中图法分类号:TP37　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)03-0811-15

论文引用格式:Li H Z, Yang Z L, Ding X, Liu Q, Yang Y and Li W. 2026. Image relighting via depth-light direction joint modeling. Journal of Image and Graphics, 31(3):0811-0825(李泓臻, 杨主伦, 丁新, 刘琼, 杨铀, 李伟. 2026. 深度—光源方向联合建模的图像重光照. 中国图象图形学报, 31(3):0811-0825)[DOI:10.11834/jig.250032]

深度—光源方向联合建模的图像重光照

李泓臻[1],杨主伦[1],丁新[1],刘琼[1],杨铀[1*],李伟[2]

1. 华中科技大学电子信息与通信学院,武汉　430074; 2. 维沃移动通信有限公司,上海　310000

摘　要:**目的**　重光照技术在元宇宙、增强现实和计算摄影中有广泛应用。当前,基于漫反射等光照反射模型的重光照方法,存在表达能力有限的问题;基于深度学习的重光照方法通过隐式建模光照过程,具有更丰富的表达能力。但端到端的重光照方法易产生错误的伪影。针对以上重光照方法存在的问题,提出一种深度—光源方向联合建模的图像重光照方法。**方法**　首先,从输入图像中提取深度、法线和漫反射反照率信息,随后将深度作为场景几何表征,使用深度—光源方向联合建模的算法计算遮挡特征,设计TransUNet与U-Net串联的注意力—卷积神经渲染器,通过注意力机制捕获长程依赖关系,并利用卷积融合本征与遮挡特征,最终生成重光照图像。**结果**　对比实验在RSR(real scene relighting)数据集和本文制作的HS(human stage)数据集上与4种重光照方法进行比较。本文方法在RSR数据集中取得了最优的峰值信噪比,结构相似性指数、可学习感知图像块相似度和平均感知得分,相比性能最优的对比方法在峰值信噪比和平均感知得分上分别提升5.45%和2.58%。本文方法在HS数据集上的可学习感知图像块相似度指标上取得了最优结果,且主观效果上更符合人类的直觉。**结论**　本文方法通过引入显式约束和非局部运算,解决了现有端到端重光照方法缺乏准确的投射阴影和表面着色的问题,有效完成了重光照任务。

关键词:图像重光照;阴影生成;注意力机制;神经渲染;深度学习

Image relighting via depth-light direction joint modeling

Li Hongzhen[1], Yang Zhulun[1], Ding Xin[1], Liu Qiong[1], Yang You[1*], Li Wei[2]

1. School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan　430074, China;
2. Vivo Mobile Communication Co., Ltd., Shanghai　310000, China

Abstract: **Objective** The photometric consistency between virtual objects and real-world scenes directly impacts user immersion in applications such as the metaverse and augmented reality. In the film industry and artistic creation, lighting plays a vital role in expressing visual aesthetics and emotional experiences. However, the arrangement of lighting has always been a technical and labor-intensive task. Image relighting, which enables the control of lighting in images, can heavily reduce the workload of professionals in related fields. Existing relighting models based on the Lambertian diffuse reflection model and depth map path tracing perform image relighting by constructing reflectance field models or leveraging computer graphics principles, but these approaches are limited by the expressive capacity of the models and the difficulty of data acquisition. The latest deep learning networks, including convolutional neural networks (CNNs), generative adversarial networks, and diffusion models, have demonstrated exceptional performance in image generation tasks and have been introduced to image relighting. Benefiting from the implicit modeling capabilities of neural networks, end-to-end

收稿日期:2025-03-05;**修回日期**:2025-10-02;**预印本日期**:2025-10-09

*通信作者:杨铀　yangyou@hust.edu.cn

基金项目:国家重点研发计划资助(2024YFC3015302);国家自然科学基金项目(62501245)

Supported by:National Key R&D Program of China(2024YFC3015302);National Natural Science Foundation of China(62501245)

learning-based relighting methods achieve remarkable results in handling complex surface reflection phenomena. However, they perform poorly in addressing cast shadows caused by pixel-to-pixel interactions. The challenges of using deep learning methods for relighting with cast shadows are due to two primary issues. First, neural networks struggle to model cast shadows without prior knowledge or shadow features matching the target lighting condition. Second, most deep learning methods rely on convolution operations, which excel at capturing local features and fusing multichannel features but are less effective at modeling the long-distance dependencies between pixels that are inherent to cast shadows. This study addresses the lack of expressiveness in modeling cast shadows, the difficulty faced by CNN in handling long-distance dependencies, and the inability of traditional reflection models to represent the complex surface reflection properties of objects. Inspired by previous approaches, including those that use computer graphics principles to model relighting, deep learning-based neural rendering methods, and attention mechanisms from natural language processing for capturing long-distance dependencies, this study analyzes past methods and resolves the issue of cast shadow generation in image relighting through a relighting method via depth-light direction joint modeling. **Method** The core features of this method are the analysis of the generation mechanism of cast shadows and the development of an explicit mathematical model, depth-light direction joint modeling algorithm, for their representation, and the design of an advanced neural renderer based on the characteristics of attention mechanisms and CNNs. First, the learning-based method is used to estimate the depth of the image, and an explicit model is constructed using the depth map to predict the feature prior where cast shadows are generated. Given that cast shadows arise from occlusion, this study designed an algorithm leveraging vector inner products to determine cast shadow regions. Additionally, considering that the scene geometry estimated by monocular methods is ill-posed, leading to inaccurate cast shadow predictions, and that explicit models describing surface reflection phenomena are limited in their ability to express complex surface reflections, the proposed neural renderer employs a two-stage, cascaded U-Net-like architecture for neural rendering. The design preserves the U-shaped structure of U-Net while incorporating a multihead attention mechanism to capture long-distance dependencies between pixels, effectively representing cast shadows in conjunction with the depth-light direction joint modeling algorithm. In the second stage, a traditional convolutional U-Net is employed to merge multichannel features such as shadows, textures, brightness, and specular highlights, thereby achieving neural rendering. To ensure effective supervision of the generated images, this study adopts common relighting loss functions and constructs an image pyramid to supervise the generated images at multiple scales. A synthetic dataset called the human stage (HS) dataset was created using the rendering software Blender to evaluate the effectiveness of the proposed method. Unlike previous datasets that used flat surfaces as shadow receivers, the human stage dataset employs horizontal ground and vertical walls as shadow receivers, better reflecting the relighting method's performance in handling cast shadows. Furthermore, training and testing were conducted on the real scene relighting (RSR) dataset to validate the generalization of the proposed method on real-world data. For the real dataset validation, depth and normal were obtained using publicly available pretrained models. Considering the method's applicability across different scenarios, a subset of the RSR dataset was used, excluding objects with colored lighting or metallic materials. **Result** This paper conducts comparative experiments on the real-world dataset and the synthetic dataset against four representative relighting methods, evaluating qualitatively and quantitatively. The qualitative evaluation of relighting images focuses on the following aspects: maintaining the texture of the input image while adjusting brightness according to the target lighting; generating clear cast shadows under the target lighting; and restoring specular highlights experimental results demonstrate that the proposed method achieves the best peak signal-to-noise ratio (PSNR), structure similarity index measure (SSIM), learned perceptual image patch similarity (LPIPS), and mean perceptual score (MPS) on the RSR dataset, with improvements of 5.45% and 2.58% in peak signal-to-noise ratio and the comprehensive metric MPS, respectively, compared with the second-best model. On the synthetic dataset, the method achieves the best LPIPS metric and produces subjectively more intuitive results aligned with human perception. Additionally, ablation studies on the cast shadow algorithm, loss function, and network architecture validate that the proposed algorithm, loss function, and network design effectively enhance the quality of relighting images. **Conclusion** The proposed method effectively addresses the cast shadow generation problem in image relighting and has been validated on synthetic and real-world datasets to produce relighted images with accurate cast shadows. This research introduces a method for predicting shadows using explicit mathematical modeling and proposes leveraging the strengths of attention

mechanisms and convolution operations in provides improvement directions for future research. For instance, in scenarios involving colored lighting and metallic objects, future work should refine the neural renderer architecture and optimize the predefined surface reflectance model on the basis of the linear nature of lighting and reflection models for metallic objects, enhancing relighting performance in more complex lighting scenarios and diverse object types.

Key words: image relighting; shadow generation; attention mechanism; neural rendering; deep learning

0 引　言

图像的光影直接影响着图像质量，同时光影还具有表达画面意境，反映创作情绪的作用。在影视行业中，光影的布置不仅影响着影片的视觉美感，还深刻地影响着观众的情感体验和故事的叙述效果。恰当的光线能够突出主体，塑造空间感，增强画面的层次感，同时也能够营造出特定的氛围和情绪，从而引导观众的情感反应。但在传统影视行业中，灯光布置一直是一个技术性和劳动密集型的环节。随着元宇宙、增强现实的日渐发展，相关技术通过混合虚拟场景和现实场景（祁佳晨 等，2024）、数字人（程龙昊等，2025）等技术为用户提供沉浸感环境体验，而虚拟场景与现实场景的光照不一致，会导致沉浸感、真实感下降。因此对图像中光照维度的控制成为多个行业的迫切需求。

图像重光照是指对给定 RGB 图像的光源特性进行编辑，改变图像中光照的方向、强度后得到与指定光照相匹配的图像的一种技术。根据是否对重光照过程进行反射模型、表面模型和光照模型等具有实际物理意义的模型建模，可以将重光照工作分为基于模型的重光照方法和基于深度学习的端到端重光照方法。

自本世纪开始，已有对基于表面反射模型和显式阴影建模的图像重光照的研究。Debevec 等人（2000）最早为人脸构建了表面反射场模型，使用 Light Stage 系统以像素为单位采集人脸反射场。在使用高动态范围图像（high dynamic range image，HDRI）作为环境光照表达时，使用反射场函数能够计算指定环境光照下的人像并将其合成到新背景中。然而，获取反射场所使用的 Light Stage 系统的构建困难，且采集过程复杂，从而限制了这种重光照方法的应用（杨主伦 等，2025）。随着深度学习的发展，对单幅图像使用深度学习方法获取图像中像素所对应的表面几何、材质等并使用计算机图形学、神经渲染等方法获得重光照图像的工作相继提出。Sengupta 等人（2018）提出的 SfSNet（structure from shading），受到朗伯渲染模型的启发，将输入的人像 RGB 照片分解为法线贴图、漫反射反照率和球谐光照，使用低阶球谐系数作为光照信息，并使用法线和光照进行渲染重建。SfSNet 对人脸表面反射模型进行建模，因此在不产生明显的投射阴影的低频光照下，能够达到较好的重建效果。而 SfSNet 对在环境中存在主导的光源，即存在明显的投射阴影时，SfSNet 的重建质量将大幅下降。此后 Pandey 等人（2021）和 Kim 等人（2024）使用了更复杂的表面反射模型。Pandey 等人（2021）在漫反射模型的基础上增加了镜面反射分量，Kim 等人（2024）则使用 Cook-Torrance 表面反射模型，能够更好地反映光线的反射结果。但上述对表面反射模型建模的方法在投射阴影上均不能取得令人满意的结果，这是因为表面反射模型不能对像素间的遮挡关系建模，因此不能产生由于遮挡关系产生的投射阴影。

在基于表面反射模型的重光照方法中，为解决由遮挡关系产生的投射阴影建模问题，Nestmeyer 等人（2020）使用 U-Net（Ronneberger 等，2015）网络从一幅输入图像中获取漫反射反照率、法线之后，再使用第 2 个 U-Net 网络估计像素对光线的可见性，根据像素的光线可见性，可以为重光照图像产生投射阴影。这项工作在对表面反射模型建模的基础上，使用神经网络中的参数对人脸的投射阴影进行隐式建模。由于这类方法产生的投射阴影是一种数据驱动的结果，缺乏对阴影的显式建模，因此缺乏几何一致性。Hou 等人（2022）提出了一种通过估计深度图对图像进行几何理解，并显式建模投射阴影的方法，该方法在人像数据集上产生的投射阴影取得令人满意的结果。尽管 Hou 等人（2022）方法显式建模了投射阴影，但正如 Tajima 等人（2025）的研究表明，Hou 等人（2022）所构建的深度和点光源显式计算的方法，在扩展到复杂结构，如人体时，该方法生成的阴影是不准确、不完整的。Griffiths 等人（2022）针对户外场

景下,为解决深度图对场景理解的不适定问题,使用路径追踪和3D卷积的方法来处理投射阴影。Nathan和Kansal(2023)将图像深度输入重光照网络,使用Res2Net Squeezed结构(Gao等,2021)捕获长距离依赖,在重光照任务中体现出有效性。Gao等人(2024)在多视点数据上,使用3D Gaussian(Kerbl等,2023)对场景进行三维重建,并设计可微的渲染方法得到令人满意的重光照图像。

基于深度学习的端到端重光照方法,指使用深度神经网络,从图像到图像的重光照方法。Isola等人(2017)提出端到端的图像到图像翻译,适用于图像重光照任务(Wang等,2020)。Sun等人(2019)使用编码器—解码器架构,实现了野外的人脸重光照,实验结论表明该架构能产生非朗伯效应的反射,但在投射阴影和高光反射方面表现较差。Wang等人(2020)将重光照网络构建为场景重建、阴影先验估计和重渲染,并在AIM 2020(advances in image manipulation challenges)(El Helou等,2020b)挑战赛路线1中,取得了最优的峰值信噪比(peak signal-to-noise ratio,PSNR)。Zhu等人(2022)在端到端的重光照方法中,提出一种光照描述模块IARB(illumination aware residual block)近似物理渲染,并在多个重光照任务中达到最优性能。Bhattad等人(2024)在生成式模型StyleGAN(style-based generator architecture for generative adversarial networks)(Karras等,2021)基础上,将图像生成过程中图像中的光照信息解耦,从而获得对生成图像光照的控制,但这种方式无法控制生成图像的内容,因此无法对用户所指定的图像重光照。Ponglertnapakorn等人(2023)使用条件扩散模型,在使用神经渲染网络简化重光照过程的同时,具有对生成内容的可控制性,但仍会表现出物理上的不准确,例如投射阴影与目标光照不匹配。Zhang等人(2024)使用类自编码器的结构,通过解耦和耦合图像本征和图像光照跳过了逆渲染过程,实现图像重光照。

因此,过往图像重光照的研究表明,基于模型的重光照方法,得益于对图像中物体表面和几何场景的理解,能一定程度反映表面反射现象,在由物体间交互产生的投射阴影上能取得更好的表现。但是,一方面现实世界中的物体往往不符合严格的朗伯体或镜面反射模型,且不同材质的反射模型差异也较大;另一方面,受限于2D图像到3D场景的不适定性,对于场景几何的理解往往是不完整的,这些都限制了基于模型的重光照方法的应用。而随着深度学习的发展,基于深度学习的端到端重光照方法得益于海量数据,将复杂的表面反射模型使用神经网络隐式表达,并取得了较好的表现,但在投射阴影这类与场景几何相关的现象上表现欠佳。

因此,本文提出一个深度—光源方向联合建模的图像重光照方法,通过显式模型联合建模场景与光源得到投射阴影特征,从而实现对投射阴影较好地表达。本文的主要贡献如下:1)提出了一种显式的深度—光源方向联合建模算法,根据投射阴影的形成机制并使用图像的深度图作为场景几何的表达,建立场景像素间遮挡的显式数学模型,并利用该模型计算与目标光照匹配的投射阴影特征,从而为重光照模型提供先验。2)提出一种使用多头注意力和卷积神经网络(convolutional neural network,CNN)的神经渲染模型,考虑到单目方法对图像几何理解的不完整性,以及卷积操作难以捕获重光照过程中阴影的长距离交互,本文在U型卷积架构中添加了注意力层,以补充非局部信息。3)为验证本文所提出模型的有效性,提出了使用Blender软件合成的,包含深度、法线和漫反射反照率信息真值的HS(human stage)数据集。

1 方 法

1.1 系统概述

本文提出一个端到端、深度—光源方向联合建模的图像重光照方法,本文方法框架如图1所示,其中,输入为一幅RGB彩色图像和一个指定的光照方向,光照方向是一个三维方向矢量。本模型重光照过程共包括3个阶段:本征分解阶段、特征显式计算阶段和神经渲染阶段。

在本征分解阶段,输入图像经过多个本征分解网络得到本征分量,包括漫反射反照率、法线图和深度图;在特征显式计算阶段,提出一个深度—光源方向联合建模算法计算重光照图像的阴影特征和使用基于Phong光照模型计算重光照图像的漫反射图和镜面反射分量,该阶段为神经渲染阶段提供与目标光照匹配的先验特征;在神经渲染阶段,提出先优化阴影特征,再进行神经渲染着色的方法。

1.2 本征分解阶段

本征分解阶段是将一幅输入 RGB 图像作为输入，从中分解出不随光照变化的图像本征，包括漫反射反照率、法线和深度。图像本征将作为后续阶段的输入变量，提供场景几何信息和色彩纹理信息。图像的漫反射反照率是一种在朗伯反射模型中描述色彩的方式，是物体表面不随光照变化色彩特征。图像的法线图是一种局部的几何表征，能与光照方向共同描述表面反射现象。图像的深度图是一种全局的几何表征，在给定光照下能够描绘像素间几何遮挡关系。

对法线和漫反射反照率，本文采用 Pix2Pix（Isola 等，2017）的方法分别训练法线预测模型和漫反射反照率预测模型。对深度图，由于单目深度估计是一个欠定问题，且获取真实数据集的深度十分困难。根据不同类型数据选取了不同的深度估计模型。具体而言，针对合成数据，采用 Pix2Pix 的方法作为深度估计网络，对真实数据集，采用 Depth Pro（Bochkovskiy 等，2024）的预训练模型获取图像深度。

1.3 特征显式计算阶段

特征显式计算阶段的目的是根据法线和深度显式计算在指定光照下场景的着色特征和投射阴影特征，为神经渲染阶段的神经渲染网络提供与目标光照下匹配的着色和投射阴影先验，进而实现更符合光线传播规律的重光照。

1.3.1 基于 Phong 模型的着色特征计算

本文使用简化的 Phong 光照模型计算着色特征，Phong 模型计算式为

$$\boldsymbol{I}_{\text{phong}} = \boldsymbol{I}_a + \boldsymbol{I}_d + \boldsymbol{I}_s \tag{1}$$

式中，$\boldsymbol{I}_a$、$\boldsymbol{I}_d$ 和 $\boldsymbol{I}_s$ 分别代表环境光分量、漫反射分量和镜面反射分量。环境光变量 $\boldsymbol{I}_a$ 为常数，漫反射和镜面反射分量的计算方法为

$$\boldsymbol{I}_d = k_d \cdot I_i \cdot \max(\boldsymbol{l} \cdot \boldsymbol{n}, 0) \tag{2}$$

$$\boldsymbol{I}_s = k_s \cdot \max(\cos(\boldsymbol{v} \cdot \boldsymbol{r}), 0)^{\alpha} \tag{3}$$

式中，$\boldsymbol{l}$、$\boldsymbol{n}$、$\boldsymbol{v}$ 和 $\boldsymbol{r}$ 分别代表光照方向、法线方向、观察方向和镜面反射方向，k_d、I_i、k_s 和 α 分别代表漫反射系数、入射光强度、镜面高光反射系数和反映物体表面光滑程度的常量。为简化模型，本文方法将环境光变量 $\boldsymbol{I}_a$ 设为 0，系数 k_d、I_i 和 k_s 设置为 1，并使用一组不同数量级的常数 $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$ 计算一组镜面反射分量 $\{\boldsymbol{I}_{s0}, \boldsymbol{I}_{s1}, \boldsymbol{I}_{s2}, \boldsymbol{I}_{s3}\}$，以反映不同材料的表面光滑程度而引起的镜面高光。

1.3.2 基于深度—光源方向联合建模计算阴影特征

阴影指光线无法照亮的暗区，根据形成的原因将阴影分为自遮挡阴影和投射阴影。自遮挡阴影是

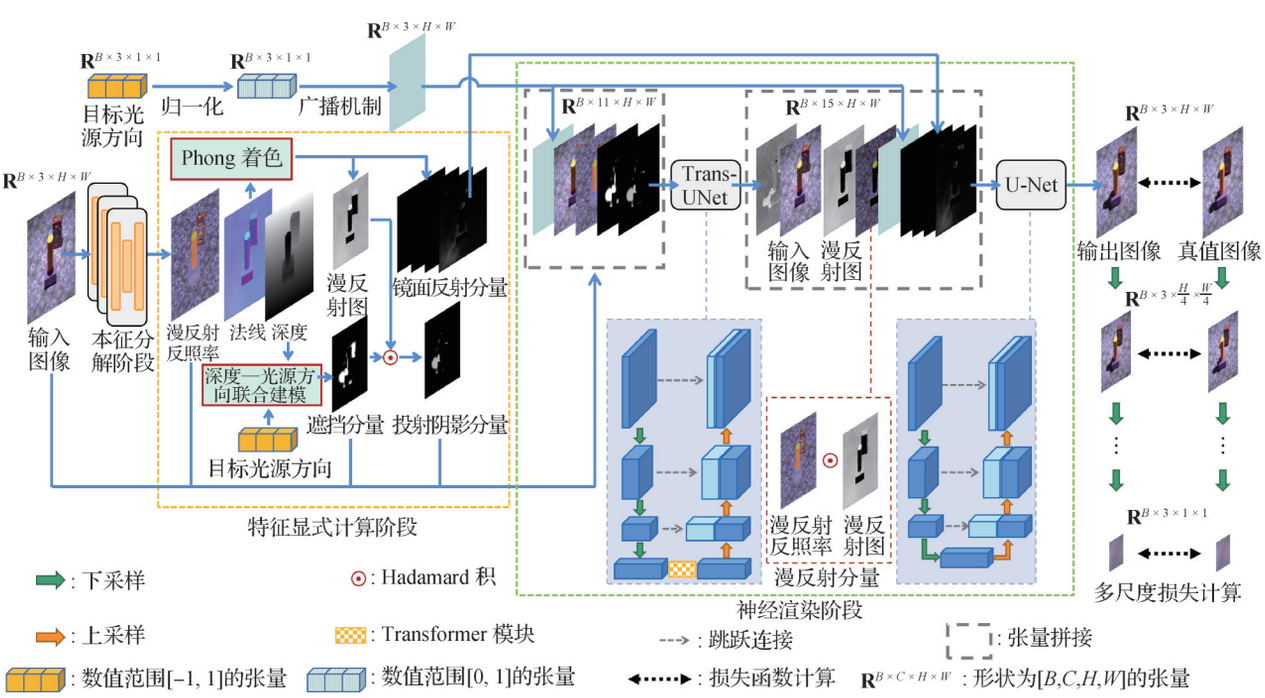

图 1 重光照模型框架

Fig. 1 Framework of relighting model

指由光照方向和物体表面法线夹角过大而形成的暗区，即式(2)中$\boldsymbol{I}_d$小于等于0的区域。投射阴影是指由阴影接收器被阴影遮蔽器遮挡光线所产生的暗区，即$\boldsymbol{I}_d$大于0而物体表面亮度远小于$\boldsymbol{I}_d$的暗区。

本文提出的深度—光源方向联合建模算法基于以下假设：图像中阴影遮蔽器$\boldsymbol{P}_o$和阴影接收器$\boldsymbol{P}_r$均为像素，将像素作为最小的几何单位。因此可以将阴影遮蔽器$\boldsymbol{P}_o$和阴影接收器$\boldsymbol{P}_r$在像素空间$\boldsymbol{I}$中表示为$(x_o, y_o, \boldsymbol{I}(x_o, y_o))$和$(x_r, y_r, \boldsymbol{I}(x_r, y_r))$，$x_*$，$y_*$表示图像中的索引，*为$o$或$r$分别表示阴影遮蔽器或接收器。因此，用一个二值函数$S_{\text{map}}$判断$\boldsymbol{P}_r$是否为$\boldsymbol{P}_i$的阴影接收器，即

$$S_{\text{map}}(\boldsymbol{P}_r, \boldsymbol{P}_i, \boldsymbol{L}_c) = \begin{cases} 1 & \text{遮蔽} \\ 0 & \text{非遮蔽} \end{cases} \tag{4}$$

式中，$\boldsymbol{L}_c$为可控光源，像素空间$\boldsymbol{I}$为不同的特征图像，如深度图。

在深度—光源方向联合建模算法中，像素空间$\boldsymbol{I}$是根据内参矩阵为$\boldsymbol{K}$的标定相机和深度图$\boldsymbol{I}_D$，根据针孔相机模型使用下式逐像素地将深度图$\boldsymbol{I}_D$中的元素映射到的像素空间$\boldsymbol{I}_{3\text{D}} = \{\boldsymbol{p}_{x,y} | (x, y) \in \boldsymbol{I}_D\}$。具体为

$$\boldsymbol{p}_{x,y} = d_{x,y}\boldsymbol{K}^{-1}(x, y, z)^{\text{T}} \tag{5}$$

式中，$(x, y, 1)^{\text{T}}$为齐次坐标下的像素坐标，$\boldsymbol{p}_{x,y} = (X, Y, Z)^{\text{T}}$为位于$(x, y, 1)^{\text{T}}$像素所对应的空间点坐标，$d_{x,y}$表示深度图在$(x, y)$像素的值。此时，$\boldsymbol{p}_o$和$\boldsymbol{p}_r$表示空间中的两点，即$(X_o, Y_o, Z_o)^{\text{T}}$和$(X_r, Y_r, Z_r)^{\text{T}}$，当产生遮蔽关系时，即光源、阴影遮蔽器、阴影接收器共线，如图2所示。而光源、阴影遮蔽器、阴影接收器共线关系可以使用向量内积(,)表达这种关系，即

$$(\boldsymbol{e}_{ro}, \boldsymbol{e}_{rl}) = 1 \tag{6}$$

式中，$\boldsymbol{e}_{ro}$和$\boldsymbol{e}_{rl}$分别表示向量$\boldsymbol{p}_r\boldsymbol{p}_o$和$\boldsymbol{p}_r\boldsymbol{L}_c$的单位化向量。此时，式(4)可以表达为

$$S_{\text{map}}(\boldsymbol{p}_r, \boldsymbol{p}_i, \boldsymbol{L}_c) = \begin{cases} 1 & (\boldsymbol{e}_{ri}, \boldsymbol{e}_{rl}) = 1 \\ 0 & (\boldsymbol{e}_{ri}, \boldsymbol{e}_{rl}) \neq 1 \end{cases} \tag{7}$$

式中，i表示点的索引。如图2所示，由于像素空间$\boldsymbol{I}_{3\text{D}}$是一系列离散点对图像所摄内容的三维表达，而三维空间中的遮挡关系如图2被遮挡光线所示，被遮挡光线穿过几个三维空间点围成的面而未穿过某个点，因此使用式(7)中的共线条件作为投射阴影的判断条件过于严苛，难以判断投射阴影。因此假设阴影遮蔽器是像素空间$\boldsymbol{I}_{3\text{D}}$中的一组点，即$\boldsymbol{P}_o = \{\boldsymbol{p}_{o1}, \boldsymbol{p}_{o2}, \boldsymbol{p}_{o3}, \cdots, \boldsymbol{p}_{on}\}$，并将式(4)中的二值函数$S_{\text{map}}$修改为

$$S_{\text{map}}(\boldsymbol{p}_r, \boldsymbol{p}_i, \boldsymbol{L}_c) = \frac{1}{n}\sum_{j=1}^{n} f_{\text{sigmoid}}((\boldsymbol{e}_{roj}, \boldsymbol{e}_{rl})) \tag{8}$$

$$f_{\text{sigmoid}}(x) = \frac{1}{1 + \text{e}^{-1\,000(x - thr)}} \tag{9}$$

式中，thr为$f_{\text{sigmoid}}(x)$函数的超参，文中值为0.995，并根据$(\boldsymbol{e}_{ro}, \boldsymbol{e}_{rl})$值的大小将$\boldsymbol{p}_o$中$\boldsymbol{p}_{oi}$降序排序，并将前$n$个$\boldsymbol{p}_o$作为$\boldsymbol{P}_o$。考虑到实际计算效率和高分辨率图像计算资源的问题，可以将宽高为W、H的$\boldsymbol{I}_{3\text{D}}$均匀下采样到宽高为$w$、$h$的$\boldsymbol{I}'_{3\text{D}}$，计算阴影特征后再使用双线性插值法还原到原分辨率。故深度—光源方向联合建模算法的流程如下：

1)将$\boldsymbol{I}_{\text{D}}$在标定相机的内参矩阵$\boldsymbol{K}$下映射到$\boldsymbol{I}_{3\text{D}}$，将宽高为$W$、$H$的$\boldsymbol{I}_{3\text{D}}$下采样到宽高为$w$、$h$的$\boldsymbol{I}'_{3\text{D}}$，创建与$\boldsymbol{I}'_{3\text{D}}$同尺寸的$\boldsymbol{I}'_{\text{occlusion}}$。

2)选取$\boldsymbol{I}'_{3\text{D}}$中元素$\boldsymbol{I}'_{3\text{D}}(x, y)$记做$\boldsymbol{p}_{x,y}$，计算$\boldsymbol{e}_{rl} = \dfrac{\boldsymbol{L}_c}{\|\boldsymbol{L}_c\|_2}$或$\boldsymbol{e}_{rl} = \dfrac{\boldsymbol{L}_c - \boldsymbol{p}_{x,y}}{\|\boldsymbol{L}_c - \boldsymbol{p}_{x,y}\|_2}$($\boldsymbol{L}_c$代表平行光方向或点光源坐标)。

3)遍历$\boldsymbol{I}'_{3\text{D}}$中元素$\boldsymbol{p}_i$，并计算向量$\boldsymbol{e}_{roi}$。具体为

$$\boldsymbol{e}_{roi} = \frac{\boldsymbol{p}_i - \boldsymbol{I}_{3\text{D}}(x,y)}{\|\boldsymbol{p}_i - \boldsymbol{I}_{3\text{D}}(x,y)\|_2} \tag{10}$$

4)计算向量内积。

$$\boldsymbol{DOT} = \{dot_i | dot_i = (\boldsymbol{e}_{roi}, \boldsymbol{e}_{rl}), i \in (1, 2, \cdots, h \cdot w)\} \tag{11}$$

5)将$\boldsymbol{DOT}$中元素降序排序并选取前n个值作为

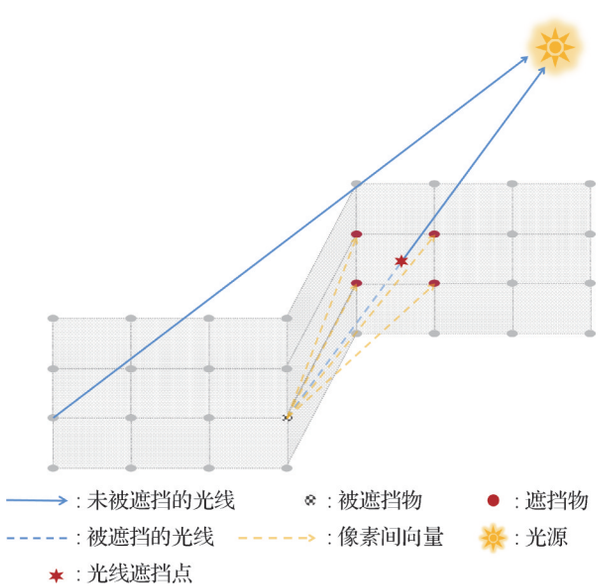

图2　光线遮挡关系示意

Fig. 2　Diagram of light occlusion relationships

集合$\boldsymbol{DOT}' = \{dot'_1, dot'_2, \cdots, dot'_n\}$。

6)计算

$$\boldsymbol{I}'_{\text{occlusion}}(x,y) = \frac{1}{n}\sum_{i=1}^{n} f_{\text{sigmoid}}(dot'_i) \tag{12}$$

7)重复步骤2)—6),直至遍历完$\boldsymbol{I}'_{\text{3D}}$。

8)将$\boldsymbol{I}'_{\text{occlusion}}$采用双线性插值法得到宽高为$W$、$H$的遮挡分量特征图$\boldsymbol{I}_{\text{occlusion}}$。

$\boldsymbol{I}_{\text{occlusion}}$是判断像素在给定光照$\boldsymbol{L}_c$下是否被其他像素遮挡的表征,而根据本文的划分,投射阴影是式(1)中$\boldsymbol{I}_d$大于0而实际物体表面远小于$\boldsymbol{I}_d$的暗区。因此使用$\boldsymbol{I}_d$与$\boldsymbol{I}_{\text{occlusion}}$共同描述投射阴影分量特征图$\boldsymbol{I}_{\text{cast}}$,具体为

$$\boldsymbol{I}_{\text{cast}} = \boldsymbol{I}_d \odot \boldsymbol{I}_{\text{occlusion}} \tag{13}$$

$\boldsymbol{I}_{\text{cast}}$是将$\boldsymbol{I}_d$与$\boldsymbol{I}_{\text{occlusion}}$求哈达玛积(⊙),将原本存在自遮挡现象的区域排除,因此相比$\boldsymbol{I}_{\text{occlusion}}$能更好地描述投射阴影。

1.4 神经渲染阶段

神经渲染阶段的目的是融合着色特征、阴影特征、图像本征和目标光照条件,得到重光照图像。本文采用深度学习模型修正阴影特征$\boldsymbol{I}_{\text{cast}}$,修正后的特征作为阴影先验参与重光照,这使重光照模型受到必要的约束并能生成完整的阴影。本文提出一个名为ShadowRefineNet的神经渲染模型,其作用是补充修正阴影特征和融合不同反射现象计算出的特征图,进而得到重光照图像。

1.4.1 网络框架

ShadowRefineNet的架构为两个串联的U形网络,第1个U形网络TransUNet(Chen等,2021)的作用是补充阴影特征;第2个U形网络U-Net(Ronneberger等,2015)的作用是融合阴影、漫反射着色、镜面反射以及目标光照,进而得到重光照图像。

由于投射阴影的产生是像素间互遮挡所致,因此投射阴影的产生是一种像素间的长距离交互。传统的CNN采用卷积核捕捉局部特征,因此难以应对像素间的长距离交互,而注意力机制能够有效地捕获全局信息,适合生成投射阴影这种像素间长距离交互产生的特征,因此本文采用TransUNet架构实现阴影补充修正的功能。该架构在保持传统U-Net下采样、上采样和跳跃连接的结构同时,在编码器编码结构和解码器解码结构之前嵌入基于多头注意力机制的Tranfomer架构,进而更好地捕捉像素间的长距离交互。一方面,投射阴影需要目标光照、投射阴影"草图",即图1中投射阴影分量和遮挡分量;另一方面,判断像素间遮挡实际上是判断像素所描述的物体间遮挡,这需要描述物体的语义信息,而输入图像和漫反射反照率均包含语义信息。因此TransUNet的输入是光照方向、漫反射反照率、输入图像、遮挡分量和投射阴影分量。其中,光照方向的输入方式是根据下式归一化的光照方向。具体为

$$\boldsymbol{I}_{lc} = 0.5 \cdot (\boldsymbol{e}_{\boldsymbol{L}_c} + \boldsymbol{1}) \tag{14}$$

式中,$\boldsymbol{e}_{\boldsymbol{L}_c}$是目标光照方向$\boldsymbol{L}_c$的单位化向量。

重光照图像是表面反射现象和阴影遮蔽现象的共同结果,因此生成的重光照图像是将多种特征融合的过程。本文采用U-Net网络融合表面反射特征和投射阴影特征。表面反射现象是场景中纹理、漫反射、镜面高光以及全局光照的综合反映,因此,U-Net的输入包括TransUNet的输出特征图、输入图像、漫反射图、漫反射分量、光照方向和镜面反射分量。为表征不同物体的表面光滑程度,镜面反射分量是由一组不同数量级的常数$\{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$计算得到的一组镜面反射分量$\{\boldsymbol{I}_{s0}, \boldsymbol{I}_{s1}, \boldsymbol{I}_{s2}, \boldsymbol{I}_{s3}\}$。

1.4.2 损失函数

考虑到神经渲染会在不同尺度上产生伪影,本文使用双线性插值方法将神经渲染输出图像$\boldsymbol{I}_{\text{pred}}$和真值图像$\boldsymbol{I}_{\text{gt}}$逐次下采样,进而得到不同尺度下的图像$\{\boldsymbol{I}^0_{\text{pred}}, \boldsymbol{I}^1_{\text{pred}}, \cdots, \boldsymbol{I}^4_{\text{pred}}\}$和$\{\boldsymbol{I}^0_{\text{gt}}, \boldsymbol{I}^1_{\text{gt}}, \cdots, \boldsymbol{I}^4_{\text{gt}}\}$,在不同尺度下分别计算损失函数$L^l$,其中,$\boldsymbol{I}^l_{\text{pred}}$和$\boldsymbol{I}^l_{\text{gt}}$的尺度为输入图像的$1/4^l$。

在第l层,神经渲染阶段输出图像$\boldsymbol{I}^l_{\text{pred}}$与目标光照下的真值图像$\boldsymbol{I}^l_{\text{gt}}$使用的损失函数$L^l(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}})$为

$$\begin{aligned} L^l(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) = {} & \lambda^l_{\text{RGB}} L1_{\text{RGB}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) + \\ & \lambda^l_{\text{HSV}} L1_{\text{HSV}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) + \lambda^l_{\text{grad}} L_{\text{grad}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) + \\ & \lambda^l_{\text{SSIM}} L_{\text{SSIM}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) + \lambda^l_{\text{lpips}} L_{\text{lpips}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) \end{aligned} \tag{15}$$

式中,$L1_{\text{RGB}}$损失和$L1_{\text{HSV}}$损失分别是$\boldsymbol{I}^l_{\text{pred}}$和$\boldsymbol{I}^l_{\text{gt}}$在RGB和HSV色彩空间下的平均绝对误差(mean absolute error,MAE),在两种颜色空间中计算损失以最小化输出图像的颜色误差和亮度误差。L_{grad}损失的计算方法为

$$L_{\text{grad}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) = \frac{1}{H^l \cdot W^l \cdot 3}\left\| \boldsymbol{C}(\boldsymbol{I}^l_{\text{pred}}) - \boldsymbol{C}(\boldsymbol{I}^l_{\text{gt}}) \right\|_1 \tag{16}$$

式中,H^l和W^l表示图像的高和宽,$\boldsymbol{C}(\boldsymbol{I})$表示图像$\boldsymbol{I}$经过Canny算子得到的梯度图,计算$\boldsymbol{I}^l_{\text{pred}}$和$\boldsymbol{I}^l_{\text{gt}}$之间的梯度损失以保持输出图像的高频细节。$L_{\text{SSIM}}$损失(Zhao等,2017)的计算式为

$$L_{\text{SSIM}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) = 1 - SSIM(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) \tag{17}$$

式中，$SSIM(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}})$为$\boldsymbol{I}^l_{\text{pred}}$和$\boldsymbol{I}^l_{\text{gt}}$之间的结构相似性指数（structural similarity index measure，SSIM），L_{SSIM}损失综合考虑了$\boldsymbol{I}^l_{\text{pred}}$和$\boldsymbol{I}^l_{\text{gt}}$之间的亮度、对比度和结构差异。$L_{\text{lpips}}$损失是采用可学习感知图像块相似度（learned perceptual image patch similarity，LPIPS）指标（Zhang等，2018）衡量图像在特征空间的相似性，计算式为

$$L_{\text{lpips}}(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) = \sum_k \tau_k (\boldsymbol{\phi}_k(\boldsymbol{I}^l_{\text{pred}}) - \boldsymbol{\phi}_k(\boldsymbol{I}^l_{\text{gt}})) \tag{18}$$

式中，$\boldsymbol{\phi}_k$指从第k层网络提取特征，τ_k是缩放每个通道差异的权重。L_{lpips}通过比较$\boldsymbol{I}^l_{\text{pred}}$和$\boldsymbol{I}^l_{\text{gt}}$在特征空间的差异计算损失，这些特征是从预训练的深度网络AlexNet（Krizhevsky等，2017）中提取的。这种方法与传统的基于像素的损失函数（如MAE）不同，更符合人类对图像的相似度感知。λ^l_{RGB}、λ^l_{HSV}、λ^l_{grad}、λ^l_{SSIM}和λ^l_{lpips}分别代表上述损失函数在第l层的权重。在本文中设置为

$$\lambda^l_{\text{SSIM}} = \lambda^l_{\text{lpips}} = \begin{cases} 1 & l = 0, 1 \\ 0 & l = 2, 3, 4 \end{cases} \tag{19}$$

其余系数均为1.0。因此，神经渲染器最终的损失函数为

$$L = \sum_{l=0}^{4} \mu^l L^l(\boldsymbol{I}^l_{\text{pred}}, \boldsymbol{I}^l_{\text{gt}}) \tag{20}$$

式中，μ^l为不同尺度下损失的权重系数，在本文中设置$\mu^l = 1.0$。

1.5 Human Stage数据集

为验证提出方法的有效性，使用Blender渲染软件渲染了一个名为HS（human stage）的独立单光（one light at a time，OLAT）数据集，其中，光源和人体保持固定的距离，并在1/4球面上均匀采样73个采样点作为点光源位置。为增加拍摄物体多样性，将每种场景布置（人物的服饰、姿势、肤色等以及多面体的大小和位姿）称为场景，每种场景的光照是从73个采样点随机采样的10个点光源，并依次点亮这10个点光源，每点亮一个光源同时渲染一幅彩色RGB图像并记录对应的光照方向，最后再导出该场景下的图像本征（深度图、法线图、漫反射反照率和前景遮罩）。HS数据集共使用了900个不同的场景，其中包括315个不同的人体模型以及5个尺度、位姿和材质随机改变的多面体，故HS数据集共包括9 000幅RGB彩色图像以及900个场景的本征图像。部分HS数据集示例如图3所示，原图像分辨率为512×512像素。

(a) RGB彩色图　(b) 法线图　(c) 前景遮罩　(d) 深度图　(e) 漫反射反照率

图3　HS数据集示例

Fig. 3　Some samples of HS dataset ((a) RGB image; (b) normal map; (c) foreground mask; (d) depth map; (e) diffuse albedo)

2　实验与结果

为验证本文方法的有效性，本文实验在合成HS数据集和真实RSR（real scene relighting）数据集（Yang等，2025）上进行对比实验。由于RSR数据集仅包括标定相机和光照方向，缺少必要的图像本征（深度、法线和漫反射反照率），对RSR数据集，本文使用Ikehata（2023）的光度立体方法获取数据的漫反射反照率和法线图作为真值数据。由于光度立体方法对金属、漆面的本征估计不理想，本文筛选出RSR数据集中不包含金属和漆面物体的子集，作为评估使用。筛选后的子集包括3 240幅真实物体的RGB彩色图像。

对真实数据集，使用Pix2Pix方法（Isola等，2017）分别训练单图像漫反射反照率估计模型和单图像法线估计模型，使用Bochkovskiy等人（2024）的单目深度估计方法作为深度估计模型。

对合成数据集，使用 Pix2Pix 方法（Isola 等，2017）分别训练单图像漫反射反照率估计模型、单目图像法线估计模型和单目深度估计模型。

2.1 对比实验

2.1.1 对比算法评价指标

Isola 等人（2017）的工作中提出 Pix2Pix 基于生成对抗网络（Goodfellow 等，2014），是一种经典的图像到图像的转换方法，该方法不仅具有普适性，还在诸多图像处理任务中表现出优秀性能。Zhu 等人（2022）提出了一个照明感知网络 IAN（illumination-aware network），其使用 IARB（illumination-aware residual block）模块近似物理渲染过程实现高质量重光照，并在多个重光照任务中取得了最优性能。Nathan 和 Kansal（2023）提出一个轻量化重光照网络，该网络在编码和解码阶段使用 Res2Net Squeezed 块（Gao 等，2021）捕捉长距离依赖，并将深度图作为输入提供几何信息，在 VIDIT（virtual image dataset for illumination transfer）（El Helou 等，2020a）数据集上与性能最优方法对比中被证明了有效性。Zhang 等人（2024）提出一种数据驱动的重光照方法，使用编码器将图像本征和光照解耦为隐式表征，再使用解码器重新耦合实现重光照，提出图像本征正则化项以提高本征估计的稳定性。

选取 Isola 等人（2017）、Zhu 等人（2022）、Nathan 和 Kansal（2023）和 Zhang 等人（2024）的方法作为对比算法。实验部分采用的评价指标，包括峰值信噪比（PSNR）、结构相似性指数（SSIM）、可学习感知图像块相似度（LPIPS）和 AIM 2020 挑战赛中使用的平均感知得分（mean perceptual score，MPS）（El Helou 等，2020b）。针对不同方法输入的光照表征方式不同的问题，引导重光照的光照条件 $\boldsymbol{I}_{\text{light}}$ 计算为

$$\boldsymbol{I}_{\text{light}} = \frac{\boldsymbol{L}_c + 1}{2} \tag{21}$$

再将 $\boldsymbol{I}_{\text{light}}$ 通过广播机制扩展到与输入图像相同尺寸，最后与输入图像以拼接的方式输入网络。特别地，Zhang 等人（2024）的方法采用参考图像的光照作为参考光源，本文采用白色漫反射球面在 $\boldsymbol{L}_c$ 下渲染图作为参考图像。在 HS 数据集和 RSR 数据集上，均在分辨率 256×256 像素下训练和测试。

2.1.2 定性对比

在 RSR 数据集上的定性结果如图 4 和图 5 所示，分别展示不同重光照在重光照图像光线不受遮挡区域和光线受遮挡区域的光度一致性对比。两者分别反映不同重光照方法在色彩纹理上的表现和重光照下生成阴影的表现。

对重光照图像光线不受遮挡的区域，定性对比结果如图 4 所示，在图 4 局部放大图中，图中物体为白色粗糙球体，对比其重光照结果反映了重光照方法是否符合朗伯定律。Isola 等人（2017）以及 Nathan 和 Kansal（2023）的方法缺乏对光照不变属性的约束，与真值图像相差较大，Zhu 等人（2022）、Zhang 等人（2024）和本文方法与真值图像更为相近。Zhu 等人（2022）的近似物理渲染模块和 Zhang 等人（2024）的本征与光照耦合方法均具有有效性，本文基于 Phong 模型计算的着色特征，在重光照过程中具有有效性。

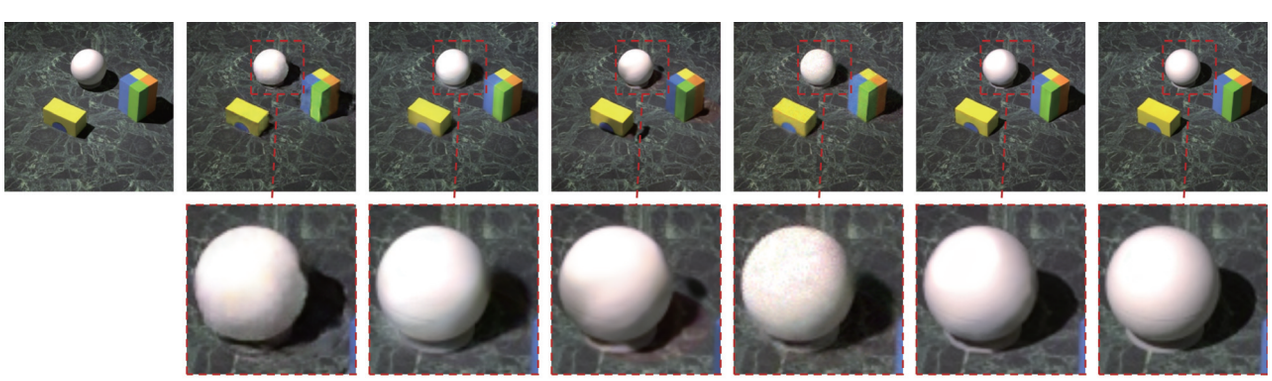

图 4 不同方法在 RSR 数据集上光线无遮挡时光度一致性对比

Fig. 4 Comparison of photometric consistency among different methods under occlusion-free conditions on the RSR dataset

图5旨在评估不同方法在重光照图像中光线受遮挡区域关于重现投射阴影的性能，以局部放大图中的积木车作为示例，能够发现对比算法均出现不同程度的阴影错误生成与边界模糊的问题，而本文方法则能够生成更贴近真值图像的投射阴影效果。

在HS数据集上的对比实验结果如图6所示。如图6(a)所示，只有本文方法在墙壁上生成与真值图像较一致的阴影。对比算法中，Isola等人(2017)和Zhang等人(2024)的方法产生了不准确的伪影，Nathan和Kansal(2023)的方法在人体模型和地面接触区域产生了较准确的阴影，Zhu等人(2022)的方法未能产生阴影。Nathan和Kansal(2023)的方法结果表明，输入深度图和使用Res2Net Squeezed使模型具备一定几何理解能力的作用，进而产生一定的投射阴影。如图6(b)所示，对比方法均生成了投射阴影，其中Isola等人(2017)、Zhu等人(2022)、Nathan和Kansal(2023)以及本文方法产生的阴影较为准确，但在更细节(如手臂遮挡光线产生的投射阴影)方面，得益于深度—光源方向联合建模算法提供先验和隐藏层注意力架构的阴影优化，本文方法与真值图像偏差最小，而对比算法由于更注重重光照图像像素级的重光照效果而非阴影，因此在阴影细节上与本文结果以及真值图像存在差异。

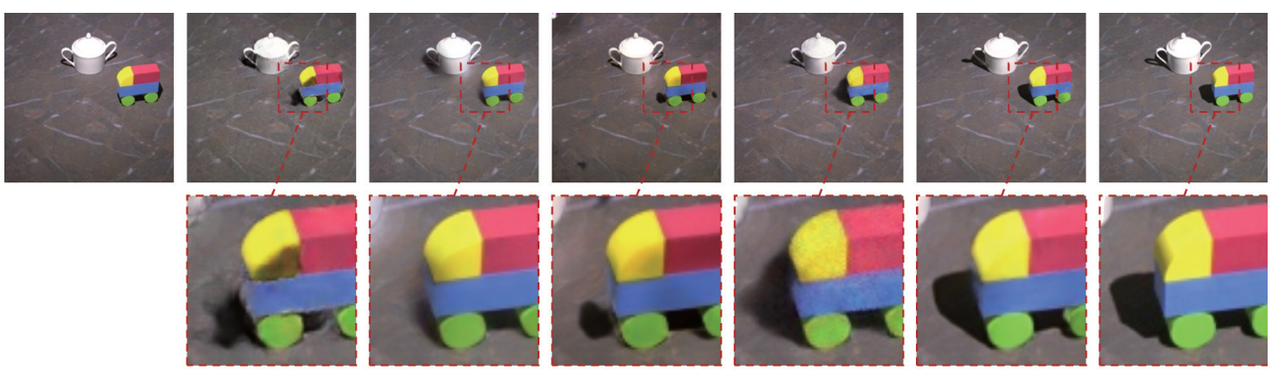

图5　不同方法在RSR数据集上光线受遮挡时光度一致性对比

Fig. 5　Comparison of photometric consistency among different methods under occlusion on the RSR dataset

为体现本文提出的使用注意力机制修正优化投射阴影的方法的有效性，将TransUNet的输入阴影特征图，包括深度—光源方向联合建模算法产生的遮挡分量特征图、经过漫反射着色修正后的投射阴影分量与TransUNet输出的阴影特征以及输出图像进行对比，结果如图7所示。由于深度图表示场景3D结构的欠缺，遮挡分量与投射阴影分量与真值图像中阴影相比，存在阴影区域不完整的问题，而TransUNet所产生的特征图显著解决了输入特征图阴影不完整的问题。以上说明本文使用注意力机制修正优化投射阴影的方法的有效性。

2.1.3　定量对比

本文方法与各对比方法分别在RSR数据集和HS数据集上进行训练和指标计算，两个数据集上的性能指标结果如表1和表2所示。

如表1所示，在RSR数据集上，本文方法在峰值信噪比PSNR、结构相似性SSIM、学习感知图像块相似度LPIPS和MPS均取得最优结果。在对比方法中，Zhang等人(2024)的方法取得了最优的PSNR和LPIPS，Zhu等人(2024)的方法取得了最优的SSIM和MPS。本文方法相比Zhang等人(2024)方法在PSNR上取得5.45%领先，相比Zhu等人(2022)方法在MPS上取得2.58%的领先。

如表2所示，在HS数据集上，本文方法在LPIPS上取得最优，Zhu等人(2022)方法在PSNR、SSIM和MPS上，略优于本文方法，取得了最优结果。从HS数据集上的定性对比中可以看出，本文方法在产生投射阴影方面优于Zhu等人(2022)方法，产生最符合直觉的投射阴影，而在重光照图像像素级准确性方面略逊于Zhu等人(2022)方法。这是由于本文方法更强调投射阴影，因此本文方法在更强调语义一致性的LPIPS上表现最好，由于ShadowRefineNet不是对渲染方程的拟合，而是通过融合特征图生成重光照图像，因此在完美符合渲染方程的合成数据集

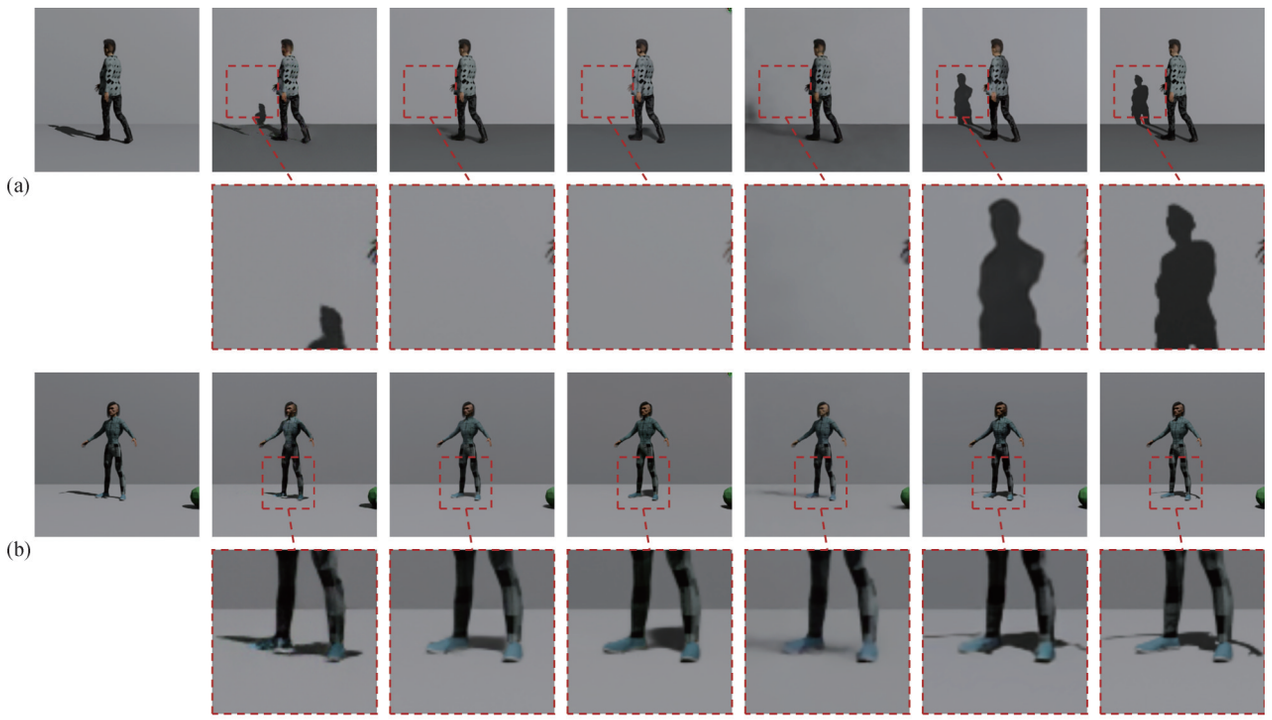

图 6　不同方法在 HS 数据集上的表现

Fig. 6　Performance of different methods on HS dataset

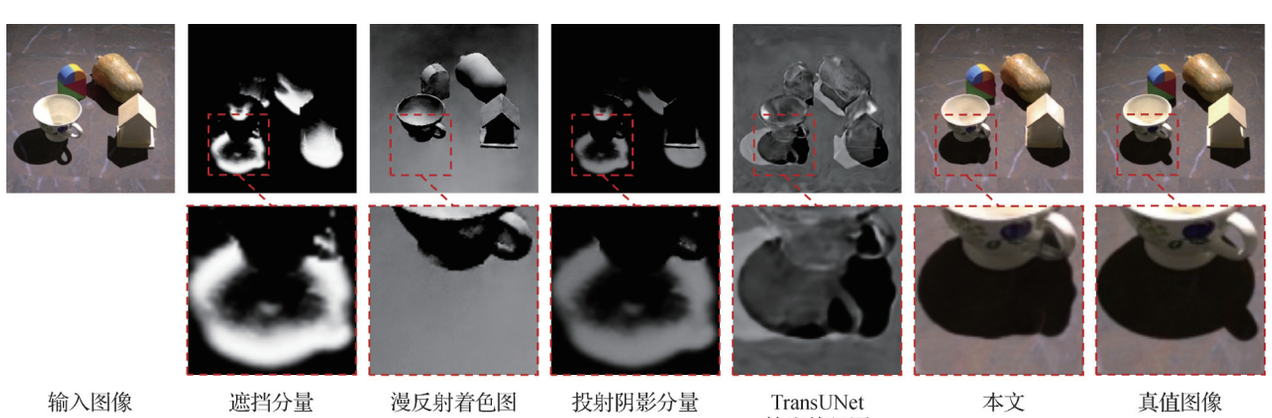

图 7　TransUNet 输入特征、输出特征与重光照图像对比

Fig. 7　Input feature, output feature and relighting image comparison of TransUNet

表 1　不同算法在 RSR 数据集上的性能

Table 1　Performance of different methods on RSR dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓	MPS ↑
Isola 等人(2017)	21.253 5	0.821 5	0.107 3	0.857 1
Zhu 等人(2022)	23.227 7	0.891 2	0.074 0	0.908 6
Nathan 和 Kansal(2023)	18.984 7	0.841 1	0.119 3	0.860 9
Zhang 等人(2024)	23.873 3	0.869 7	0.072 0	0.898 9
本文	**25.174 1**	**0.912 9**	**0.048 7**	**0.932 0**

注：加粗字体表示各列最优结果。“↑”表示值越大越好，“↓”表示值越小越好。

表2　不同算法在HS数据集上的性能指标

Table 2　Performance of different methods on HS dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓	MPS ↑
Isola等人(2017)	23.830 9	0.912 1	0.052 9.	0.929 6
Zhu等人(2022)	**26.182 7**	**0.953 7**	0.037 5	**0.958 1**
Nathan和Kansal(2023)	25.512 4	0.940 9	0.062 5	0.938 5
Zhang等人(2024)	23.307 4	0.926 8	0.061 8	0.932 5
本文	25.558 6	0.936 9	**0.036 3**	0.950 3

注：加粗字体表示各列最优结果。“↑”表示值越大越好，“↓”表示值越小越好。

上像素级对比指标PSNR和SSIM略低于Zhu等人(2022)的方法。

值得注意的是，Nathan和Kansal(2023)方法在RSR数据集和HS数据集上性能差异较大，该方法的输出在输入图像阴影处产生颜色偏差，这是由于RSR数据集色彩纹理更加丰富导致该轻量化方法难以预测阴影区域原本的色彩，因此在色彩风格更单一的HS数据集上，该方法指标明显提升，说明该方法在色彩丰富且存在明显阴影的数据上泛化性能低。

2.2　消融实验

2.2.1　消融实验配置

为验证本文所提出的深度—光源方向联合建模算法、多尺度损失函数和使用TransUNet的对提升重光照效果的有效性，设置了4种配置并在HS数据集上进行消融实验。

1)配置1。在原网络架构基础上，将图1中遮挡分量设置为0，达到消除深度—光源方向联合建模模块的目的，以验证深度—光源方向联合建模算法的有效性。

2)配置2。在原网络架构基础上，将图1中多尺度损失计算保留最顶层，即原分辨率下预测图像与真值图像的损失，也即仅保留式(20)中$l=0$的损失，以验证多尺度损失计算的有效性。

3)配置3。在原网络架构的基础上，将图1中TransUNet对隐藏层特征进行注意力计算的Transformer模块去除，将隐藏层特征直接输入解码器，以验证使用注意力机制对提升重光照效果的有效性。

4)配置4。在原网络架构的基础上，将图1中TransUNet对隐藏层特征进行注意力计算的Transformer模块去除，替换为3×3卷积层，以验证不同架构，即卷积架构和注意力架构，处理隐藏层特征提升阴影细节的差异。

2.2.2　结果分析

本文在保持训练超参数不变的前提下，根据2.2.1节中4种消融实验配置以及本文方法在HS数据集上重新训练模型至收敛。

定性效果对比如图8所示，配置1、配置2产生的阴影与真值图像相差较大，配置3、配置4、本文方法与真值图像相比，本文方法的阴影边界更加准确。

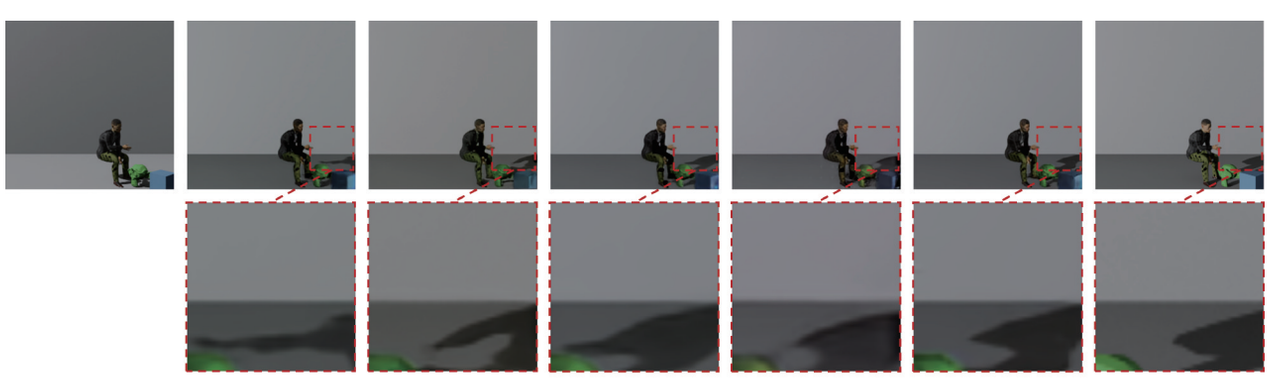

图8　不同配置在HS数据集上的表现

Fig. 8　Performance of different configurations on HS dataset

与配置3的差异表明，注意力模块有助于产生更高质量的阴影。与配置4的差异表明，注意力架构相比卷积架构能够产生更准确的阴影细节。

定量指标对比如表3所示，在HS数据集上，本文方法取得了最优的PSNR，相比第2优的配置2有6.94%的提升，配置2在SSIM、LPIPS和MPS指标上表现最优，但与本文方法相差不大，本文方法的指标综合性能最优。

本文方法在HS数据集上相较所设置的4组不同配置的消融实验，在主观效果和客观指标上均有较大优势。定性和定量结果表明，本文提出的深度—光源方向联合建模算法、使用的多尺度损失函数计算和TransUNet网络中注意力架构在提升重光照图像质量上具有有效性。

表3　在HS数据集上的消融实验指标对比

Table 3　Ablation study on HS dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓	MPS ↑
配置1	22.444 5	0.917 8	0.044 2	0.936 8
配置2	23.900 5	**0.937 5**	**0.029 2**	**0.954 1**
配置3	22.949 8	0.921 4	0.039 5	0.9409
配置4	23.497 6	0.917 5	0.049 8	0.933 9
本文	**25.558 6**	0.936 9	0.036 3	0.950 3

注：加粗字体表示各列最优结果。“↑”表示值越大越好，“↓”表示值越小越好。

2.3　讨论和展望

在研究和实验过程中，遇到了如下限制。1）本文为点光源或平行光源建立了投射阴影模型，但不适用于HDRI和球谐光照这类光照表示。2）在真实数据集上，获取图像本征的困难度大，难以获取图像的深度以及金属物体、透明物体、漆面物体和强镜面反射物体的反照率和法线。这类数据的缺失限制了训练更具泛化性的模型。针对以上限制，展望了未来研究的方向，以解决以上问题。1）使用适当的基函数将HDRI表达的海量光照模式简化为若干个参数。2）采用自监督的方式，通过分解重构图像的方式，训练本征分解网络，减少对标记数据的依赖；另一方面，可以采用合成数据，利用合成数据集的本征信息更易获取的优势，扩充重光照方法的训练数据。

3　结　论

本文提出一种深度—光源方向联合建模的图像重光照方法，可以生成具有高质量阴影的重光照图像。核心优势在于提出的独特的深度—光源方向联合建模算法，该算法不直接为重光照图像计算阴影，而是利用深度图显式计算投射阴影特征，使用神经渲染器ShadowRefineNet融合到重光照图像中。使用神经渲染器融合特征的方式，既能使神经渲染器保留必要的显式约束，弥补神经渲染器难以学习光线传播物理规律的不足，又能利用数据驱动的方式，补充显式方法产生的不完整的阴影特征。本文还提出一个合成数据集HS数据集，有利于评估重光照方法产生的投射阴影。在实验验证方面，本文选取真实RSR数据集和合成HS数据集，通过PSNR，SSIM，LPIPS和MPS指标对本文提出的重光照模型的性能进行全面评估。同时，通过可视化ShadowRefineNet重光照过程中阴影特征图，进一步说明了ShadowRefineNet补充修正阴影特征的有效性。

虽然本文在点光源重光照的投射阴影上取得了较好的性能，但未来还存在拓展空间。如对HDRI的光照表示下重光照建立模型，使用自监督方法提取图像本征，这将有助于扩大重光照方法的适用性和减少对标记数据的依赖。

参考文献（References）

Bhattad A, Soole J and Forsyth D A. 2024. StyLitGAN: image-based relighting via latent control//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4231-4240 [DOI: 10.1109/CVPR52733.2024.00405]

Bochkovskiy A, Delaunoy A, Germain H, Santos M, Zhou Y C, Richter S R, et al. 2024. Depth Pro: sharp monocular metric depth in less than a second//Proceedings of the 13th International Conference on Learning Representations. Singapore, Singapore: OpenReview.net

Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2025-02-20]. https://arxiv.org/pdf/2102.04306.pdf

Cheng L H, Li C H, Hu R Z and Liu L G. 2025. 3D Gaussian splatting model for low-light enhancement. Journal of Image and Graphics, 30(10): 3289-3301 (程龙昊, 李常颢, 胡瑞珍, 刘利刚. 2025.

面向低光增强的三维高斯泼溅模型. 中国图象图形学报, 30(10): 3289-3301)[DOI: 10.11834/jig.240598]

Debevec P, Hawkins T, Tchou C, Duiker H P, Sarokin W and Sagar M. 2000. Acquiring the reflectance field of a human face//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. [s.l.]: ACM Press: 145-156 [DOI: 10.1145/344779.344855]

El Helou M, Zhou R F, Barthas J and Süsstrunk S. 2020a. VIDIT: virtual image dataset for illumination transfer [EB/OL]. [2025-03-05]. https://arxiv.org/pdf/2005.05460.pdf

El Helou M, Zhou R F, Süsstrunk S, Timofte R, Afifi M, Brown M S, et al. 2020b. AIM 2020: scene relighting and illumination estimation challenge//Proceedings of 2020 Workshops on Computer Vision. Graz, Austria: Springer: 499-518 [DOI: 10.1007/978-3-030-67070-2_30]

Gao J, Gu C, Lin Y T, Li Z H, Zhu H, Cao X, et al. 2024. Relightable 3D Gaussians: realistic point cloud relighting with BRDF decomposition and ray tracing//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 73-89 [DOI: 10.1007/978-3-031-72995-9_5]

Gao S H, Cheng M M, Zhao K, Zhang X Y, Yang M H and Torr P. 2021. Res2Net: a new multi-scale backbone architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(2): 652-662 [DOI: 10.1109/TPAMI.2019.2938758]

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. 2014. Generative adversarial nets//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 2672-2680

Griffiths D, Ritschel T and Philip J. 2022. OutCast: outdoor single-image relighting with cast shadows. Computer Graphics Forum, 41(2): 179-193 [DOI: 10.1111/cgf.14467]

Hou A, Sarkis M, Bi N, Tong Y Y and Liu X M. 2022. Face relighting with geometrically consistent shadows//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4207-4216 [DOI: 10.1109/CVPR52688.2022.00418]

Ikehata S. 2023. Scalable, detailed and mask-free universal photometric stereo//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 13198-13207 [DOI: 10.1109/CVPR52729.2023.01268]

Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]

Karras T, Laine S and Aila T. 2021. A style-based generator architecture for generative adversarial networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(12): 4217-4228 [DOI: 10.1109/TPAMI.2020.2970919]

Kerbl B, Kopanas G, Leimkühler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4): #139 [DOI: 10.1145/3592433]

Kim H, Jang M, Yoon W, Lee J, Na D and Woo S. 2024. SwitchLight: co-design of physics-driven architecture and pre-training framework for human portrait relighting//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 25096-25106 [DOI: 10.1109/CVPR52733.2024.02371]

Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90 [DOI: 10.1145/3065386]

Nathan S and Kansal P. 2023. End-to-end depth-guided relighting using lightweight deep learning-based method. Journal of Imaging, 9(9): #175 [DOI: 10.3390/jimaging9090175]

Nestmeyer T, Lalonde J F, Matthews I and Lehrmann A. 2020. Learning physics-guided face relighting under directional light//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5123-5132 [DOI: 10.1109/CVPR42600.2020.00517]

Pandey R, Escolano S O, Legendre C, Häne C, Bouaziz S, Rhemann C, et al. 2021. Total relighting: learning to relight portraits for background replacement. ACM Transactions on Graphics, 40(4): #43 [DOI: 10.1145/3450626.3459872]

Ponglertnapakorn P, Tritrong N and Suwajanakorn S. 2023. DiFaReli: diffusion face relighting//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 22589-22600 [DOI: 10.1109/ICCV51070.2023.02070]

Qi J C, Xie L J, Ruan W K and Wang X Q. 2024. Neural relighting methods for mixed reality flight simulators. Journal of Image and Graphics, 29(10): 3008-3021 (祁佳晨, 解利军, 阮文凯, 王孝强. 2024. 混合现实飞行模拟器的神经重光照方法. 中国图象图形学报, 29(10): 3008-3021) [DOI: 10.11834/jig.230817]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Sengupta S, Kanazawa A, Castillo C D and Jacobs D W. 2018. SfSNet: learning shape, reflectance and illuminance of faces "in the Wild"//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6296-6305 [DOI: 10.1109/CVPR.2018.00659]

Sun T C, Barron J T, Tsai Y T, Xu Z X, Yu X M, Fyffe G, et al. 2019. Single image portrait relighting. ACM Transactions on Graphics, 38(4): #79 [DOI: 10.1145/3306346.3323008]

Tajima D, Kanamori Y and Endo Y. 2025. All-frequency full-body human image relighting. Computer Graphics Forum, 44(2): #e70007 [DOI: 10.1111/cgf.70007]

Wang L W, Siu W C, Liu Z S, Li C T and Lun D P K. 2020. Deep relighting networks for image light source manipulation//Proceedings of 2020 ECCV Workshops on Computer Vision. Glasgow, UK: Springer: 550-567 [DOI: 10.1007/978-3-030-67070-2_33]

Yang Y X, Sial H A, Baldrich R and Vanrell M. 2025. Relighting from a single image: datasets and deep intrinsic-based architecture. IEEE Transactions on Multimedia, 27: 2608-2622 [DOI: 10.1109/TMM.2025.3535397]

Yang Z L, Liu Y B, Ju Y K, Liu Q, Li X T, Yin Y G, et al. 2025. Review of scene relighting methods. Journal of Image and Graphics, 30(6): 1543-1575 (杨主伦, 刘烨斌, 举雅琨, 刘琼, 李旭涛, 尹亚光, 等. 2025. 场景重光照研究综述. 中国图象图形学报, 30(6): 1543-1575) [DOI: 10.11834/jig.240772]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]

Zhang X, Gao W, Jain S, Maire M, Forsyth D A and Bhattad A. 2024. Latent intrinsics emerge from training to relight//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #3068

Zhao H, Gallo O, Frosio I and Kautz J. 2017. Loss functions for image restoration with neural networks. IEEE Transactions on Computational Imaging, 3(1): 47-57 [DOI: 10.1109/TCI.2016.2644865]

Zhu Z L, Li Z, Zhang R X, Guo C L and Cheng M M. 2022. Designing an illumination-aware network for deep image relighting. IEEE Transactions on Image Processing, 31: 5396-5411 [DOI: 10.1109/TIP.2022.3195366]

作者简介

李泓臻，男，硕士研究生，主要研究方向为图像重光照。
E-mail: hongzhen@hust.edu.cn

杨铀，通信作者，男，教授，主要研究方向为视觉感知与计算为核心的计算机视觉、计算摄像学、立体视频系统。
E-mail: yangyou@hust.edu.cn

杨主伦，男，博士研究生，主要研究方向为图像重光照。
E-mail: zlyang3@hust.edu.cn

丁新，男，博士后，主要研究方向为人像重光照。
E-mail: xding07@163.com

刘琼，女，教授，主要研究方向为3D视频处理和理解、交互式多媒体、3D计算机视觉。E-mail: q.liu@hust.edu.cn

李伟，男，工程师，主要研究方向为图像处理和计算机视觉。
E-mail: liwei.yxgh@vivo.com