2024). This process is fundamentally similar to light effects transfer, which aims to replicate intrinsic attributes from a reference image onto a target image while preserving the target's content consistency. A well-studied challenge in this task is disentangling content and style in the reference image to prevent content leakage. StyelAligned (Hertz et al., 2024) proposes replacing full attention with shared attention during style injection to reduce content leakage. FreeTuner (Xu et al., 2024) employs a disentanglement strategy that separates the generation process into two stages to effectively mitigate concept entanglement. InstantStyle (Wang et al., 2024a) and Swapping self-attention (Jeong et al., 2024) reduce the introduction of content information by injecting style features into specific attention layers. The aforementioned methods are designed to disentangle abstract stylistic attributes and effectively mitigate content leakage. However, they cannot accurately capture the geometric and structural characteristics of light effects, making light extraction difficult and rendering them unsuitable for light effects transfer.

2.2. Portrait Relighting and Lighting Control

Diffusion models have recently been widely applied to highly challenging tasks requiring precise control over image illumination, such as portrait relighting and indoor lighting adjustment. For portrait relighting, methods such as DiFaReli (Ponglertnapakorn et al., 2023), Lite2Relight (Rao et al., 2024), and Holo-Relighting (Mei et al., 2024) specifically target the relighting of human faces. Relightful Harmonization (Ren et al., 2024) and Total Relighting (Pandey et al., 2021) manipulate the illumination of the foreground portrait using background conditions. SwitchLight (Kim et al., 2024) controls portrait illumination by using high dynamic range (HDR) images as conditions. Text2Relight (Cha et al., 2025) generates relighted portrait using text prompt while keeping the original contents. Furthermore, IC-Light (Zhang et al., 2025), trained extensively on large-scale datasets using the principle of light transport consistency, achieves impressive capabilities in text prompt and background controllable portrait illumination generation. For indoor lighting adjustment, common operations include introducing new light sources or regulating the illumination intensity of those already present in a scene (Zhang et al., 2024; Wang et al., 2022; Magar et al., 2025). SpotLight (Fortier-Chouinard et al.) control the local lighting of an object through a coarse shadow. Similar to the structure-guided generation methods (Zhang et al., 2023a; Li et al., 2025), ScribbleLight (Choi et al., 2025) controls image light effects based on user scribbles. Latent Intrinsics(Zhang et al., 2024) and LumiNet(Xing et al., 2025) are capable of modifying the illumination states of input images based on the light effects from a reference image. DiffusionRenderer (Liang et al., 2025) use specified environment map to accurately estimates geometry and material buffers and generates photorealistic images. In addition, in the field of video generation, Light-A-Video (Zhou et al., 2025) and RelightVid (Fang et al., 2025) use IC-Light to add artistic lighting and shadow effects to the video generation process.While the aforementioned light control methods achieve good results in specific scenarios, their customized control over light effects is highly limited. They cannot precisely and comprehensively adjust details such as light position, and shape while preserving image content. Our TransLight aims to provide a solution.

3. Method

In this section, we first provide a clear task definition for the light effect transfer task: *Given a user input image I, a reference image R and user control parameters p (e.g., translation, flip, intensity), finally obtain the edited result image I_L.* The most ideal generation result is that I_L has the same light effects as the reference image R, mainly manifested in the consistency of the geometric structure and the type of the light effect in the two images. Moreover, I_L should maintain complete content consistency with the input image I. The light effect transfer task we define is formally similar to the illumination condition transfer task in LumiNet (Xing et al., 2025), yet exhibits clear distinctions in many aspects. We provide a discussion in the appendix.

We systematically illustrate the overall framework for achieving the task of light effect transfer in Figure 2. This framework comprises three key modules: the model training of generative decoupling, the image-content-light triplet construction pipeline, and the training of our TransLight.

3.1. Generative Decoupling

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

Source data collection. We first employ InternVL2.5 (Chen et al., 2024) to select 1M images without light effects from a massive database. Then, we collect 100K light material images, all featuring pure black backgrounds. Specifically, we generate approximately 90K light material images using FLUX.1-schnell (Labs, 2024) and collect an additional 10K real images from public datasets (Dai et al., 2022; Wu et al., 2021). The details of the prompts used in the image collection process are given in the appendix.

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., 2015) architecture using the weights from IC-Light (Zhang et al., 2025). The training input of the U-net (Ronneberger et al., 2015) 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 light extraction model achieves good performance in real-world scenarios.

3.2. Image-content-light Triplet Construction Pipeline

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.

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.

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.

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., 2024), and compute the cosine similarity of these features. The specific calculation method is as follows:

$$cos(\alpha) = Sim_{cos}(\mathcal{D}(I_L), \mathcal{D}(I)), \tag{1}$$

$$cos(\beta) = Sim_{cos}(\mathcal{D}(I_L), \mathcal{D}(I_S)), \tag{2}$$

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:

$$(cos(\beta) > cos(\alpha)) \wedge (cos(\alpha) < \gamma), \tag{3}$$

where γ is set to 0.98 to balance data quality and quantity. A larger γ retains more data after filtering but with lower average quality, while a smaller γ yields fewer but higher-quality samples.

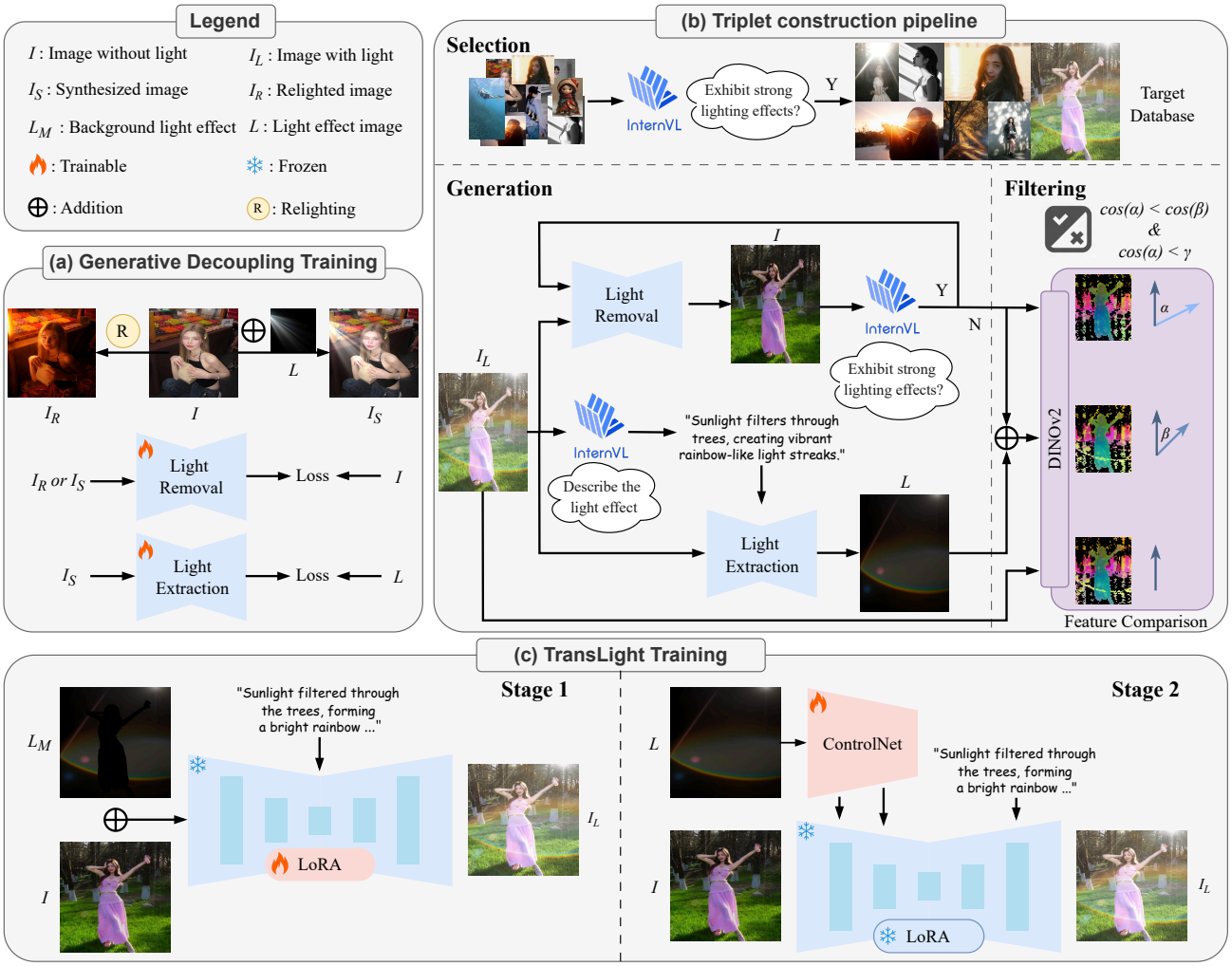

Figure 2. **Overall framwork.** (a) We fine-tune two diffusion models initialized with the weights of IC-Light. During training, the input synthesized image I_S is the direct addition result of no light image I and light material image L. (b) This pipeline involves selecting relevant data using a vision 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.

3.3. TransLight

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.

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. 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 $L \times (1 - M)$.

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 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.

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 & Welling, 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.

3.4. Scope of Application

Our TransLight is specifically designed for the transfer of geometric-structured light effects in portrait scenarios. We envision the ideal application scenario as follows: A user wishes to enhance the visual quality of a captured portrait by incorporating striking light effects. After retrieving a satisfactory reference image from the internet, the user employs our TransLight to transfer and replicate the light effect from the template onto their own photograph. Throughout this process, our TransLight is capable of effectively transferring a diverse range of light effects with explicit geometric structures, such as the Tyndall effect, dappled light and lens flare. The core architecture of TransLight is designed to learn the physical mapping between a decoupled light effect L and arbitrary surface geometries, rather than being category-specific. Our focus on visually salient light effects (beams, halos, flares) is a conscious design choice to address the most challenging and artistically relevant tasks in portrait photography, which general relighting models often fail to capture. While our method utilizes 2D spatial transformations of the light effect L , this 2D effect compositing is inherently content-aware and generative rather than a simple linear blend. Our TransLight framework is not a narrow 2D tool but a robust lighting control engine capable of generalizing to diverse scenes by interpreting the underlying geometric structure of light. Moreover, for cases where the input image already contains strong lighting, a pre-processing light removal step can be applied to obtain a neutral input, ensuring that the subsequently transferred light effects from the reference image do not conflict with original textures. Our decoupled workflow preserves physical consistency and content fidelity.

Table 1. **Quantitative results.** IC-Light employs its background-conditioned model, using the ground truth as the background input.

METHOD	PSNR ↑	SSIM ↑	LPIPS ↓
IC-LIGHT	15.34	0.6784	0.2756
OURS	19.58	0.7931	0.1982

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.

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.

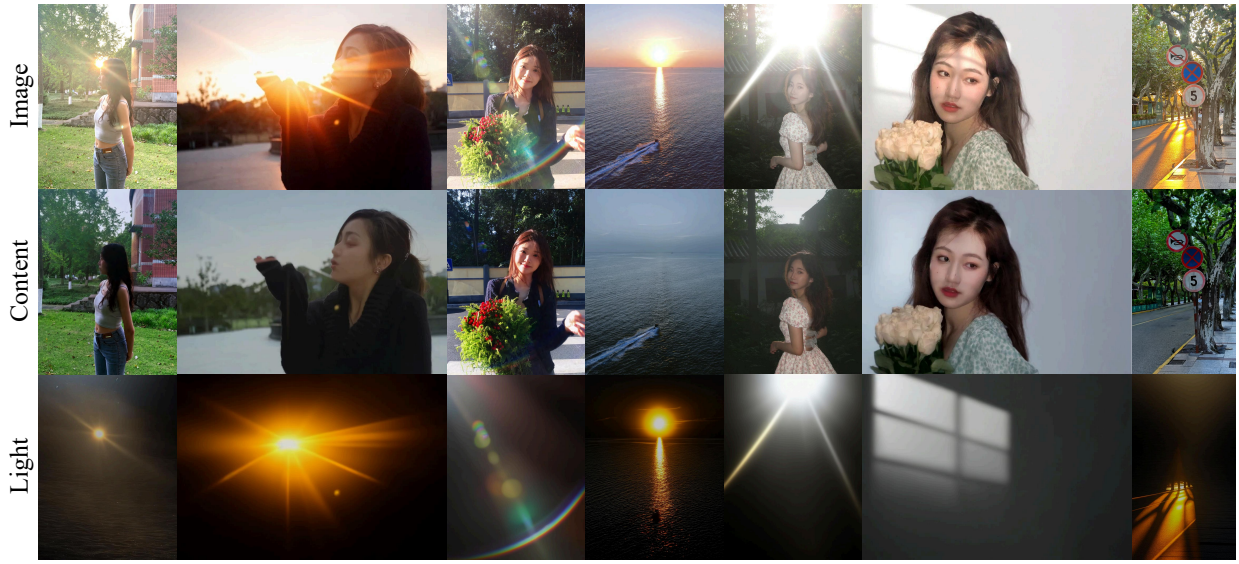

Figure 3. **Visualizations of our generative decoupling.** Our light removal model can eliminate light effects from images while preserving the underlying content unchanged as shown in the second line. Additionally, our light extraction model is capable of isolating the light effects from the image without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

4.2. Performance of Generative Decoupling

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

4.3. Performance of TransLight

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

In addition to the similarity metrics computed on individual images, we argue that measuring the distributional distance between generated and real images offers a more objective assessment of overall model performance. Therefore, we sample 12,000 triplets that are not used during training. We compute the Fréchet Inception Distance (FID) (Heusel et al., 2017) to evaluate the light transfer capability of the model, which we refer to as **Light-FID**. The Light-FID reflects the distance of the data distribution between the transfer result images and the real images with obvious light effects. We report the result in Table 2. When performing inference using the background conditioned model of IC-Light, the ground truth is directly used as the background input. Our method demonstrates significant advantages over IC-Light.

For subjective comparison in real-world application sce-

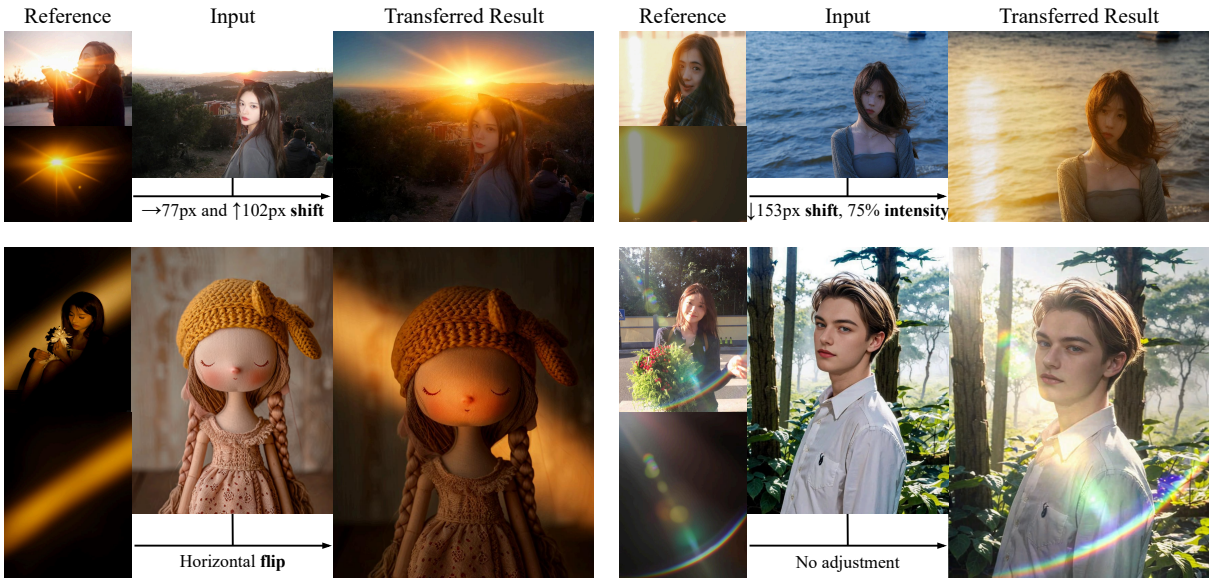

Figure 4. **Visualizations of the specific details about the light transfer process.** Our TransLight tackles this challenging task through generative decoupling. We extract the light effect from the reference image using our fine-tuned model and composite it onto the target image, enabling flexible control over position and direction for realistic, high-freedom results.

Table 2. **Light-FID.** "Content" in the first row refers to result of lighting-removed content image. "Content+Light" refers to the result formed by directly adding content and light effect maps.IC-Light uses its background conditioned model, with 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

narios, we show some visualizations in Figure 1. To fully highlight the advantages of our TransLight, we also show the result of the SOTA relighting method IC-Light, alongside the closed-source SOTA image editing methods SeedEdit3.0 (Wang et al., 2025), GPT-Image-1 (OpenAI, 2025) and Nano Banana (Google DeepMind, 2025). The reference image is used as background input for IC-Light's background conditioned model. For SeedEdit3.0, GPT-Image-1 and Nano Banana, we directly provide both input image and reference image to the model simultaneously. We experiment with different prompts to instruct the model and select the best output from the results. As illustrated in Figure 1, current relighting and image editing methods are unable to fulfill our expected light effect transfer requirements. IC-Light leaks reference content into results and tends to produce darker outputs. SeedEdit3.0 modifies lighting but cannot transfer user-specified effects from a reference. GPT-Image-1 alters facial appearance noticeably. Nano Banana best captures light type and structure from the reference, yet its results lack realistic interaction between lighting and subjects. Moreover, freely adjusting the position and direction of light effects enables more natural integration with the target image, yielding visually superior results as shown in Figure 4. Our TransLight achieves image-guided customized lighting control by transferring light effects from a reference image to a target image. We have also provided the results of the user study in appendix.

4.4. Ablation Study

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

Training data. We conduct an experiment in which TransLight is trained on relighted data instead. Specifically, we employ the text-conditioned model of IC-Light to relight real images that originally contain no light effects, and then

Table 3. **Ablation of filtering threshold γ.** We report training sample count and the Light-FID for different 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

extract the corresponding light effects images from the generated results to form the triplet data. Using this approach, we also generate one million training samples. The Light-FID is reported in Table 2. This result further justify the rationality of our motivation to construct triplets for training by leveraging real-world data.

Data filtering. To validate the filtering step in the recipe construction pipeline, we train our model on one million unfiltered triplets and report the Light-FID score in Table 2. Training with filtered data enables our model to perform better. Moreover, the threshold γ significantly affects both the quantity and quality of the data. We perform filtering under different values of γ and train the model using the resulting eligible images, and the experimental results are presented in Table 3. As shown in the table, we set $\gamma = 0.98$ for the trade off between data quality and data quantity.

5. Conclusion

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

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 62376102).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Cha, J., Ren, M., Singh, K. K., Zhang, H., Hold-Geoffroy, Y., Yoon, S., Jung, H., Yoon, J. S., and Baek, S. Text2relight: Creative portrait relighting with text guidance. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, pp. 1980–1988, 2025.

Chaturvedi, S., Ren, M., Hold-Geoffroy, Y., Liu, J., Dorsey, J., and Shu, Z. Synthlight: Portrait relighting with diffusion model by learning to re-render synthetic faces. In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 369–379, 2025.

Chen, Z., Wu, J., Wang, W., Su, W., Chen, G., Xing, S., Zhong, M., Zhang, Q., Zhu, X., Lu, L., et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 24185–24198, 2024.

Choi, J. M., Wang, A., Peers, P., Bhattad, A., and Sengupta, R. Scribblelight: Single image indoor relighting with scribbles. In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 5720–5731, 2025.

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

Dai, Y., Li, C., Zhou, S., Feng, R., and Loy, C. C. Flare7k: A phenomenological nighttime flare removal dataset. Advances in Neural Information Processing Systems, 35: 3926–3937, 2022.

Dastjerdi, M. R. K., Eisenmann, J., Hold-Geoffroy, Y., and Lalonde, J.-F. Everlight: Indoor-outdoor editable hdr lighting estimation. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 7420–7429, 2023.

Fang, Y., Sun, Z., Zhang, S., Wu, T., Xu, Y., Zhang, P., Wang, J., Wetzstein, G., and Lin, D. Relightvid: Temporal-consistent diffusion model for video relighting. arXiv preprint arXiv:2501.16330, 2025.

Fortier-Chouinard, F., Zhang, Z., Messier, L.-E., Garon, M., Bhattad, A., and Lalonde, J.-F. Spotlight: Shadow-guided

object relighting via diffusion. In Thirteenth International Conference on 3D Vision.

Frenkel, Y., Vinker, Y., Shamir, A., and Cohen-Or, D. Implicit style-content separation using b-lora. In European Conference on Computer Vision, pp. 181–198. Springer, 2024.

Gao, J., Sun, Y., Liu, Y., Tang, Y., Zeng, Y., Qi, D., Chen, K., and Zhao, C. Styleshot: A snapshot on any style. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025.

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

Hertz, A., Voynov, A., Fruchter, S., and Cohen-Or, D. Style aligned image generation via shared attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4775–4785, 2024.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30, 2017.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. Lora: Low-rank adaptation of large language models. ICLR, 1(2):3, 2022.

Hu, Z., Huang, X., Li, Y., and Wang, Q. Sa-ae for any-to-any relighting. In European Conference on Computer Vision, pp. 535–549. Springer, 2020.

Jeong, J., Kim, J., Choi, Y., Lee, G., and Uh, Y. Visual style prompting with swapping self-attention. arXiv preprint arXiv:2402.12974, 2024.

Jin, H., Li, Y., Luan, F., Xiangli, Y., Bi, S., Zhang, K., Xu, Z., Sun, J., and Snavely, N. Neural gaffer: Relighting any object via diffusion. In The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024.

Kim, H., Jang, M., Yoon, W., Lee, J., Na, D., and Woo, S. Switchlight: Co-design of physics-driven architecture and pre-training framework for human portrait relighting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 25096–25106, 2024.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

Labs, B. F. Flux. https://github.com/black-forest-labs/flux, 2024.

Labs, B. F., Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., Kulal, S., Lacey, K., Levi, Y., Li, C., Lorenz, D., Müller, J., Podell, D., Rombach, R., Saini, H., Sauer, A., and Smith, L. Flux.1 kontext: Flow matching for in-context image generation and editing in latent space, 2025. URL https://arxiv.org/abs/2506.15742.

Lei, M., Song, X., Zhu, B., Wang, H., and Zhang, C. Stylestudio: Text-driven style transfer with selective control of style elements. In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 23443–23452, 2025.

Li, Z., Cheng, T., Chen, S., Sun, P., Shen, H., Ran, L., Chen, X., Liu, W., and Wang, X. Controlar: Controllable image generation with autoregressive models. In International Conference on Learning Representations, 2025.

Liang, R., Gojcic, Z., Ling, H., Munkberg, J., Hasselgren, J., Lin, C.-H., Gao, J., Keller, A., Vijaykumar, N., Fidler, S., et al. Diffusion renderer: Neural inverse and forward rendering with video diffusion models. In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 26069–26080, 2025.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101, 2017.

Magar, N., Hertz, A., Tabellion, E., Pritch, Y., Rav-Acha, A., Shamir, A., and Hoshen, Y. Lightlab: Controlling light sources in images with diffusion models. In Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers, pp. 1–11, 2025.

Mei, Y., Zeng, Y., Zhang, H., Shu, Z., Zhang, X., Bi, S., Zhang, J., Jung, H., and Patel, V. M. Holo-relighting: Controllable volumetric portrait relighting from a single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4263–4273, 2024.

Murmann, L., Gharbi, M., Aittala, M., and Durand, F. A multi-illumination dataset of indoor object appearance. In 2019 IEEE international conference on computer vision (ICCV), 2019.

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

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. Dinov2: Learning robust visual features without supervision. Transactions on Machine Learning Research Journal, 2024.

Pandey, R., Orts-Escolano, S., Legendre, C., Haene, C., Bouaziz, S., Rhemann, C., Debevec, P. E., and Fanello, S. R. Total relighting: learning to relight portraits for background replacement. ACM Trans. Graph., 40(4): 43–1, 2021.

Phongthawee, P., Chinchuthakun, W., Sinsunthithet, N., Jampani, V., Raj, A., Khungurn, P., and Suwajanakorn, S. Diffusionlight: Light probes for free by painting a chrome ball. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 98–108, 2024.

Ponglertnapakorn, P., Tritrong, N., and Suwajanakorn, S. Difareli: Diffusion face relighting. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 22646–22657, 2023.

Qi, T., Fang, S., Wu, Y., Xie, H., Liu, J., Chen, L., He, Q., and Zhang, Y. Deadiff: An efficient stylization diffusion model with disentangled representations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8693–8702, 2024.

Rao, P., Fox, G., Meka, A., BR, M., Zhan, F., Weyrich, T., Bickel, B., Pfister, H., Matusik, W., Elgharib, M., et al. Lite2relight: 3d-aware single image portrait relighting. In ACM SIGGRAPH 2024 Conference Papers, pp. 1–12, 2024.

Ren, M., Xiong, W., Yoon, J. S., Shu, Z., Zhang, J., Jung, H., Gerig, G., and Zhang, H. Relightful harmonization: Lighting-aware portrait background replacement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6452–6462, 2024.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10684–10695, 2022.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer, 2015.

Seedream, T., Chen, Y., Gao, Y., Gong, L., Guo, M., Guo, Q., Guo, Z., Hou, X., Huang, W., Huang, Y., et al. Seedream 4.0: Toward next-generation multimodal image generation. arXiv preprint arXiv:2509.20427, 2025.

Wang, G., Yang, Y., Loy, C. C., and Liu, Z. Stylelight: Hdr panorama generation for lighting estimation and editing. In European conference on computer vision, pp. 477–492. Springer, 2022.

Wang, H., Spinelli, M., Wang, Q., Bai, X., Qin, Z., and Chen, A. Instantstyle: Free lunch towards style-preserving in text-to-image generation. arXiv preprint arXiv:2404.02733, 2024a.

Wang, H., Xing, P., Huang, R., Ai, H., Wang, Q., and Bai, X. Instantstyle-plus: Style transfer with content-preserving in text-to-image generation. arXiv preprint arXiv:2407.00788, 2024b.

Wang, P., Shi, Y., Lian, X., Zhai, Z., Xia, X., Xiao, X., Huang, W., and Yang, J. Seededit 3.0: Fast and high-quality generative image editing. arXiv preprint arXiv:2506.05083, 2025.

Wang, Z., Zhao, L., and Xing, W. Styledifusion: Controllable disentangled style transfer via diffusion models. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 7677–7689, 2023.

Wu, C., Li, J., Zhou, J., Lin, J., Gao, K., Yan, K., ming Yin, S., Bai, S., Xu, X., Chen, Y., Chen, Y., Tang, Z., Zhang, Z., Wang, Z., Yang, A., Yu, B., Cheng, C., Liu, D., Li, D., Zhang, H., Meng, H., Wei, H., Ni, J., Chen, K., Cao, K., Peng, L., Qu, L., Wu, M., Wang, P., Yu, S., Wen, T., Feng, W., Xu, X., Wang, Y., Zhang, Y., Zhu, Y., Wu, Y., Cai, Y., and Liu, Z. Qwen-image technical report, 2025. URL https://arxiv.org/abs/2508.02324.

Wu, Y., He, Q., Xue, T., Garg, R., Chen, J., Veeraraghavan, A., and Barron, J. T. How to train neural networks for flare removal. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2239–2247, 2021.

Xing, P., Wang, H., Sun, Y., Wang, Q., Bai, X., Ai, H., Huang, R., and Li, Z. Csgo: Content-style composition in text-to-image generation. arXiv preprint arXiv:2408.16766, 2024.

Xing, X., Groh, K., Karaoglu, S., Gevers, T., and Bhattad, A. Luminet: Latent intrinsics meets diffusion models for indoor scene relighting. In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 442–452, 2025.

Xu, Y., Wang, Z., Li, K., Xiao, J., and Chen, L. Freetuner: Any subject in any style with training-free diffusion. ACM Transactions on Multimedia Computing, Communications and Applications, 2024.

Yang, H.-H., Chen, W.-T., and Kuo, S.-Y. S3net: A single stream structure for depth guided image relighting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 276–283, 2021.

Yang, Y., Sial, H. A., Baldrich, R., and Vanrell, M. Relighting from a single image: Datasets and deep intrinsic-based architecture. IEEE Transactions on Multimedia, 27:2608–2622, 2025.

Zeng, C., Dong, Y., Peers, P., Kong, Y., Wu, H., and Tong, X. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In ACM SIGGRAPH 2024 Conference Papers, pp. 1–12, 2024a.

Zeng, Z., Deschaintre, V., Georgiev, I., Hold-Geoffroy, Y., Hu, Y., Luan, F., Yan, L.-Q., and Hašan, M. Rgb-x: Image decomposition and synthesis using material- and lighting-aware diffusion models. In ACM SIGGRAPH 2024 Conference Papers, SIGGRAPH '24, New York, NY, USA, 2024b. Association for Computing Machinery. ISBN 9798400705250. doi: 10.1145/3641519.3657445. URL https://doi.org/10.1145/3641519.3657445.

Zhang, L., Rao, A., and Agrawala, M. Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 3836–3847, 2023a.

Zhang, L., Rao, A., and Agrawala, M. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In The Thirteenth International Conference on Learning Representations, 2025.

Zhang, X., Fanello, S., Tsai, Y.-T., Sun, T., Xue, T., Pandey, R., Orts-Escolano, S., Davidson, P., Rhemann, C., Debevec, P., et al. Neural light transport for relighting and view synthesis. ACM Transactions on Graphics (TOG), 40(1):1–17, 2021.

Zhang, X., Gao, W., Jain, S., Maire, M., Forsyth, D., and Bhattad, A. Latent intrinsics emerge from training to relight. Advances in Neural Information Processing Systems, 37:96775–96796, 2024.

Zhang, Y., Huang, N., Tang, F., Huang, H., Ma, C., Dong, W., and Xu, C. Inversion-based style transfer with diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10146–10156, 2023b.

Zheng, P., Gao, D., Fan, D.-P., Liu, L., Laaksonen, J., Ouyang, W., and Sebe, N. Bilateral reference for high-resolution dichotomous image segmentation. CAAI Artificial Intelligence Research, 2024.

Zhou, Y., Bu, J., Ling, P., Zhang, P., Wu, T., Huang, Q., Li, J., Dong, X., Zang, Y., Cao, Y., et al. Light-a-video: Training-free video relighting via progressive light fusion. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 13315–13325, 2025.

Zhu, Z.-L., Li, Z., Zhang, R.-X., Guo, C.-L., and Cheng, M.-M. Designing an illumination-aware network for deep image relighting. IEEE Transactions on Image Processing, 31:5396–5411, 2022.

A. Dataset Details

A.1. Light Material

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

- Mystical light beams traversing a clear yet particulate-rich environment, showcasing the diffusion and beauty of light.
- Organic patterns of random color and random shape light forming soft, diffused spots with varied shapes and sizes, as if passing through a complex, uneven medium.
- A single, intense light source creating a bright spot. The light should have a soft glow around it, with subtle gradients transitioning from bright white to deep black. Include lens flares and light streaks for added realism.
- Abstract a small light with curved rainbow effect, no objects, vibrant light streaks, dynamic glow, detailed texture, photorealistic.
- An abstract underwater scene featuring dynamic light rays piercing through a rippling water surface. The surface is textured with subtle waves and ripples. The overall atmosphere is ethereal, focusing solely on the interplay of light and water.
- Abstract warm oval patches of light beams casting soft shadows on a dark background wall, mimicking the dramatic lighting often used in portrait photography. Focus on the interplay of light and shadow without any objects or figures.
- The glow reflected on the lake surface, with its floating light leaping like gold.
- The golden side light projects regular light and dark patterns on the deep black background, suggesting the warmth and depth of the hidden light source.

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

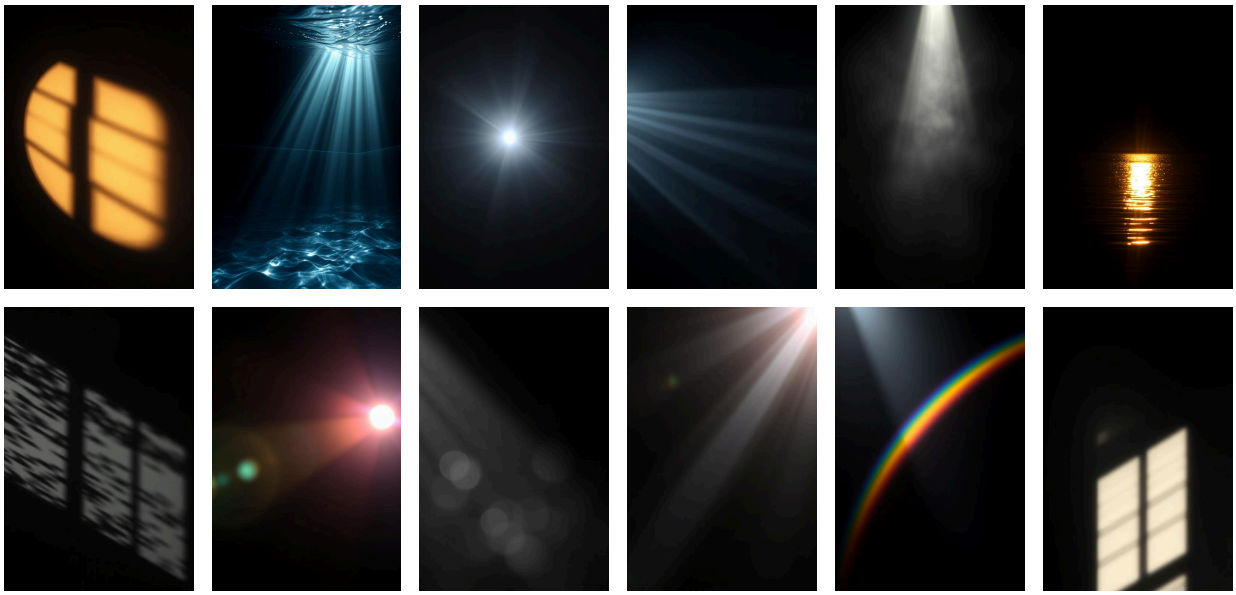

Figure 5. **Visualizations of the light material generated by Flux.**

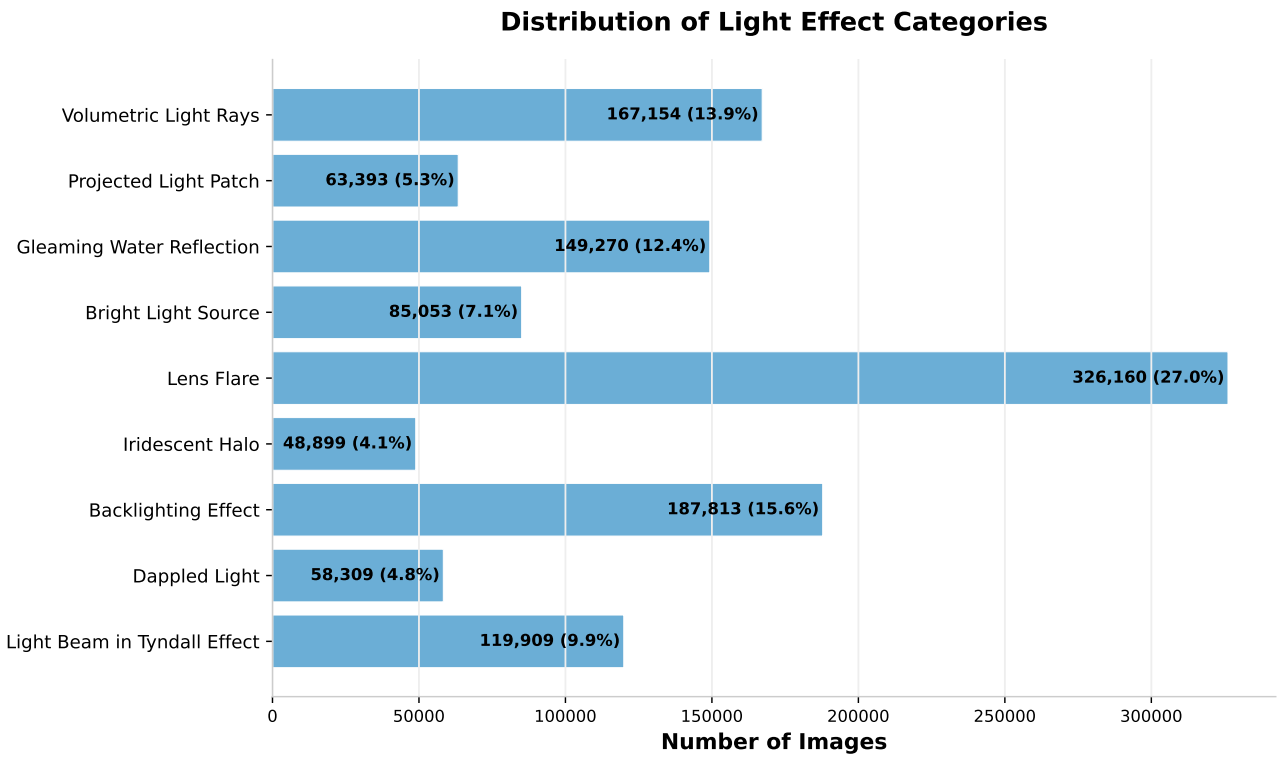

Figure 6. **Light effect distribution.** We categorize light effects into nine classes and perform a statistical analysis on our million-scale triplet dataset.

A.2. Image-content-light Triplets

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

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

A.3. The Usage of Vision Language Model

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

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

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

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.

B. Comparison with Indoor Scene Relighting Methods

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

- **Design motivation:** Our TransLight aims to enable the flexible transfer of geometric-structured lighting effects that are difficult to describe with simple text to other images, whereas physically-based indoor relighting methods primarily seek to achieve improved illumination consistency or coherence within a scene.
- **Application scenario:** Our TransLight focuses on portrait photography scenarios with broad practical applicability, aiming to enhance the aesthetic quality of images, whereas relighting methods such as LumiNet are primarily targeted at indoor scenes, such as bedrooms and living rooms.
- **Prediction targets:** Our TransLight can add customized light effects with distinct geometric structures to the input image. Moreover, our TransLight can also adjust the illumination state of the image by extracting light effects from the reference image to make the transfer result more realistic. For example, it can modify the shadow of the person's face

Table 4. **Quantitative results on the MIIW dataset.** Although our method is not specifically designed for indoor scenes, it remains highly competitive.

Methods	Labels	RMSE↓	SSIM↑
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)(σ=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
IAN (Zhu et al., 2022)	-	-	0.755
DIR (Yang et al., 2025)	-	-	0.797
Ours	-	0.212	0.627

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

according to the direction and position of the light effect, as shown in the example in the second sample of Figure 1. The target of previous physically-based methods is to render the input image to a new illumination status, such as converting an "off" desk lamp to an "on" state based on the state of the reference image, or modifying the direction of the shadow from left to right according to a specific Env.Map.

We follow the evaluation method of Latent Intrinsics(Zhang et al., 2024) and conduct zero-shot evaluation of our TransLight on the MIIW dataset(Murmann et al., 2019) to enable a thorough comparison. We randomly select an image and its 12 reference image for light effect transfer 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 4. We also provide visualization comparison on MIIW(Murmann et al., 2019) dataset in Figure 8. Although our method is not specifically designed for indoor scenes, it nevertheless demonstrates strong generalization capability.

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 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.

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

Table 5. **User study.** For "IC-Light (bg)", we use the reference image as the background input to the background-conditioned IC-Light. For "IC-Light (prompt)", we use the description derived from the reference image. For Nano Banana and Seedream 4.0, we provide both the input image and the reference image to the image editing model simultaneously and use appropriate prompt to complete the light effect transfer 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%**

in eliminating complex light effects. To intuitively illustrate the role of I_R, we present in Figure .10 the generated results of the 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.

D. User Study

We conduct a user study to evaluate the performance of our model on the light effect transfer task. In total, we collect voting responses from 15 participants. We asked the participants to vote respectively from the dimension of light effect consistency and content consistency, and select the results that meet the requirements in each samples. Good light effect consistency means that the light effect in the generated image closely matches that in the reference image. Good content consistency indicates that neither the subject nor the background in the generated image undergoes noticeable changes. For each sample, we present the generation results of five methods: prompt-conditioned IC-Light (Zhang et al., 2025), background-conditioned IC-Light, Nano Banana (Google DeepMind, 2025), Seedream 4.0 (Seedream et al., 2025), and our TransLight, displayed in random order. We show the result in Table 5. Our TransLight significantly outperforms all competing methods in both light effect consistency and content consistency.

E. 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 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 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.

F. 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 geometric-structured 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.

(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.

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.

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

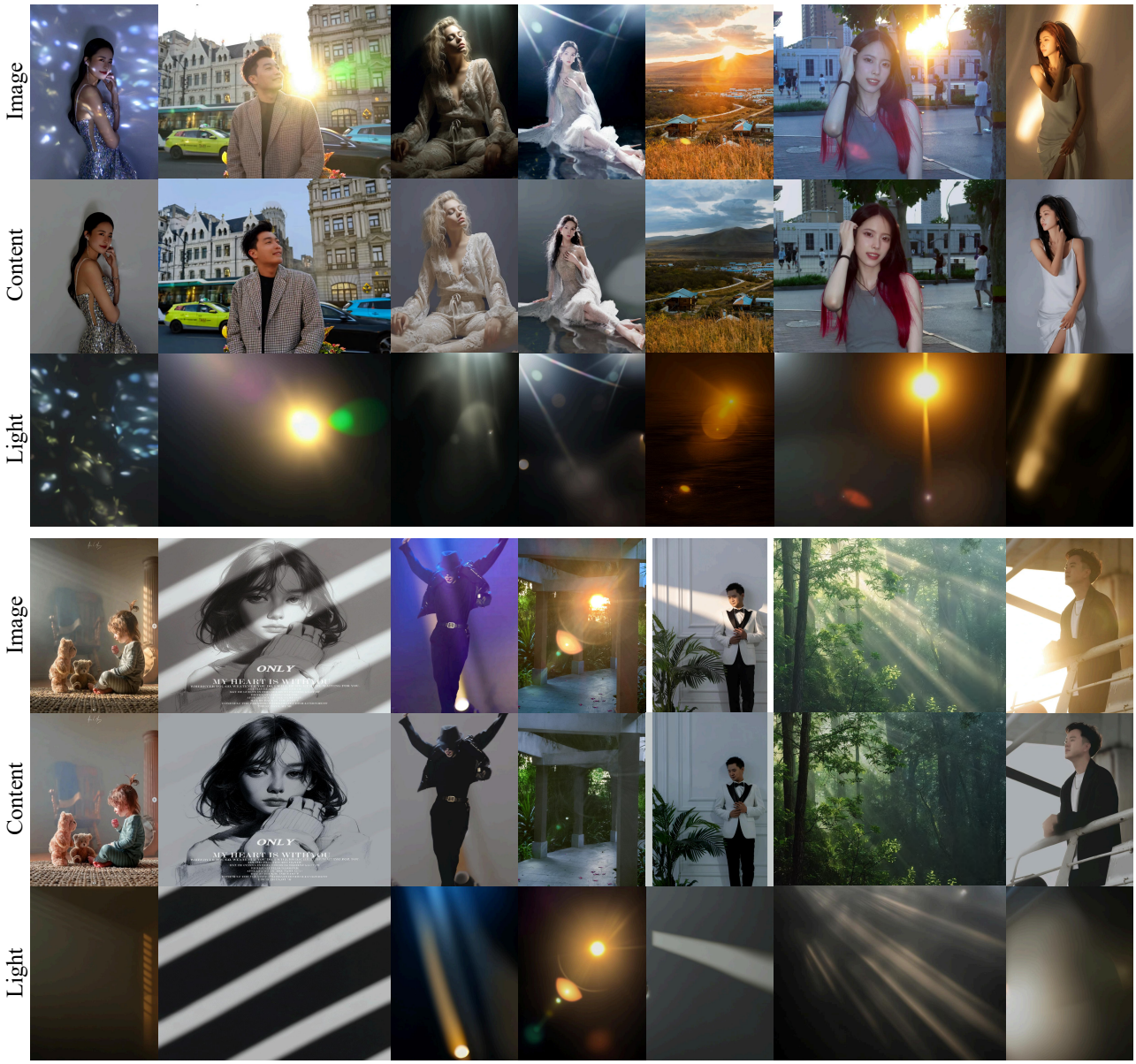

Figure 12. **Visualizations of our generative decoupling.** Our light removal model can eliminate light effects from images while preserving the underlying content unchanged as shown in the second line. Additionally, our light extraction model is capable of isolating the light effects from the image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

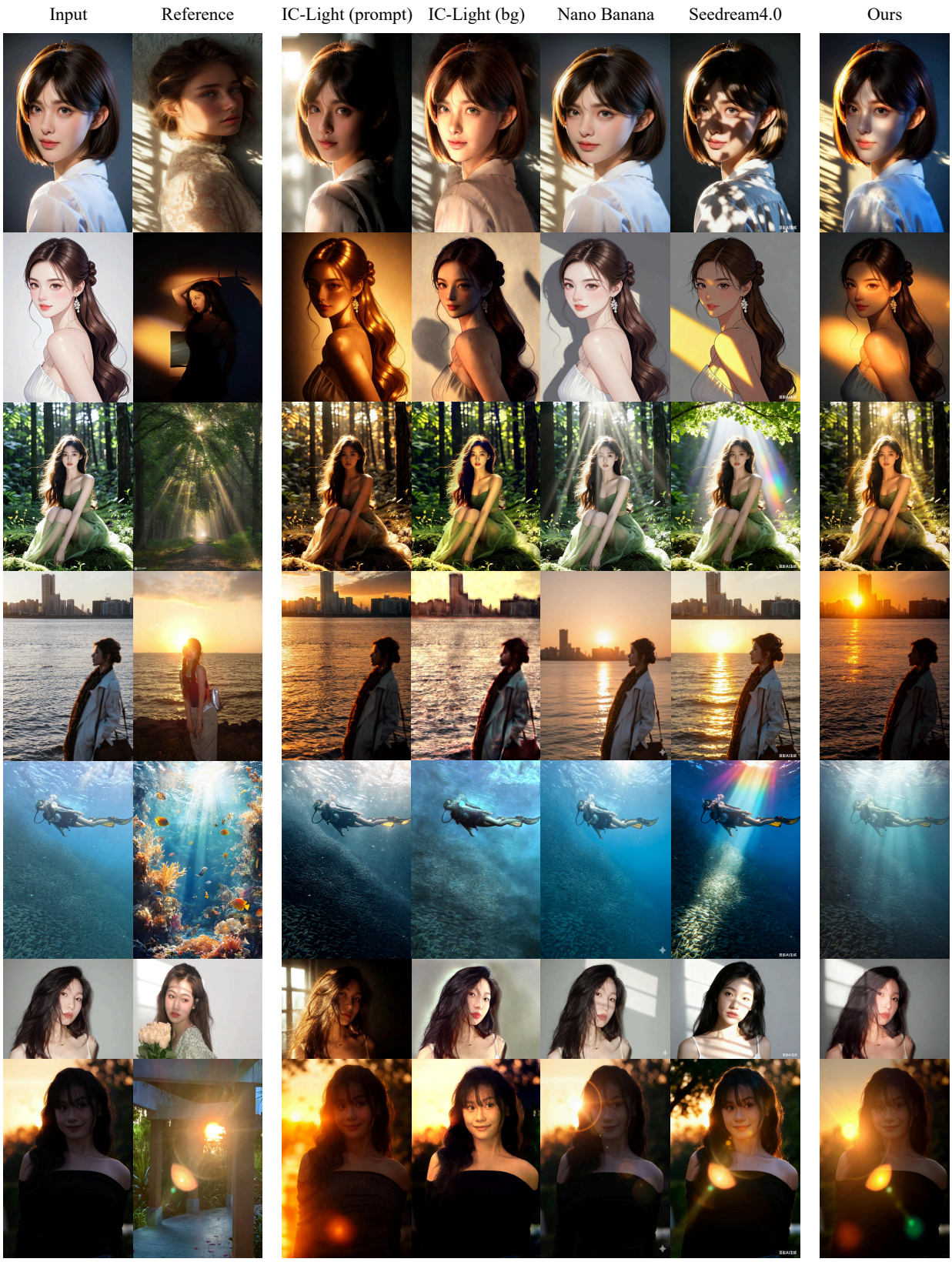

Figure 13. **Visualization comparison.** IC-Light (prompt) uses its prompt-conditioned model, and the prompt is derived from the reference image. IC-Light (bg) uses its background-conditioned model. For Nano Banana and Seedream 4.0, we provide both the input image and the reference image to the image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

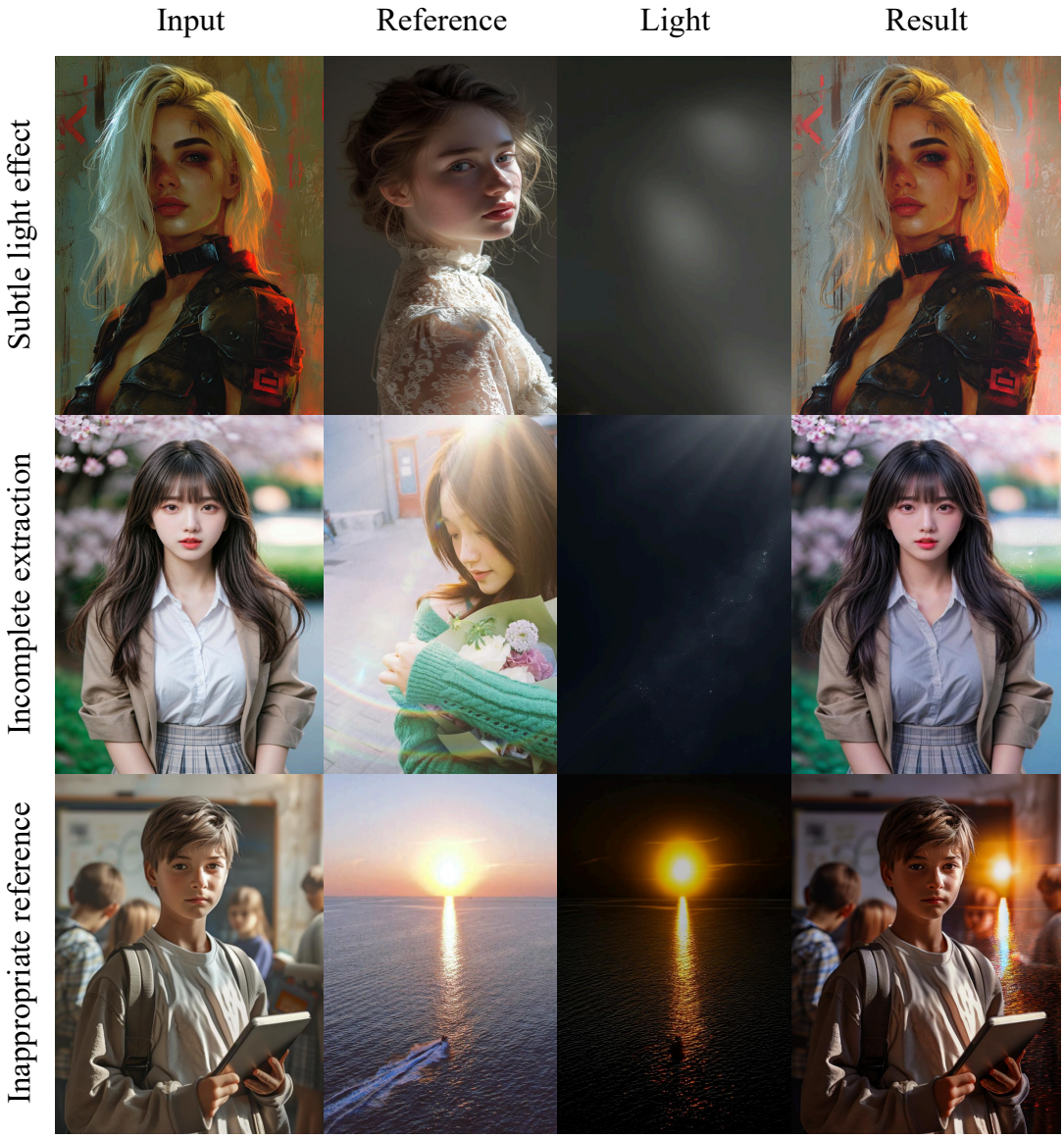

Figure 14. **Failure cases.** For subtle light effect case, the transfer degree is not obvious. For incomplete extraction case, since it is impossible to ensure 100% perfect disentanglement of light effects, the generated results may fall short of expectations. For inappropriate reference case, large 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.