Check for updates

ACM DIGITAL LIBRARY | Association for Computing Machinery | acm open

Latest updates: https://dl.acm.org/doi/10.1145/3770081

SURVEY

Learning-based Human Relighting: A Survey

SHUMIN ZHU, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Hong Kong

WAIKEUNG WONG, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Hong Kong

XINGXING ZOU, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Hong Kong

Open Access Support provided by:

The Hong Kong Polytechnic University

PDF Download
3770081.pdf
05 February 2026
Total Citations: 0
Total Downloads: 767

Published: 21 November 2025
Online AM: 02 October 2025
Accepted: 22 September 2025
Revised: 04 August 2025
Received: 16 January 2025

Citation in BibTeX format

Learning-based Human Relighting: A Survey

SHUMIN ZHU, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China and Laboratory for Artificial Intelligence in Design, Hong Kong, China
WAI KEUNG WONG, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China and Laboratory for Artificial Intelligence in Design, Hong Kong, China
XINGXING ZOU, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China and Laboratory for Artificial Intelligence in Design, Hong Kong, China

Human relighting refers to the process of adjusting the lighting effects on human subjects in digital images, 3D scenes, and videos to simulate various lighting scenarios, ultimately achieving realistic visual outcomes. This review provides a comprehensive examination of learning-based human relighting techniques. In doing so, it explores mainstream approaches while systematically documenting the development of related hardware and algorithms. Furthermore, it offers a detailed analysis of how learning-based human relighting methods have evolved across image-based, 3D-based, and video-based contexts. In addition, the review presents an in-depth evaluation of the respective advantages and limitations of these approaches, comparing them across key dimensions such as performance, robustness, and functional capabilities. Finally, it discusses current challenges and future research trends in learning-based human relighting. The goal of this review is to serve as a concise reference guide, offering practical support for both human relighting research and its real-world applications.

CCS Concepts: • **Computing methodologies → Appearance and texture representations**

Additional Key Words and Phrases: Learning-based human relighting, inverse rendering, image harmonization

ACM Reference Format:
Shumin Zhu, Wai Keung Wong, and Xingxing Zou. 2025. Learning-based Human Relighting: A Survey. *ACM Comput. Surv.* 58, 5, Article 133 (November 2025), 36 pages. https://doi.org/10.1145/3770081

1 Introduction

Human relighting seeks to adjust the lighting effects on human subjects in digital images, 3D scenes, and videos to align with diverse lighting scenarios, ultimately delivering realistic visual outcomes. With the rapid advancement of generative artificial intelligence, human relighting has garnered significant attention, emerging as a highly significant and compelling research area. Historically,

This work is partially supported by a grant from the Research Grants Council of the Hong Kong, SAR. (Project No. PolyU/RGC Project PolyU 25211424). This study is also partially supported by the Laboratory for Artificial Intelligence in Design (Project Code: Project 1.1), the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government.
Authors' Contact Information: Shumin Zhu, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China and Laboratory for Artificial Intelligence in Design, Hong Kong, China; e-mail: shumin.zhu@connect.polyu.hk; Wai Keung Wong, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China and Laboratory for Artificial Intelligence in Design, Hong Kong, China; e-mail: calvin.wong@polyu.edu.hk; Xingxing Zou (corresponding author), School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China and Laboratory for Artificial Intelligence in Design, Hong Kong, China; e-mail: xingxing.zou@polyu.edu.hk.

This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2025 Copyright held by the owner/author(s).
ACM 0360-0300/2025/11-ART133
https://doi.org/10.1145/3770081

technical advances in this field have primarily centered on the development of hardware and algorithms, where relighting equipment plays a pivotal role in capturing data and supporting the advancement of related algorithms.

For data capture, Debevec et al. [31] pioneered Light Stage technology, employing the "**one light at a time**" (**OLAT**) method to capture the 4D reflectance field of human faces. This innovation enables the rendering of faces under diverse lighting conditions and viewpoints. Later Light Stage iterations [32, 50] improved its performance by adding more controllable light sources. Wenger et al. [164] further advanced this technology by integrating high-speed cameras for recording actors' performances. Subsequently, a larger-scale Light Stage was developed [16] to capture the full-body reflectance fields of actors–this advancement facilitated the creation of realistic fill-lighting effects for post-production of their performances. Turning to human relighting algorithms, they can be categorized into three main types: image-based relighting methods, physics-based rendering methods, and learning-based relighting methods.

Image-based relighting methods [16, 31, 35, 82, 87, 113, 131, 137, 164] leverage existing images to adjust the lighting of a target object. Among these, OLAT-based methods [31] utilize a Light Stage to capture the reflectance field of a human face under dense incident lighting, then relight the face using light's linear additivity [13]. However, this approach requires capturing hundreds of OLAT images via complex and costly hardware setups, rendering it impractical for relighting single-object images in uncontrolled environments. Another key technique in image-based relighting is the quotient image method, proposed by Riklin-Raviv and Shashua. It uses two objects' reflectance properties and a small bootstrap dataset to render a new object under different lighting [119]. While extended to more applications [82, 104, 137], its effectiveness depends heavily on the quantity and quality of the bootstrap dataset's sample. Shifting to **physically based rendering** (**PBR**) models: Rooted in the rendering equation [66], PBR models simulate interactions between object materials, geometry, and light to create photorealistic lighting. For material modeling, empirical models-including the Lambert model [72], Phong model [11], and Blinn-Phong model [10] are widely adopted in relighting, due to their ease of parameter adjustment and implementation. A limitation, however, is that these models do not always adhere to the principle of energy conservation, leading to non-physical lighting artifacts. In contrast, the Cook-Torrance model [27] enables a more accurate relighting by incorporating microfacet theory and simulating the interaction between light and the microstructural features of object surfaces. Complementing this, the **Bidirectional Reflectance Distribution Function** (**BRDF**) [98], a core physical model to describe light-surface interactions, quantifies the light reflection process of opaque objects. These PBR models [79, 88, 110] generate highly realistic images and accurately replicate the interactions between light and material in the real world. However, their high computational demands and reliance on extensive hardware resources restrict them primarily to offline rendering scenarios, making them unsuitable for real-time applications. Additionally, physical models for specific materials are often unavailable, further restricting their practical application.

Deep learning has advanced rapidly, greatly expanding the scope of human relighting applications. These data-driven neural networks learn lighting patterns to effectively relight input images. Using implicit field relighting or inverse rendering techniques, learning-based relighting methods have been explored for 2D image relighting [91, 102, 106, 128, 139, 143, 163, 180, 194, 195], 3D model relighting [22, 124, 129, 140, 145, 150, 166, 169, 178, 191], and video relighting [14, 18, 44, 111, 183]. Figure 1 outlines the scope of this field. Notably, reconstructing relightable dynamic 3D representations from videos has become a research hotspot [78, 124, 141, 166, 167, 170, 177, 178, 190], opening new avenues for more realistic, high-quality visual effects. Despite these advancements, existing methods have limitations: the human body's geometric complexity and material diversity hinder accurate lighting simulation, and scenarios like online meetings require real-time relighting.

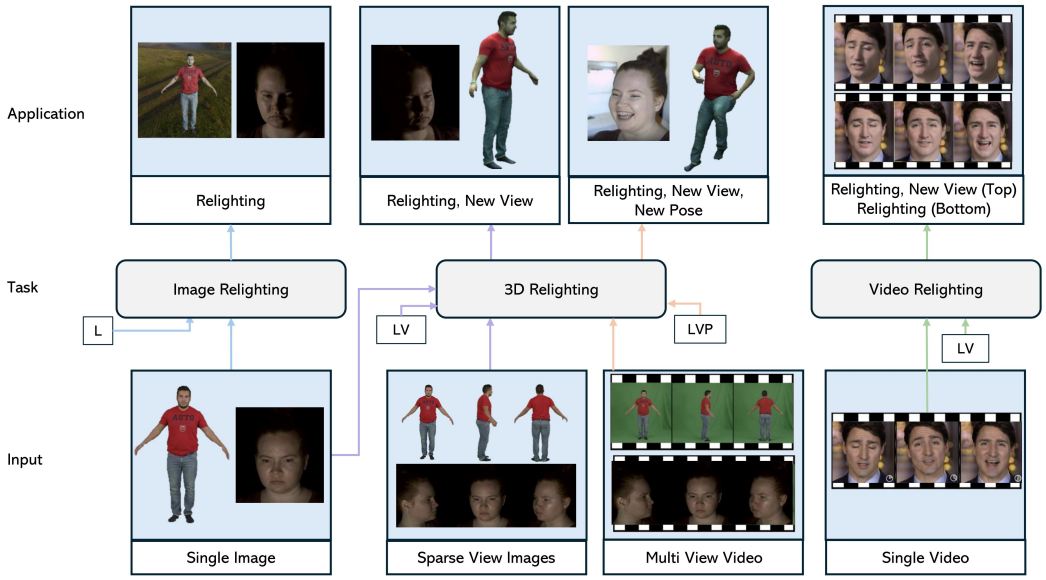

Fig. 1. Scope of learning-based human relighting in this work. Elements like input, task and applications are demonstrated. L denotes lighting information, V denotes view information, and P denotes pose information. Image source [14, 124, 167].

To support related research, there are several surveys [24, 37, 39]: Choudhury et al. [24] discuss image-based relighting technologies (e.g., reflection-based, basis function-based methods) for both human and object lighting, but lack comprehensive coverage of human-specific techniques. It is also outdated and lacks recent advances. Tewari et al. [149] summarize neural rendering trends, with only a subset focusing on human relighting. Einabadi et al. [37] explore deep learning in lighting estimation, relighting, and inverse rendering, but only part of it addresses human relighting. Likewise, the more recent survey by Garces et al. [39] focuses on intrinsic image decomposition and deep learning-based relighting, but also only partially covers human relighting. **In contrast to these existing works, this study offers a comprehensive overview of relevant techniques, with three key contributions**: (1). We systematically organize related technologies (hardware, image-based, physics-based, and learning-based relighting) and thoroughly discuss learning-based methods for 2D, 3D and video relighting. (2). We conduct in-depth analysis of the advantages and limitations of these methods and compare them in terms of performance and capability dimensions. (3). We summarize the core challenges of learning-based human relighting and outline key directions for future work. The remainder of this article is structured as follows: Section 2 introduces the background of human relighting; Section 3 details the foundational preparations for learning-based methods; Section 4 provides a detailed taxonomy of these learning-based methods; and finally, Section 5 presents the key challenges and future research directions. Section 6 summarizes this review. All license/permission information for reused images or tables from other papers in this paper can be found in Table 1 of the supplementary materials.

2 Background

2.1 Problem Definition

Human relighting entails adjusting the lighting effects on human subjects in digital images, videos, and 3D scenes to align with new lighting conditions, thereby achieving realistic visual outcomes. A

foundational formula for modeling light interaction in a scene is the rendering equation, proposed by Jim Kajiya in 1986 [66].

$$L_o(x, \omega_o) = L_e(x, \omega_o) + \int_{\Omega^+} f_r(x, \omega_i, \omega_o) L_i(x, \omega_i)(\omega_i \cdot n) d\omega_i, \tag{1}$$

where $L_o(x, \omega_o)$ represents the intensity of light reflected from point x on the surface of an object in the direction ω_o. Meanwhile, $L_e(x, \omega_o)$ denotes the surface's direct emission. The right part refers to the effect of incoming light interacting with the surface's reflectivity. $f_r(x, \omega_i, \omega_o)$ is the reflectance value for the incident light ω_i, which is irradiated on x and then emitted in the direction ω_o; $L_i(x, \omega_i)$ is the intensity of the incident light with the incident direction ω_i irradiating point x. n represents the normal vector at point x.

In this process, geometry, materials, and light are the core factors. **Geometry** refers to the shape of objects and their spatial positions within a scene. It is regarded as a critical parameter, which is represented via 3D vertices and edges, surface normals (N), or depth maps (D), all of which determine how objects appear in the scene. The construction of 3D models starts with 3D vertices, which form the corners of polygons. Edges connect these vertices, delineating the boundaries of surfaces and shaping the overall form of objects. Surface normals regulate how light interacts with object surfaces, directly influencing texture rendering, reflections, and shading effects. Depth maps illustrate the distance from the viewer to each point on the surface, which is crucial for creating perspective and depth effects. Geometry can be recovered using **Multi-View Stereo** (**MVS**) techniques [20, 28, 42, 49, 93, 95, 97, 122, 126, 158] and **Photometric Stereo** (**PS**) [41, 86, 91, 92, 181] techniques, or a combination of these two technologies [195].

Material dictates how an object's surface interacts with light, including key phenomena like reflection, refraction, and absorption. For physics-based methods, researchers have developed various models to simulate light behavior on different materials. The Lambertian model [72] is an idealized diffuse reflection model that assumes surfaces reflect light uniformly in all directions, independent of the viewing angle. However, it performs poorly for materials like glass or metal since it ignores specular reflection. To address this limitation, Phong [11] integrated ambient light, diffuse reflection, and specular reflection to compute the overall surface color of an object. Its enhanced variant, the Blinn-Phong model [10], employs half-vector calculations to compute specular highlights, resulting in both smoother rendering effects and greater computational efficiency. However, these models still don't fully capture fine material details, limiting their ability to simulate many real-world materials. Thus, the Cook-Torrance model [27] was proposed: it includes material details to simulate microfacet-based surfaces, better capturing visual differences between surfaces of different roughness.

The classic BRDF [98] quantifies light reflection from non-transparent objects, it describes the probability of light reflecting in a specific direction as a function of incident angle. To extend this framework beyond reflection, the **Bidirectional Scattering Distribution Function** (**BSDF**) [110] integrates the **Bidirectional Transmission Distribution Function** (**BTDF**) which models light transmission, thus covering both reflection and transmission. It can also be expanded to the **Bidirectional Scattering-Surface Distribution Function** (**BSSDF**), which incorporates surface scattering effects. Building on the BSDF, the Microfacet BRDF [155] focuses on microfacet models for rough surface refraction. It highlights the key role of sampling methods in microfacet simulations, allowing accurate modeling of more materials (e.g., skin, marble, paint). The **Spatially Varying Bidirectional Reflectance Distribution Function** (**svBRDF**) [88] addresses spatial variations in reflective properties across a surface, supporting more precise modeling of complex, non-uniform surfaces. In contrast, the Disney BRDF [12] adopts more robust parameterization principles and optimizes the simulation of material characteristics, further refining the realism of

light-material interactions. In summary, these material models have evolved to simulate realistic light-material interactions. They not only boost visual realism in computer graphics but also lay a foundation for deep learning-based relighting methods.

Light is the core element of the relighting process, directly dictating the visual quality of object surfaces and the overall realism of the scene. Specifically, incident light in real-world scenes can be categorized into two primary types: *Far-field illumination*: Light emitted or reflected by distant objects; *Near-field illumination*: Light emitted or reflected by surfaces in close proximity to the observation point. In computer graphics, far-field illumination is typically simulated via environmental map lighting–a technique that involves mapping **high-dynamic-range** (**HDR**) images onto a hypothetical infinite sphere or cube enclosing the scene. In contrast, near-field illumination is simulated by modeling the light reflected from local surfaces, where this reflected light is itself influenced by far-field illumination. Currently, common approaches for simulating near-field illumination include environment mapping, **Spherical Harmonics** (**SH**) [113], and **Spherical Gaussians** (**SG**) [156].

2.2 Relighting Devices

Cameras and lighting equipment are essential for capturing data required for relighting. As demonstrated in Figure 1 of Debevec et al. [30], the successful integration of these two components enables the generation of controllable lighting to illuminate target objects. Debevec [31] developed a Light Stage equipped with a spotlight (with adjustable azimuth and elevation angles) paired with two cameras. Using the OLAT method, this system captures the 4D reflectance field of human faces–supporting the rendering of facial appearances under novel lighting conditions and viewpoints. To accelerate the reflectance field capture process, Light Stage 2 [50] integrates 30 strobe lights mounted on a semicircular arm. Building on this, Light Stage 3 [32] incorporates 156 controllable RGB LED light sources, enabling it to replicate the **High Dynamic Range Imaging** (**HDRI**) lighting conditions experienced by real actors. Notably, the captures from these early Light Stage iterations were limited to static content. A key technological advancement came with the advent of high-brightness white LEDs, which enabled time-division multiplexing lighting and high-speed video capture. This breakthrough allowed for dynamic recording of actors' surface reflectivity, laying the groundwork for post-production relighting of their performances [164]. Light Stage 5 [164] built on the design of Light Stage 3, incorporating an array of 156 white LEDs, further supporting the post-production relighting of actors' performances. To address the need for illuminating the entire moving human body, Light Stage 6 [16] was developed. By integrating view interpolation, light field rendering, and image-based relighting techniques, this system enables actors to be placed in new virtual environments with dynamic lighting and camera movements.

In addition to the traditional Light Stage systems, Sevastopolsky et al. [129] developed a more user-friendly setup that uses a smartphone camera to capture relighting data–functionally replacing the need for complex Light Stage hardware. Building on the trend of simplifying capture setups, Wang et al. [161] further proposed that the sun itself can serve as a unique, natural "light stage" for relighting data acquisition. Similarly, Sengupta et al. [127] discussed how specific real-world light sources, such as dynamic images displayed on desktop monitors, can similarly function as makeshift light stages for relighting purposes.

Light Stage systems and related technologies provide robust data support for developing relighting algorithms. By capturing precise reflectance responses of human subjects under different lighting, these technologies lay a solid data foundation for later generating high-fidelity, high-resolution geometric and material info. This data includes, but isn't limited to, vertex coordinates, surface normals, depth maps, and reflectivity properties, all key for simulating real-world lighting effects. Using Light Stage tech lets researchers accurately reconstruct human subjects' true reflective

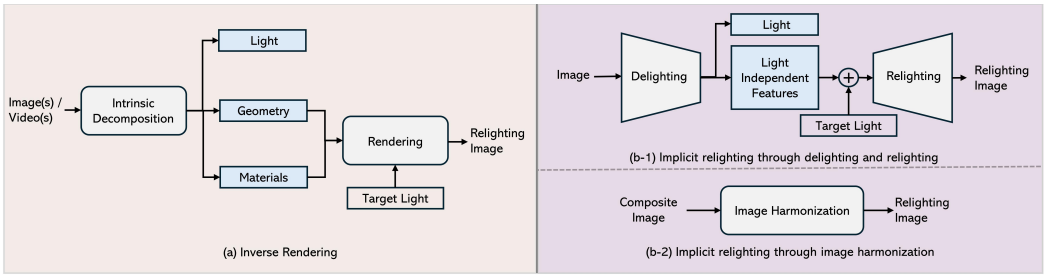

Fig. 2. Structures for learning-based human relighting.

properties under different light sources, thus reproducing photorealistic facial lighting in digital environments. Furthermore, these data can be leveraged to optimize and train deep learning-based human relighting algorithms, further advancing the performance and realism of such methods.

3 Preliminaries

Learning-based relighting algorithms, which build on traditional relighting algorithms and corresponding datasets, have expanded the application scope of human relighting, covering 2D image relighting, 3D model relighting, and video relighting. To formalize the task of learning-based human relighting within a unified mathematical framework, an ideal relighting network aims at learning the mapping: $\hat{I}_t = \psi((I_i, L_t, \theta)$. In this equation, I_i denotes an input image or a frame of the video or a view of the 3D model. L_t denotes the lighting of the target. θ represents the learnable parameter in the network. $\hat{I}_t$ is the predicted result. A well-designed relighting network effectively preserves the shape and texture details of the input while capturing the lighting information from the target source. Consequently, it can generate new results with realistic lighting effects. Formally speaking, with I_i, L_t denoted before the problem can be shown as

$$\begin{aligned} \theta_\psi^* &= \arg\min_{\theta_\psi} \ell(\psi((I_i, L_t), \theta_\psi), I_t) \\ &= \arg\min_{\theta_\psi} \ell(\hat{I}_t, I_t)' \end{aligned} \tag{2}$$

where the $\psi(\cdot)$ is the neural network for the transfer and ℓ denotes the spatial distance between the images. And θ_ψ represents the learnable parameters of neural network ψ.

3.1 Categories of the Approaches

The methods widely used by learning-based methods to illuminate humans can be divided into inverse rendering and implicit relighting. The inverse rendering method explicitly estimates the geometry, material and lighting from the observed input, then re-renders the result under the new lighting based on the estimated results. In contrast, the implicit relighting method does not require strict decomposition of material geometry and lighting information. Instead, it directly estimates the relighted result based on the input and the new lighting. These two structures are illustrated in Figure 2.

3.1.1 Inverse Rendering. The core principle of inverse rendering is to infer a scene's geometry, materials, lighting, and other key rendering parameters from known image data (e.g., 2D images or videos). By contrast, in traditional forward rendering, the scene's geometry, material, and lighting are predefined. The pixel colors of the final image on the image plane are then computed via algorithms such as ray tracing. Inverse rendering operates in the reverse manner. The inverse rendering method explicitly estimates the geometry, material and lighting from the observed input, and then re-renders the result under the new lighting based on the estimated results [8, 55, 61, 92, 96, 102, 145, 173]. In

learning-based relighting, the objective function of inverse rendering is shown in Equation (3).

$$\theta_{\psi}^{*} = \arg\min_{\theta_{\psi}} \ell(\hat{I}_t, I_t) + \ell(\hat{G}_t, G_t) + \ell(\hat{M}_t, M_t) + \ell(\hat{L}_i, L_i), \tag{3}$$

where the $\hat{G}_t$ denoted predicted geometry, M_t denotes predicted material and L_i denotes the predicted illumination of the input. And I_t, M_t, and L_t represent their ground truth value, respectively. Geometry G refers to a normal map, a depth map, and other related information. Material M refers to the albedo map, specular reflection map, specular roughness map and other related information.

Learning-based inverse rendering offers notable advantages, with high computational efficiency. Once trained, it can quickly process new input images and generate approximate results in real time, providing robust support for scenarios demanding real-time feedback. However, this method is highly dependent on intrinsic images. Specifically, the quality of intrinsic image estimation directly impacts the final rendering result. An intrinsic image contains the fundamental information of scene objects, free from distortion by complex lighting and other interferences. It acts as the foundation for inverse rendering to infer various scene elements' if intrinsic image estimation deviates, the final rendering can easily differ significantly from the real scene. Additionally, learning-based inverse rendering places strict demands on the quality and quantity of training data. It requires a comprehensive, detailed dataset with annotations covering key aspects (e.g., geometry, materials, lighting). Only then can the model fully learn the complex correlation patterns between image features and scene intrinsic information. In summary, while learning-based inverse rendering benefits from learning models' efficiency, it is limited by two key issues: reliance on intrinsic images and strict training data requirements. For practical application and development, ongoing exploration of optimization paths is needed to overcome these bottlenecks, thereby expanding its application scope and effectiveness.

3.1.2 Implicit Relighting. The implicit relighting method bypasses the tedious intrinsic image modeling process in inverse rendering (e.g., geometry, materials, lighting). Instead, it uses the strong learning and fitting abilities of deep neural networks to automatically extract the underlying patterns of light and shadow changes from large volumes of image data. It understands and adjusts lighting effects in a more abstract, nuanced way, embedding the complex mapping between scene images and target relighting results into the network's parameters.

One category of implicit relighting methods follows a "delight first, then relight" workflow [58, 71, 80, 139, 183, 184, 194]. Delighting seeks to remove lighting information from the original image and recover the person's light-invariant features. Relighting then applies suitable lighting effects to the person, based on newly configured lighting conditions and the light-shadow transformation rules learned by the neural network. The objective function is given by Equation (4):

$$\theta_{\psi}^{*} = \arg\min_{\theta_{\psi}} \ell(\hat{I}_t, I_t) + \ell(\hat{L}_i, L_i). \tag{4}$$

This method only requires a dataset containing human images under different lighting conditions and their corresponding lighting information, no additional geometric or material annotations are needed.

Another category of implicit relighting methods adopts an image harmonization approach [21, 25, 26, 45, 46, 62, 84, 118, 147, 151, 157, 168, 180, 196]. These methods treat the human subject as the foreground and place it in a background environment with lighting. The goal of image harmonization is to make the foreground human blend naturally into the background, achieving a cohesive unity between the two. Notably, image harmonization-based methods involve relatively simple dataset annotation and optimization processes. The dataset only needs to include image pairs: foreground and background images (before blending), and the final blended result images.

Table 1. List of Public Human Datasets Suitable for Relighting

Dataset	Year	Type	Scale	Resolution	Region	Lighting	Light Source	S or R
YaleB [40]	2001	Image	4050 images, from 10 people	640 × 480	Portrait	45	Point Light	R
Multi-PIE [43]	2010	Image	750,000+ images, from 337 people	3072 × 2048	Portrait	19	Point Light	R
DPR [194]	2019	Image	138, 135 images with 27,627 items	1024× 1024	Portrait	5	illumination prior	S
Wang et al. [163]	2020	Image	272,000 image groups from 80 people	4096 × 4096	Portrait	391	Environment Light	S
ICT-3DRFE [138]	2011	3D model	345 3D models with textures from 23 people	1296 × 1944	Portrait	9	Point Light and Environment Light	R
UltraStage [195]	2023	3D model	100 file groups with different people	7680 × 4320	Full Body	2965	Environment Light	S
Ego3DHands [77]	2021	Video and Image	55,000 images and 110 videos (500 frames/per)	≥ 256 × 256	Hand	100	Environment Light	S
Re:InterHand [94]	2024	3D model and Image	400 3D models and 1.5M images from 10 people	4096 × 2668	Hand	2144	Environment Light	S
iHarmony4 [26]	2020	Image	73146 image pairs from 4 sub-dataset	≈ 640 × 480	-	-	-	S

The abbreviation "S" refers to "Synthetic", and "R" refers to "Real".

3.2 Deep Neural Network Architecture

Neural network architectures used in relighting vary by task. For image relighting tasks, three types are widely adopted to generate relit images: encoder-decoder architectures (based on convolutional neural networks, CNNs) [2, 74], UNet [121], and UNet-like networks [194]. The encoder-decoder architecture excels at capturing global features and multi-level semantic information. Through end-to-end training, it directly learns the input-output image mapping, thus offering high flexibility. Building on this architecture, methods for portrait and full-body image relighting have been developed [55, 56, 71, 75, 128, 134, 135, 139, 143, 163, 173]. UNet builds on the encoder-decoder architecture, using skip connections to connect the encoder's feature maps with the decoder's corresponding layers. This lets the decoder leverage both high-level semantic info and low-level details, better preserving the image's edges and textures. UNet-like networks enhance UNet with deeper architectures or more complex skip connections, improving the efficiency of feature extraction and utilization. Owning to UNet and UNet-like networks' effectiveness in preserving detailed features, a large part of relighting methods were developed based on them [69, 91, 102, 118, 171, 194, 197].

For 3D reconstruction and relighting tasks, methods primarily rely on the neural volume rendering framework. Mildenhall et al. [93] proposed **neural radiance fields** (**NeRF**)–a method where the learned radiance fields can then be rendered via traditional volume rendering techniques. In this work, an object's 3D appearance is encoded into a neural implicit field realized via a **multi-layer perceptron** (**MLP**). This MLP takes two inputs: the 3D coordinate $x \in R^3$ (via a sinusoidal-based position encoding mapping [93, 148]) and the viewing direction $d \in S^2$. It outputs the volume density $\sigma \in R^+$ and the view-dependent color $c \in R^3$. To render an image, the pixel color C is accumulated along each camera ray $r(t) = \mathbf{o} + t\mathbf{d}$ as follows:

$$\mathbf{C}(\mathbf{r},\ \mathbf{d}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \tag{5}$$

where $T(t) = exp(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds)$ and bounds t_n and t_f. Compared with surface-based rendering, volumetric rendering more naturally handles translucent materials and complex geometric regions (e.g., thin structures). Most relighting methods [61, 125, 169] rely on this framework to decompose the materials required for relighting (e.g., albedo maps, specular components) and geometry (e.g., normal maps), which are further used for relighting.

3.3 Datasets

Learning-based relighting methods use task-specific datasets for training and validation, tailored to process requirements. Table 1 lists publicly available human-centric relighting datasets, while Figure 3 shows visual samples of mainstream datasets. Most of these datasets focus on portrait relighting, including YaleB [40], Multi-PIE [43], DPR [194], Wang et al. [163], and ICT-3DRFE [138]. YaleB [40] were captured in controlled lab environments. It contains 4,050 face images of 10 individuals–each captured under 45 fixed lighting conditions and 9 facial poses. Learning-based

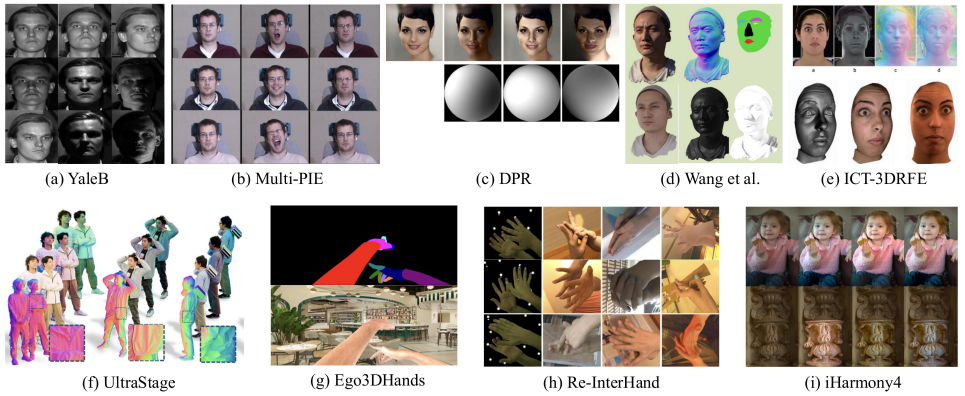

Fig. 3. Example images from public human dataset [26, 40, 43, 77, 94, 138, 163, 194, 195].

relighting methods [1, 152] have used it to validate single-image relighting performance. Later, a larger dataset, Multi-PIE [43], was developed. It includes over 750,000 images of 337 subjects captured across 15 views, 19 lighting conditions, and diverse expressions. Similarly, it has been used to assess single-image relighting performance [56, 106, 109]. To address the limitation that existing portrait datasets lack support for multi-reflection channel modeling, Wang et al. [163] proposed a large-scale dataset that includes ground truths for the channels (including facial albedo, geometry, and two lighting components), enabling supervised training of deep neural networks. ICT-3DRFE [138] is a re-lightable facial 3D dataset for 3D graphics. It comprises 3D models of 23 subjects with 15 distinct expressions–supporting photorealistic rendering with lighting information. Several synthetic datasets have also been proposed to support relighting algorithm research. For example, the DPR dataset [194] is a large-scale, high-resolution collection of "in-the-wild" portrait relighting images. Built on the CelebA-HQ dataset [81], it was expanded to 138,135 relit images by applying 5 random lighting conditions (from a lighting prior dataset [36]) to each of 30,000 original face images. This dataset has been used to train portrait relighting models [56, 194].

For full-body relighting, the only available dataset is UltraStage [195]. Though it provides NeRF parameters, normal maps, and multi-view images, it currently only includes 100 samples. Recently, relightable hand datasets have been proposed to advance hand relighting research, such as Ego3DHands [77] and the latest Re: InterHand [94]: Ego3DHands [77] is a large-scale synthetic dataset designed for training neural networks to estimate hands' global 3D joint positions from monocular RGB input images. The latest dataset, Re: InterHand [94], is a 3D interactive hand dataset designed for relighting. Its goal is to deliver diverse, realistic image appearances and a large, varied set of ground-truth 3D poses. Additionally, iHarmony4 [26] was introduced, comprising 73,146 image pairs across four sub-datasets. It is designed for image harmonization tasks, including (but not limited to) fusing human subjects with background scenes.

3.4 Evaluation Method

To evaluate relighting methods, the relit images generated by models are typically compared against real images. Evaluation approaches include standard evaluation metrics and user studies.

3.4.1 Standard Evaluation Metrics. Currently, multiple standard evaluation metrics have been proposed to assess the quality of relighting results from different perspectives: **(a) Overall Quality Measurement**: **Mean Absolute Error** (**MAE**), **Mean Squared Error** (**MSE**), **Scale-Invariant Mean Squared Error** (**Si-MSE**) [3], as well as **Root Mean Square Error** (**RMSE**) are commonly

used to measure the overall quality of relit images. **(b) Fidelity Measurement**: **Peak Signal-to-Noise Ratio** (**PSNR**) [54], **Structural Similarity Index** (**SSIM**) [162], SSIM difference metric $DSSIM = \frac{1-SSIM}{2}$ are used to assess fidelity. Moreover, relighting methods based on generative models also employ **Fréchet Inception Distance** (**FID**) [52] and **Kernel Inception Distance** (**KID**) [9] to evaluate the image fidelity. **(c) Perceptual Quality Measurement**: **Learned Perceptual Image Patch Similarity** (**LPIPS**) [185] and **Natural Image Quality Evaluator** (**NIQE**) [182] are used to assess the perceptual quality of images. **(d) Identity Preservation Measurement**: In portrait and full-body relighting tasks, the identity similarity measurement (based on the cosine similarity between LightCNN [165] or ArcFace [34] features) is used to evaluate the degree of identity preservation. **(e) Semantic-level Video Coherence**: Recent text-to-video methods [189] leverage the similarity of frame-wise CLIP image embeddings to evaluate semantic-level video coherence.

3.4.2 User Study. While there is a wealth of evaluation metrics for relighting quality, user studies remain critical for validating the perceptual quality of relighting results [19, 69, 144]. For instance, SwitchLight [69] conducted a user study to evaluate three key aspects: (1) the consistency of lighting with the environment map, (2) the preservation of facial details, and (3) the preservation of the original identity. Tajima et al. [144] asked participants to select the most natural relighting effect for each input image, with a focus on facial highlights and shadows on the arms and clothing. In conclusion, while existing evaluation metrics can measure the technical quality of the relighting results, user studies are equally essential to assess their perceptual quality and user experience.

4 Learning-based Human Relighting

Traditional rendering techniques focus on reconstructing scenes and simulating light behavior to generate images. In contrast, deep learning-based methods approach the same task from a statistical perspective. As data-driven approaches, deep learning methods leverage large-scale real-world or synthetic datasets to train models that learn patterns in images. Unlike traditional rendering techniques, often limited to a small set of images, deep generative models can learn data distributions from extensive images.

4.1 Single Image Relighting

4.1.1 Inverse Rendering. Inverse rendering approaches have shown that explicitly incorporating physical processes into neural architectures offers advantages for robust estimation with limited data and enhanced generalization [38, 60, 69, 71, 96, 102, 143, 163]. Such approaches first decompose an image into its intrinsic properties, then generate a relit target image by combining these properties with a predefined physical model. As shown in Figure 2(a), inverse rendering first uses multiple networks to decompose intrinsic image components, including geometric details (e.g., normal map N) and material properties (e.g., albedo map A, specular reflection map S), before rendering these components into a relit image with the target lighting.

The key advancement of this category of methods lies in using appropriate network models to parse images into more comprehensive, detailed geometric and material representations, then leveraging more accurate rendering models to improve relighting results. Representative cases include: SfSNet [128] uses residual blocks to produce normal and albedo features, then combines these with image features to estimate lighting–drawing inspiration from the Lambertian rendering model. Total Relighting [102] further estimates specular light maps via Phong reflectance lobes, enabling non-Lambertian effects (e.g., specular highlights). Finally, it fuses the derived diffuse light map, specular light map, albedo, and normal via a U-Net to output the relit target image. SwitchLight [69] Predicts intrinsic information (normal, illumination, albedo, specular reflection,

roughness, and more) using multiple U-Nets, then integrates the Cook-Torrance physical model to achieve more realistic lighting. While inverse rendering-integrated networks deliver realistic relighting effects, they face two critical limitations: (1). High reliance on ground-truth intrinsic data: These methods require ground-truth intrinsic images to supervise network training, yet acquiring such data incurs significant costs. (2). Vulnerability to inaccurate component estimation: The scheme relies on the accuracy of estimated intrinsic components, but these are hard to predict accurately in real-world scenarios. For example, if an input image includes cast shadows requiring removal, these methods often leave shadow residuals in the predicted albedo map, causing artifacts in the final output. Estimating geometry for fine details (e.g., hair, ears) is highly challenging; such regions are often excluded from relighting pipelines, leading to unrealistic final composites.

4.1.2 Implicit Relighting. Using a single network to achieve human image relighting mainly includes two parts: delighting and relighting [71, 80, 139, 194]. Figure 2(b-1) shows the general process of this type of algorithm. For a given input image I, the encoder first separates the image into light-independent features F and estimated illumination L_e. Then, the decoder takes F and the target illumination information L_t as input and outputs the target re-lit image I^*. In supervised learning, the input image I, the target image I_t, the input image L_i, and the target image L_t are often known. During network training, the illumination information of the input image L_i and the target image I_t are used as ground truth to calculate errors with the network's estimated illumination information L_e and the output image I^*, thereby supervising the network's training. Utilizing off-the-self estimators, DiFaReLi [109] based on a conditional DDIM that produces photorealistic shading without requiring accurate intrinsic decomposition or 3D and lighting ground truth.

Image Harmonization aims at compositing a subject into a new background, adjusting its lighting and color to ensure harmony with the background scene. Figure 2(b-2) illustrates this progress. In deep learning, this task is typically formulated as an end-to-end image-to-image translation problem [21, 25, 26, 45, 46, 62, 84, 118, 147, 151, 157, 168, 196], where the network is trained to predict a harmonized image from the input composite (foreground + background). Let H_f be the input foreground human and S_b be the input target background scene. Let $\hat{I}_t$ be the output of the network $\psi((H_f, S_b, \theta)$. The objective function of image harmonization is shown in Equation (6).

$$\begin{aligned} \theta_\psi^* &= \arg\min_{\theta_\psi} \ell(\psi((H_f, S_b), \theta_\psi), I_t) \\ &= \arg\min_{\theta_\psi} \ell(\hat{I}_t, I_t). \end{aligned} \tag{6}$$

Pixel-aligned datasets are created by modifying the foreground color and brightness in real images using either predesigned [26] or learned [99] augmentations. These methods primarily focus on changing the overall color, brightness, and lighting direction of the foreground. However, many harmonization-based methods only prioritize the foreground's overall color and brightness–overlooking lighting characteristics and leaving key foreground illumination effects (e.g., lighting direction, shadows) unchanged. This can lead to unnatural composites when the background has distinct lighting conditions. To generate realistic lighting-aware composites within a straightforward harmonization framework, Relightful Harmonization [118] introduces a lighting-aware diffusion model, designed to use any background image to seamlessly harmonize complex lighting effects for the foreground portrait. Building on large-scale text-to-image models [107, 120], IC-Light [180] combines text-generated backgrounds with input foreground images to generate more realistic and creative relighting effects.

Despite significant advancements in image synthesis technology, enhancing the realism and naturalness of synthesized images, particularly in complex scenes, continues to pose a challenge.

In this context, core challenges in image synthesis fall into two primary categories: foreground adaptability issues and visual harmony issues. The former encompasses geometric consistency problems, such as mismatches in the size, position, and geometric angle of the foreground relative to the background and appearance consistency issues, for example, mutual occlusion between the foreground and background as well as blurred edge details. The latter, by contrast, involves color consistency mismatches between the foreground and background, variations in contrast and saturation between these two components, and the lack of matching shadows for the foreground (i.e., no shadows corresponding to the background lighting). Addressing these challenges is critical for advancing the field of image synthesis and producing more authentic synthetic images.

4.2 3D Relighting

3D reconstruction and relighting play a crucial role in human digitization, particularly in enabling realistic rendering across diverse virtual environments. These technologies have wide-ranging applications in AR/VR [85, 100], holographic communication [73, 188], and the film and game industries [29].

Video-based 3D Relighting: Sevastopolsky et al. [129] propose a 3D head portrait reconstruction system that analyzes flickering flash frame sequences from smartphone videos. The system combines **Structure-from-Motion** (**SfM**) technology with multi-view denoising to reconstruct point clouds, then trains a deep rendering network to enable high-quality image synthesis across diverse viewpoints and lighting conditions. To our knowledge, this is the only existing work that uses a single individual's video sequence as input to reconstruct their 3D facial expressions and render the portrait from multiple viewpoints under new lighting conditions. Separately, Wang et al. [161] simultaneously reconstructs geometry, skin reflectance, lighting, and camera pose from a selfie video of a person rotating under sunlight. While the aforementioned face modeling methods derived from videos are static, recent research has turned attention to animated, re-illuminated 3D human modeling, a more dynamic and practical direction. Saito et al. [124] were the first to extend 3D Gaussian Splatting [68] to animatable, relightable facial avatar modeling, laying the groundwork for dynamic facial reconstruction with lighting adaptability.

For 3D-aware full-body human video relighting, the **Skinned Multi-Person Linear Model** (**SMPL**) [83] and its extension SMPL-X [103] are widely used for 3D pose and shape estimation. Their strength in accurately simulating human posture and body shape changes makes them foundational for many state-of-the-art methods [22, 78, 141, 166, 170, 177, 178, 190], which fall into two main branches: those integrating NeRF [22, 141, 166, 170, 178] and those combining Gaussian Splatting [177, 190]. NeRF is a technique that learns continuous volumetric representations from sparse images to synthesize high-quality new views. Methods integrating NeRF include: Relighting4D [22], the first to achieve free-viewpoint relighting of dynamic characters using only video. To do this, It uses a 4D neural field (based on SMPL/SMPL-X) to represent dynamic characters, separates geometric and reflectance information, and employs a physically based renderer. Sun et al. [141], who used neural fields to model geometry and materials, enabling dynamic body mapping via posture-driven deformation. They adopt Disney BRDF, spherical Gaussians, and physical rendering for relighting under new poses/lighting. Lin et al. [78] introduced a reversible neural network for more accurate simulation of non-rigid motion and body geometry reconstruction. By estimating posture-aware light visibility and separating material/lighting parameters, they achieved rendering under arbitrary poses, lighting, and views via differentiable rendering. Xu et al. [170] proposed a method to generate a relightable avatar using a hierarchical distance query algorithm for approximating world space distances, sphere tracing for light visibility calculations, the Microfacet BRDF model for material representation, and integrating all properties for physically based rendering. Zhang et al. [178] proposed a method for integrating real human bodies with virtual

lighting environments using multi-view video analysis, enabling the re-illumination of human bodies in virtual settings under new postures through precise pose mapping and joint estimation of multiple neural fields. Xiao et al. [166] introduced the NECA, a framework for learning customizable neural avatars, supporting realistic rendering across poses/views/lighting and editing of shape, texture, and shadows. Gaussian Splatting models human geometry by fusing Gaussian distributions with body models, excelling at capturing wrinkles, loose clothing deformations, and complex motions [177, 190]. Key examples are: Zhan et al. [177] proposed creating relightable and animated avatars from multi-view or monocular videos using Gaussian Splatting, which enables interactive rendering under various viewpoints, postures, and lighting conditions for digital human creation and virtual reality applications. Zhao et al. [190] utilized 2D Gaussian Splatting and linear blended skinning to model dressed human avatars in canonical space, combined with physical rendering and image-based lighting for efficient light calculation, and employed occlusion approximation and progressive training to decouple material lighting while reconstructing geometry.

While current 3D-aware full-body human relighting methods deliver promising results, they still have notable limitations: lengthy training durations, insufficient fidelity in capturing facial expressions, and room for improvement in rendering humans with loose clothing. Additionally, these methods typically require individual optimization for each subject.

Discrete Image-based 3D Relighting: The core approach of sparse view reconstruction methods involves projecting query points onto individual views to interpolate local features, followed by feature aggregation and input into a **Multi-Layer Perceptron** (**MLP**) for inference [23, 57, 123, 130, 140, 142, 174]. However, this approach suffers from occlusion ambiguity: some views may be occluded, and mixing their features with those from visible views leads to inefficient feature utilization, ultimately degrading reconstruction quality [123]. While human templates like SMPL [83] can act as effective guidance [6, 105, 154, 192], they introduce additional template alignment errors and thus cannot ensure complete occlusion awareness. Additionally, Function4D [174] uses the truncated **Projective Signed Distance Function** (**PSDF**) to indicate visibility, but its detail level is sensitive to depth noise. Despite their strong performance, these works are less accessible to average users and often fail on in-the-wild images, since their design is tailored to more controlled experimental settings.

Single Image-based 3D Relighting: 3D GAN models provide a method for generating new views from a single image or latent code. For instance, ShadeGAN [101]–built on neural radiance fields, incorporates various lighting constraints to generate precise 3D shapes and enable relighting. However, its reliance on a simplified Lambertian model and support for only one light direction restricts it to diffuse lighting effects, limiting versatility. Next, VoLux-GAN [145] proposes a volumetric HDRI relighting method to enable more diverse lighting effects: it accumulates albedo, diffuse, and specular reflections along each 3D ray for any HDR environment map. Yet its high computational cost limits the quality and resolution of the images it generates. EG3D [17] marks a major breakthrough by pioneering the tri-plane representation, boosting 3D GANs' computational efficiency and image quality. Building on it, NeRFFaceLighting [61] leverages a pre-trained EG3D to develop an implicit facial lighting representation. This work addresses key challenges, like limited lighting scenarios (e.g., directional light) and the difficulty of acquiring datasets for intrinsic component supervision (e.g., albedo). However, its lighting control is tightly coupled to the underlying 3D shape, shape errors can lead to issues like inaccurate shadows for details such as beards or hats. Complementing this, Holo-Relighting [90] uses pre-trained EG3D models to extract 3D-perceptual features from input portraits. Its relighting module predicts a tri-plane 3D representation, enabling rendering from any viewpoint with controllable lighting and head posture. A key advantage is its ability to produce more realistic non-Lambertian effects, such as specular highlights and cast shadows. Meanwhile, LumiGAN [33] employs a self-supervised,

efficient visibility formula. It generates realistic shadow effects while reconstructing diffuse albedo and specular shading, striking a balance between realism and efficiency.

While generative model-based relighting algorithms have achieved promising results in generating new views and relighting facial regions, extending these algorithms to full-body or upper-body relighting remains challenging–primarily due to limitations of the pre-trained models employed. Additionally, they share the common GAN inversion limitation in image editing: imperfect inversion causes identity drift and loss of fine details. Developing better inversion techniques [7, 172, 175] is a promising way to tackle these problems.

4.3 Video Relighting

The core requirements for 3D-aware portrait video relighting include realistic relighting effects, temporal consistency, and real-time performance. For instance, Zhang et al. [183] proposed a video portrait relighting method that enables natural, consistent relighting of portraits from monocular RGB video streams under dynamic lighting conditions. It achieves this by jointly modeling semantic, temporal, and lighting consistency, leveraging a newly constructed dynamic OLAT dataset, and ensuring real-time performance on mobile devices. Similarly, Huynh et al. [58] used OLAT data to train a U-Net and a differentiable renderer, allowing realistic relighting of dynamic human performances. Additionally, Chandran et al. [18] built on the DPR framework to propose an end-to-end differentiable pipeline for video portrait relighting. This pipeline delivers high-quality, temporally consistent relighting by jointly optimizing for target lighting conditions and temporal coherence, no ground-truth lighting annotations required. Separately, ReliTalk [111] was proposed for generating relightable audio-driven talking portraits from monocular videos. Cai et al. [14] also proposed a real-time 3D perception method, enabling effective portrait video relighting and new perspective synthesis while opening new possibilities in related areas.

However, the above video relighting methods confine facial relighting to the input viewpoint, restricting users from adjusting camera angles or perspectives. This limits their practical application, especially in creative and AR/VR scenarios. To enable free-viewpoint relighting, 3D-aware video relighting methods have been proposed for human portraits [14, 111]. For example, Qiu et al. introduced the ReliTalk framework [111], which generates realistic, relightable audio-driven portrait videos from single view footage. It leverages 3D facial priors, reflectance decomposition, identity consistency supervision, and audio-to-expression guidance, thus enabling realistic rendering under various lighting and viewing conditions. However, the framework requires 3 days of training for a 2-minute video clip, and each frame takes 0.2 seconds to relight. In pursuit of high-quality, efficient 3D-aware video relighting, Cai et al. [14] proposed a real-time technique based on NeRF. This method enables realistic rendering across different viewpoints and lighting conditions, while ensuring temporal consistency across video frames by integrating a dual encoder and temporal consistency networks. Finally, current 3D-aware portrait video relighting methods still struggle to model glare on glasses. Future improvements could involve incorporating advanced reflection and refraction modeling techniques.

4.4 Overview

To provide an overview of existing human relighting methods and their achievements, we draw on the terminology from the review by Einabadi et al. [37] and summarize these methods across multiple dimensions, including:

— **Task**: Application scenarios of the algorithm.
Possible values include: Single **Image** relighting, **3D** reconstruction and relighting, **Video** relighting.
— **Region**: The part of the human body for which the algorithm is designed for relighting.

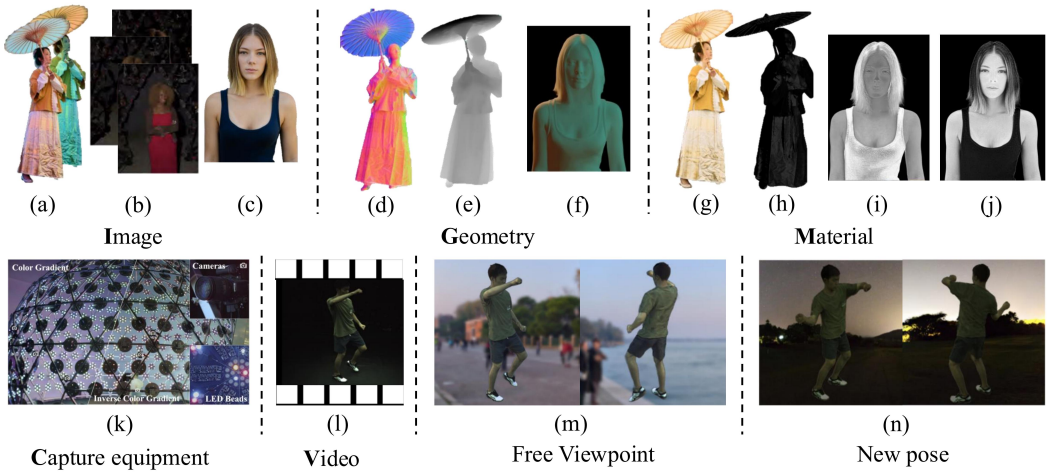

Fig. 4. Example images from [31, 141, 195].

Possible values include: **Portrait**, **Full Body**, **Hand**

- **Required Data**: All data required to reproduce the relighting system. For example, in a deep neural network-based relighting system, this includes both input data and labeled data needed to supervise network training.
- **System Inputs**: The data is directly inputted into the relighting system.
- **System Outputs**: This refers to all content output by the system, excluding intermediate derivative data. For example, facial keypoint data, automatically generated by the network, do not have a corresponding ground truth.
  Possible values for Required Data, System Inputs and System Outputs: Single **I**mage (Figure 4(c)), I_s (OLAT images (Figure 4(b)), Color gradient images (Figure 4(a)), Sparse set of views), **V**ideos (Figure 4(l)), **A**udio, latent c**O**de, **C**apture equipments (Cameras, Light Stage (Figure 4(k))), **L**ighting (Environment Light, SH coefficients, Reference light style of human image), **G**eometry (Normal Figure 4(d), Depth (Figure 4(e)), Mesh, and the derived definition containing geometric information such as shading, Light transport map (Figure 4(f)), **M**aterials Diffuse Albedo (Figure 4(g)), Specular Albedo (Figure 4(h)), Roughness (Figure 4 (i))), Reflectivity (Figure 4(j))) **P**rior (Subject ID, Keypoints, Mask)
- **Lighting**: The type of lighting used in the system.
  Possible values include: **Spherical Harmonics** (**SH**), **E**nvironment Light, **R**eference light style of human image, **Directional Light** (**DirL**), **SG**.
- **Material**: This refers to the material model employed by the relighting system.
  Possible values include: Lambertian, BRDF, Phong, Cook-Torrance, Disney BRDF, Blinn-Phong, BSDF, BSSDFs.
- **Relighting effects**: The lighting effects achieved by the experimentally verified relighting system.
  Possible values include: **D**iffuse interreflection, **S**pecular highlights, s**U**bsurface scattering, s**H**adow
- **Control**: This refers to the system's capability to generate images by controlling new **L**ighting, new **vi**ew and new **Po**se.
  Possible values include: new **L**ighting, new **Vi**ew (Figure 4(m)), and new **Po**se (Figure 4(n)).
- **CGmodule**: Embedded **Computer Graphics** (**CG**) Knowledge Levels in the System.

Possible values include: ✗: No CG Module. **N**on-differentiable CG Module: Refers to the computer graphics modules in the system that do not support differentiation operations. **D**ifferentiable CG Module: Refers to the computer graphics modules in the system that support differentiation operations.

— **Generality** refers to whether the system needs to be retrained when faced with different instances during inference.
Possible values include: ✗(Instance Specific): The system necessitates separate training for each new input instance. For instance, a system capable of synthesizing images of specific individuals must be retrained whenever a new person is introduced. ✓(General): The system can be applied to various input instances with only a single training session, exhibiting strong universality.

Table 2 provides a comprehensive comparison of learning-based human relighting methods. Most research methods focus on portrait relighting, facilitated by the relative abundance and accessibility of public portrait datasets. From the perspective of data requirements, most image relighting methods [48, 55, 60, 69, 71, 76, 102, 128, 135, 143, 163, 171, 173, 195, 197], all 3D relighting methods, and some video relighting methods [14, 111] employ inverse rendering to estimate geometry, materials, and other intrinsic information from inputs for subsequent relighting. Due to the lack of large-scale, high-quality public datasets, many relighting methods focus on constructing datasets using image capture devices, such as Light Stages or camera arrays, for supervised model training. This is particularly common in portrait image relighting [69, 75, 91, 102, 139, 163] and 3D portrait relighting [90, 117, 124, 125, 132, 161, 169, 186], as facial data capture setups are simpler compared to full-body datasets. Only a handful of full-body relighting studies [186, 195] construct their own datasets, with most methods relying on existing datasets or purchasing 3D data from commercial websites, which are then rendered to obtain relighting data for model training. Additionally, certain image relighting methods aim at achieving realistic relighting implicitly, without predicting intrinsic properties [56, 75, 80, 91, 106, 134, 136, 139, 187, 194].

Regarding input and output, early methods in 3D portrait reconstruction involved constructing 3D point clouds from videos [129] to render results from multiple perspectives. Subsequent methods explored building 3D representations from sparse viewpoint images to achieve relighting from new viewpoints [115, 117, 125, 140]. The advent of pre-trained 3D GANs has facilitated research into generating relit images from different views based on a single portrait image [90, 116] or latent codes [33, 145, 169]. To achieve high-frequency lighting effects, some methods focus on modeling OLAT images that include facial reflectance fields from input data [115, 150], while others rely on more accurate physical rendering models [125, 169]. However, these methods only model static 3D faces. The recent popularity of Gaussian splatting methods [124] has enabled the rapid construction of animatable and relightable portrait avatars from videos.

In the task of 3D full-body relighting, while a few early methods focused on constructing static 3D representations from multi-view data [186, 191], most methods aim at building animatable and relightable 3D avatars from videos [22, 78, 141, 166, 167, 170, 177, 178, 190]. These methods, which rely on the SMPL framework, typically require individual training for each person's video to achieve more accurate 3D reconstructions. Video relighting algorithms primarily concentrate on human portraits. Beyond developing methods for continuous relighting across time series [183], some approaches aim at reconstructing 3D portraits from videos, enabling relighting from novel perspectives [14, 44]. Additionally, certain methods generate audio-driven relighting videos [44, 111].

Regarding lighting, spherical harmonics and environment maps are widely adopted lighting formats. Moreover, assumptions about material significantly impact lighting effects. Early methods that assumed Lambertian models [60, 128, 129, 173, 191] were often limited to modeling diffuse

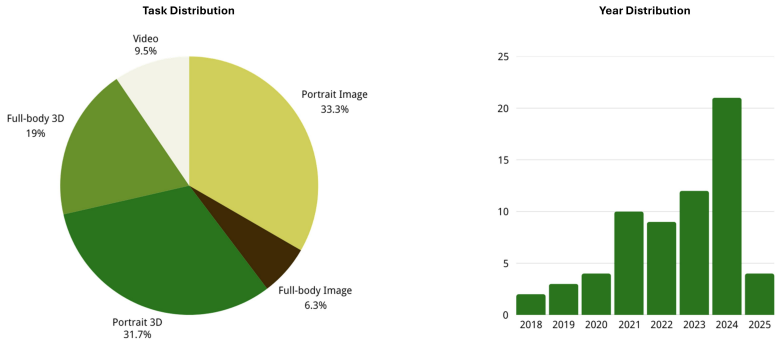

Fig. 5. Statistic overview of the existing methods by task and year.

effects and some shadow effects. Subsequent methods [15, 22, 33, 48, 56, 69, 75, 78, 102, 124, 125, 132, 140, 141, 161, 167, 169, 170, 177, 178, 186, 190], by incorporating more accurate models, have achieved relatively realistic lighting effects.

Based on Table 2, we analyzed the distribution of recent relighting works across different tasks, including "Portrait image", "Full-body image", "Portrait 3D", "Full-body 3D", and "Video" relighting. The Figure 5 pie chart reveals a clear focus on portrait-related relighting tasks. "Portrait image" relighting is the most prominent, accounting for 32.3% of the works, while "Portrait 3D" relighting follows closely with 30.8%. In contrast, "Full-body image" relighting (9.2%) and 'Full-body 3D' relighting (18.5%), along with "Video" relighting (9.2%), occupy smaller proportions. This distribution highlights the significant emphasis on portrait-centric tasks and suggests promising opportunities for development in full-body and video relighting areas. In addition, we tallied the relighting works from Table 2 spanning 2018 - 2025. As shown in the bar chart (Figure 5), there's a clear upward trajectory in the number of relighting works from 2018 to 2024, with a sharp spike in 2024. Notably, four new works had already emerged by early 2025. This trend highlights the increasing importance of human relighting as a research focus.

4.5 Evaluation

4.5.1 Evaluation Metrics and Quantitative Results. Building on Table 2's comparisons, Table 3 presents the quantitative metrics and numerical results of relevant methods. Its goal is to comprehensively showcase both the evaluation metrics for image quality assessment and the performance of various relighting methods across different tasks, as reflected by these metrics. Similar to Tables 2 and 3 also encompasses image quality assessment for relighting tasks involving portrait images, full-body images, 3D portraits, 3D full-body scans, and videos. The table's data is derived from the articles associated with the bolded methods in each evaluation group. Since each method uses different test datasets, there are differences in the results despite the consistent evaluation criteria.

As illustrated in Table 3, the portrait image relighting task employs the most diverse evaluation metrics. For example, metrics such as **Mean Absolute Error** (**MAE**), **Mean Squared Error** (**MSE**), and Si-MSE [3] are used to assess the overall quality of relit images. The PSNR [54], SSIM [162], and **Dissimilarity Index** (**DSSIM**) are employed to evaluate the fidelity of these images. Additionally, the LPIPS [185] measures the perceptual quality, while ID similarity assesses the level of identity preservation. It is worth noting that Tajima et al.'s 2024 [144] method for full-body image relighting utilizes SSIM and LPIPS. Furthermore, it incorporates the RMSE to evaluate the quality of relit full-body images, offering a comprehensive assessment approach for this specific task.

Table 2. Summary of Learning-based Human Relighting Methods Presented in this Survey

Method	Task	Region	Year	Required Data	Inputs	Outputs	Lighting	Material	Relighting Effects	Control	CG Module	Generality
SfSNet [128]	Image	Portrait	2018	IGMLP	I	GML	SH	Lambertian	DH	L	D	✓
Zhou et al. [194]	Image	Portrait	2019	IL	IL	I	SH	-	DH	L	✗	✓
Sun et al. [139]	Image	Portrait	2019	CIL	IL	I	E	-	DSH	L	✗	✓
Meka et al. [91]	Image	Portrait	2019	CIL	IL	I	E	-	DSH	L	✗	✓
Wang et al. [163]	Image	Portrait	2020	CIGML	IL	I	E	-	DSH	L	✗	✓
LeGendre [75]	Image	Portrait	2020	CIL	I	IL	SH	BRDF	DSH	L	✗	✓
Liu et al. [80]	Image	Portrait	2020	IL	IL	IL	SH	-	-	L	✗	✓
Hou et al. [56]	Image	Portrait	2021	ILP	ILP	I	SH	-	DH	L	✗	✓
Song et al. [134]	Image	Portrait	2021	IL	IL	I	R	-	DSH	L	✗	✓
Total Relighting [102]	Image	Portrait	2021	CIGML	IL	I	E	Phong	DSH	L	✗	✓
PR-RL [187]	Image	Portrait	2021	IL	IL	I	SH/R	Lambertian	DSH	L	✗	✓
Hou et al. [55]	Image	Portrait	2022	IGML	IL	I	SH	BRDF	DSH	L	D	✓
Zhu et al. [197]	Image	Portrait	2022	IGL	IL	L	SH	-	DSH	L	✗	✓
YEH et al. [171]	Image	Portrait	2022	IGML	IL	L	E	-	DSH	L	✗	✓
Song et al. [135]	Image	Portrait	2022	IGL	IL	I	E	-	DSH	L	✗	✓
Han et al. [48]	Image	Portrait	2023	IGML	IL	L	SH	svBRDF	DSH	L	D	✓
Pidaparthy et al. [106]	Image	Portrait	2024	IL	IL	I	SH	-	DSH	L	✗	✓
LEIFR-Net [76]	Image	Portrait	2024	IPML	I/ML	I	E	-	DSH	L	✗	✓
BiPR-RL [136]	Image	Portrait	2024	IL	IL	I	SH/R	-	DH	L	✗	✓
SwichLight [69]	Image	Portrait	2024	CIGML	IL	I	E	Cook-Torrance	DSH	L	✗	✓
SynthLight [19]	Image	Portrait	2025	TIL	TIL	I	E	-	DSH	L	✗	✓
Kanamori and Endo [173]	Image	Full Body	2018	IGML	IL	I	SH	Lambertian	DH	L	✗	✓
Ji et al. [60]	Image	Full Body	2022	IGML	IL	I	E	Lambertian	DH	L	✗	✓
Lagunas et al. [71]	Image	Full Body	2021	IGML	I	MGI	SH	-	DSH	L	✗	✓
Tajima et al. [143]	Image	Full Body	2021	IGML	IL	I	SH	Lambertian	DSH	L	✗	✓
UltraStage. [195]	Image	Full Body	2023	CIGML	I	GM	E	-	DSH	L	✗	✓
Tajima et al. [144]	Image	Full Body	2024	IPL	IPL	I	E	Disney BRDF	DSH	L	✗	✓
Relightful Harmonization [118]	Image	-	2024	IL	IL	I	E	-	DSH	L	✗	✓
IC-Light [180]	Image	-	2024	TIGPL	TI	I	-	-	DSH	L	✗	✓
Sevastopolsky et al. [129]	3D	Portrait	2020	CVGMLPW	VLW	I	dirL/SH	Lambertian	DH	L/Vi	✗	✓
SunStage [161]	3D	Portrait	2023	CVGMLW	VLW	GML	E	Blinn-Phong	DSH	L/Vi	D	✓
NeLF [140]	3D	Portrait	2021	I_sGMLW	I_sLW	I	E	BSDF	DSH	L/Vi	D	✓
Reddy et al. [150]	3D	Portrait	2021	I_sIGLW	ILW	I_s I	E	-	DSUH	L/Vi	D	✓
Tan et al. [145]	3D	Portrait	2022	OGMLW	OLW	I	E	-	DSH	L/Vi	✗	✓
VoRF [115]	3D	Portrait	2022	I_sILW	I_sILW	I_s I	E	-	DSUH	L/Vi	✗	✓
NeRFFaceLighting [61]	3D	Portrait	2023	OIGMLW	OLW	IGM	SH	-	DSH	L/Vi	✗	✓
FaceLit (2023) [114]	3D	Portrait	2023	OILW	OLW	I	SH	-	DSH	L/Vi	✗	✓
SIRA [15]	3D	Portrait	2023	IGMLW	ILW	IGM	E	svBRDF	DSH	L/Vi	D	✓
ReNeRF [169]	3D	Portrait	2023	CI_sILW	ILW	I_s I	SH	BSDF	DSUH	L/Vi	✗	✗
LitNeRF [125]	3D	Portrait	2023	CI_sIGMLW	I_sLW	IGM	E	BSDF	DSUH	L/Vi	D	✓
PN-Relighting[160]	3D	Portrait	2023	I_sIGML	II_sL	IGML	E	Phong	DSH	L/Vi	D	✓
Lite2Relight [116]	3D	Portrait	2024	ILW	ILW	I	E	-	DSUH	L/Vi	✗	✓
Shen et al [132]	3D	Portrait	2024	CVGMLW	VLW	IGM	SH	BRDF	DSUH	L/Vi	✗	✓
Rao et al. [117]	3D	Portrait	2024	CI_sILW	ILW	Is I	E	-	DSUH	L/Vi	✗	✓
LumiGAN [33]	3D	Portrait	2024	OIGMLW	OLW	IGM	SH	svBRDF	DSH	L/Vi	✗	✓
Holo-Relighting [90]	3D	Portrait	2024	CIGMLW	ILW	I	E	-	DSH	L/Vi	✗	✓
Saito [124]	3D	Portrait	2024	CVGMLW	VLW	IGM	SH	BRDF	DSUH	L/Vi/Po	✗	✓
FLARE [5]	3D	Portrait	2023	VIWP	VIW	IP	E	Cook-Torrance	DSH	L/Vi/Po	D	✗
HRAvatar [179]	3D	Portrait	2025	VIGMLW	VLW	IGM	E	Disney BRDF	DSH	L/Vi/Po	✗	✗
Zhang et al. [186]	3D	Full Body	2021	CI_sIGMLW	I_sLGMW	I	E	BSSRDFs	DSUH	L/Vi	D	✗
Zheng et al. [191]	3D	Full Body	2023	I_sIGMLW	I_sGLW	IGM	SH	Lambertian	DH	L/Vi	✗	✓
Relighting4D [22]	3D	Full Body	2022	VIGMLW	VLW	IGM	E	BRDF	DSH	L/Vi/Po	D	✗
Sun et al. [141]	3D	Full Body	2023	VIGMLW	VLW	IGM	SG	Disney BRDF	DSH	L/Vi/Po	✗	✗
NECA [166]	3D	Full Body	2024	VGMLW	VW	GML	E	-	DH	L/Vi/Po	✗	✗
Zhang et al. [178]	3D	Full Body	2024	VIGMLW	VLW	IGM	E	BRDF	DSH	L/Vi/Po	D	✗
Zhao et al. [190]	3D	Full Body	2024	VIGMLW	VLW	IGM	E	Disney BRDF	DSH	L/Vi/Po	✗	✗
Lin et al. [78]	3D	Full Body	2024	VIGMLW	VLW	IGM	E	Disney BRDF	DSH	L/Vi/Po	D	✗
Xu et al. [170]	3D	Full Body	2024	VIGMLW	VLW	IGM	E	Microfacet BRDF	DSH	L/Vi/Po	D	✗
Zhan et al. [177]	3D	Full Body	2024	VIGMLW	VLW	IGM	E	Disney BRDF	DSH	L/Vi/Po	✗	✗
ReN Human [167]	3D	Full Body	2024	VIGMLW	VLW	IGM	E	BRDF	DSH	L/Vi/Po	D	✗
DNF-Avatar [63]	3D	Full Body	2025	VLW	VLW	I	E	Cook-Torrance BRDF	DSH	L/Vi/Po	✗	✗
Zhang [183]	Video	Portrait	2021	CVL	VL	V	E	-	DS	L	✗	✓
Chandran [18]	Video	Portrait	2022	VLP	VLP	V	SH	-	-	L	✗	✓
Relitalk [111]	Video	Portrait	2024	VAGML	VAL	VGM	SH	Blinn-Phong	DSH	L/Vi/Po	✗	✓
Cai et al. [14]	Video	Portrait	2024	VGML	VLW	VGM	SH	-	DH	L/Vi/Po	D	✓
LumiSculpt [189]	Video	Portrait	2024	TL	TL	V	-	-	DSH	L	✗	✓
Lux Post Facto [89]	Video	Portrait	2025	VIL	V/IL	V/I	E	-	DSH	L	✗	✓

In the field of 3D portrait relighting, the Holo-Relighting [90] method innovatively integrates the NIQE [182] to evaluate the perceptual quality of relit images more effectively. NIQE can capture features in images that align with human perception, such as texture and contrast. This alignment ensures that its evaluation results closely match human subjective assessments of image quality,

Table 3. Quantitative Results of Feasible Methods on Multiple Evaluation Metrics

Task	Method	Eva. Dataset	MAE↓	MSE↓	Si-MSE↓	RMSE↓	PSNR↑	SSIM↑	DSSIM↓	LPIPS↓	NIQE↓	ID Similarity ↑	FID↓	KID ↓	CLIP ↑
Image Portrait	SfSNet (2018) [128]		-	0.0275	0.0184	-	16.25	-	0.1562	0.3594	-	-	-	-	-
	Zhou et al. (2019) [194]		-	0.0134	0.0108	-	19.39	-	0.0708	0.0858	-	-	-	-	-
	PR-RL (2021) [187]		-	0.0098	0.0085	-	21.00	-	0.0584	0.0808	-	-	-	-	-
	BiRP-RL (2024) [136]	CelebAMask-HQ	-	0.0084	0.0073	-	21.70	-	0.0554	0.0772	-	-	-	-	-
	Sun et al. (2019) [139]		0.1715	0.0748	-	-	-	0.8432	-	3.648	-	-	-	-	
	Total Relighting (2021) [102]		0.1643	0.0658	-	-	-	0.8465	-	3.425	-	-	-	-	
	SwichLight (2024) [69]	Self-created	0.1023	0.0235	-	-	-	0.9002	-	2.137	-	-	-	-	
	Neural gaffer (2024) [64]		-	-	-	-	25.327	0.9	-	0.102	-	1-0.357	-	-	-
	DiLightNet (2024) [176]		-	-	-	-	22.991	0.86	-	0.128	-	1-0.333	-	-	-
	IC Light (2024) [180]		-	-	-	-	20.283	0.874	-	0.108	-	1-0.284	-	-	-
	SwichLight (2024) [69]		-	-	-	-	21.432	0.911	-	0.088	-	1-0.198	-	-	-
	SynthLight (2025) [19]	Self-created	-	-	-	-	29.572	0.945	-	0.063	-	1-0.165	-	-	-
Image Full Body	Lagunas et al. (2021) [71]		-	-	-	0.224	-	0.483	-	0.162	-	-	-	-	-
	Tajima et al. (2021) [143]		-	-	-	0.173	-	0.586	-	0.0737	-	-	-	-	-
	Ji et al.(2022) [60]		-	-	-	0.121	-	0.686	-	0.0819	-	-	-	-	-
	Tajima (2024) [144]	Self-created	-	-	-	0.0744	-	0.787	-	0.0493	-	-	-	-	-
3D Portrait	Tan et al. (2022) [145]		-	-	-	-	-	-	-	-	-	0.831	59.79	4.124	-
	FaceLit (2023) [114]		-	-	-	-	-	-	-	-	-	0.901	4.01	0.115	-
	LumiGAN (2024) [33]	FFHQ	-	-	-	-	-	-	-	-	-	0.947	3.86	0.147	-
	NeRFFaceLighting (2023) [61]		-	-	-	-	19.4	0.7214	-	0.2179	8.132	0.6953	-	-	-
	FaceLit (2023) [114]		-	-	-	-	20.86	0.7281	-	0.1542	8.017	0.7821	-	-	-
	Holo-Relighting (2024) [90]	Self-created	-	-	-	-	27.35	0.8511	-	0.0917	5.275	0.8978	-	-	-
3D Full Body	FLARE (2023) [5]		1.715	-	-	-	25.82	0.8576	-	0.123	-	-	-	-	-
	HRAvatar (2025) [179]	Self-created	1.123	-	-	-	28.97	0.9054	-	0.1059	-	-	-	-	-
	Relighting4D (2022) [22]		-	-	-	-	24.5498	0.7989	-	0.1178	-	-	-	-	-
	Lin et al. (2024) [78]		-	-	-	-	28.6252	0.829	-	0.0775	-	-	-	-	-
	Xu et al. (2024) [170]		-	-	-	-	28.4215	0.8258	-	0.0828	-	-	-	-	-
	Zhan et al. (2024) [177]	Self-created	-	-	-	-	28.7703	0.8304	-	0.074	-	-	-	-	-
	Relighting4D (2022) [22]		-	-	-	-	16.62	0.837	-	0.1726	-	-	-	-	-
	Intrinsicavatar (2024) [159]		-	-	-	-	18.18	0.8722	-	0.1279	-	-	-	-	-
	DNF-Avatar (2025) [63]	RANA	-	-	-	-	19.04	0.8772	-	0.1307	-	-	-	-	-
Video	Open-Sora (2024) [193]		-	-	-	0.4542	-	-	-	1.3503	-	-	-	-	0.9845
	IC Light (2024) [180]		-	-	-	0.5264	-	-	-	2.5329	-	-	-	-	0.9703
	LumiSculpt (2024) [189]	LumiHuman	-	-	-	0.3597	-	-	-	1.1312	-	-	-	-	0.9951
	Total Relighting (2021) [102] (2021)		-	-	-	-	22.44	0.7793	-	0.1794	6.458	-	-	-	-
	SwichLight (2024) [69] (2024)		-	-	-	-	19.87	0.7481	-	0.2129	7.166	-	-	-	-
	Lux Post Facto (2025) [89]	Self-created	-	-	-	-	24.62	0.8278	-	0.1158	5.653	-	-	-	-

The abbreviation 'Eva.' refers to 'Evaluation'.

thereby enhancing the credibility of its evaluation results. Meanwhile, LumiGAN [33] utilizes the **Fréchet Inception Distance** (**FID**) [52] and **Kernel Inception Distance** (**KID**) [9] as pivotal evaluation metrics. By measuring the distance between the feature distributions of generated and real-world images within the feature space of a pre-trained Inception network, FID and KID can effectively evaluate the fidelity of relit images. Table 3 presents the comparison results of new poses from the article by Zhan et al.

In the video relighting task, LumiSculpt [189] employs conventional metrics, (e.g., RMSE, LPIPS) to evaluate the quality, and innovatively introduces the similarity of frame-wise CLIP image embeddings for assessment. This novel approach focuses on semantic-level coherence of the video. By calculating the similarity of these embeddings, LumiSculpt [189] effectively measures the alignment of semantic content across consecutive video frames, which is crucial for ensuring semantically coherent visual flow. Moreover, this innovative metric complements traditional quality metrics, providing a more holistic view of the overall quality and coherence of relit videos.

Overall, the evaluation metrics employed across diverse relighting tasks tend to vary. This variance is primarily attributed to the distinct characteristics inherent in each task and the unique features of different methods. In addition, the numerical results presented in the table offer valuable insights into the effectiveness of these methods, empowering readers to develop a more comprehensive understanding of their performance.

4.5.2 Qualitative Results. To visualize the lighting effects of recent relighting methods, Figure 6 presents a qualitative comparison of the methods detailed in Table 3. The results are sourced from the articles associated with the bolded methods in Table 3.

In the "Portrait Image Relighting Methods" section, the first row presents the relighting results of BIPR-RL, a reinforcement learning-based method, alongside those of other comparative methods. SfSNet [128], which estimates based on 3D information, generates blurry renders. DPR [194], as a one-step black-box approach, is prone to local errors and inaccurate lighting edits. PR-RL [187] shows incorrect details, such as in the chin area of the example. In contrast, BIPR-RL [136]

Fig. 6. Qualitative results of recent relighting methods.

produces a more natural relighting effect. However, it struggles with shadows on facial accessories, such as hair and hats, causing unnatural relighting in partially occluded areas. In the second row, SwitchLight surpasses previous methods with notable strengths. It effectively balances light direction and softness, avoiding harsh highlights and inaccurate lighting. It also efficiently captures high-frequency details, such as specular highlights and hard shadows. As the example shows, it retains facial details and identity information, ensuring high fidelity and minimizing distortion compared to earlier approaches. However, it has difficulty neutralizing strong shadows. This leads to residual shadow artifacts in the albedo and relit images. Moreover, SwitchLight [69] mistakenly

treats reflective surfaces, such as sunglasses, as opaque objects in normal images, preventing it from effectively removing reflective highlights from albedo and relit images. The third row illustrates the relighting performance of SynthLight [19] and other methods on in-the-wild images. DiLightNet [176] has artifacts from the failed 3D reconstruction. The Neural Gaffer, untrained in human portraits, produces inaccurate facial shadow contours. IC-Light utilizes background images for relighting, addressing the challenges of face relighting. Total Relighting [102] and SwitchLight [69], trained on Light-Stage data, produce soft shadows even under strong sunlight and alter skin tones. In contrast, SynthLight [19] achieves excellent relighting while preserving subject identity. However, it has some flaws: when relighting half-body portraits, it fails to retain precise facial details, and it faces challenges in accurately capturing the texture of fabric.

Overall, portrait image relighting still faces several challenges, including accurately simulating shadows from accessories like hats, glasses, and facial hair that obscure the face; realistically rendering reflections on accessories such as glasses under new lighting conditions; preserving precise facial details; and faithfully capturing fabric textures. These challenges could be addressed by expanding the diversity of training datasets. This can be achieved by adding more camera pose variations and enhancing the variety and quality of fabric materials included in the datasets. Additionally, adopting a more precise lighting model could further help address these challenges.

The full-body image relighting task illustrates the relighting performance of Tajima et al. (2024) [144] and other methods on real-world images. Sun et al. [139] and Nestmeyer et al. [96] disregard physical laws, resulting in blurry outputs. Lagunas et al. [71] and Tajima et al. (2021) [143] rely on low-order SH lighting approximations. This makes it difficult for them to simulate high-frequency illumination, resulting in strong shadow contrast. For the same reason, these methods fail to reproduce high-frequency highlights and shadows. Total Relighting [102] can handle low-frequency highlights but cannot reproduce high-frequency ones or shadows. Hou et al. [55] employ directional light sources for lighting representation. This makes it challenging to manage low-frequency shadows properly and leads to noticeable artifacts around shadow boundaries. Ji et al.'s [60] method can reproduce high-frequency shadows via ray tracing. However, due to the many defects in PIFuHD's 3D reconstruction, it has significant artifacts. In addition, the methods proposed by Ji et al. [60] and Hou et al. [55] are both limited to diffuse reflection and cannot reproduce gloss. By contrast, Tajima et al. (2024) [144] employ physically-based relighting with specular reflection components and 3D geometry. This method can generate high-frequency shadows from light occlusion by hands and legs, as well as natural skin highlights. However, it doesn't explicitly model complex lighting effects such as anisotropic reflection and subsurface scattering, so it may not accurately render the relighting impact of materials that exhibit these properties.

Overall, current methods for full-body image relighting can generate high-frequency shadows from hand and leg occlusion and natural skin highlights, achieving more realistic lighting than previous methods. However, simulating complex light effects, such as anisotropic reflection and subsurface scattering, remains a significant challenge. These challenges may be alleviated in the future by using precise physics-aware neural networks or approaches that bridge the synthetic-to-real domain gap.

In the 3D portrait relighting task, the first row shows the lighting performance of Holo-Relighting [90] and the comparison methods across multiple views. It can be seen that NeRFFace-Lighting [61] and FaceLit [114] generate visual artifacts and fail to reproduce the desired lighting effects fully. Holo-Relighting [90], on the other hand, generates the most realistic lighting, with shading and specular specularity closely matching those in the reference image. Holo-Relighting [90] relies on pretrained EG3D and GAN inversion to extract 3D information from 2D images. However, this approach has limitations: It struggles with upper-body portrait relighting, as this exceeds EG3D's capabilities. The imperfections of GAN-based inversion can lead to identity shifts and detail

loss. Moreover, Holo-Relighting [90] cannot generate complex lighting effects, such as shadows from foreign objects and glare from eyeglasses. Since it approximates light transport from data without adhering to physical constraints, it may not perfectly handle challenging cases, such as view-dependent effects. In the second row, ShadeGAN [101] is limited to simple directional lights with changing hues and utilizes an oversimplified Lambertian lighting model. This prevents it from generating realistic lighting effects under complex conditions. VoluxGAN [145] disregards visibility and self-occlusion, resulting in disturbing artifacts, physically inaccurate reflections, and multi-view inconsistencies. By contrast, LumiGAN produces more realistic, multi-view-consistent faces with credible shadows, even under complex lighting conditions. Like Holo-Relighting, LumiGAN [33] also uses pretrained EG3D, facing the same challenges in upper-body portrait relighting. Similarly to the original precomputed radiance transfer, extending the Neural Radiance Transfer of LumiGAN [33] to dynamic scenes remains a challenging task. The third row compares FLARE [5] and HRAvatar [179] in relighting. FLARE [5] exhibits some artifacts due to partially corrupted normal reconstruction. In contrast, the smoother normals produced by HRAvatar [179] produce more consistent and realistic relighting results. Although HRAvatar [179] effectively models individual deformations, it is constrained by FLAME priors, which can limit its control over elements such as hair and accessories when insufficient training data is available. Additionally, the robust texture modeling of 3DGS and limitations in current albedo estimation models may cause incorrect coupling of shadows or wrinkles with albedo or reflectance. This can lead to relighting defects, particularly in specular reflection and shadow regions. Moreover, HRAvatar [179] cannot reconstruct the entire head from monocular video if the camera pose is unknown, even if the back of the head is visible.

In general, current single-image 3D portrait reconstruction faces challenges: 3D GAN limitations prevent upper-body portrait relighting; GAN inversion flaws cause identity shifts and detail loss; and relighting dynamic 3D scenes from a single portrait remains difficult. Future directions could expand 3DGAN-based relighting to more specialized human generation models, develop advanced inversion techniques, and combine with animatability to create fully editable 3D human portrait assets. Current video-based animatable 3D Portrait relighting has limitations, especially in specular reflection and shadows. One potential solution is to integrate advanced albedo estimation models (which help associate shadows or wrinkles with albedo or reflectance) to address this issue.

In the part of 3D full-body human relighting methods, the first row presents the qualitative results of DNF Avatar and other comparative methods. Relighting4D [22] fails to generalize to new poses, producing unreasonable results. IntrinsicAvatar [159] uses iNGP, making it prone to high-frequency noise in certain areas (red box). Unlike surface-based student models, the volumetric scattering-based teacher model can sample secondary rays within surfaces, resulting in darker shadows. Moreover, its limited sample counts leads to noisy or inaccurately estimated materials (orange box). DNF-Avatar [63] (using 2DGS) produces smoother geometry. Its split-sum-based appearance model also prevents the noise that is common in Monte Carlo estimation. However, the final quality of DNF-Avatar [63] depends heavily on the stability of the teacher model. In addition, the ambient occlusion assumption in the DNF-Avatar [63] method may not hold when there are strong point light sources. This might prevent the shading model from capturing the correct shadow effects. The second row compares the method proposed by Zhan et al. with others. Relighting4D [22] shows poor generalization to new poses. The results generated by IntrinsicAvatar [159] are blurry, as it fails to model pose-dependent non-rigid deformation. Due to its signed distance function design, the method proposed by Xu et al. [170] tends to generate hollow artifacts under the armpits (red box). In contrast, the method proposed by Zhan et al. [177] outperforms those of Lin et al. [78] and Xu et al. [170] in detail preservation, this is evident in the facial details of the ZJU-377 dataset (blue box). The specular reflection on Jody's leggings, as generated by Zhan et al. [177], is more realistic. However, similar to the methods, the approach proposed by Zhan et al. [177] learns

consistent materials from video frames and cannot model pose-dependent wrinkle deformations. In addition, it cannot reconstruct loose clothing, such as dresses. Reconstructing loose clothing (e.g., dresses) remains a challenge for current 3D full-body relighting methods. Future work could focus on better modeling pose-dependent wrinkle deformation and lighting effects.

In the first row of the video relighting part, LumiSculpt [189], a text-to-video illumination control generation method, is compared with the state-of-the-art text-to-video generation method Open Sora and the relighting method IC-Light [180]. The comparison focuses on the strong and soft lighting effects generated under explicit lighting trajectories. For details, please refer to the LumiSculpt [189] for generating video text descriptions. IC-Light [180] can generate single-image strong lighting with precise direction (left image) and soft lighting (right image). However, due to the lack of interframe perception, the output video has poor coherence and noticeable background flickering. Open Sora [193] generates coherent and visually pleasing videos, but struggles to control lighting direction via text, resulting in relatively fixed lighting throughout the video. LumiSculpt [189] ensures video coherence and quality while precisely controlling the trajectory and intensity of lighting. The second row compares Lux Post Facto [89] with top video relighting methods, such as NVPR [183] and SwitchLight [69]. Lux Post Facto [89] produces more realistic highlights, natural skin tones, finer details, and better shadows. It delivers more accurate lighting effects and shows strong robustness to changes in facial expressions and head movements. However, Lux Post Facto [89] has several limitations. First, it may fail to preserve the details of rare accessories such as decorative hairpieces, partially occluding the face. Second, since it learns to synthesize lighting from OLAT renderings, it can only generate lighting effects represented in the lighting stage, not challenging ones such as foreign shadows. Lastly, its reliance on a video diffusion model and the iterative diffusion process make it difficult to apply to real-time applications.

In general, text-to-video relighting methods, such as LumiSculpt [189] and Open-Sora [193], cannot modify the lighting in existing videos. Models that can edit lighting on existing videos, such as Lux Post Facto [89], face challenges in preserving details of rare accessories, simulating extreme lighting conditions (e.g., foreign shadows) and real-time applications. Potential solutions include expanding the training dataset for greater diversity, designing more efficient architectures, or exploring distillation techniques to reduce inference time.

4.5.3 Resource Consumption. This subsection presents a statistical analysis of the GPU and time requirements for both the training and inference phases of the methods summarized in Table 4. By detailing their computational resource and time demands, this analysis enables readers to evaluate the efficiency and practicality of each method, and supports their selection for real-world applications.

In the image-based human portrait relighting task, SfSNet [128], PR-RL [187], and Sun et al.'s method [139] stand out for their relatively fast rendering speeds, each capable of rendering a single image in under a second. Of these, PR-RL [187] is notably the fastest, achieving a rapid rendering time of just 0.09 seconds per image, which translates to approximately 11.11 **frames per second** (**fps**). Yet, this speed is insufficient to meet the 30 fps threshold necessary for smooth, real-time rendering. Furthermore, the images produced by PR-RL [187] have a lower resolution of 256×256. In contrast, newer methods such as SwitchLight [69], IC Light [180], and SynthLight [19] demand extensive computational resources during training. However, details regarding their testing resource requirements and time consumption remain undisclosed. In full-body human image relighting, the methods by Lagunas et al. [71], Tajima et al. (2021) [143], and Tajima et al. (2024) [144] take roughly 1 second to render an image. Tajima et al. (2021) [143] render the fastest at 0.18 seconds per image (5.56 fps), yet this speed is still much lower than the real-time rendering requirement. In the task of

Table 4. Resource Consumption Statistics of Feasible Methods

Task	Method	Training Device	Training Time	Rendering Device	Rendering Time	Rendering Resolution
Image Portrait	SFSNet (2018) [128]	-	-	-	0.7442 seconds	128 × 128
	DPR (2019) [194]	-	-	-	-	1024 × 1024
	PR-RL (2021) [187]	1 NVIDIA Tesla P100 GPU	-	-	0.09 seconds	256 × 256
	BIRP-RL (2024) [136]	1 NVIDIA Tesla P100 GPU	-	-	-	-
	Sun et al. (2019) [139]	4 NVIDIA Tesla V100 GPUs	≈ 26 hours	-	0.1616 seconds	640 × 640
	Total Relighting (2021) [102]	8 NVIDIA Tesla V100 GPUs (16 GB)	-	-	-	2048 × 1536
	SwitchLight (2024) [69]	32 NVIDIA A6000 GPUs	168 hours	-	-	512 × 512
	Neural Gaffer (2024) [64]	8 A6000 GPUs	600 hours	1 A6000 GPU	≈ 480 seconds	256 × 256
	DilightNet (2024) [176]	8 NVIDIA A100 GPUs	30 hours	-	-	512 × 512
	IC Light (2024) [180]	8 H100 NVLink GPUs (80 GB)	140 hours	-	-	1024 × 1024
	SwitchLight (2024) [69]	32 NVIDIA A6000 GPUs	168 hours	-	-	512 × 512
	SynthLight (2025) [89]	32 A100 GPUs (40 GB)	≈ 24 hours	-	-	512 × 512
Image Full Body	Lagunas et al. (2021) [71]	8 Tesla V100-SXM2 (16 GB)	55 hours	1 NVIDIA GeForce GTX 1080 Ti	1.18 seconds	768 × 768
	Tajima et al. (2021) [143]	1 NVIDIA GeForce GTX 1080 Ti	90 hours	1 NVIDIA GeForce GTX 1080 Ti	0.18 seconds	1024 × 1024
	Ji et al. (2022) [60]	4 RTX2080Ti-server	240 hours	1 NVIDIA RTX2080Ti	17 seconds	512 × 512
	Tajima (2024) [144]	1 NVIDIA Tesla A5000	-	1 NVIDIA RTX A5000	1.01s	1024 × 1024
3D Portrait	Tan et al. (2022) [145]	8 NVIDIA V100 GPUs	192 hours	-	0.272 seconds	256 × 256
	FaceLit (2023) [114]	8 NVIDIA A100 GPUs	≈115 hours	-	-	512 × 512
	LumiGAN (2024) [33]	8 NVIDIA RTX A6000 GPUs	-	-	-	512 × 512
	NeRFLight (2023) [61]	4 NVIDIA V100 GPUs	-	1 NVIDIA V100 GPU	≈ 0.043	512 × 512
	FaceLit (2023) [114]	8 NVIDIA A100 GPUs	152 hours	-	-	512 × 512
	Holo-Relighting (2024) [90]	8 NVIDIA A100 GPUs	≈74 hours	-	-	512 × 512
	FLARE (2023) [5]	1 NVIDIA RTX 3090 GPU	-	1 NVIDIA RTX 3090 GPU	≈ 1.714 seconds	512 × 512
	HRAvatar (2025) [179]	1 NVIDIA RTX 3090 GPU	-	1 NVIDIA RTX 3090 GPU	0.006 seconds	512 × 512
3D Full Body	Relighting4D (2022) [22]	1 NVIDIA RTX3090 GPU	49hours	1 NVIDIA RTX 4090 GPU	4 seconds	500 × 500
	Lin et al. (2024) [78]	1 NVIDIA RTX3090 GPU	44hours	1 NVIDIA RTX 4090 GPU	15 seconds	500 × 500
	Xu et al. (2024) [170]	1 NVIDIA RTX3090 GPU	33hours	1 NVIDIA RTX 4090 GPU	4.9 seconds	500 × 500
	Zhan et al. (2024) [177]	1 NVIDIA RTX3090 GPU	≈ 7 hours	1 NVIDIA RTX 4090 GPU	0.145 seconds	500 × 500
	Relighting4D (2022) [22]	1 Tesla V100 GPU	-	1 NVIDIA RTX 1080 Ti	>4 seconds	1080 × 1080
	Intrinsicavatar (2024) [159]	1 NVIDIA RTX 3090 GPU	4 hours	1 NVIDIA RTX 3090 GPU	20 seconds	540 × 540
	DNF-Avatar (2025) [63]	-	-	1 NVIDIA RTX 4090 GPU	0.896 seconds	540 × 540
Video	Open-Sora (2024) [193]	96 H100 GPUs	36 hours	-	-	2560 × 1440
	IC Light (2024) [180]	8 H100 NVLink GPUs (80 GB)	140 hours	-	-	1024 × 1024
	LumiSculpt (2023) [189]	8 NVIDIA A100 GPUs	-	-	-	640 × 480
	Total Relighting (2021) [102]	8 NVIDIA Tesla V100 GPUs (16G)	-	-	-	2048 × 1536
	SwitchLight (2024) [69]	32 NVIDIA A6000 GPUs	168 hours	-	-	512 × 512
	Lux Post Facto (2025) [89]	8 NVIDIA A100 GPUs	-	-	-	512 × 512

3D human portrait relighting, the method by Tan et al. [145] and NeRFLighting [61] can generate relighting effects from various angles using a single portrait image, rendering images in under a second. NeRFLighting [61] has a rendering speed of 0.043 seconds per image (23.25 fps), nearly meeting real-time requirements. However, it requires an NVIDIA V100 GPU (32 GB GPU memory) and isn't suitable for real-time systems with limited hardware resources. The animatable 3D portrait relighting method, HRAvatar [179], also achieves a relatively quick rendering speed of 0.5 seconds per frame, equivalent to 2 fps. However, this speed remains much lower than required for real-time rendering. In the task of 3D full-body human relighting, recent methods like DNF-Avatar [63] and the one proposed by Zhan et al. [177] can render an avatar in under a second. DNF-Avatar [63] notably achieves 67 fps, surpassing the 30 fps requirement for smooth real-time rendering. However, it requires an NVIDIA RTX 4090 GPU (possibly with 24 GB memory) and a separate model trained for each avatar. This makes DNF-Avatar [63] not well-suited for real-time platforms with limited resources. Moreover, the trained models lack generalizability and cannot be adapted for relighting new avatars. Recent video relighting methods, LumiSculpt [189] and Lux Post Facto [89] require extensive GPU resources for training, using 8 A100 GPUs each. Although their exact training and rendering times are not public, their inference is likely time-consuming. This is due to their reliance on base models needing multi-step inference. For example, LumiSculpt [189] (based on Open Sora) requires 50 iterative inference steps. Lux Post Facto [89] is based on SD 1.5.

In general, most existing methods still have limitations in real-time rendering. Those that approach or meet real-time requirements, such as NerfFaceLight and DNF-Avatar [63], depend on high-end GPUs, making deployment difficult for resource-constrained systems. Additionally, some methods have not disclosed their training or rendering resource and time requirements, which limits their usability. Despite these challenges, the field continues to advance. Future research is expected to focus on optimizing models to minimize computational and time demands while maintaining result quality.

5 Challenges and Future Directions

This chapter summarizes the key challenges of current relighting methods regarding relighting performance, efficiency, 3D human relighting limitation, and ethical implications and outlines future research directions.

Precise Simulation of Complex Lighting Effects: Current methods struggle to simulate complex lighting effects accurately, with key challenges outlined below: ***(i) Accessory Shadows***: Accessories such as hats, glasses, and hair, whose anisotropic filament structures complicate occlusion, generate dynamic shadows with jagged or distorted boundaries [176]. The geometric complexity of these accessories, particularly the fibrous microstructure of hair, challenges traditional rendering pipelines, leading to aliasing artifacts or inaccurate shadow edges. ***(ii) Reflections and Glare***: Transparent or glossy accessories (e.g., eyeglass lenses, metallic frames) exhibit complex specular reflections, refractions, and glare effects that simplified models fail to capture [69]. Existing approaches often neglect anisotropic reflection properties or polarization effects, resulting in unrealistic rendering of high-gloss surfaces under varying illumination. ***(iii) Extreme Lighting Conditions***: Extreme lighting conditions, such as strong external shadows or HDR scenes, pose significant challenges [69, 102]. Additionally, simulating subsurface scattering in skin and hair or anisotropic highlights on materials like silk or metal requires computationally intensive techniques (e.g., Monte Carlo path tracing or advanced BRDF models) that are impractical for real-time applications. To address these challenges, future works may include: ***(a) Data Augmentation***: Construct multimodal datasets using high-resolution 3D scanning, 4D light-field capture to encompass diverse accessories (e.g., hats, glasses, jewelry) and extreme lighting conditions (e.g., point sources, ambient light, HDR scenes). Leverage generative models like Stable Diffusion [4, 108] to synthesize realistic accessory-lighting combinations, including boundary lighting and dynamic shadows, reducing reliance on costly real-world data collection while enhancing dataset diversity. ***(b) Semantic-region Modeling***: Employ deep learning-based semantic segmentation (e.g., U-Net [121] or Mask R-CNN [51]) to precisely detect accessory regions, such as lenses or hair. Assign region-specific geometry and material descriptors, such as Fresnel models for lens refractions or Kajiya-Kay models [67] for hair's anisotropic highlights. This layered rendering approach ensures locally consistent shadows and reflections, improving visual fidelity in complex scenes. ***(c) Integrate More Precise Physically Based Rendering***: Incorporate a real-time physically based rendering pipeline that integrates GGX microfacet BRDF models [155], ambient occlusion, and precomputed subsurface scattering. This allows for accurate simulation of specular, diffuse, and translucent effects of diverse materials while maintaining computational efficiency. Optimized techniques, such as precomputed radiance transfer [133], further reduce rendering latency to support real-time applications.

Integrating Relighting Algorithms into Real-time Systems: According to statistics in subsection 4.5.3, integrating relighting algorithms into real-time systems such as video conferencing platforms or AR devices presents two key challenges: ***(i) Computing Performance Bottleneck:*** Most current relighting algorithms cannot meet the requirements of real-time rendering [63, 136, 144, 177], which means achieving a minimum frame rate of 30 fps on consumer devices. Complex lighting calculations (such as global illumination, shadow casting, or subsurface scattering) involve high-dimensional integration and large-scale matrix operations, resulting in significant inference latency. ***(ii) Hardware Requirements:*** High-performance methods like NeRFLight [61] and DNF-Avatar [63] need high-end GPU, such as an RTX 4090 with 24 GB GDDR6X memory and over 16k CUDA cores. This blocks use on mobile AR glasses or mid-range PCs. To overcome these limitations, future work may focus on: ***(a) Lightweight Models Developments:*** The construction of lightweight model architectures is achieved primarily through two methods: knowledge distillation [53] and quantization-pruning [47]. On the one hand, using knowledge distillation technology, high-quality

lighting knowledge from high-performance models is transferred to small models. A teacher-student framework is adopted, combined with BRDF-based and other lighting loss functions, retaining most of the rendering quality while reducing computational requirements. On the other hand, by applying weight quantization and network pruning methods such as L1 norm-based channel pruning, the number of model parameters can be compressed, thereby enabling the model to run smoothly on edge devices at a rate of 25-30 fps. ***(b) Hardware Adaptation and Heterogeneous Computing:*** In terms of hardware adaptation and heterogeneous computing, cross-platform optimization is required: specifically, leveraging heterogeneous computing frameworks such as CUDA, Vulkan, or Metal API to enhance inference performance on consumer-grade GPUs or embedded chips, while reducing memory usage through block-wise computation and memory-efficient tensor operations [70]. Meanwhile, neural accelerator support should be provided for AR devices by integrating dedicated **neural processing units** (**NPUs**) or **tensor processing units** (**TPUs**) to accelerate matrix operations and convolution operations, thereby achieving real-time rendering at no less than 30 fps [65]. Additionally, in resource-constrained scenarios, a cloud-edge collaborative computing model should be adopted, where computationally intensive initial lighting estimation tasks are offloaded to the cloud, while edge devices only handle real-time inference and post-processing, thus reducing local computation [146].

Limitations in 3D human relighting: Current 3D relighting tasks are limited in in two aspects: ***(i) Single-image 3D Portrait Complexity:*** Specifically, models based on 3DGAN are limited by the geometric diversity and texture complexity of training data, making it difficult to accurately capture non-rigid structures (such as clothing hems) when reconstructing upper-body portraits [33, 90]. Meanwhile, GAN inversion technology tends to cause identity shifts (e.g., distortion of facial features) and loss of details (e.g., skin textures, clothing, and hair strand details) when inferring 3D geometry from a single image [33, 61, 90, 114]. ***(ii) Loose-cloth Relighting:*** Current methods [78, 170, 177] struggle to capture pose-related dynamic wrinkle deformations and illumination effects of non-uniform materials when reconstructing loose clothing. Traditional mesh-based models require high-density triangular meshes to simulate clothing hems or the anisotropic reflections of fabrics, leading to a significant increase in computational costs. Additionally, modeling of light interactions (e.g., clothing shadows, refraction, or subsurface scattering) is insufficient, and existing methods cannot generate realistic light and shadow changes in real time, especially under dynamic light sources. Future works may include: ***(a) Multimodal Data Fusion:*** Integrating multiview images, depth maps, and skeletal pose data to expand the geometric and texture diversity of training datasets, while using synthetic data generators (e.g., Blender) to simulate diverse human poses and clothing types, reducing reliance on real-world data [153]. ***(b) Improving Inversion Technology to Preserve Identity Details:*** On one hand, adopting high-performance inversion frameworks such as Stable Diffusion [4, 108] to enhance the fidelity of facial features and skin textures and reduce the identity shift rate; on the other hand, applying layered inversion to different regions of the portrait (e.g., face, hair, clothing), combining Mask R-CNN-based semantic segmentation [51] to separate regional features, and optimizing geometric and texture details respectively to reduce detail loss in global inversion. ***(c) Developing Physics-based Simulation Models for Loose Clothing Reconstruction:*** For loose clothing reconstruction and lighting optimization, developing physics-based simulation models for loose clothing reconstruction (e.g., cloth simulation based on the finite element method or neural cloth networks), combined with high-fidelity 4D scanning data to accurately capture pose-related wrinkle deformations, while integrating a physically based rendering pipeline combined with microfacet BRDF models (e.g., Disney BRDF [12]) and precomputed radiance transfer [133] to simulate in real time the anisotropic reflections, refractions, and subsurface scattering of clothing, and using real-time ray tracing technology to enhance the realism of loose clothing under dynamic lighting.

Ethical and Social Impact: Realistic human relighting technology poses ethical challenges and risks of misuse. ***(i) The Generation and Dissemination of Deep Fake Content:*** Specifically, advanced generative models and realistic human reconstruction technologies can generate high-resolution fake images, videos, or virtual avatars [63, 89, 124, 179]. These contents, with pixel-level realism and temporal consistency (e.g., lip synchronization, micro-expression simulation), can mislead the public on **virtual reality** (**VR**), **augmented reality** (**AR**), or social media platforms, and may even be used for malicious purposes such as spreading disinformation, identity impersonation, or public opinion manipulation. ***(ii) Privacy and Security of Personal Biometric Data:*** Human reconstruction technologies need to process massive amounts of biometric data (e.g., facial geometry, iris texture, voice features, or motion capture data), involving the collection, storage, and transmission of high-dimensional images, depth maps, and 3D meshes. Improper data management (e.g., unencrypted transmission or storage in insecure cloud environments) can lead to data leakage, triggering threats such as identity theft, privacy infringement, or extortion. The corresponding solutions include: ***(a) Developing Advanced AI-based Detection Algorithms:*** Utilize multimodal analysis [59, 112] (e.g., combining visual, audio and temporal features) and leverage deep learning models such as ResNet classifiers or temporal Transformers to analyze pixel-level anomalies (e.g., boundary artifacts), inconsistent lighting, or lip synchronization deviations, thereby improving detection accuracy. ***(b) Promote Public Education:*** Developing interactive tools such as browser plugins or mobile applications to mark potential fake content in real time, use visual explanations like heatmaps to enhance the public's ability to identify deepfakes, and popularize technical principles and risk knowledge through online education platforms. ***(c) Privacy-preserving End-to-end Security:*** For data privacy and security, it is necessary to adopt end-to-end encryption algorithms throughout the entire data lifecycle and use **secure multiparty computation** (**SMPC**) to achieve distributed data processing, thereby preventing a single node from leaking full data. Meanwhile, implement **role-based access control** (**RBAC**) to restrict access to sensitive data and utilize blockchain technologies, such as Hyperledger Fabric, to record data access logs, thereby enabling traceable and tamper-proof audits that ensure transparency in data usage.

Possibility of Multidisciplinary Research: These aforementioned challenges highlight the need to develop an interdisciplinary roadmap for human relighting. To enhance photorealism, insights from neuroscience and psychology could be integrated to systematically decode the human visual-cognitive pipeline. By quantifying perceptual thresholds and modeling attentional dynamics, an interpretable mapping between pixel-level fidelity and subjective realism might be established–guiding relit results toward a "what you see is what you believe" standard. To address potential ethical and societal risks (e.g., privacy breaches and possible misuse) posed by human relighting, an "ethics-by-design" paradigm could be adopted. Through in-depth collaboration with ethicists and science and technology sociologists, concepts such as informed consent, traceability, and risk controllability can be embedded into both algorithmic architectures and data governance workflows. This would help ensure that technological progress remains grounded in an ethical framework that prioritizes human well-being. Cross-disciplinary synergy thus offers a potential dual safeguard for the robust evolution of human relighting technology, supporting its healthy development from both technical and societal perspectives.

6 Conclusion

This review systematically synthesizes the development context and core challenges in the field of learning-based human relighting. It first provides an overview of the progress of hardware and algorithms in human relighting. Subsequently, it delves into the development of learning-based relighting methods as well as their applications in single-image, 3D, and video human relighting tasks. Building on the above analysis, this review identifies four key challenges in current learning-based

human relighting: Precise simulation of complex lighting effects; Integration of relighting algorithms into real-time systems; 3D human relighting challenges; Ethical and social impacts. To address these issues, this review suggests developing high-quality training datasets and more targeted advanced technologies in the future. In addition, two future trends are anticipated: developing controllable multimodal interactive relighting methods and advancing multidisciplinary collaboration.

References

[1] Amr Almaddah, Sadi Vural, Yasushi Mae, Kenichi Ohara, and Tatsuo Arai. 2012. 2D spherical spaces for face relighting under harsh illumination. *International Journal of Computer and Information Engineering* 6, 9 (2012), 1138–1143.

[2] Dzmitry Bahdanau. 2014. Neural machine translation by jointly learning to align and translate.

[3] Jonathan T. Barron and Jitendra Malik. 2014. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 8 (2014), 1670–1687.

[4] Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, and others. 2025. FLUX.1 Kontext: Flow matching for in-context image generation and editing in latent space.

[5] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. 2023. FLARE: Fast learning of animatable and relightable mesh avatars.

[6] Bharat Lal Bhatnagar, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. 2020. Combining implicit function learning and parametric models for 3d human reconstruction. In *Proceedings of the European Conference on Computer Vision*. Springer, 311–329.

[7] Ananta R. Bhattarai, Matthias Nießner, and Artem Sevastopolsky. 2024. Triplanenet: An encoder for eg3d inversion. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3055–3065.

[8] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, David Kriegman, and Ravi Ramamoorthi. 2020. Deep 3d capture: Geometry and reflectance from sparse multi-view images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5960–5969.

[9] Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. 2018. Demystifying MMD GANs.

[10] James F. Blinn. 1977. Models of light reflection for computer synthesized pictures. In *Proceedings of the Computer Graphics and Interactive Techniques*. 192–198.

[11] Phong Bui-Tuong. 1975. Illumination for computer generated pictures. *CACM* 18, 6 (1975), 311–317.

[12] Brent Burley and Walt Disney Animation Studios. 2012. Physically-based shading at disney. In *ACM SIGGRAPH* 2012, 2012 (2012), 1–7.

[13] I. W. Busbridge. 1960. *The Mathematics of Radiative Transfer*. Cambridge University Press.

[14] Ziqi Cai, Kaiwen Jiang, Shu-Yu Chen, Yu-Kun Lai, Hongbo Fu, Boxin Shi, and Lin Gao. 2024. Real-time 3D-aware portrait video relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6221–6231.

[15] Pol Caselles, Eduard Ramon, Jaime Garcia, Xavier Giro-i Nieto, Francesc Moreno-Noguer, and Gil Triginer. 2023. Sira: Relightable avatars from a single image. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, 775–784.

[16] Per Einarsson, Charles-Felix Chabert, Andrew Jones, Wan-Chun Ma, Bruce Lamond, Tim Hawkins, Mark Bolas, Sebastian Sylwan, and Paul Debevec. 2006. Relighting human locomotion with flowed reflectance fields. In *Proceedings of the 17th Eurographics Conference on Rendering Techniques (EGSR '06)*. Eurographics Association, Nicosia, Cyprus, 183–194.

[17] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J. Guibas, Jonathan Tremblay, Sameh Khamis, et al. 2022. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16123–16133.

[18] Sreenithy Chandran, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Zhixin Shu, and Suren Jayasuriya. 2022. Temporally consistent relighting for portrait videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, 719–728.

[19] Sumit Chaturvedi, Mengwei Ren, Yannick Hold-Geoffroy, Jingyuan Liu, Julie Dorsey, and Zhixin Shu. 2025. SynthLight: Portrait relighting with diffusion model by learning to re-render synthetic faces.

[20] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. 2021. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14124–14133.

[21] Jianqi Chen, Yilan Zhang, Zhengxia Zou, Keyan Chen, and Zhenwei Shi. 2023. Dense pixel-to-pixel harmonization via continuous image representation. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 5 (2023), 3879–3890.

[22] Zhaoxi Chen and Ziwei Liu. 2022. Relighting4d: Neural relightable human from videos. In *Proceedings of the European Conference on Computer Vision*. Springer, Springer, 606–623.
[23] Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. 2020. Implicit functions in feature space for 3d shape reconstruction and completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6970–6981.
[24] Biswarup Choudhury and Sharat Chandran. 2009. A survey of image-based relighting techniques. *Journal of Virtual Reality and Broadcasting* 4, 7 (2009), 176–183.
[25] Wenyan Cong, Xinhao Tao, Li Niu, Jing Liang, Xuesong Gao, Qihao Sun, and Liqing Zhang. 2022. High-resolution image harmonization via collaborative dual transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18470–18479.
[26] Wenyan Cong, Jianfu Zhang, Li Niu, Liu Liu, Zhixin Ling, Weiyuan Li, and Liqing Zhang. 2020. Dovenet: Deep image harmonization via domain verification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8394–8403.
[27] Robert L. Cook and Kenneth E. Torrance. 1982. A reflectance model for computer graphics. *ACM Transactions on Graphics* 1, 1 (1982), 7–24.
[28] Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 886–893.
[29] Paul Debevec. 2000. Pursuing reality with image-based modeling, rendering, and lighting. In *Proceedings of the European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*. Springer, 1–16.
[30] Paul Debevec. 2012. The light stages and their applications to photoreal digital actors. *Proceedings of the SIGGRAPH Asia* 2, 4 (2012), 1–6.
[31] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the Annual Conference on Computer graphics and Interactive Techniques*. 145–156.
[32] Paul Debevec, Andreas Wenger, Chris Tchou, Andrew Gardner, Jamie Waese, and Tim Hawkins. 2002. A lighting reproduction approach to live-action compositing. *ACM Transactions on Graphics* 21, 3 (2002), 547–556.
[33] Boyang Deng, Yifan Wang, and Gordon Wetzstein. 2024. Lumigan: Unconditional generation of relightable 3d human faces. In *Proceedings of the International Conference on 3D Vision*. IEEE, IEEE, 302–312.
[34] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4690–4699.
[35] Julie Dorsey, James Arvo, and Donald Greenberg. 1995. Interactive design of complex time dependent lighting. *IEEE Computer Graphics and Applications* 15, 2 (1995), 26–36.
[36] Bernhard Egger, Sandro Schönborn, Andreas Schneider, Adam Kortylewski, Andreas Morel-Forster, Clemens Blumer, and Thomas Vetter. 2018. Occlusion-aware 3d morphable models and an illumination prior for face image analysis. *International Journal of Computer Vision* 126, 12 (2018), 1269–1287.
[37] Farshad Einabadi, Jean-Yves Guillemaut, and Adrian Hilton. 2021. Deep neural models for illumination estimation and relighting: A survey. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 315–331.
[38] David Futschik, Kelvin Ritland, James Vecore, Sean Fanello, Sergio Orts-Escolano, Brian Curless, Daniel Sỳkora, and Rohit Pandey. 2023. Controllable light diffusion for portraits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8412–8421.
[39] Elena Garces, Carlos Rodriguez-Pardo, Dan Casas, and Jorge Lopez-Moreno. 2022. A survey on intrinsic images: Delving deep into lambert and beyond. *International Journal of Computer Vision* 130, 3 (2022), 836–868.
[40] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 6 (2001), 643–660.
[41] Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. In *Proceedings of the SIGGRAPH Asia*. 1–10.
[42] Michael Goesele, Brian Curless, and Steven M. Seitz. 2006. Multi-view stereo revisited. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2402–2409.
[43] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. 2010. Multi-pie. *Image and Vision Computing* 28, 5 (2010), 807–813.
[44] Yudong Guo, Keyu Chen, Sen Liang, Yong-Jin Liu, Hujun Bao, and Juyong Zhang. 2021. Ad-nerf: Audio driven neural radiance fields for talking head synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5784–5794.
[45] Zonghui Guo, Dongsheng Guo, Haiyong Zheng, Zhaorui Gu, Bing Zheng, and Junyu Dong. 2021. Image harmonization with transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14870–14879.

[46] Zonghui Guo, Haiyong Zheng, Yufeng Jiang, Zhaorui Gu, and Bing Zheng. 2021. Intrinsic image harmonization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16367–16376.
[47] Song Han, Huizi Mao, and William J Dally. 2015. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding.
[48] Yuxuan Han, Zhibo Wang, and Feng Xu. 2023. Learning a 3d morphable face reflectance model from low-cost data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8598–8608.
[49] Richard Hartley and Andrew Zisserman. 2003. *Multiple View Geometry in Computer Vision*. Cambridge university press.
[50] Tim Hawkins, Andreas Wenger, Chris Tchou, Andrew Gardner, Fredrik Göransson, and Paul E. Debevec. 2004. Animatable facial reflectance fields. *Rendering Techniques* 4 (2004), 309–320.
[51] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*. 2961–2969.
[52] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems* 30 (2017), 6629–6640.
[53] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network.
[54] Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *Proceedings of the International Conference on Pattern Recognition*. IEEE, 2366–2369.
[55] Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. 2022. Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4217–4226.
[56] Andrew Hou, Ze Zhang, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. 2021. Towards high fidelity face relighting with realistic shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Computer Vision Foundation / IEEE, 14719–14728.
[57] Buzhen Huang, Yuan Shu, Tianshu Zhang, and Yangang Wang. 2021. Dynamic multi-person mesh recovery from uncalibrated multi-view cameras. In *Proceedings of the International Conference on 3D Vision*. IEEE, 710–720.
[58] Loc Huynh, Bipin Kishore, and Paul Debevec. 2021. A new dimension in testimony: Relighting video with reflectance field exemplars.
[59] Muhammad Javed, Zhaohui Zhang, Fida Hussain Dahri, and Teerath Kumar. 2025. Enhancing multimodal deepfake detection with local–global feature integration and diffusion models. *Signal, Image and Video Processing* 19, 5 (2025), 1–9.
[60] Chaonan Ji, Tao Yu, Kaiwen Guo, Jingxin Liu, and Yebin Liu. 2022. Geometry-aware single-image full-body human relighting. In *Proceedings of the European Conference on Computer Vision*. Springer, 388–405.
[61] Kaiwen Jiang, Shu-Yu Chen, Hongbo Fu, and Lin Gao. 2023. Nerffacelighting: Implicit and disentangled face lighting representation leveraging generative prior in neural radiance fields. *ACM Transactions on Graphics* 42, 3 (2023), 1–18.
[62] Yifan Jiang, He Zhang, Jianming Zhang, Yilin Wang, Zhe Lin, Kalyan Sunkavalli, Simon Chen, Sohrab Amirghodsi, Sarah Kong, and Zhangyang Wang. 2021. Ssh: A self-supervised framework for image harmonization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4832–4841.
[63] Zeren Jiang, Shaofei Wang, and Siyu Tang. 2025. DNF-Avatar: Distilling neural fields for real-time animatable avatar relighting.
[64] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. 2024. Neural gaffer: Relighting any object via diffusion. In *Proceedings of the Annual Conference on Neural Information Processing Systems*.
[65] Norman P. Jouppi, Cliff Young, Nishant Patil, David Patterson, Gaurav Agrawal, Raminder Bajwa, Sarah Bates, Suresh Bhatia, Nan Boden, Al Borchers, et al. 2017. In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th Annual International Symposium on Computer Architecture*. 1–12.
[66] James T. Kajiya. 1986. The rendering equation. In *Proceedings of Computer Graphics and Interactive Techniques*. 143–150.
[67] James T. Kajiya and Timothy L. Kay. 1989. Rendering fur with three dimensional textures. *ACM Siggraph Computer Graphics* 23, 3 (1989), 271–280.
[68] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* 42, 4 (2023), 139–1.
[69] Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, and Sanghyun Woo. 2024. 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*. IEEE, 25096–25106.
[70] Jungwon Kim, Seyong Lee, Beau Johnston, and Jeffrey S. Vetter. 2024. IRIS: A performance-portable framework for cross-platform heterogeneous computing. *IEEE Transactions on Parallel and Distributed Systems* 35, 10 (2024), 1796–1809.

[71] Manuel Lagunas, Xin Sun, Jimei Yang, Ruben Villegas, Jianming Zhang, Zhixin Shu, Belén Masiá, and Diego Gutierrez. 2021. Single-image full-body human relighting. In *Proceedings of the 32nd Eurographics Symposium on Rendering, EGSR 2021 - Digital Library Only Track, Saarbrücken, Germany, June 29 - July 2, 2021*. Adrien Bousseau and Morgan McGuire (Eds.), Eurographics Association, 167–177. DOI: https://doi.org/10.2312/SR.20211300

[72] Jean-Henri Lambert. 1760. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Sumptibus viduae Eberhardi Klett, typis Christophori Petri Detleffsen.

[73] Jason Lawrence, Ryan Overbeck, Todd Prives, Tommy Fortes, Nikki Roth, and Brett Newman. 2024. Project starline: A high-fidelity telepresence system. In *Proceedings of the ACM SIGGRAPH*. 1–2.

[74] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[75] Chloe LeGendre, Wan-Chun Ma, Rohit Pandey, Sean Fanello, Christoph Rhemann, Jason Dourgarian, Jay Busch, and Paul Debevec. 2020. Learning illumination from diverse portraits. In *Proceedings of the SIGGRAPH Asia*. 1–4.

[76] Zhiru Li, Chenchu Rong, and Yuanqing Wang. 2024. LEIFR-Net: light estimation for implicit face relight network. *Optics Express* 32, 4 (2024), 4827–4838.

[77] Fanqing Lin, Connor Wilhelm, and Tony Martinez. 2021. Two-hand global 3d pose estimation using monocular rgb. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2373–2381.

[78] Wenbin Lin, Chengwei Zheng, Jun-Hai Yong, and Feng Xu. 2024. Relightable and animatable neural avatars from videos. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 3486–3494.

[79] Xinguo Liu, Yizhou Yu, and Heung-Yeung Shum. 2001. Synthesizing bidirectional texture functions for real-world surfaces. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*. 97–106.

[80] Yang Liu, Alexandros Neophytou, Sunando Sengupta, and Eric Sommerlade. 2021. Relighting images in the wild with a self-supervised siamese auto-encoder. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. IEEE, 32–40. DOI: https://doi.org/10.1109/WACV48630.2021.00008

[81] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2015. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*. 3730–3738.

[82] Zicheng Liu, Ying Shan, and Zhengyou Zhang. 2001. Expressive expression mapping with ratio images. In *Proceedings of the Computer Graphics and Interactive Techniques*. 271–276.

[83] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2023. SMPL: A skinned multi-person linear model. In *Proceedings of the Seminal Graphics Papers: Pushing the Boundaries*. 851–866.

[84] Lingxiao Lu, Jiangtong Li, Junyan Cao, Li Niu, and Liqing Zhang. 2023. Painterly image harmonization using diffusion model. In *Proceedings of the ACM International Conference on Multimedia*. 233–241.

[85] Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. 2021. Pixel codec avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 64–73.

[86] Wan-Chun Ma, Tim Hawkins, Pieter Peers and Charles-F´elix Chabert, Malte Weiss, and Paul E. Debevec. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques* 9, 10 (2007), 2.

[87] Stephen R. Marschner and Donald P. Greenberg. 1997. Inverse lighting for photography. In *Proceedings of the Color and Imaging Conference*. Society of Imaging Science and Technology, 262–265.

[88] Wojciech Matusik, Boris Ajdin, Jinwei Gu, Jason Lawrence, Hendrik P. A. Lensch, Fabio Pellacini, and Szymon Rusinkiewicz. 2009. Printing spatially-varying reflectance. *ACM Transactions on Graphics* 28, 5 (2009), 1–9.

[89] Yiqun Mei, Mingming He, Li Ma, Julien Philip, Wenqi Xian, David M. George, Xueming Yu, Gabriel Dedic, Ahmet Levent Taşel, Ning Yu, et al. 2025. Lux post facto: Learning portrait performance relighting with conditional video diffusion and a hybrid dataset. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 5510–5522.

[90] Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, and Vishal M. Patel. 2024. Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4263–4273.

[91] Abhimitra Meka, Christian Haene, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dourgarian, et al. 2019. Deep reflectance fields: High-quality facial reflectance field inference from color gradient illumination. *ACM Transactions on Graphics* 38, 4 (2019), 1–12.

[92] Abhimitra Meka, Rohit Pandey, Christian Haene, Sergio Orts-Escolano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, et al. 2020. Deep relightable textures: Volumetric performance capture with neural rendering. *ACM Transactions on Graphics* 39, 6 (2020), 1–21.

[93] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* 65, 1 (2021), 99–106.

[94] Gyeongsik Moon, Shunsuke Saito, Weipeng Xu, Rohan Joshi, Julia Buffalini, Harley Bellan, Nicholas Rosen, Jesse Richardson, Mallorie Mize, Philippe De Bree, et al. 2024. A dataset of relighted 3D interacting hands. *Advances in Neural Information Processing Systems* 36 (2024), 17689–17701.

[95] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics* 41, 4 (2022), 1–15.

[96] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas Lehrmann. 2020. Learning physics-guided face relighting under directional light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5124–5133.

[97] Pauline C. Ng and Steven Henikoff. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* 31, 13 (2003), 3812–3814.

[98] Fred E. Nicodemus. 1965. Directional reflectance and emissivity of an opaque surface. *Applied Optics* 4, 7 (1965), 767–775.

[99] Li Niu, Junyan Cao, Wenyan Cong, and Liqing Zhang. 2023. Deep image harmonization with learnable augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7482–7491.

[100] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the Annual Aymposium on User Interface Software and Technology*. 741–754.

[101] Xingang Pan, Xudong Xu, Chen Change Loy, Christian Theobalt, and Bo Dai. 2021. A shading-guided generative implicit model for shape-accurate 3d-aware image synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 20002–20013.

[102] Rohit Pandey, Sergio Orts-Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul E. Debevec, and Sean Ryan Fanello. 2021. Total relighting: Learning to relight portraits for background replacement. *ACM Transactions on Graphics* 40, 4 (2021), 43–1.

[103] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive body capture: 3D hands, face, and body from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10975–10985.

[104] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. 2007. Post-production facial performance relighting using reflectance transfer. *ACM Transactions on Graphics* 26, 3 (2007), 52–es.

[105] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. 2021. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9054–9063.

[106] Hemanth Pidaparthy, Abhay Chauhan, and Pavan Sudheendra. 2024. Multi-level Attention Aggregation for Aesthetic Face Relighting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 4057–4066.

[107] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. 2023. SDXL: improving latent diffusion models for high-resolution image synthesis.

[108] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *Proceedings of the 12th International Conference on Learning Representations*.

[109] Puntawat Ponglertnapakorn, Nontawat Tritrong, and Supasorn Suwajanakorn. 2023. DiFaReli: Diffusion face relighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 22589–22600. DOI: https://doi.org/10.1109/ICCV51070.2023.02070

[110] Arcot J. Preetham, Peter Shirley, and Brian Smits. 1999. A practical analytic model for daylight. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*. 91–100.

[111] Haonan Qiu, Zhaoxi Chen, Yuming Jiang, Hang Zhou, Xiangyu Fan, Lei Yang, Wayne Wu, and Ziwei Liu. 2024. Relitalk: Relightable talking portrait generation from a single video. *International Journal of Computer Vision* 132, 8 (2024), 1–16.

[112] Shavez Mushtaq Qureshi, Atif Saeed, Sultan H. Almotiri, Farooq Ahmad, and Mohammed A. Al Ghamdi. 2024. Deepfake forensics: A survey of digital forensic methods for multimodal deepfake identification on social media. *PeerJ Computer Science* 10 (2024), e2037.

[113] Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*. 497–500.

[114] Anurag Ranjan, Kwang Moo Yi, Jen-Hao Rick Chang, and Oncel Tuzel. 2023. Facelit: Neural 3d relightable faces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8619–8628.

[115] Pramod Rao, Mallikarjun BR, Gereon Fox, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Ayush Tewari, Christian Theobalt, et al. 2022. Vorf: Volumetric relightable faces. In *Proceedings of the British Machine Vision Conference*.

[116] Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun BR, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, et al. 2024. Lite2Relight: 3D-aware Single Image Portrait Relighting. In *Proceedings of the ACM SIGGRAPH*. 1–12.

[117] Pramod Rao, BR Mallikarjun, Gereon Fox, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Fangneng Zhan, Ayush Tewari, Christian Theobalt, et al. 2024. A deeper analysis of volumetric relightable faces. *International Journal of Computer Vision* 132, 4 (2024), 1148–1166.

[118] Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, HyunJoon Jung, Guido Gerig, and He Zhang. 2024. Relightful harmonization: Lighting-aware portrait background replacement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6452–6462.

[119] Tammy Riklin-Raviv and Amnon Shashua. 1999. The quotient image: Class based recognition and synthesis under varying illumination conditions. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 566–571.

[120] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.

[121] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer-assisted Intervention*. Springer, 234–241.

[122] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the International Conference on Computer Vision*. IEEE, 2564–2571.

[123] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. 2019. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2304–2314.

[124] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable gaussian codec avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 130–141.

[125] Kripasindhu Sarkar, Marcel C. Bühler, Gengyan Li, Daoye Wang, Delio Vicini, Jérémy Riviere, Yinda Zhang, Sergio Orts-Escolano, Paulo Gotardo, Thabo Beeler, et al. 2023. LitNeRF: Intrinsic radiance decomposition for high-quality view synthesis and relighting of faces. In *Proceedings of the SIGGRAPH Asia*. ACM, 1–11.

[126] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 519–528.

[127] Soumyadip Sengupta, Brian Curless, Ira Kemelmacher-Shlizerman, and Steven M. Seitz. 2021. A light stage on every desk. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2420–2429.

[128] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. 2018. Sfsnet: Learning shape, reflectance and illuminance of faces in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Computer Vision Foundation / IEEE Computer Society, 6296–6305.

[129] Artem Sevastopolsky, Savva Ignatiev, Gonzalo Ferrer, Evgeny Burnaev, and Victor Lempitsky. 2020. Relightable 3D head portraits from a smartphone video.

[130] Ruizhi Shao, Zerong Zheng, Hongwen Zhang, Jingxiang Sun, and Yebin Liu. 2022. Diffustereo: High quality human reconstruction via diffusion-based stereo using sparse cameras. In *Proceedings of the European Conference on Computer Vision*. Springer, 702–720.

[131] Amnon Shashua and Tammy Riklin-Raviv. 2001. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 2 (2001), 129–139. DOI: https://doi.org/10.1109/34.908964

[132] Sheng Shen, Shujun Xing, Xinzhu Sang, Binbin Yan, Shuang Zhang, Xinhui Xie, and Jiahui Yang. 2024. Portrait relighting for 3D light-field display based on radiance fields. *Optics Communications* 572 (2024), 130920.

[133] Peter-Pike Sloan, Jan Kautz, and John Snyder. 2023. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *Proceedings of the Seminal Graphics Papers: Pushing the Boundaries*. 339–348.

[134] Guoxian Song, Tat-Jen Cham, Jianfei Cai, and Jianmin Zheng. 2021. Half-body portrait relighting with overcomplete lighting representation. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 371–381.

[135] Guoxian Song, Tat-Jen Cham, Jianfei Cai, and Jianmin Zheng. 2022. Real-time shadow-aware portrait relighting in virtual backgrounds for realistic telepresence. In *Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 729–738.

[136] Yukai Song, Guangxin Xu, Xiaoyan Zhang, and Zhijun Zhang. 2024. BiPR-RL: Portrait relighting via bi-directional consistent deep reinforcement learning. *Computer Vision and Image Understanding* 239 (2024), 103889.

[137] Arne Stoschek. 2000. Image-based re-rendering of faces for continuous pose and illumination directions. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 582–587.

[138] Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. 2011. Effect of illumination on automatic expression recognition: A novel 3D relightable facial database. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 611–618.

[139] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. 2019. Single image portrait relighting. *ACM Transactions on Graphics* 38, 4 (2019), 1–12.

[140] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. 2021. NeLF: Neural light-transport field for portrait view synthesis and relighting.

[141] Wenzhang Sun, Yunlong Che, Yandong Guo, and Han Huang. 2023. Neural reconstruction of relightable human model from monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 397–407. DOI: https://doi.org/10.1109/ICCV51070.2023.00043

[142] Xin Suo, Yuheng Jiang, Pei Lin, Yingliang Zhang, Minye Wu, Kaiwen Guo, and Lan Xu. 2021. Neuralhumanfvv: Real-time neural volumetric human performance rendering using rgb cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6226–6237.

[143] Daichi Tajima, Yoshihiro Kanamori, and Yuki Endo. 2021. Relighting humans in the wild: Monocular full-body human relighting with domain adaptation. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 205–216.

[144] Daichi Tajima, Yoshihiro Kanamori, and Yuki Endo. 2024. All-frequency Full-body human image relighting. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, e70007.

[145] Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escolano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. 2022. Volux-gan: A generative model for 3d face synthesis with hdri relighting. In *Proceedings of the ACM SIGGRAPH*. 1–9.

[146] Jingwei Tan, Fagui Liu, Bin Wang, Qingbo Wu, and C. L. Philip Chen. 2025. EC5: Edge–cloud collaborative computing framework with compressive communication. *Future Generation Computer Systems* 166 (2025), 107715.

[147] Linfeng Tan, Jiangtong Li, Li Niu, and Liqing Zhang. 2023. Deep image harmonization in dual color spaces. In *Proceedings of the ACM International Conference on Multimedia*. 2159–2167.

[148] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems* 33 (2020), 7537–7547.

[149] Ayush Tewari, Ohad Fried, Justus Thies, Vincent Sitzmann, Stephen Lombardi, Kalyan Sunkavalli, Ricardo Martin-Brualla, Tomas Simon, Jason Saragih, Matthias Nießner, et al. 2020. State of the art on neural rendering. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 701–727.

[150] Ayush Tewari, Tae-Hyun Oh, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, Christian Theobalt, et al. 2021. Monocular reconstruction of neural face reflectance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4791–4800.

[151] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang. 2017. Deep image harmonization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3789–3797.

[152] Ching-Ting Tu, Hwei-Jen Lin, and Chin-Yu Chang. 2022. Lighting-and personal characteristic-aware markov random field model for facial image relighting system. *IEEE Access* 10 (2022), 20432–20444.

[153] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. 2017. Learning from synthetic humans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 109–117.

[154] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. 2022. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 5481–5490.

[155] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. 2007. Microfacet models for refraction through rough surfaces. *Rendering Techniques* 2007 (2007), 18th.

[156] Jiaping Wang, Peiran Ren, Minmin Gong, John Snyder, and Baining Guo. 2009. All-frequency rendering of dynamic, spatially-varying reflectance. In *Proceedings of the SIGGRAPH Asia*. 1–10.

[157] Ke Wang, Michaël Gharbi, He Zhang, Zhihao Xia, and Eli Shechtman. 2023. Semi-supervised parametric real-world image harmonization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5927–5936.

[158] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction.

[159] Shaofei Wang, Bozidar Antic, Andreas Geiger, and Siyu Tang. 2024. Intrinsicavatar: Physically based inverse rendering of dynamic humans from monocular videos via explicit ray tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1877–1888.

[160] Youjia Wang, Kai He, Taotao Zhou, Kaixin Yao, Nianyi Li, Lan Xu, and Jingyi Yu. 2023. Free-view face relighting using a hybrid parametric neural model on a small-olat dataset. *International Journal of Computer Vision* 131, 4 (2023), 1002–1021.

[161] Yifan Wang, Aleksander Holynski, Xiuming Zhang, and Xuaner Zhang. 2023. Sunstage: Portrait reconstruction and relighting using the sun as a light stage. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20792–20802.

[162] Zhou Wang, Alan C. Bovik, Hamid R Sheikh, and Eero P. Simoncelli. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.

[163] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. 2020. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM Transactions on Graphics* 39, 6 (2020), 1–13.

[164] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics* 24, 3 (2005), 756–764.

[165] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. 2018. A light CNN for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security* 13, 11 (2018), 2884–2896.

[166] Junjin Xiao, Qing Zhang, Zhan Xu, and Wei-Shi Zheng. 2024. NECA: Neural customizable human avatar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20091–20101.

[167] Rengan Xie, Kai Huang, In-Young Cho, Sen Yang, Wei Chen, Hujun Bao, Wenting Zheng, Rong Li, and Yuchi Huo. 2024. ReN human: Learning relightable neural implicit surfaces for animatable human rendering. *ACM Transactions on Graphics* 43, 5 (2024), 1–22.

[168] Yazhou Xing, Yu Li, Xintao Wang, Ye Zhu, and Qifeng Chen. 2022. Composite photograph harmonization with complete background cues. In *Proceedings of the ACM International Conference on Multimedia*. 2296–2304.

[169] Yingyan Xu, Gaspard Zoss, Prashanth Chandran, Markus Gross, Derek Bradley, and Paulo Gotardo. 2023. Renerf: Relightable neural radiance fields with nearfield lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 22581–22591.

[170] Zhen Xu, Sida Peng, Chen Geng, Linzhan Mou, Zihan Yan, Jiaming Sun, Hujun Bao, and Xiaowei Zhou. 2024. Relightable and animatable neural avatar from sparse-view video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 990–1000.

[171] Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. 2022. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *ACM Transactions on Graphics* 41, 6 (2022), 1–21.

[172] Fei Yin, Yong Zhang, Xuan Wang, Tengfei Wang, Xiaoyu Li, Yuan Gong, Yanbo Fan, Xiaodong Cun, Ying Shan, Cengiz Oztireli, et al. 2023. 3d gan inversion with facial symmetry prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 342–351.

[173] Kanamori Yoshihiro. 2018. Relighting humans: Occlusion-aware inverse rendering for full-body human images. *ACM Transactions on Graphics* 37, 6 (2018), 270–1.

[174] Tao Yu, Zerong Zheng, Kaiwen Guo, Pengpeng Liu, Qionghai Dai, and Yebin Liu. 2021. Function4d: Real-time human volumetric capture from very sparse consumer rgbd sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5746–5756.

[175] Ziyang Yuan, Yiming Zhu, Yu Li, Hongyu Liu, and Chun Yuan. 2023. Make encoder great again in 3d gan inversion through geometry and occlusion-aware encoding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2437–2447.

[176] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. 2024. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In *Proceedings of the ACM SIGGRAPH*. 1–12.

[177] Youyi Zhan, Tianjia Shao, He Wang, Yin Yang, and Kun Zhou. 2025. Interactive rendering of relightable and animatable gaussian avatars. *IEEE Transactions on Visualization and Computer Graphics* 31, 10 (2025), 8491–8502.

[178] Aijia Zhang, Weiqiang Jia, Zhiguo Wan, Wei Hua, and Zisong Zhao. 2024. Virtual lighting environment and real human fusion based on multiview videos. *Information Fusion* 103, C (2024), 102090.

[179] Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Kangjie Chen, Minghan Qin, Yu Li, and Haoqian Wang. 2025. HRAvatar: High-quality and relightable gaussian head avatar. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 26285–26296.

[180] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2025. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *Proceedings of the 13th International Conference on Learning Representations*.

[181] Longwen Zhang, Chuxiao Zeng, Qixuan Zhang, Hongyang Lin, Ruixiang Cao, Wei Yang, Lan Xu, and Jingyi Yu. 2022. Video-driven neural physically-based facial asset for production. *ACM Transactions on Graphics* 41, 6 (2022), 1–16.

[182] Lin Zhang, Lei Zhang, and Alan C. Bovik. 2015. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing* 24, 8 (2015), 2579–2591.

[183] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. 2021. Neural video portrait relighting in real-time via consistency modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 802–812.

[184] Qian Zhang, Qing Guo, Ruijun Gao, Felix Juefei-Xu, Hongkai Yu, and Wei Feng. 2024. Adversarial relighting against face recognition. *IEEE Transactions on Information Forensics and Security* 19 (2024), 9145–9157.

[185] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 586–595.

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

[187] Xiaoyan Zhang, Yukai Song, Zhuopeng Li, and Jianmin Jiang. 2021. PR-RL: Portrait relighting via deep reinforcement learning. *IEEE Transactions on Multimedia* 24 (2021), 3240–3255.

[188] Yizhong Zhang, Jiaolong Yang, Zhen Liu, Ruicheng Wang, Guojun Chen, Xin Tong, and Baining Guo. 2022. Virtualcube: An immersive 3d video communication system. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2146–2156.

[189] Yuxin Zhang, Dandan Zheng, Biao Gong, Jingdong Chen, Ming Yang, Weiming Dong, and Changsheng Xu. 2024. LumiSculpt: enabling consistent portrait lighting in video generation.

[190] Yiqun Zhao, Chenming Wu, Binbin Huang, Yihao Zhi, Chen Zhao, Jingdong Wang, and Shenghua Gao. 2024. Surfel-based gaussian inverse rendering for fast and relightable dynamic human reconstruction from monocular video.

[191] Ruichen Zheng, Peng Li, Haoqian Wang, and Tao Yu. 2023. Learning visibility field for detailed 3D human reconstruction and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 216–226. DOI: https://doi.org/10.1109/CVPR52729.2023.00029

[192] Yang Zheng, Ruizhi Shao, Yuxiang Zhang, Tao Yu, Zerong Zheng, Qionghai Dai, and Yebin Liu. 2021. Deepmulticap: Performance capture of multiple characters using sparse multiview cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6239–6249.

[193] Zangwei Zheng, Xiangyu Peng, Tianji Yang, Chenhui Shen, Shenggui Li, Hongxin Liu, Yukun Zhou, Tianyi Li, and Yang You. 2024. Open-sora: Democratizing efficient video production for all. arXiv:2412.20404. Retrieved from https://arxiv.org/abs/2412.20404

[194] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W. Jacobs. 2019. Deep single-image portrait relighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7194–7202.

[195] Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yu. 2023. Relightable neural human assets from multi-view gradient illuminations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4315–4327.

[196] Jun-Yan Zhu, Philipp Krahenbuhl, Eli Shechtman, and Alexei A. Efros. 2015. Learning a discriminative model for the perception of realism in composite images. In *Proceedings of the IEEE International Conference on Computer Vision*. 3943–3951.

[197] Zuo-Liang Zhu, Zhen Li, Rui-Xun Zhang, Chun-Le Guo, and Ming-Ming Cheng. 2022. Designing an illumination-aware network for deep image relighting. *IEEE Transactions on Image Processing* 31 (2022), 5396–5411.

Received 16 January 2025; revised 4 August 2025; accepted 22 September 2025