基于深度学习的真实感渲染研究

(申请清华大学工学博士学位论文)

培 养 单 位： 计算机科学与技术系

学　　科： 计算机科学与技术

研 究 生： 高　　端

指 导 教 师： 徐　　昆　副教授

二〇二二年三月

Neural Rendering: Deep Learning Based Photorealistic Rendering

Dissertation Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

in

Computer Science and Technology

by

Gao Duan

Dissertation Supervisor: Associate Professor Xu Kun

March, 2022

摘 要

真实感渲染是计算机图形学中的核心研究方向之一，在虚拟现实、影视特效制作、电子游戏和设计可视化等领域发挥着重要作用。随着各行各业对三维内容需求的日益增长，研究面向一般用户的轻量化真实感渲染方法变得愈加重要。传统真实感渲染算法主要面向具有丰富领域知识的专家用户且依赖于繁琐的手工操作，因此存在易用性差和自动化程度低等不足。神经渲染，即基于深度学习的真实感渲染，通过将计算机图形学领域知识和基于大数据的深度学习技术相结合，可以实现更加轻量化的解决方案。

神经渲染研究的重点和难点在于如何合理地将深度学习和真实感渲染的领域知识进行结合。本文围绕真实感渲染中的采集和建模、存储和表达以及绘制和可视化三个核心领域进行研究，通过充分挖掘问题本身的特性以找到将深度学习有效融入的方式，并提出了一系列全新的神经渲染方法，具体包括：

1. 提出了一种基于逆渲染和数据驱动的表观建模方法。该方法以轻量化的采集设备（手机相机）所拍摄的若干图片作为输入，其核心思路是在深度自编码器网络构建的隐空间中进行逆渲染优化。基于深度学习的数据先验信息的融入使得整个逆渲染优化过程中无需任何手工设计的启发式约束。该方法可以支持任意分辨率和任意数量的输入图片，并且其重建质量随输入图片数量的增加而不断提升，大大提升了方法适用范围和表观建模效率。

2. 提出了一种基于深度场景表达的重光照方法。该方法以两个手持设备拍摄的无结构化图片为输入，可以实现真实世界复杂场景的自由视点重光照渲染。该方法以神经纹理和辐射亮度信息作为深度场景表达并利用神经渲染网络完成重光照渲染。该方法还提出了光照增强策略来扩展神经渲染网络所支持的光源类型。相较于基于建模的传统方法，该方法大大降低了采集工作量和复杂度，并且可以支持包含复杂表观和全局光照效果的复杂真实场景。

3. 提出了一种基于深度绘制管线的全局光照绘制方法。该方法提出使用深度全连接网络来建模着色点到全局光照的复杂映射，可支持动态面光源下全频率全局光照的快速渲染。得益于该方法提出的神经网络友好的输入表达，深度全连接网络可以在保证紧凑性的同时有效学习复杂的全局光照效果。该方法还提出了一种基于材质划分的加速策略以进一步提升运行效率和降低存储开销。

关键词：神经渲染；真实感渲染；表观建模；重光照；全局光照

Abstract

Photorealistic rendering plays a crucial role in many computer graphics applications such as movie visual effects, 3D video games, virtual reality, design visualization. With the increasing demand for 3D content, it is important to develop lightweight photorealistic rendering approaches for general users. Classic photorealistic rendering methods mainly focus on expert users with rich domain knowledge and require expensive manual intervention, and thus have difficulty in applying to lightweight applications. Neural rendering, deep-learning-based photorealistic rendering, is able to achieve lightweight solutions by combining domain knowledge of computer graphics with deep learning techniques.

How to combine deep learning with knowledge of photorealistic rendering in an efficient way is the main challenge in neural rendering. This dissertation focuses on the three major directions of photorealistic rendering, i.e. acquisition and modeling, storage and representation, and rendering and visualization. This dissertation proposes several efficient neural rendering approaches by exploiting the characteristics of the specific problems first and then integrating suitable deep learning modules. Specifically, this dissertation proposes:

1. A unified framework for estimating high-resolution surface reflectance properties of a spatially-varying planar material sample from an arbitrary number of photographs. This method combines deep learning and inverse rendering in a flexible and easy-to-implement framework that performs the inverse rendering optimization without any handcrafted heuristics in a learned SVBRDF latent space characterized by a fully convolutional auto-encoder. The precision of the estimated appearance scales from plausible approximations when the input images fail to reveal all the reflectance details to accurate reproductions for sufficiently large sets of input images. The proposed unified framework is suitable for estimating high-resolution SVBRDFs from an arbitrary number of input photographs with a lightweight acquisition setup.

2. A novel image-based method for 360˚ free-viewpoint relighting from unstructured photographs of a scene captured with double handheld devices. This method leverages a scene-dependent neural rendering network for relighting a rough geometric proxy with learnable neural textures. The key to making the rendering network lighting-aware are radiance cues: global illumination renderings of a rough proxy geometry of the scene for

a small set of basis materials and lit by the target lighting. This method introduces a novel lighting augmentation strategy that exploits the linearity of light transport to extend the relighting capabilities of the neural rendering network to support other lighting types beyond the lighting used during acquisition. Compared to classic model-based approaches, this method can handle more intricate scenes with a wide variety of material properties and global light transport effects and reduce the data acquisition cost.

3. A carefully designed framework for interactively rendering full global illumination with dynamic area light sources. The complex mapping from the input of each shading point to global illumination is modeled by a deep fully-connected network that is well-suited to approximate such complex mapping. The neural-network-friendly input representation plays a crucial role in reducing the requirement of network size without affecting fitting quality. This method supports many realistic global illumination effects such as glossy interreflection, caustics, and color bleeding. This method proposes a material-based partition strategy to further improve the run-time performance and reduce the storage cost.

Keywords: Neural rendering; photorealistic rendering; appearance modeling; relighting; global illumination

目 录

插图清单

附表清单

符号和缩略语说明

BRDF	双向反射分布函数（Bidirectional Reflectance Distribution Function）
BSDF	双向散射分布函数（Bidirectional Scattering Distribution Function）
SVBRDF	空间变化双向反射分布函数（Spatially Varying Bidirectional Reflectance Distribution Function）
ABRDF	表观双向反射分布函数（Apparent Bidirectional Reflectance Distribution Function）
BTF	双向纹理函数（Bidirectional Texture Function）
GGX	未知毛玻璃模型（Ground Glass X(unknown)）[1]
N	输入图片数量
Gamma	伽马，指图像伽马矫正中幂函数中的指数
LDR	低动态范围（Low Dynamic Range）
HDR	高动态范围（High Dynamic Range）
Radiance cues	本文第 4 章所述方法中提出的辐射亮度信息，并在第 5 章方法中使用
Mipmap	多层级贴图（Mip 是拉丁语 multum in parvo 首字母缩写）
Trimap	三值图，每个像素取值为 0、128 或 255，分别代表背景、未知区域和前景
DSLR	数码单反（Digital Single-Lens Reflex）
MLP	多层感知机（Multi-Layer Perceptron）
CNN	卷积神经网络（Convolutional Neural Network）
BN	批归一化（Batch Normalization）
MAE	平均绝对误差（Mean Absolute Error）
MSE	均方误差（Mean Squared Error）
SSIM	结构相似性（Structural Similarity Index Measure）
PSNR	峰值信噪比（Peak Signal-to-Noise Ratio）
LPIPS	基于学习的感知图像片相似性（Learned Perceptual Image Patch Similarity）[2]

第 1 章　引言

1.1　课题背景

真实感渲染是计算机图形学中的核心研究领域之一，高质量的数字化渲染技术在生活娱乐和工业生产当中均发挥着重要的作用。近几年来，随着深度学习的蓬勃发展，神经渲染逐渐成为真实感渲染领域的研究热点。本节将首先介绍传统真实感渲染的概况和其存在的主要挑战，接着介绍神经渲染的含义，最后讨论神经渲染和真实感渲染间的联系以及神经渲染的重要意义。

1.1.1　真实感渲染

真实感渲染是指借助计算机技术数字化地合成高真实感的二维图片或视频的过程。真实感渲染已经被广泛应用于各个领域中，具体包括动画/电影特效制作、电子游戏制作、艺术创作、虚拟/增强现实、以数字博物馆和数字购物为代表的线下场景和以建筑设计和工业产品设计为代表的计算机辅助设计领域等（图 1.1 展示了一些典型的应用场景）。

(a) 电影《狮子王 · 重制版》[3]　(b) 电子游戏《黑神话：悟空》[4]　(c) 计算机生成的光影艺术[5]

(d) Meta 虚拟视频会议[6]　(e) 数字化《蒙娜丽莎》[7]　(f) 建筑设计渲染预览[8]

图 1.1　真实感渲染的典型应用举例

真实感渲染同其他计算机图形学研究方向一样可以分为采集和建模、存储和表达以及绘制和可视化三个主要领域。采集和建模是指通过人工设计或自动重建的方式实现数据准备和场景建模，其中，场景自动重建是首先通过相机或其他传感器完成场景的原始数据采集，然后基于采集数据来对场景中相机参数、几何表达、材质属性等信息进行重建和恢复。采集和建模是真实感渲染的基础步骤，采集

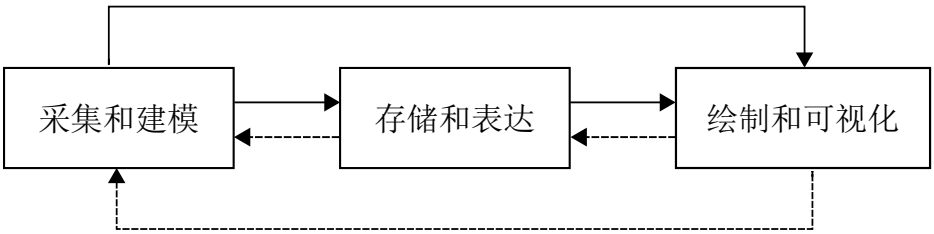


图 1.2　真实感渲染概览

数据和重建场景的质量将直接影响最终的可视化效果。存储与表达则主要研究面向采集数据的高效压缩方法和适合于后续绘制步骤的高效场景表达，其中高效压缩方法主要是针对图形学中广泛存在的高维测量数据（例如定义在六维空间的双向纹理函数或空间变化双向反射分布函数等材质数据），传统上所采用的制表存储方法需要耗费大量存储空间，结合渲染数据性质的高效压缩方法可以在不影响表达力的前提下大大降低存储开销使之更加适合于轻量化应用场景。此外，场景表达与后续的绘制和可视化方法息息相关，场景表达形式往往与具体的绘制方法相适应。绘制和可视化是指基于采集数据和合适的场景表达，通过某种绘制流水线完成最终二维图片或视频的生成。绘制算法的选择不仅取决于场景表达形式也同样取决于实际应用需求：不同的场景几何表达（典型的几何表达包括三角形网格、点云、体素、隐式曲面等）往往需要不同的绘制方法，而对于相同的几何表达也需要根据应用需求来选择合适的绘制方法，例如同样是针对三角形网格表达也需要根据渲染效率和渲染质量间的权衡来分别选择光栅化方法或路径跟踪方法。

采集和建模、存储和表达以及绘制和可视化三个步骤之间既顺次递进同时也互相依赖（如图 1.2 所示）：其一，从逻辑上，三者是真实感渲染流程中从数据准备到数据表达再到最终可视化展示的递进步骤，因此后续步骤自然地依赖于先前步骤；其二，当前步骤也同样依赖于后续步骤，例如原始数据采集过程中需要根据后续的存储和表达以及绘制和可视化过程来进行相应的优化和适应，存储和表达的高效性和合理性也取决于后续绘制算法。总之，采集和建模、存储和表达以及绘制和可视化三个步骤互相依赖，因此在真实感渲染算法研究中，既要专注于提升每个步骤的效率和质量，也要兼顾整个系统不同步骤间的依赖关系以实现整体的性能提升。

真实感渲染可以看作是物理学光传输过程在计算机图形学领域的数字化模拟和求解。形式化地，真实感渲染可以使用辐射传输方程[9]（Radiative Transfer Equation）来建模：

$$L_i(p,\omega) = T_r(p_0 \to p)L_o(p_0,-\omega) + \int_0^t T_r(p' \to p)L_s(p',-\omega)\,\mathrm{d}t' \qquad (1.1)$$

其中，$L_i(p,\omega)$ 代表空间中一点 p 在 ω 方向的入射辐射亮度，$L_o(p_0,-\omega)$ 代表物体表面一点 p_0 在 $-\omega$ 方向的出射辐射亮度，$L_s(p',-\omega)$ 代表光线传播到 $p' = p + t\omega$ 点在 $-\omega$ 方向接收到的源辐射亮度，$T_r(p_0 \to p)$ 代表光线从物体表面的 p_0 点传输到 p 点过程中的透射率，$T_r(p' \to p)$ 代表光线从 p' 点传输到 p 点过程中的透射率。

在不考虑空间中的参与性介质的情况下，以上辐射传输方程可以简化为渲染方程[10]：

$$L(p,\omega_v) = L_e(p,\omega_v) + \int_{S^2} f_p(\omega_v,\omega_i)L_i(p,\omega_i)|n_p \cdot \omega_i|\,\mathrm{d}\omega_i, \qquad (1.2)$$

其中，p, f_p, n_p 是着色点的位置、双向散射分布函数和法向量，ω_v 是视角方向，$L_e(p,\omega_v)$ 是发光辐射亮度而 $L_i(p,\omega_i)$ 代表 p 点在入射方向 ω_i 的入射辐射亮度。

数学上，以上方程不存在解析形式的解。计算机图形学中的一系列真实感渲染算法可以看作是在对真实世界进行数字化建模的基础上，实现以上复杂方程的近似求解：数字化的真实世界建模过程包含前述的采集和建模及存储和表达步骤，在数字化过程会引入一定的人为启发式约束和相应的模型简化；绘制算法则是在数字化建模基础上的一种近似求解算法。例如，在典型的离线渲染流程中，首先，需要通过场景重建或美术人员手工设计的方式获取三维资产数据，然后，针对大量复杂的渲染数据进行针对性压缩存储并转换为适合于后续使用的表达，最后，通过路径跟踪等绘制算法以蒙特卡洛采样的方式完成对渲染方程的数值求解。

虽然传统的真实感渲染算法已经涵盖了采集和建模、存储和表达以及绘制和可视化等各个步骤并在多种实际应用中获得广泛使用。然而，随着各行各业对数字化三维内容的需求的日益增长，传统真实感渲染算法已经逐渐难以满足现实需求，其主要的不足和挑战包括：

自动化程度低、使用门槛高　在采集和建模过程中，传统真实感渲染方法往往需要大量繁琐且依赖于专业人员的手工操作，这严重制约了三维图形数据的大规模自动化生成，导致相关数字资产存在成本高昂和难以高效复用的问题。此外，传统真实感渲染算法主要面向具有丰富领域知识的专业用户，导致其使用门槛高，进而限制了其他相关行业对于真实感渲染方法的使用和传播。

模型和真实世界存在差距　传统真实感渲染方法依赖于对真实世界的人工建模和模拟。由于精确而完美的重建和模拟不可能实现，因此真实感渲染流程中不可避免地需要引入简化或近似，进而导致模型和真实世界存在差距，而这些差距会直

接影响现有方法在复杂场景中的生成质量，从而制约了相关创作或应用的发展。

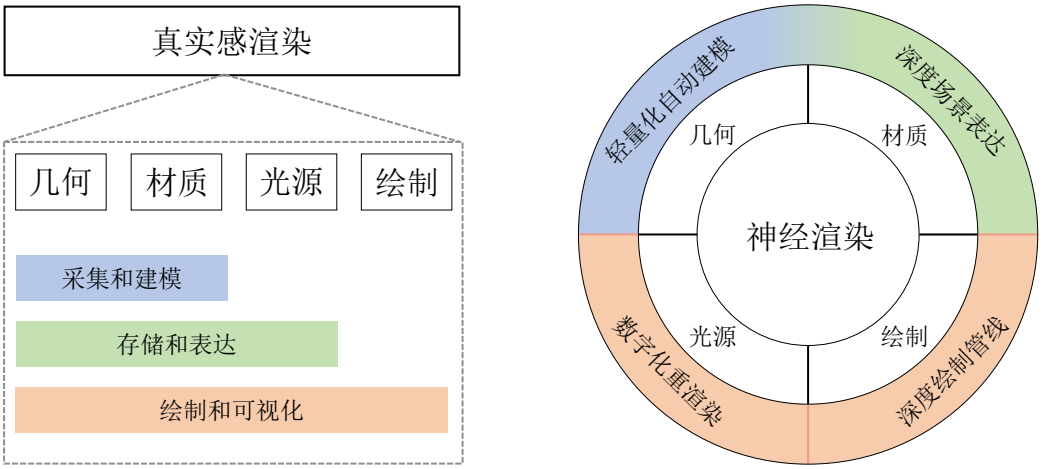
效率和质量难以兼顾　传统真实感渲染方法一般可以分为实时方法和离线方法两类，其中实时渲染方法可以实现极高的运行效率以满足电子游戏或其他交互级应用场景的效率需求，然而其生成质量难以达到照片级真实感，体现出较强的“CG”感；离线渲染方法可以实现照片级真实感的渲染，因而被广泛应用于动画制作、电影特效制作等应用当中，然而其运行效率往往很低，难以达到交互级或实时的运行效率。

1.1.2　神经渲染

近年来，以神经网络为典型代表的深度学习方法在计算机相关领域中被广泛使用。数据驱动的自动化特征学习逐渐取代了传统的人工特征设计，一方面大大降低了用户的使用门槛，另一方面扩大了传统方法的适用范围并提升了结果质量。基于深度学习的方法已经成为图像识别、目标跟踪、机器翻译等领域中的主流方法。

将深度学习和真实感渲染问题进行结合是非常自然的思路，近年来也逐渐成为计算机图形学中的热点研究方向。真实感渲染问题的自身特性使之非常适合于结合深度学习来进行求解，具体体现在：

- 真实感渲染问题一般具有明确的输入和输出但其中间过程非常复杂。一方面，明确的输入和输出使之易于通过深度学习模型进行描述和建模；另一方面，深度学习擅长于从大量数据中学习从输入到输出的复杂映射。
  例如：在表观建模问题中，输入是若干二维拍摄图片而其输出是用于描述物体表观信息的材质贴图，输入到输出的映射非常复杂且充满二义性。
- 真实感渲染问题一般可以借助仿真或模拟的方法自动化地生成大量带有明确标注的训练数据，因而适合于通过有监督的深度学习方法进行求解。
  例如：在实时光线跟踪的后处理降噪中，输入和输出分别是低采样的噪声图片和高采样的无噪声图片，因此训练数据对可以通过离线路径跟踪方法自动化生成。
- 数据驱动的方法已经被广泛应用于传统真实感渲染问题中，深度学习作为一种更加有效的数据驱动技术，自然地可以应用到相关问题中，以更好地利用渲染数据本身的特性，从而实现更加轻量化且高质量的解决方案。
  例如：在测量材质数据的压缩存储中，传统方法采用矩阵分解或主成分分析等方法实现压缩，基于深度学习的技术可以更好地利用材质数据本身的特性实现更加高效的压缩。



(a) 真实感渲染的基本组件和研究领域　　(b) 神经渲染的基本组件和典型研究方向

图 1.3　神经渲染概览及其与真实感渲染的研究领域间的关系展示

- 真实感渲染问题一般与二维图片或视频具有密切关联，而深度学习方法擅长基于二维图片或视频进行特征提取或内容生成，因此真实感渲染和深度学习可以将图片或视频作为桥梁来实现基于感知或理解的信息传递。

  例如：基于深度学习构造的感知损失函数可以更好地衡量方法图片间的误差，适合于在以图片为输出的任务中提升图片质量或在以图片为输入的任务中提升输入信息提取效率。

基于深度学习的真实感渲染也被称作神经渲染，参照 Tewari 等人[11]给出的定义，神经渲染是指结合深度学习的高质量图片或视频合成技术。

真实感渲染的过程可以概括为从场景的几何、材质和光源等信息出发，通过某种给定的绘制方法，实现特定视角下高逼真图片的生成。真实感渲染包括四个基本组件，分别是物体的三维几何、物体的材质属性、场景光源以及具体的绘制方法（如图 1.3 (a) 所示）。神经渲染是将以神经网络为代表的深度学习方法与真实感渲染相结合，对真实感渲染的四个基本组件中的某一个或某几个进行改进或更换。神经渲染的典型研究领域同真实感渲染一致，也可以概括性地分为采集和建模、存储和表达以及绘制和可视化三个领域（如图 1.3 (b) 所示）。具体而言，在神经渲染研究中，采集和建模领域主要研究针对几何和材质的轻量化自动建模；存储和表达领域主要研究针对场景几何、材质和光源等各个组分的高效深度场景表达；绘制和可视化领域的主要研究方向包括针对真实世界复杂场景的数字化重渲染（即自由视点重光照渲染）以及兼顾效率和质量的全新深度绘制管线等。

神经渲染不仅有助于解决上一节中提到的真实感渲染面临的困难和挑战，而且还具有额外的优势和潜力。具体而言，在解决真实感渲染所面临的挑战方面：首

先，深度学习的引入使得在采集和建模的过程当中可以借助大数据先验来降低采集成本，使得基于消费级相机或其他轻量化传感器所采集数据进行高质量场景重建成为可能。另外，深度学习方法的使用门槛更低且更容易部署到其他行业应用场景当中。其次，针对人工模型和真实世界之间存在差距的问题，基于深度学习的图片绘制方法可以在无需精确场景重建的情况下实现重渲染，因此可以避免因模型表达力限制而导致的重渲染瑕疵。最后，神经渲染方法可以在一定程度上兼顾效率和质量，复杂渲染效果的拟合和学习可以在神经网络的离线训练过程中完成，在运行时通过神经网络前向推理即可快速实现高真实感结果的生成。考虑到近年来显卡算力的不断增长和神经网络推理效率的不断提升，神经渲染方法潜在的运行效率还有进一步的提升潜力。神经渲染的其他优势和潜力体现在：其一，全新的深度场景表达和绘制管线使得我们可以不再依赖于传统的三维内容生产流程，为我们带来了全新的看待计算机图形学问题的视角；其二，神经渲染有助于多模态内容的理解和创作。目前主流的自然语言处理和计算机视觉方法均使用神经网络模型，而将三维内容和一维的自然语言、二维的图片或视频进行多模态融合时，神经渲染同这些方法共享相似的底层架构和模块（如卷积神经网络、自注意力网络等），有助于多模态信息的共享和整合，从而实现高效的多模态内容理解和生成。

1.2　研究现状

正如前一节所述，神经渲染同真实感渲染一样可以概括性地分为采集和建模、存储和表达以及绘制和可视化三个领域。本节仅对以下几个同本文研究内容直接相关的研究方向进行阐述，更全面和详细的文献综述参见本文第 2 章的介绍。

表观建模　表观建模是指以给定样本的若干拍摄图片为输入，通过算法自动化地重建该样本的材质数据的过程。由于不同的材质数据可能对应相似的图片，因此表观建模问题存在高度二义性。传统基于优化的策略必须依赖大量输入图片和人工设计的启发式约束，才可以解决上述二义性问题，因此无法满足轻量化应用需求。近年来，基于深度学习的表观建模方法不断涌现，该类方法可以从单张输入图片重建出较为合理的材质数据，然而受限于单张输入图片信息不足的缺陷，该类方法往往无法有效解决二义性问题。给定更多的观测图片是解决二义性问题的一种有效方式，但是现有的基于深度学习的表观建模方法无法轻易扩展到多张输入图片的情况。此外，现有基于深度学习的方法均采用回归策略，无法充分利用观测图片中的数据约束信息。因此，针对任意数量输入图片的轻量化表观建模是亟待研究的重要研究问题，此外，如何更加有效地将大数据先验和观测数据中的

数据约束信息进行结合也是富有挑战的研究方向。

深度场景表达　传统上，计算机图形学中的几何表达一般采用三角形网格、点云和体素等形式，材质表达一般采用双向反射分布函数、双向纹理函数等形式，光源表达则采用点光源、方向光源、环境光照等形式。近些年来，研究人员提出了多种基于深度学习的三维场景表达，以更加神经网络友好的形式对场景各个组分进行编码。典型的深度场景表达包括深度体素、深度平面扫描体、深度点云、神经辐射场等，其共同思路是在充分借鉴传统计算机图形学领域知识和传统表达的基础上提出适应于特定问题的深度表达。现有的深度场景表达主要集中在几何信息的编码上，而针对材质和光源的有效编码以及适合于特定神经渲染管线的全场景表达仍是待解决的重要研究问题。此外，深度场景表达的可编辑性和动态性也是有潜力的探索方向。

基于图像的绘制　基于图像的绘制是指以若干真实场景的观测图片作为输入，在不显式重建场景的情况下，直接合成真实场景在全新视角或全新光照下高逼真结果的过程。基于图像的绘制方法被广泛应用于真实世界复杂场景的数字化展览应用中，例如数字博物馆、虚拟购物、虚拟现实等。近年来，深度学习的融入使得基于轻量化采集设备和少量输入图片的数字化重渲染逐渐成为可能。基于图像的绘制方法可以分为全新视角生成和重光照渲染两类，其中全新视角生成仅考虑视角的动态变化，其研究重点和难点在于如何尽量降低采集成本和所需的视角数量，以及保持视角间插值的连续性。而重光照渲染则需要考虑光源的动态变化，其研究重点在于如何高效拟合动态光照下复杂且高频的表观变化和全局光照效果。现有方法大多关注于全新视角生成或针对特定物体或特定光源类型的重光照渲染，针对一般性场景的自由视点重光照渲染仍是尚未解决的重要研究问题，其主要挑战包括：输入光源信息难以有效地输入到神经网络当中；光源和视角各自的采样空间组合后复杂度大大增加。

基于物理的绘制　基于物理的绘制是指以场景描述作为输入，通过模拟和近似物理上的光传输过程来合成高逼真渲染结果的过程。近年来，深度学习也融入到基于物理的绘制方法当中：一方面，深度学习被广泛应用于离线绘制方法中以提升方法的运行效率，主要研究方向包括探索针对光线跟踪渲染结果的高效降噪以及研究如何将神经网络以子模块的形式嵌入到经典绘制管线中（例如使用神经网络替换原管线中的重要性采样、辐射场缓存等组件）。另一方面，深度学习的融入也可以提升实时绘制方法的质量，主要研究方向包括屏幕空间着色信息预测以及全

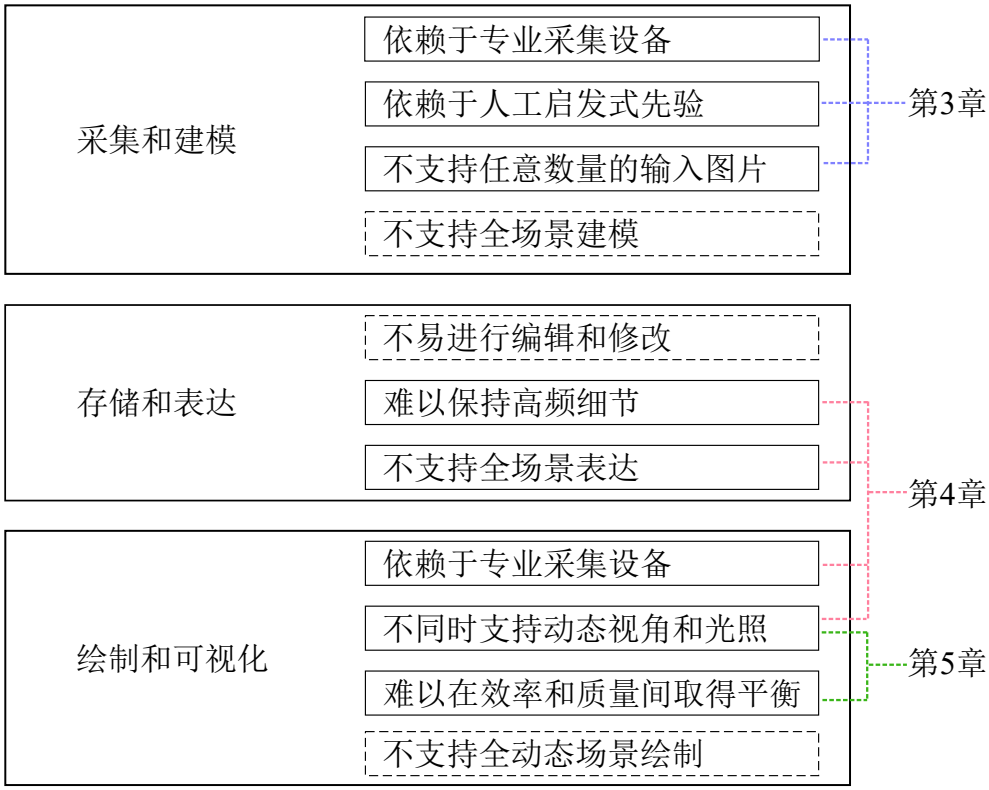


图 1.4　本文研究内容和神经渲染中的主要挑战间的关系

局光照的快速绘制等。然而，离线绘制方法的运行效率仍旧是其主要瓶颈，实时绘制方法的渲染质量仍然和离线绘制方法存在较大差距，换言之，基于物理的绘制领域的主要挑战仍是质量和效率间的平衡。快速且高质量的全局光照效果合成仍旧是尚未解决的重要研究问题。

1.3　研究内容与主要贡献

针对前一节中提到的神经渲染所存在的种种困难和挑战，本文提出了一系列全新的神经渲染算法来解决这些问题。图 1.4 列举了本文研究内容和神经渲染领域中的挑战间的关系。

在采集和建模领域，本文提出一种基于深度学习和逆渲染的表观建模方法，有效解决了传统逆渲染方法中存在的依赖专业采集设备及依赖人工设计启发式先验等不足，并且支持基于任意数量输入图片实现鲁棒的表观建模。在存储和表达领域，本文提出一种深度场景表达并将其应用于基于图片的重光照渲染当中。该深度场景表达同时考虑了物体几何、材质和场景光照信息，并且利用多层次结构来更好地保留高频表观细节。在绘制和可视化领域，在基于图片的绘制方向，上述基于深度场景表达的重光照渲染方法有效解决了传统方法依赖于专业采集设备的

缺陷，并且可以支持 360 度自由视点以及多种不同类型光源下的重光照渲染。在基于物理的绘制方向，本文提出一种深度绘制管线，可以实现动态视角和动态面光源下的全频率全局光照快速绘制，有效地实现了效率和质量间的平衡。

本文所提出的神经渲染算法的技术内容和主要贡献总结如下：

基于逆渲染和数据驱动的表观建模　现有的基于深度学习的表观建模方法均采用回归策略，即对于给定的材质样本的单张观测图片，通过神经网络模型直接预测其对应的材质贴图。该类方法存在材质重建精度不够准确和难以扩展到多张输入图片的情况等不足。经典逆渲染方法是另一类典型的表观建模方法，其核心思路是对于给定的材质样本的大量观测图片，通过优化材质贴图本身使其渲染后的图片和输入观测图片间的误差尽可能小，以实现材质贴图的恢复。该类方法虽然重建精度高，但其依赖于大量输入图片或对材质样本的强假设，此外，在其逆渲染优化中还依赖于人工设计的启发式约束或其他先验知识。本文提出一种结合深度学习和逆渲染优化策略的表观建模方法，其核心思路是在深度学习构造的材质数据隐空间中进行逆渲染优化来实现高质量的表观建模。本方法针对如何构造光滑且易于优化的深度隐空间、如何进行逆渲染优化的初始化和探索适合于本任务的神经网络架构等难点进行了深入分析并提出了有效的解决方法。本方法的优势主要体现在：首先，本方法适用于从单张到多张的任意数量输入图片，其重建精度随输入图片数量增加而不断提升并最终收敛到准确结果；其次，本方法的逆渲染优化过程是在紧凑的低维深度学习隐空间中进行，得益于大量材质数据提供的数据先验，整个优化过程无需任何手工设计的启发式约束；最后，本方法自然地可以支持高分辨率的表观建模。通过在大量合成数据和真实采集数据上的实验分析，本文验证了所提方法可以实现基于任意数量输入图片的高质量表观建模，并且其在单张输入图片情况下的重建质量相较于之前的基于深度学习的回归方法亦有提升。

基于深度场景表达的重光照　针对真实世界复杂场景的数字化重渲染问题，其中一种典型思路是基于建模的方法，即首先通过场景重建技术对场景中各个组分进行显式重建和恢复，随后使用传统真实感渲染管线完成新视角或新光源下的重渲染。该类方法受限于采集误差或模型表达力不足等因素，无法做到精确的无误差的场景重建，而重建误差会导致其最终重渲染结果存在瑕疵。另一类典型思路是基于图像的绘制方法，其优势是无需显式场景重建，而是直接利用输入图片来合成新视角或新光源下的渲染结果。然而，该类方法依赖于专业且复杂的设备进行图片采集并且往往需要大量输入图片。本文提出一种基于深度场景表达的自由视点重光照方法。一方面，该方法属于基于图像的绘制方法，其渲染流程通过神经

渲染网络来实现；另一方面，本文也借鉴了基于建模的方法，使用重建的粗糙几何作为多视角间的一致性先验，并以定义在该粗糙几何上的神经纹理和辐射亮度信息作为深度场景表达。本文提出的辐射亮度信息是一种图片友好的、适用于多种不同类型光源的高效光源编码方式，将其与本文提出的光照增强策略相结合后，本文方法不仅可以处理点光源还可以处理环境光照等更为复杂的光照下的重光照渲染。通过在多个包含复杂材质属性、几何特征以及全局光照效果的真实场景中的实验分析，本文验证了所提方法可以实现高质量的自由视点重光照渲染。

基于深度绘制管线的全局光照绘制　全局光照快速绘制是真实感渲染中至关重要的研究方向。现有方法中，实时方法运行效率高，但往往仅适用于特定的场景且整体渲染质量不高；离线方法虽然可以准确地捕捉复杂的全局光照效果，但其渲染效率较低无法满足交互式应用的需求。本文提出一种可以支持动态视角、动态面光源下全频率、高质量的全局光照快速绘制的深度绘制管线。本方法的核心观察是对于静态场景而言，每个着色点的全局光照被该着色点位置、视角方向和输入光照三者完全决定，因此本文提出使用深度神经网络来建模从着色点位置、视角以及光照到全局光照的复杂映射。本文方法可以视为一种基于预计算策略的全局光照渲染方法：其中训练数据生成和神经网络训练过程可以视为场景辐射亮度的预计算过程，而运行时只需要将若干屏幕空间的贴图输入到训练好的神经网络当中，即可实现高质量的全局光照的快速渲染。为了提升神经网络学习效率并保持紧凑的神经网络规模，本文提出神经网络友好的输入表达和深度全连接渲染网络。此外，本文还提出一种基于材质划分的加速方案，可以在不影响最终渲染质量的前提下降低运行时计算开销并降低存储代价。通过在多个包含色溢、镜面反射、多次高光互反射等复杂全局光照效果的室内场景中的实验分析，本文验证了所提方法可以实现全频率全局光照的快速渲染。

1.4　本文组织结构

本文后续内容按照如下方式组织：第 2 章介绍真实感渲染和神经渲染领域的研究现状，并着重整理和分析同本文内容紧密关联的相关工作。第 3 章提出一种基于逆渲染和数据驱动的表观建模方法，可以支持任意数量输入图片和任意分辨率。第 4 章提出一种基于深度场景表达的重光照算法，支持真实世界复杂场景在自由视点和多种类型光照下的重光照渲染。第 5 章提出一种基于深度绘制管线的全局光照渲染方法，支持动态视角和动态面光源下的全频率全局光照快速渲染。第 6 章总结和梳理本文内容，并对神经渲染领域未来工作方向进行展望。

第 2 章　相关工作

神经渲染是将以神经网络为代表的深度学习方法与计算机图形学中的真实感渲染问题相结合的一类方法的统称。神经渲染的主要意义体现在：一方面，神经渲染算法可替换传统真实感渲染流程中某一个或某几个组件，提升整体运行效率或降低成本；另一方面，神经渲染方法可以作为一种全新的深度绘制管线，以深度场景表达代替传统场景表达、以神经渲染网络代替传统渲染器，扩展真实感渲染所适用的场景范围。近年来，研究人员提出了一系列神经渲染方法，通过将深度学习技术与特定的真实感渲染问题相融合取得了众多突破和进展，神经渲染方法目前已经在计算机图形学相关的科学研究和实际应用中被广泛使用。

本章将从采集和建模、存储和表达以及绘制和可视化三个领域介绍神经渲染的相关工作。本章重点介绍和本文方法相关联方向的现有工作，关于神经渲染领域的全面调研可参考 Tewari 等人[11]和 赵烨梓等人[12]的文献综述。

2.1　采集和建模

采集和建模的主要研究对象是物体的表观和几何，其研究内容包括表观建模、几何建模以及表观和几何协同建模三个方向。本节主要关注基于消费级相机的轻量化采集和建模方法，其他的基于主动式传感器等专业采集设备进行几何或表观重建的工作本节不再赘述，读者可以自行参阅相关文献综述[13-14]。

2.1.1　表观建模

表观建模方法根据其输入图片数量的多少，可以分为基于大量图片、基于少量图片和基于单张图片的表观建模方法三类。

基于大量图片的表观建模方法　该类方法一般假定物体几何或光源等场景信息已知，以大量图片（通常是几十到几百张）或视频序列作为输入，通过逆渲染优化实现物体表观重建，其主要目标是在尽可能提高重建准确性的情况下兼顾数据采集过程的轻量性和易用性。Palma 等人[15]提出一种利用自然光照下拍摄的视频序列实现物体空间变化表观属性重建的方法，该方法还提出一种启发式约束以分别恢复物体材质的漫反射部分和高光反射部分。Dong 等人[16]假定物体几何已知而拍摄所处的自然光照和物体本身的材质属性未知，该方法以旋转物体的视频序列为输入，实现物体表面逐点的法向量和材质属性的重建，其核心思路是对材质参数

和光照参数进行迭代优化，并在优化过程中使用自然光照在梯度域的稀疏性作为启发式先验。Riviere 等人[17]使用带有闪光灯的手机相机来获取被采集物体的视频序列，并利用和 Palma 等人[15]方法中类似的启发式先验来实现漫反射和高光反射的分解和恢复。Hui 等人[18]的方法同样以带有闪光灯的手机相机来拍摄视频序列，其核心思路是迭代地优化物体表面法向量和材质模型参数。该方法假定物体的材质具有稀疏性，因此提出采用基于字典的材质模型来描述物体的表观信息，将未知物体的表观属性表达为字典中有限个已知材质的线性组合。

该类方法尽管依赖于大量输入图片或视频序列，但在表观建模的逆渲染优化过程中仍然依赖于正则化约束，例如人为给定的启发式先验信息或是表观属性自身的稀疏性。此外，该类方法可以给出合理的材质重建结果的前提是输入图片数量必须超过某个给定的下限值。与该类方法不同，本文第 3 章所提方法的目标是针对任意数量的输入图片，均实现鲁棒而高质量的表观建模且整个过程无需任何人工给定的正则化约束。

基于少量图片的表观建模方法　该类方法的主要目标是在尽量减少所需输入图片数量和降低采集方法复杂性的同时确保可以重建出合理的表观属性。Aittala 等人[19]提出一种以两张手机拍摄图片为输入，针对类纹理材质的表观建模方法。类纹理材质是指假定材质本身具有自相似性，即其不同局部分块之间存在某种重复性的相似结构。该方法的两张输入图片中的其中一张是在环境光照加上闪光灯下拍摄而成，另一张则是仅在环境光照下拍摄而成，其中环境光照下拍摄的图片用于不同局部分块之间相似性判定和对应关系的建立，而带有闪光灯的输入图片则用于最终每个局部分块内的表观建模。此外，该方法同样利用了大量人工设计的正则化约束，例如光滑性约束、法向量约束等。冯洁等人[20]的方法以类似拍摄条件得到的两张图片作为输入，提出基于像素聚类的策略进行材质参数的拟合，可适用于非类纹理材质的表观建模。Xu 等人[21]的方法也通过利用材质样本在空间维度上的联系，实现基于两张近距离透视相机拍摄图片的高质量材质重建，该方法主要面向匀质材质，也可以扩展到分段空间变化材质。Zhou 等人[22]提出一种支持任意数量输入图片的表观建模方法，可以实现除表面法向量外的其他材质参数的重建。针对单张图片的表观建模，该方法依赖于材质符合分段空间变化性质的强假设，而随着输入图片数量的增加，该方法可以逐步放松假设并可以支持更加细致的空间变化材质。

该类方法虽然大大降低了表观建模所需的输入图片数量，但该类方法依赖于人为引入的材质参数在空间维度的变化具有稀疏性这样的强假设。此外，该类方法往往也需要在表观建模的优化过程中使用人工给定的正则化约束。本文第 3 章

所提方法则可以应用于空间维度存在复杂变化的材质样本。

基于单张图片的表观建模方法　由于表观建模问题存在很强的二义性，而单张输入图片包含的信息无法使逆渲染优化收敛，因此传统表观建模方法无法有效处理单张图片的情况。深度学习的引入使得基于单张图片重建物体表观成为可能，因此目前基于单张图片的表观建模方法均依赖于深度学习。Li 等人[23]提出一种以单张图片（在未知自然光照下拍摄）为输入，可以实现近平面物体的表观建模的卷积神经网络模型。针对标注材质数据不足的问题，该方法的提出了自增强训练策略。在引入自增强训练策略后，该方法所需的训练数据除少量带有标注的材质数据外还包括大量无标注的材质样本图片。该方法分为两个阶段，分别是依赖少量标注数据的预训练阶段以及依赖大量无标注数据的自增强训练阶段。Ye 等人[24]对该方法进行了进一步改进，使之可以在完全不依赖任何带有标注的材质数据的情况下完成表观建模。Deschaintre 等人[25]和 Li 等人[26]的方法则基于通过带有闪光灯的手机拍摄的单张图片，利用卷积神经网络来预测物体表观属性。这两个工作均提出基于现有的少量高分辨率材质数据进行数据增广（包括裁剪、混合等）的方式来构建大量低分辨率材质数据以完成神经网络的训练，二者的主要区别在于神经网络架构方面：Deschaintre 等人[25]采用了双分支的结构，其中卷积神经网络分支负责提取光照和纹理相关的局部特征，而全连接分支负责提取全局信息；Li 等人[26]则在卷积神经网络基础上增加了密连接条件随机场（Densely-connected Conditional Random Fields, DCRFs)）作为后处理步骤以进一步提升重建材质质量。

该类方法设计的目标是针对单张输入图片实现材质建模，然而，单张输入图片所包含的信息往往并不足以解决二义性问题，因此自然需要考虑增加输入图片的数量以提升表观建模的结果质量。现有的基于深度学习的表观建模方法均采用回归策略，将其扩展到多张输入图片的直接思路包括：其一，使用固定数量的多个输入图片作为神经网络输入；其二，基于以池化为代表的信息聚合方法将多张输入图片中提取信息进行整合，使得神经网络可以接收任意数量的输入图片。例如，Deschaintre 等人[27]直接扩展了 Deschaintre 等人[25]的方法，使用最大池化实现多张输入图片特征的聚合，实现了基于多张图片的表观建模。然而，该类基于深度学习的回归方法及其直接扩展方法均严重依赖于在大量训练数据上训练好的神经网络模型，对于给定的输入图片均是通过神经网络前向推理直接得到预测的表观属性，无法有效地利用输入图片中蕴含的约束信息。本文第 3 章所提方法则将深度学习和逆渲染策略相结合，不仅可以处理任意数量输入图片，而且在多张输入图片的情况下可以更加高效地利用输入图片中蕴含的数据约束。本文 3.2.3 节会针对回归策略和本文第 3 章方法所采用的优化策略进行更加详细的讨论和分析。

2.1.2　几何建模

三维几何建模是计算机图形学和视觉领域的经典研究问题，已有几何建模方法根据其采集方式可以分为主动式几何建模方法和被动式几何建模方法两类。主动式几何建模方法一般需要借助专业级的主动式传感器向被采集物体发射结构光或激光等信号来完成三维物体的重建，使用场景较为受限。本节主要关注以相机拍摄的若干图片为输入的被动式几何建模方法。被动式几何建模方法可以进一步分为传统几何建模方法和数据驱动的几何建模方法两类：

传统几何建模方法　该类方法按照其原理的不同可以进一步分为光度立体法和多视角立体视觉法。其中，光度立体法[28]是根据给定物体在固定视角和多个不同光源下拍摄的一系列图片来恢复物体的几何信息。该方法假定物体材质为漫反射材质，可以实现物体几何的准确求解。后续，研究人员在该方法基础上进行了很多探索和改进，使之可以处理具有更加复杂表观属性的物体以及非固定视角或未知光源等更加一般化的情况。多视角立体视觉方法[29]的输入则是物体在固定光照条件下拍摄的多视角图片，该类方法首先针对多视角图片进行关键特征提取，其常用的特征包括 SIFT（scale-invariant feature transform）、DOG（difference-of-Gaussians）等，然后根据关键点的特征匹配关系来建立多视角图片间的对应关系，最后利用多视角的一致性约束来优化求解物体的空间位置。后续有研究人员在多视角立体视觉方法的基础上，引入额外的先验信息或场景假设来实现更高质量的几何重建或更低的采集成本，典型的工作包括：基于轮廓信息的方法[30-32]，该方法通过引入物体在二维图片中的轮廓信息来帮助三维几何重建；基于着色过程的方法[33]，该方法假定场景物体是漫反射材质并利用方向光下漫反射材质的着色过程来帮助物体几何重建。

该类方法受限于采集条件和模型表达力导致其重建几何存在一定误差（一般体现为细节模糊或缺失、存在空洞或拓扑关系错误等瑕疵），进而导致基于该模型的重渲染方法的渲染结果存在瑕疵。本文第 4 章所提方法使用了基于多视角立体视觉领域的经典方法 COLMAP[34]方法所重建的三维几何，然而该方法采用基于图像的绘制策略并提出使用深度渲染网络来完成渲染过程，因此该方法并不依赖于完美的重建几何，仅需要粗糙的三维几何来为渲染过程提供指导。

数据驱动的几何建模方法　研究人员提出基于三维模型库的交互式几何建模方法[35-36]，该类方法的输入是用户给定的二维参考图片或手工绘制的草图等，其基本思路是首先在三维模型库中检索和输入中的物体最为接近的模型，然后利用检索到的模型来构建完整的场景。虽然该类方法所恢复的场景中的几何体均没有

任何瑕疵，但是由于模型库覆盖面有限，因此其重建准确度难以保证。

近年来，随着大规模几何模型库的提出，基于深度学习的几何建模方法也随之涌现。Wu 等人[37]提出了一个大规模基于 CAD（computer-aided design）模型的几何数据集，被称为 ModelNet。该数据集包含多个物体类别（例如飞机、桌子等），其总模型数量超过 15 万个。Chang 等人[38]提出 ShapeNet，是一个规模更为庞大（索引了超过 300 万个模型）、具有丰富标注信息（例如物体的物理尺寸、基于语义的关键词及其他对称性标注等）的基于 CAD 模型的几何数据集。基于深度学习的几何重建方法[39-40]可以实现基于单张图片的三维几何重建。后续，研究人员提出一系列改进工作：例如基于 ShapeNet 数据集和体素表达的深度多视角几何重建方法[41]以及结合 ShapeNet 数据集和大量二维图片数据集的三维几何重建方法[42-44]。

数据驱动的几何建模方法的优势在于降低采集成本，可以实现基于单张或少量拍摄图片进行合理的三维几何重建。然而，目前的数据驱动的几何建模方法的重建精度有限，难以满足高真实感重渲染应用的要求，因此本文第 4 章所提的重光照方法中采用了传统的多视角立体视觉方法来重建场景粗糙几何。需要说明的是，本文第 4 章的方法并不依赖于某种特定的几何重建算法，其他可以实现鲁棒几何重建的方法也均可以应用于本文方法当中。

2.1.3　表观和几何协同建模

表观和几何协同建模是极富挑战的研究问题。输入的观测图片中每个像素的颜色值均代表对应着色点处的辐射亮度，其耦合了物体几何、表观属性和光照效果，因此协同建模算法必须将三者进行解耦并消除二义性。Holroyd 等人[45]提出一种基于双臂同轴光学扫描仪的协同建模方法，可实现物体几何和材质的协同恢复。然而，该方法的采集设备过于复杂，限制了其应用场景。Xia 等人[46]提出一种轻量化的协同建模方法，该方法通过扩展 Dong 等人[16]的方法使之可以同时恢复物体的几何和材质，该方法仅依赖于未知光照下旋转物体的视频序列而无需其他复杂的采集设备。Nam 等人[47]也提出一种轻量化的协同建模方法，其输入是带有共位点光源的相机所拍摄的若干观测图片，该方法简化采集条件的代价是其优化过程计算量更大，整体算法更加复杂。Li 等人[48]提出一种基于深度学习的协同建模方法，可以从单张输入图片中同时恢复物体的几何和材质。该方法的核心思路是采用级联神经网络来迭代优化预测结果，此外该方法还提出基于神经网络的间接光照预测模块以提升渲染结果的真实感，从而更好地实现物体表观和着色过程之间的解耦。Kang 等人[49]提出一种基于 LED 立方体装置的最优主动光照学习算法，并成功应用于物体几何和表观协同建模中。Bi 等人[50]提出使用深度多视角

立体视觉网络和多视角表观估计网络来分别估计物体的深度和材质，该方法在训练过程中通过协同优化两个网络的特征空间变量来最小化渲染误差，多视角信息的有效融合有助于提升该方法的重建质量。

在真实世界数字化重渲染中，以上基于建模的方法均依赖于显式地恢复场景几何模型和参数化材质模型，因此其重渲染质量受限于重建模型的质量。在实际采集过程中，重建几何的精度和材质模型的质量均受到输入观测图片完整性、相机标定精度以及模型本身表达力等因素影响，因此基于建模的方法往往难以实现高质量且无瑕疵的重渲染。本文第 4 章所提方法通过引入神经渲染网络来修正几何和材质模型不精确导致的误差，实现了针对复杂场景的高质量重光照渲染。

2.2　存储和表达

存储和表达领域的核心研究目标是探索更加紧凑而高效的面向神经渲染管线的场景表达，其研究内容对象包括几何、材质和光源。

2.2.1　几何表达

传统计算机图形学中的经典几何表达包括三角形网格、体素、点云以及隐式曲面等。在神经渲染中，研究人员对传统表达进行了适当扩展和改进，使之更加适合于同深度神经网络进行结合。体素表达是将整个三维空间均匀地切分为三维栅格，深度体素表达[41,51-53]和传统体素表达的不同之处在于其中每个栅格（也叫做体素）中不再存储具有明确物理含义的几何和材质信息，而是存储基于深度神经网络提取的高维特征。Xu 等人[54]提出一种基于扫描平面体的深度场景表达，其将三维场景表达为给定视角下一系列预定义深度平面的集合。Thies 等人[55]提出将传统纹理贴图扩展为神经纹理以作为场景表达。传统纹理贴图（例如漫反射贴图、法线贴图等）在计算机图形学中被广泛使用，神经纹理同传统纹理贴图一样也是定义在物体三维几何的二维参数化展开平面上，而与传统纹理贴图不同之处在于其每个纹素中存储的是深度可学习向量而非具有明确物理含义的物理量（如漫反射颜色、法向量等）。Park 等人[56]将深度学习和传统的距离符号函数进行结合并提出了深度距离符号函数作为场景几何表达，并应用于几何补全等实际应用当中。Mescheder 等人[57]提出深度占用场表达，深度占用场是定义在三维空间各个点上的连续函数。若空间中某个点位于几何体的内部，则该点的深度占用场取值为一，否则其取值为零。Sitzmann 等人[58]提出场景表达网格（scene representation networks,SRNs），该表达是一种同时考虑场景几何和表观信息的连续隐式场景表达。Mildenhall 等人[59]提出一种隐式体表达，该表达不需要像传统的显式体表达

（如体素表达）一样依赖于大量的空间存储，而是采用定义在连续空间的紧凑的隐函数表达来描述三维空间。Granskog 等人[60]提出一种解耦的深度场景表达，利用隐向量的不同通道对场景的几何、材质和光源信息进行独立编码。

2.2.2　光源表达

高效的光源表达对于深度重光照渲染而言至关重要。光源表达相关研究的主要目标是为神经渲染网络提供表达力充足且易用性强的输入光照信号。根据光源类型和采集环境不同，研究人员提出了一系列不同的光源表达方法。Ren 等人[61]提出使用点光源的位置信息来编码入射点光源。Sun 等人[62]提出使用低分辨率的环境光照贴图对自然光照进行建模并应用于单张人脸图片重光照任务中。Granskog 等人[60]提出基于隐向量表达来对输入光源信息进行编码并支持光源的动态编辑。

然而，目前在神经渲染领域，尚没有统一的光源表达可以实现针对多种不同类型光源的高效编码。本文第 4 章提出的辐射亮度信息表达是一种适用于多种类型光源（包括点光源、环境光照、面光源等常见光源类型）的统一光源表达。此外，针对动态面光源的编码，本文第 5 章提出了一种全新的组合式光源表达，该表达综合考虑了屏幕空间光照信息和全局光照信息，可以有效应用于动态面光源下高质量全局光照预测任务中。

2.2.3　材质数据压缩与表达

由于解析材质模型难以准确地捕捉真实世界中的复杂表面散射行为，因此测量材质模型被广泛应用于对渲染质量具有极高要求的应用当中。测量材质模型是一种数据驱动的材质模型，其构建过程可总结如下：首先需要对表观模型所处高维空间进行密集采样，其次将以上采集数据通过高维空间制表的方式进行存储。由于高维空间制表存储的存储代价极为高昂，因此测量材质模型的高效压缩是非常有价值的研究方向。测量材质数据压缩的主要研究目标是降低其存储代价并尽可能提升重建后表观的准确性。在真实感渲染中，典型的测量材质模型包括双向纹理函数 (Bidirectional Texture Function, 下文简称为 BTF) 和测量双向反射分布函数（下文简称为测量 BRDF）。

BTF 表达和数据压缩　BTF 是一种描述光线和物体表面所发生的复杂散射行为的模型。与渲染中常用的 BRDF 不同，BTF 不仅包含单个散射点处的单次散射，还包含全局性的次表面散射、自阴影、多次散射等复杂的全局光照效果。生活中常见的纤维织物和皮革等带有复杂微结构的材质均适合于使用 BTF 进行描述。

形式化地，BTF $f(p,\omega_i,\omega_o)$ 是关于物体表面某个着色点 p、入射方向 ω_i 以及

出射方向 ω_o 的 6 维函数。现有方法主要研究 BTF 数据在固定位置维度后的 4 维数据切片 $f_p(\omega_i, \omega_o)$ 的高效压缩和存储。经典 BTF 数据压缩方法的核心思路是使用主成分分析（Principal component analysis, PCA）技术来实现高维数据的降维，其优势是压缩操作简单且运行时解压缩的计算代价也相对低廉，然而基于主成分分析的压缩方法的不足之处在于其未考虑渲染数据本身的特性，导致重建的 BTF 数据通常会包含偏色或模糊等瑕疵，从而影响最终的渲染质量。

Rainer 等人[63]提出了一种非对称自编码器神经网络结构用于 BTF 数据的压缩。非对称自编码器结构的主要优势在于运行时可以独立查询某个入射和出射方向对所对应的 BTF 数据，无需每次都恢复整个 BTF 数据切片。该方法可以做到在保持和 PCA 方法相同压缩比的情况下大大提升重建结果的视觉质量，此外神经网络本身的连续性使得该方法可以自然地支持入射方向、出射方向上的连续插值。Rainer 等人[64]后续又提出一种适用于多种类型 BTF 数据压缩的统一框架，可以将多种类型的 BTF 数据统一编码到公共的特征空间，通过充分利用不同材质间的共有信息来提高 BTF 数据压缩的鲁棒性。Kuznetsov 等人[65]提出了一种支持多尺度 BTF 数据的 BTF 表达，该表达通过深度纹理金字塔和全连接网络来实现多尺度 BTF 数据的高效压缩和存储。此外，该方法还可以在无监督的情况下模拟视差等复杂的表观现象。

测量 BRDF 表达和数据压缩　除 BTF 数据外，测量 BRDF 数据的高效压缩和表达也是研究的重点方向。Matusik[66]提出了包含 100 个各向同性真实材质的测量 BRDF 数据集（该数据集被称为 MERL 数据集），其包含了漫反射和高光反射等多种类型的材质数据。后续，Filip 等人[67]提出了包含各向异性材质的 UTIA 数据集。Dupuy 等人[68]提出一种提升测量数据采集效率的自适应参数化方法，并根据该方法构建了一个包含 51 个各向同性和 11 个各向异性材质的测量 BRDF 数据集（该数据集被称为 EPFL 数据集）。Sun 等人[69]提出了一种基于 PCA 的测量 BRDF 数据表达和压缩方法，该方法的核心观察是分别对 BRDF 的漫反射和高光反射进行独立地降维压缩，可以大大提高整体的压缩效率和重建质量。得益于漫反射和高光反射的解耦合，该方法可以支持部分材质参数的编辑。后续，Hu 等人[70]提出一种基于深度自编码器网络的 BRDF 数据压缩方法，相较于传统方法进一步提升了重建质量和鲁棒性。Zheng 等人[71]提出一种面向测量 BRDF 数据的基于神经过程的高效压缩方法，该方法将测量 BRDF 数据视为连续函数并在函数空间完成测量 BRDF 数据的高效压缩和存储。

2.3　绘制和可视化

绘制和可视化负责将输入场景转化为高真实感的二维图片或视频。根据输入信息维度的不同，绘制和可视化相关工作可以分为基于图像的绘制方法和基于物理的绘制方法两类，其中基于图像的绘制方法以场景的二维观测图片作为输入，而基于物理的绘制方法则以三维场景作为输入。

2.3.1　基于图像的绘制方法

该类方法以若干场景的二维观测图片作为输入，整个过程无需进行显式场景重建，而是直接利用输入图片中蕴含的信息来合成原场景在全新视角或全新光照下的重渲染结果。

2.3.1.1　动态视角

基于光场的方法[72-73]是传统方法中的典型代表，该类方法以密集采样的场景拍摄图片作为输入，其核心思路是针对某个给定的全新视角通过对输入数据进行视角间插值的方式合成该视角下的渲染结果。通过融入物体几何先验知识[74-75]或视角相关的几何估计信息[76-78]可以进一步提升视角插值的质量。研究人员提出表面光场[79-80]方法来捕捉高频光照（例如点光源）下视角相关的表观变化。

基于深度学习的全新视角合成方法能够大大降低数据采集代价并提升视角合成质量，其核心思路是基于多视角图片间的一致性关系来实现三维深度场景表达的重建，在训练过程中并不依赖于精确的三维几何作为监督信号。现有工作根据其几何表达形式不同，可分为基于深度纹理表达、基于体素表达、基于多平面图片表达以及基于隐式曲面表达等四类方法。Thies 等人[55]提出了延迟神经渲染方法，该方法使用神经纹理和深度渲染器来分别代替传统延迟渲染中的纹理贴图和光栅化渲染器，可以实现高质量的全新视角合成。 Sitzmann 等人[81]提出的深度体素方法是基于体素表达的方法[82-85]中的典型工作，其表达中每个体素中存储可以表达局部几何特征和表观属性的深度特征向量。此外，该方法还提出深度遮挡模块来预测场景深度信息并用于模拟真实感渲染中的遮挡剔除过程。多平面图片表达[86-88]被广泛应用于基于深度学习的全新视角合成中。Xu 等人[54]提出一种基于扫描平面体的全新方法，该方法以 6 张特定视角下的观测图片作为输入，支持生成在输入视角范围内的任意新视角下的渲染结果。基于隐式曲面的方法中，Saito 等人[89]将 Mescheder 等人[57]提出的深度占用场表达扩展到绘制任务中，可以基于单张或多张人体图片完成场景几何和材质表达的构建并支持全新视角生成。Mildenhall 等人[59]提出了神经辐射场（Neural radiance fields，NeRF）表达，该表

达可以支持非常复杂的真实场景的全新视角合成。该工作采用深度全连接网络对场景隐式体表达进行建模，并采用基于物理的光线步进算法完成体渲染。此外，该方法还提出使用位置编码技术对低维输入进行编码和使用层次化表达来提升运行效率。后续，研究人员从提升运行时效率、支持重光照渲染、支持动态场景、提升泛用性以及扩展到人体渲染等方向对 NeRF 进行改进，感兴趣的读者可以参考常远等人[90]总结的关于神经辐射场相关方法的文献综述。

此外，其他结合深度学习来提升全新视角合成质量的方法包括：基于流的方法[91-95]、视角内插方法[96]、视角外推方法[97-98]以及基于图像的解耦学习方法[99-102]等。

全新视角合成方法仅支持静态光照下的重渲染，因而其使用场景受限。本文第 4 章提出的将神经纹理和辐射亮度信息相结合的方法，则可以同时支持动态视角和动态光源下的重渲染。

2.3.1.2　动态光源

Debevec 等人[103]开创性地提出了基于图像的重光照方法，该方法利用了光传输过程的线性性质，以一系列受控光源下的观测图片作为输入，支持多种全新光源下的重光照渲染。后续，研究人员从降低存储开销[104]、降低输入图片数量[105]、提升运行效率[106]以及基于非受控光源进行采集[107-111]等角度对其进行改进。

深度学习的融入可以进一步降低对输入图片数量的需求，相关工作可分为基于少量图片的重光照方法[54,112-115]和基于单张图片的重光照方法[62,116-118]。

Meka 等人[118]提出一种神经网络模型来学习从两张人脸图片到其完整四维反射场的复杂映射关系，该方法中作为输入的人脸图片需在颜色渐变的光源下进行拍摄。Kanamori 等人[115]提出一种针对完整人体渲染的遮挡感知的逆渲染方法，该方法假定输入光照满足低频性质以及物体表观满足漫反射假设。以上方法局限于人脸或人体的渲染，难以直接扩展到其他具有复杂几何和材质的普通物体的重光照渲染中。

Meshry 等人[112]提出一种针对地标性建筑的基于大规模互联网图片的重光照方法，该方法首先根据大量互联网图片来重建地标性建筑的三维点云，随后基于点云和图片信息进行渲染并得到深度贴图、颜色贴图和语义标签图，最后利用神经渲染网络俩预测新光照下的渲染结果。该方法仅可以支持粗粒度光照变化（例如不同日照时间或不同天气等），无法提供更加直接和精细的光照控制。Philip 等人[114]提出一种几何感知的神经渲染网络结构，可以实现基于单张图片的重光照渲染。该方法的核心贡献是提出一种基于粗糙几何的阴影细化模块，可以很好地消除原光照下的阴影并合成新光照下的阴影。由于该方法假定场景中仅包含单个

方向光和简单的环境雾，因而该方法难以在室内场景和其他一般性的重光照任务中使用。

Xu 等人[116]提出一种基于稀疏输入的重光照方法，该方法以 5 张在预定义方向的方向光下拍摄的图片作为输入，支持半球面内任意方向的方向光下的重光照渲染。该方法的神经网络结构主要包括采样网络和重光照网络两个模块，其中采样网络负责学习最佳的预定义光源方向组合，而重光照网络以预定义光源方向所对应的图片来生成全新光源下的渲染结果。后续，Xu 等人[54]将固定视角的全新视角生成方法和以上单视角重光照算法进行结合，可以支持多视角下的重光照渲染。然而，由于该方法仅支持半球面范围内下的方向光源，因此无法应用于遍布于整个球面的环境光照下的重光照结果生成。此外，由于该方法依赖于固定方向的预定义光源，因此需要专用的采集设备进行图片拍摄。本文第 4 章所提方法虽然依赖更加密集的视角采样，但可以同时支持 360 度自由视点和环境光照等复杂光源下的重渲染，并且其数据采集过程仅依赖于轻量化的手持移动设备。

Chen 等人[113]在 Thies 等人[55]方法的基础上进行了扩展，该方法提出将从未知自然光照下拍摄的多视角图片中推理得到的“光传输函数”编码到神经纹理当中，以支持重光照渲染。由于以上问题严重欠约束，因此该方法提出了多种人工设计的启发式约束并假定场景的自然光照可以使用 10 阶球面谐波函数来进行表达。受限于以上假设和约束，该方法无法处理高光反射和自阴影等全局光照效果。本文第 4 章所提方法则无需任何启发式约束并且支持带有复杂表观属性和全局光照效果的场景在多种类型光源下的重光照渲染。

Bi 等人[119]提出一种基于体表达的重光照方法，该方法可以基于输入图片来恢复深度反射体表达并将其应用于新视角、新光照下渲染结果的生成。扩展 NeRF 方法以支持重光照渲染的工作[120-121]也是近年来研究的热点。例如，Bi 等人[120]在 NeRF 提出的神经辐射场表达的基础上提出神经反射场表达，该表达显式地将物体材质属性和输入光照信息进行解耦，可以支持单个点光源下直接光照的重渲染。Srinivasan 等人[121]在上述神经反射场表达的基础上增加了可见性场，该方法可支持环境光照下包含单次间接光照效果的重光照渲染。

2.3.2　基于物理的绘制方法

该类方法以三维场景信息作为输入，通过模拟或近似光传输的物理过程实现高真实感结果渲染，其核心研究目标是实现快速且高质量的全局光照渲染。

2.3.2.1　基于屏幕空间的渲染算法

该类渲染方法以屏幕空间的若干贴图作为输入，利用单独的后处理步骤来合成全局光照。Dachsbacher 等人[122]提出反射阴影贴图方法，通过扩展经典的阴影映射技术使之可以合成漫反射物体间单次散射的间接光照。Ritschel 等人[123]通过扩展经典的屏幕空间环境光遮蔽算法使之支持包含方向阴影和色溢在内的多种间接光照效果的生成。Robison 等人[124]提出对完美镜面反射结果进行运行时模糊的全新思路，以合成不同粗糙度下的高光反射效果和软阴影等全局光照效果。然而，以上传统方法适用场景范围有限且生成结果质量不高。

基于深度学习的屏幕空间方法通过端到端的神经网络实现全局光照的预测。Nalbach 等人[125]提出深度着色框架，该方法可以支持包含间接光照在内的多种着色效果生成。近期，Xin 等人[126]提出一种轻量化的神经网络模型，可以支持漫反射场景中包含单次散射的间接光照的快速渲染。

实际上，基于屏幕空间信息实现全局光照合成是一个高度欠约束的问题。从屏幕空间信息到最终全局光照渲染结果间的映射是典型的一对多映射，神经网络无法高效地学习该类映射。由于基于屏幕空间的方法具有二义性问题，因此其生成质量和适用范围往往受限。

2.3.2.2　基于预计算的渲染算法

现有的离线方法效率低而实时方法质量较差，均难以满足高质量且快速的渲染任务的需求。基于预计算的渲染算法的策略是将计算开销转移到预计算步骤当中，以实现运行时快速且高质量的结果生成。

基于预计算的渲染方法可以大致分为预计算辐射传输方法、光照贴图/探针方法以及基于回归的方法三类。

预计算辐射传输方法[127-130]在预计算阶段计算每个着色点处的光传输结果并将其以某种基函数（例如球面谐波函数）系数的形式进行存储，而在运行时阶段复杂的全局光照渲染可以使用简单的点乘操作实现。早期的预计算辐射传输方法由于采用球面谐波函数作为光源表达因此仅支持低频的全局光照效果。后续，研究人员[131-132]将球面谐波函数替换为小波函数或球面高斯函数以支持全频率的全局光照。现有大多数预计算辐射传输方法仅可以处理远距离光照下的全局光照渲染，而可以支持局部光源的方法也往往依赖于非常复杂的数据结构并且无法处理高频的全局光照。

光照贴图方法适用于漫反射场景的全局光照渲染。在预计算阶段，使用离线的辐射度算法[133]或路径跟踪算法[10]实现漫反射物体间的全局光照的预计算并将

其烘焙到光照贴图中，在运行时阶段，通过对烘焙好的光照贴图进行插值查询即可完成全局光照渲染。起初 Greger 等人[134]提出光照探针方法并将其应用于动态漫反射物体的全局光照渲染，后续该方法被广泛应用于高光反射场景的全局光照渲染中。近期，McGuire 等人[135]提出一种光场探针方法，通过在预计算阶段计算并存储静态场景的完整光场和可见性信息，可以支持运行时的实时渲染。Rodriguez 等人[136]则通过扩展光照探针表达使其可以存储包含高光反射的光路信息。该方法将着色计算中漫反射部分和高光反射部分分开考虑，其中高光反射部分通过将预计算得到的高光光照探针表达重投影来得到，而漫反射部分则通过传统的光照贴图方法得到。然而，以上基于光照贴图/探针类的方法均局限于静态光照，不支持动态光源下的全局光照生成。

Ren 等人[61]提出辐射亮度回归函数来对静态场景中每个着色点处的辐射亮度场进行建模。刘晓芸等人[137]提出使用基于径向基函数的神经网络模型对静态场景在点光源下的间接光照进行拟合。以上方法均仅适用于点光源而无法轻易扩展到其他更加复杂的光照类型。

总之，现有的基于预计算的渲染算法的适用范围均存在明显限制，动态面光源下全频率全局光照的快速绘制仍是尚未解决的重要研究方向。

2.3.2.3　基于光线跟踪的渲染算法

在离线渲染中，基于光线跟踪的算法是使用最为广泛的全局光照渲染算法。近年来，得益于 GPU 算力的提升和其新增的硬件层次的光线跟踪支持，基于 GPU 的光线跟踪方法相较于传统的基于 CPU 的光线跟踪方法而言，其效率取得明显提升。然而，目前的硬件资源和算法仍然难以实现实时光线跟踪渲染。将深度学习和光线跟踪进行结合是近年来被广受关注的研究领域。其中，实时光线跟踪（real time ray tracing，RTRT）是研究的热点方向，受限于实时性要求，光线跟踪算法只能使用极低的采样数，导致其渲染结果存在明显的噪声，因此需要借助基于深度学习的降噪方法[138-141]来进行后处理降噪。除降噪外，光传输过程中的采样优化[142-143]也是具有潜力的深度学习和光线跟踪方法的结合方向。然而，目前而言，光线跟踪方法仍然无法针对复杂场景实现实时的全频率全局光照渲染，其原因主要包括：一方面，极低的采样数难以捕捉复杂而高频的全局光照效果，另一方面，目前的降噪方法也并不适用于高频全局光照的降噪。

Bitterli 等人[144]提出一种基于水塘抽样的时空重要性重采样（Reservoir-based SpatioTemporal Importance Resampling，ReSTIR）方法，该方法利用在时间和空间两个维度复用邻域采样的方式来提高光线跟踪渲染的收敛速度。该方法可以支持多光源下直接光照的快速绘制，但无法轻易扩展到全局光照渲染中。Müller 等

人[145]提出神经辐射亮度缓存（neural radiance cache，NRC）方法，该方法可针对动态场景实现全局光照快速绘制。该方法的核心思路是在线训练神经辐射亮度缓存表达，即在光线跟踪渲染的同时进行该神经表达的训练，光线跟踪渲染过程不仅负责最终渲染结果的生成还需要为神经网络训练提供训练数据，而神经辐射亮度缓存则可以为光线跟踪提供指导以加快其收敛速度。然而，该方法无法有效捕捉和局部表面信息无关的高频全局光照效果，例如阴影和焦散等。此外，为实现实时的运行效率，该方法依赖于 Hasselgren 等人[146]提出的实时降噪方法对渲染结果进行后处理降噪。由于降噪方法倾向于模糊高频细节，因此 NRC 方法在降噪后也倾向于得到较为模糊的全局光照效果。此外，NRC 的整体算法流程依赖于大量繁琐的工程实现，限制了其使用范围。本文第 5 章所提方法可以有效生成包含焦散在内的高频全局光照效果，同时具有易于实现以及易于和现有实时渲染管线集成的优势。

第 3 章　基于逆渲染和数据驱动的表观建模

正如第 1 章所述，采集与建模是计算机图形学算法的基础步骤，其主要任务是通过对真实世界进行观测和离散采样来生成后续图形学算法所需要的数据和模型。在真实感渲染中，高质量的表观属性是保证渲染结果逼真的重要因素。传统上，高质量的表观数据需要专业的美术人员利用复杂的专业设计软件通过交互式绘制得到，该传统流程的主要缺陷包括：其一，该过程依赖于专业美术人员的领域知识和熟练的专业软件操作经验；其二，该过程中繁琐的交互式绘制需要耗费大量的时间和精力。随着各行各业对表观数据需求的日益增长，费时费力的传统流程已经逐渐难以满足需求，因此研究自动化表观重建方法迫在眉睫。表观建模是以若干张物体的观测图片作为输入通过算法自动化地重建物体表观属性的一类方法的统称，由于其完全自动化而无需手工交互的特性可以大大降低人力成本和时间开销，因此被广泛应用于电子游戏制作、影视动画的特效制作等传统领域以及电子商品展示、面向普通用户的人脸特效生成等其他领域当中。

图 3.1　基于逆渲染和数据驱动的表观建模方法的材质重建结果可视化

图形学中通常使用空间变化双向反射分布函数（下文简称为 SVBRDF）来描述物体的表观属性。由于 SVBRDF 是一个高维函数，因此对其进行采集和建模均是非常富有挑战性的问题。如本文 2.1.1 节所述，前人工作主要可以分为经典逆渲染方法和基于深度学习的方法两类。经典逆渲染方法[16,18]的核心思路是通过优化材质贴图来最小化材质贴图对应的渲染图片和输入图片间的差距，该类方法只有在给定大量输入图片的情况下才可以给出准确的表观估计，而当输入图片数量不足时，经典逆渲染方法无法给出可信的结果。近年来，基于深度学习的表观建模方法[23-26,48]不断涌现，该类方法可以从单张输入图片中回归出可信的表观估计，但由于该类方法无法鲁棒地求解表观估计中常见的二义性问题，因此其重建结果不够准确，难以达到实际应用的质量需求。给定更多的输入图片可以有效地解决二义性问题，然而目前尚不存在有效的策略可以将基于深度学习的表观建模方法扩

展到多张输入图片。因此，无论是传统的经典逆渲染方法还是基于深度学习的回归方法均无法有效处理任意数量输入图片下的表观建模问题。

3.1　本章引言

针对上述表观建模问题，本章提出一种基于逆渲染和数据驱动的全新方法，支持从任意数量输入图片（从单张到多张）中估计平面物体的高分辨率的空间变化表观属性（结果示例参见图 3.1 ）。具体而言，本章方法的目标是：当输入图片数量仅为单张或者少量的时候，本章方法可以给出可信的表观估计结果；而随着输入图片逐渐增多，本章方法重建的表观估计的质量也不断提高，最终会收敛到精确的结果。为实现以上目标，本章提出的表观建模框架将深度学习和逆渲染策略以一种灵活且易于实现的方式结合在一起，其核心思路和亮点是在深度学习构造的 SVBRDF 隐空间中进行逆渲染优化来估计物体表观属性。此外，本章方法还提出细节增强策略来进一步增加重建贴图中的高频细节。本章方法在整个优化过程无需任何其他手工设计的启发式正则约束。上述 SVBRDF 隐空间是通过专为优化任务而设计的全卷积的自编码器网络[147]来进行构造，而逆渲染优化的变量是 SVBRDF 隐空间中的隐向量，优化过程中通过计算隐向量对应的 SVBRDF 贴图的渲染误差并将该误差梯度经由可微分渲染层反向传播到隐向量完成更新。

经典逆渲染方法采用直接在 SVBRDF 贴图空间进行逆渲染优化的方式实现材质贴图的重建，而本章方法的核心策略则是在深度学习构造的隐空间中进行逆渲染优化，因此该核心策略也叫做深度逆渲染策略。深度逆渲染策略的核心优势在于可以充分利用大量表观数据中的先验信息来约束逆渲染优化过程。然而，构建适合于优化的 SVBRDF 隐空间仍然存在以下三点核心挑战：首先，如何设计自编码器网络架构，使之更加适合于逆渲染优化问题？其次，如何训练自编码器网络，使得构造好的隐空间具有光滑且适合于基于梯度的优化算法进行逆渲染优化的良好性质？最后，如何找到一种简单但高效的策略进行初始化？

本章提出了以下策略来解决上述三点核心挑战：

1. 在神经网络架构设计方面，本章提出在自编码器网络的解码器中不使用批归一化[148]（batch normalization，简称 BN）。批归一化是包含自编码器网络在内的卷积神经网络中一种常用的正则化策略。然而，本章方法中的 SVBRDF 自编码器网络所起的作用和传统的自编码器网络不同，传统的自编码网络的训练目标是使得输出结果尽可能和网络输入接近，而本章方法中训练 SVBRDF 自编码器网络的主要目的是构建一个适合于优化的 SVBRDF 的隐空间。我们发现批归一化会影响梯度从损失函数到隐空间变量的反向传播，从而降低

逆渲染优化的效率和质量，因此，本章提出在自编码器网络的解码器中不使用批归一化；

2. 在自编码器网络训练方面，本章提出在传统贴图损失函数的基础上增加渲染损失函数和光滑性约束。其中，渲染损失函数可以保证材质贴图间的误差度量和最终渲染结果间的误差度量具有一致性，而光滑性约束则保证了隐空间中的微小扰动对应着解码后 SVBRDF 贴图间的微小区别，进而提升了优化过程的鲁棒性；
3. 在逆渲染优化的初始化方面，本章提出以基于单张图片的深度表观重建方法[25-26]的预测结果来作为本章方法的优化起点。本章方法并不局限于某一种特定的方法来进行初始化，可以由单张输入图片完成鲁棒表观重建的方法均可以为本章方法中的优化过程提供可靠的初始化。

由于 SVBRDF 自编码器网络采用全卷积的神经网络架构，因此本章方法可以在无需任何重新训练的情况下支持更高分辨率下的表观重建。通过将自编码器网络的隐空间向量的分辨率扩大，解码器输出的 SVBRDF 贴图的分辨率也随之扩大。我们在大量公开的 SVBRDF 数据集以及真实拍摄的材质数据集上验证了本章方法可以从任意数量输入图片中得到高质量的材质估计。此外，实验表明即使在仅有单张输入图片的情况下，本章方法相较于前人方法[25-26]也取得了进一步的提升。

总之，本章方法的主要贡献包括：

- 可以针对从单张到多张的任意数量的输入图片给出高质量的表观估计；
- 支持任意分辨率下的表观估计，并不局限于训练数据中的固定分辨率；
- 相较于之前的基于单张图片的深度表观建模方法，进一步提升了单张输入图片下表观估计的质量和鲁棒性。

3.2　方法概览

本节首先介绍本章方法适用的场景假设和采集拍摄设置（ 3.2.1 节），然后介绍深度逆渲染策略的核心思路和形式化定义（ 3.2.2 节），最后讨论基于深度学习的表观建模方法中逆渲染优化策略和回归策略的区别（ 3.2.3 节）。

3.2.1　场景假设

正如之前所述，本章方法的目标是从任意数量输入图片中估计物体的空间变化表观属性。几何方面，我们假定拍摄样本的几何满足近平面假设，即宏观上物体的几何是一个平面而其微观几何细节由法线贴图来建模。材质方面，我们假定拍摄样本的表观属性包含漫反射部分和高光反射部分，其中漫反射部分使用 Lambertian

(a) 贺卡　(b) 皮革　(c) 乐高　(d) 金属　(e) 生锈金属　(f) 木头

图 3.2　真实拍摄的材质样本示例

材质模型来描述，高光反射部分使用基于 GGX 法向分布函数的 Cook-Torrance 微面元模型来描述，物体表面一点 p 处的反射属性包含以下参数：p 点处的法向量 $n(p)$，漫反射颜色 $k_d(p)$，高光反射颜色 $k_s(p)$ 以及粗糙度 $\alpha(p)$。

数据采集方面，我们假定每张输入图片 $\{I_i\}_i$ 均由闪光灯与相机共位的手机拍摄得到。在拍摄过程中，相机在距拍摄样本恒定距离的平面上移动来完成多张图片的拍摄，我们对于拍摄每张图片时相机的位置没有严格的要求，即用户在手持拍摄过程中可以自由移动相机而无需遵守某种设定好的模式或者路径。我们假定每张图片的相机内参和外参 C_i 均是已知的，且所有输入图片均已对齐。在实际采集时，我们采用相机标定技术完成相机参数的估计和多视角图片的对齐。此外，本章方法既可以支持高动态范围的输入图片，也可以支持低动态范围的输入图片。图 3.2 展示了通过上述拍摄流程所得到的部分真实材质样本的观测图片，图中所示材质样本包括多种类型：(a) 贺卡样本，在叶子部分和其他部分展现出截然不同的复杂反射特性；(b) 皮革样本，展示出细致的高光反射和丰富的纹理细节；(c) 乐高样本，展示出复杂的表面微几何；(d) 金属样本，展示出强烈的高光反射和复杂的表面微几何；(e) 生锈金属样本，展示出复杂的表面纹理；(f) 木头样本，展示出规律性的纹理样式和复杂的高光反射。

3.2.2　深度逆渲染策略的核心思路

由于本章方法中的深度逆渲染策略同经典逆渲染策略类似，二者均是采用基于优化的逆渲染策略，因此在介绍深度逆渲染策略前，我们先简单回顾一下经典逆渲染策略。经典逆渲染方法的核心思路是通过优化材质参数 $s = (n, k_d, \alpha, k_s)$ 来最小化输入图片 $\{I_i\}_i$ 和渲染图片 $R(s, C_i)$ (渲染图片根据材质参数 s 和相机参数 C_i 生成) 间的损失函数 $\mathcal{L}(\cdot,\cdot)$：

$$\underset{s}{\operatorname{argmin}} \sum_i \mathcal{L}(I_i, R(s, C_i)). \tag{3.1}$$

其中损失函数采用 log 空间的 L_1 距离，log 编码可以有效降低高光处可能存在的极端大的像素值对于整体训练的影响，损失函数的具体形式为：

$$\mathcal{L}(x, y) = ||\log(x + 0.01) - \log(y + 0.01)||_1. \tag{3.2}$$

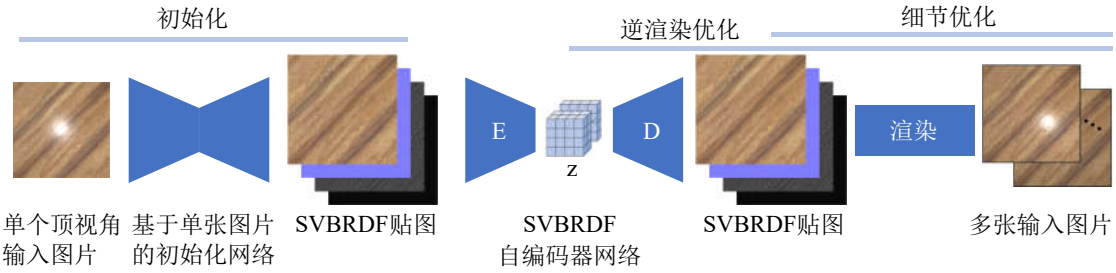


图 3.3　基于逆渲染和数据驱动的表观建模方法的流程示意图

与以上经典逆渲染策略不同，本章提出的深度逆渲染策略并不直接优化材质参数 s，而是优化隐空间中的隐向量 z：

$$\underset{z}{\operatorname{argmin}} \sum_{i} \mathcal{L}(I_i, R(D(z), C_i)). \tag{3.3}$$

式（3.1）和式（3.3）的主要区别在于：前者优化的是逐像素的材质贴图 s，而后者则优化整张图片对应的隐空间向量 z。实际上，SVBRDF 隐空间包含了从大量材质数据学习得到的关于 SVBRDF 属性的先验知识，可以为优化过程提供正则化约束。在本章方法中，SVBRDF 隐空间通过全卷积的自编码器神经网络来建模，该自编码器网络包含负责将 SVBRDF 贴图 s 映射为隐空间向量 z 的编码器网络 $E(\cdot)$ 以及负责将隐空间向量 z 映射回对应的 SVBRDF 贴图 s 的解码器网络 $D(\cdot)$:

$$z = E(s), \tag{3.4}$$

$$s = D(z). \tag{3.5}$$

图 3.3 概述了本章提出的基于逆渲染和数据驱动的表观建模框架的整体流程。概括而言，专为逆渲染优化而设计的 SVBRDF 自编码器网络是本章方法的核心组件，本章方法通过优化 SVBRDF 隐空间向量来最小化渲染损失函数以实现材质重建，其优势是在整个优化过程无需任何其他手工设计的启发式正则约束。另外，逆渲染优化的初始化通过前人提出的基于单张输入图片的 SVBRDF 估计算法[25]给出，而深度逆渲染优化之后会通过细节增强这一后处理步骤来进一步细化材质贴图中的细节。SVBRDF 自编码器的细节将在 3.3 节中介绍，逆渲染优化的细节将在 3.4 节中介绍。

3.2.3　讨论：回归策略和优化策略的分析

基于深度学习的表观建模方法中，之前的工作均采用回归策略，即神经网络以观测图片为输入直接预测材质贴图，而本文的深度逆渲染策略则属于优化策略，即通过在深度学习构造的隐空间进行逆渲染优化来重建材质贴图。本章方法选择使用优化策略而非回归策略有理论和实践两方面的考量。

理论上，基于深度学习的回归策略可以视为将表观优化过程提前到了训练阶段。由于回归网络的训练过程是通过优化神经网络可学习参数来最小化其在整个训练数据集上的误差，因此，训练后该类网络的精度会随训练数据的变化而变化并且对训练数据集覆盖范围外的样本没有精度保证。在测试阶段，对于用户给定的单个物体的若干输入图片而言，回归网络往往无法找到最优解。深度逆渲染策略则不直接训练从输入图片到材质贴图的回归网络，而是将训练好的隐空间作为优化的正则化约束。对于测试阶段的单个物体的若干输入图片而言，本章方法可以充分利用输入数据中的信息以更好地处理不在训练数据集覆盖范围内的样本。

实践中，回归网络在处理任意视角下拍摄的多张输入图片上存在困难。测试阶段多张输入图片对应的视角组合和训练时的视角组合可能不同，导致以卷积神经网络为代表的回归网络无法有效地利用多张图片的信息。深度逆渲染策略则将神经网络模块嵌入到逆渲染流程中，其可以有效继承逆渲染方法在兼容不同数量和不同拍摄条件下的输入图片方面的优势。本章的神经网络模块的输入和输出均为材质贴图，并不涉及输入图片，因此训练过程中并不需要考虑视角信息未知等情况。在深度学习构造的隐空间中进行逆渲染优化可以鲁棒地处理欠约束和二义性问题且无需提供手工设计的正则化约束。

3.3　SVBRDF 自编码器网络

本章方法中深度逆渲染策略所优化的隐空间是通过 SVBRDF 自编码器网络来构造。本节将具体介绍 SVBRDF 自编码器网络的设计和训练方面的细节。

3.3.1　神经网络架构

本章方法中的 SVBRDF 自编码器网络主要有两个作用：首先是构造适合于优化的且紧凑的隐空间，其次是同其他自编码器网络一样保证解码后的输出仍然可以保留输入中的复杂空间变化。

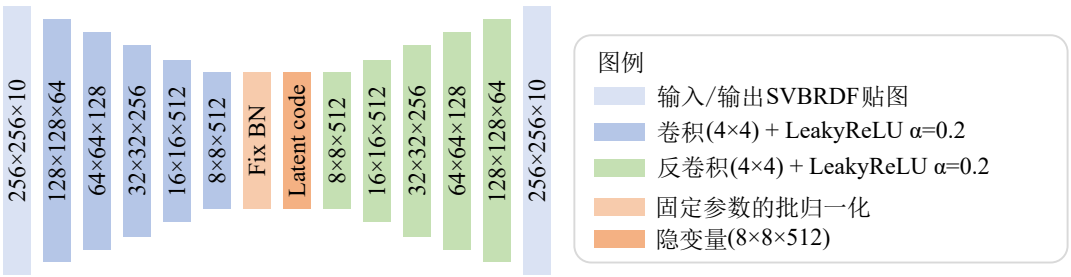


图 3.4　SVBRDF 自编码器网络架构

本章方法选择使用编码器加解码器的经典架构，其中编码器网络的输入是分

辨率为 256×256 的材质贴图 $s = (n, k_d, \alpha, k_s)$，该输入经过一系列卷积操作后得到的分辨率为 $8 \times 8 \times 512$ 的特征贴图所在空间即是逆渲染优化的隐空间，解码器网络则以隐空间向量作为输入，经过一系列反卷积操作恢复原分辨率。SVBRDF 自编码器网络的网络架构和卷积层参数等细节参见图 3.4 。

批归一化设计　本章提出的自编码器网络架构中关于批归一化的设计和使用同标准实践不同。在神经网络的标准实践中，每个卷积层后均会添加批归一化，归一化层一方面可以提升训练的表现和稳定性，另一方面还可以提供一定的正则化约束。批归一化的主要包含两个步骤，针对每个批次的输入数据，首先计算其均值和方差并据此对原始数据进行归一化，然后通过全局可学习的缩放和偏置参数对数据进行线性变换。实践中人们通常认为批归一化可以起到惩罚不常见特征（例如噪声）的作用。

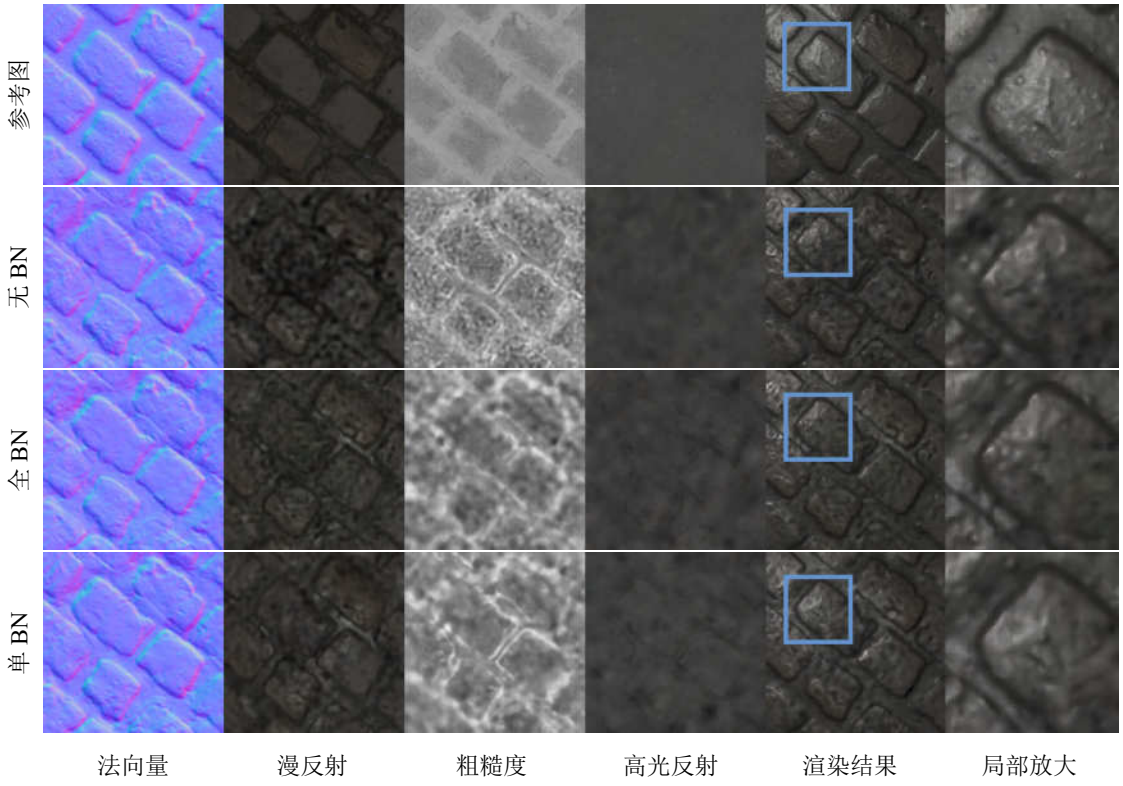


图 3.5　不同批归一化设计的结果对比，本实验中输入图片数量为 1。

为了探索批归一化在表观建模中的作用，我们训练了两个版本的自编码器网络：其中一个不包含任何批归一化，而另一个则在每个卷积层后面均添加批归一化。给定单张顶视角的输入图片，我们分别在两个版本的自编码器网络的隐空间进行逆渲染优化来重建 SVBRDF（结果参见图 3.5 ）。基于无批归一化自编码器所重建的材质贴图存在明显的噪声（图 3.5 第二行），而每个卷积层后均带有批归一

化的自编码器所重建的材质贴图（图 3.5 第三行）虽然更加光滑，但存在过度光滑的问题，即很多材质细节也和噪声一同被模糊。总之，批归一化在表观建模中具有正则化效果，可以有效地消除结果中的噪声，但过度正则化也会导致过度模糊。

在逆渲染优化中，我们期望解码器网络在梯度从渲染图片反向传播到隐空间向量的过程中可以尽量保留渲染图片中的细节信息。由于解码器中的批归一化中会过度正则化，因此会导致细节相关的梯度信息在反向传播中丢失。编码器中的批归一化可以在正则化隐空间的同时不影响逆渲染优化的梯度反向传播。基于以上结论，本章提出在解码器网络中不使用批归一化而只在编码器网络中加入批归一化的独特网络设计。另外实验发现，编码器网络中每个卷积层后均添加批归一化和只在编码器网络最后添加批归一化的结果非常相似，因此简单起见，我们采用只在编码器网络最后添加单个批归一化的方案。图 3.5 第四行展示了该方案的重建结果，结果显示该方案比每个卷积层均带有批归一化的方案拥有更多的细节，而同时又不像完全没有批归一化的方案一样带有明显的噪声。

3.3.2　训练损失函数

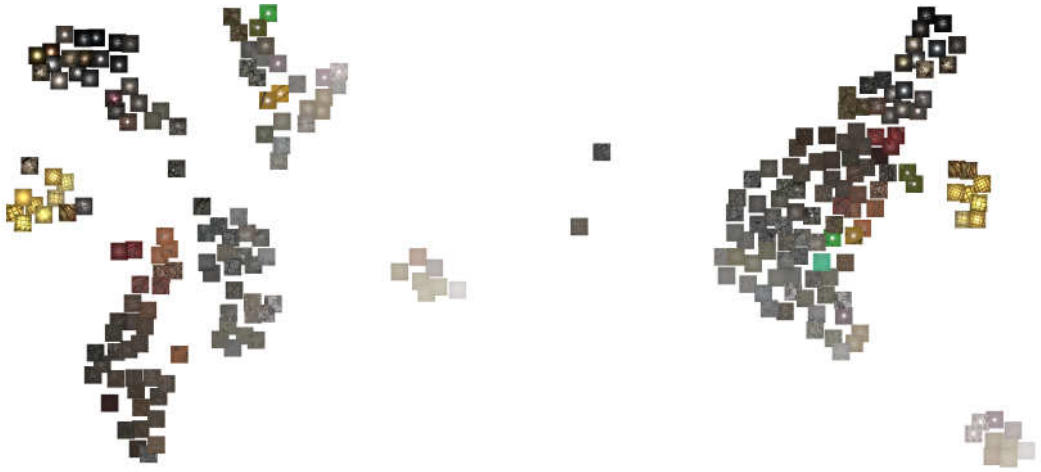
本章方法所使用的训练数据集是 Deschaintre 等人[25]所提供的 SVBRDF 数据集。使用相同的训练数据集，可以使得后续结果对比更加公平。

本章方法所使用的训练损失函数则是三项之和：

$$\mathcal{L}_{train} = \mathcal{L}_{map} + \frac{1}{9}\mathcal{L}_{render} + \mathcal{L}_{smooth}, \tag{3.6}$$

其中，$\mathcal{L}_{map}$ 代表材质贴图间的 L_1 损失函数，$\mathcal{L}_{render}$ 则是 9 张渲染图片间的 log 空间的 L_1 损失函数，$\mathcal{L}_{smooth}$ 是光滑性约束。渲染损失函数中 9 张渲染图片的视角和光照选择参照了 Deschaintre 等人[25]提出的方案：其中 3 张图片对应的视角和光源假定在无限远处，因此其视角和光源方向均从上半球面的余弦加权分布中独立采样得到；而其他 6 张图片则假定其相机和光源在近处，视点位置通过在样本平面内随机采样得到，相机和光源距离视点的 log 距离服从正态分布（其均值为 0.5，标准差为 0.75，这里假定样本平面是大小为 2 的正方形），光源方向则从上半球余弦加权分布中采样，视角方向则设置为光源反向的镜像方向。

在逆渲染优化隐空间变量的过程中，我们依赖于渲染图片梯度信息的反向传播来指引整个优化过程。理想情况下，渲染图片中表观的微小差别应该导致反向传播后隐空间变量的微小更新。然而，训练过程中难以有效评估渲染图片的表观变化，因此在没有任何正则化约束的情况下无法保证该性质成立。本章提出在自编码器网络的训练过程中引入一个比上述性质更强的约束，即要求隐空间变量的微小改变对应于 SVBRDF 贴图的微小改变。



(a) 无光滑性约束　　(b) 有光滑性约束

图 3.6　有无光滑性约束的隐空间 t-SNE 可视化对比

上述光滑性约束的具体定义如下：

$$\mathcal{L}_{smooth} = \lambda_{smooth} ||D(z) - D(z+\xi)||_1, \tag{3.7}$$

其中，ξ 是服从均值为 0、方差为 0.2 的正态分布的随机变量，λ_{smooth} 是用于控制光滑性程度的权重因子（实验发现 $\lambda_{smooth} = 2$ 可以取得很好的结果）。

光滑性约束通过重组隐空间的分布使得邻近的隐空间变量对应于相似的 SVBRDF 贴图。图 3.6 展示了有无光滑性约束两种情况下的隐空间 t-SNE[149]可视化对比，从图中可以看出，隐空间的流形在加上光滑性约束后整体更加连续，因而更加适合于插值和优化等任务。

3.4　逆渲染优化

输入样本的表观属性需要在通过自编码器网络构造的隐变量空间中执行逆渲染优化来完成重建。如图 3.3 所示，逆渲染优化的变量是隐空间向量 z，z 通过固定参数的解码器后得到其对应的 SVBRDF 贴图，再经由可微分渲染器得到渲染图片，最后计算渲染图片和输入图片间的损失函数并将梯度反向传播到隐空间向量 z 完成更新。然而，为鲁棒地完成以上过程，逆渲染优化中还需要仔细考虑初始化、细节增强、高分辨率贴图支持以及优化加速等问题，本节将具体介绍这些内容。

3.4.1　初始化分析和策略

与其他基于逆渲染的方法一样，初始化对于本章方法最终收敛的结果质量有重要的影响。

概率论视角的分析　为了更好地理解自编码器网络在逆渲染优化中的作用，进而可以更好地了解逆渲染优化所需的初始化策略，我们以概率论的视角重新表达该问题：给定一系列输入图片 $\{I_i\}_i$，逆渲染问题可以视为最大化 SVBRDF s 的条件概率 $\text{argmax}_s\, P(s|\{I_i\}_i)$。假定每个 I_i 是相互独立的，那么根据贝叶斯定理该式可写作：

$$P(s|\{I\}_i) = \Pi_i \left(\frac{P(I_i|s)P(s)}{P(I_i)} \right). \tag{3.8}$$

假定每张图片等可能，那么上式中 $P(I_i)$ 可以忽略：

$$P(s|\{I\}_i) = \Pi_i \left(P(I_i|s)P(s) \right). \tag{3.9}$$

式（3.9）中第一项 $P(I_i|s)$ 代表给定 SVBRDF s 的情况下某图片 I_i 出现的概率，该项对应于优化过程中的渲染损失函数；第二项 $P(s)$ 代表 SVBRDF s 出现的概率。当输入图片足够多的时候，数据可以提供充分的约束，此时式（3.9）中第一项将占主导。所有图片的联合概率将是各自概率空间的交集，输入图片越多其相交的区域就会越小，极限情况下只有单个 SVBRDF 可以符合所有输入图片的约束。因此，这种情况下 $P(s)$ 的作用相对较小。反之，当输入图片数量不足时，很多 SVBRDF 的条件概率都很大，这种情况下 $P(s)$ 作为正则项的作用愈发明显。需要说明的是，自编码器网络本身并不能提供该概率，其作用仅是提供一种 SVBRDF 数据空间的隐空间嵌入，并不能区分嵌入该空间的不同 SVBRDF 的概率。前面介绍的光滑性约束（参见式（3.7））会在训练过程中对隐空间进行重组，使得邻近的隐变量解码后得到相似的材质贴图。假定相似的 SVBRDF 具有相似的概率，那么可以认为这种情况下 $P(s)$ 在局部是常量：

$$P(D(z)) \approx P(D(z+\xi)). \tag{3.10}$$

如果优化的初始点落在满足上式性质的区域（即 $P(s)$ 接近常量，概率分布的形状由渲染损失函数主导的区域），则优化后容易收敛到好的解。虽然我们无法直接找到满足上述条件的初始点 z_0，但光滑性约束可以增大空间中满足 $P(s)$ 接近常量的区域，降低了初始化的难度，进而提升了方法的鲁棒性。

初始化策略　本章使用前人提出的基于单张图片的深度表观建模方法[25-26]来进行初始化。该类方法训练的目标是最大化概率 $P(s|I_0)$，对于很多材质而言，该概率可以视为 $P(s|\{I_i\}_i)$ 的一个可接受的近似。实验表明本章方法可以基于该类方法所提供的 SVBRDF 作为初值而最终收敛到合理的解。图 3.7 展示了一个木头材质在不同初始化策略下的材质重建结果对比，其中采用随机初始化策略无法收敛到合理的解，其重建结果存在明显瑕疵，而基于前人提出的基于单张图片的深度表观

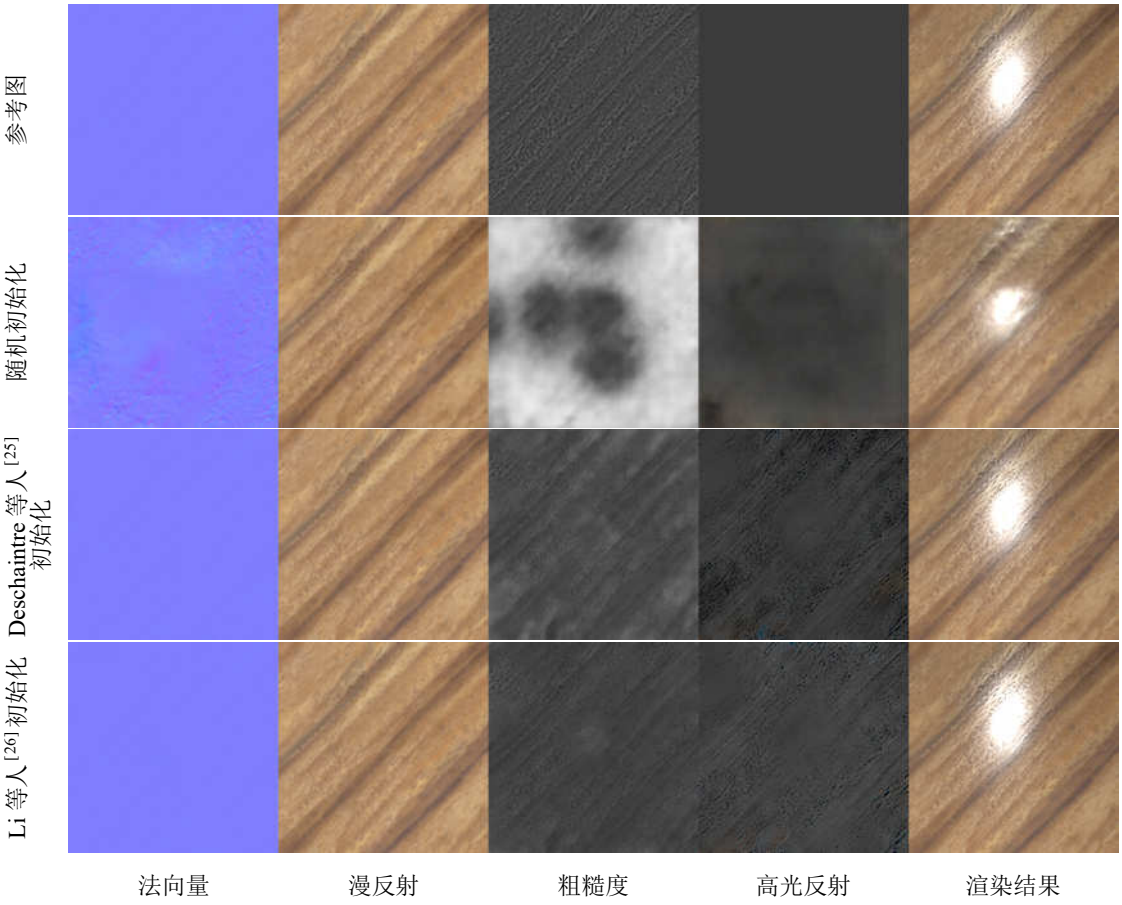


图 3.7　不同初始化策略的结果对比，本实验中输入图片数量为 5。

建模方法（包括图中第三行的 Deschaintre 等人[25]的方法和第四行的 Li 等人[26]的方法）则都可以得到合理的结果。本章后续的所有实验默认使用 Deschaintre 等人[25]的方法进行初始化。

3.4.2　细节增强

通常而言，针对给定的输入，自编码器神经网络首先会利用编码器网络将其映射到低维隐空间（也被叫做瓶颈层），然后再经由解码器网络将其重新映射回原输入空间。由于输入空间往往是非常复杂的，超出了自编码器瓶颈层的表达范围，因此在自编码器的编码-解码过程中会损失部分细节。在神经网络的标准实践中，解决自编码器细节丢失问题的常用方法是在编码器网络与解码器网络的同分辨率层之间添加跳跃连接，从而将输入中包含的细节信息直接传递到解码过程中。然而，本章方法在逆渲染优化过程中仅依赖解码器网络而不使用编码器网络，因而无法使用跳跃连接技术。另一种直接的思路是通过扩展隐空间的维度使之可以表达更多细节，但该思路会降低自编码器网络的正则化效果，导致优化后得到的材质贴图带有更多瑕疵。

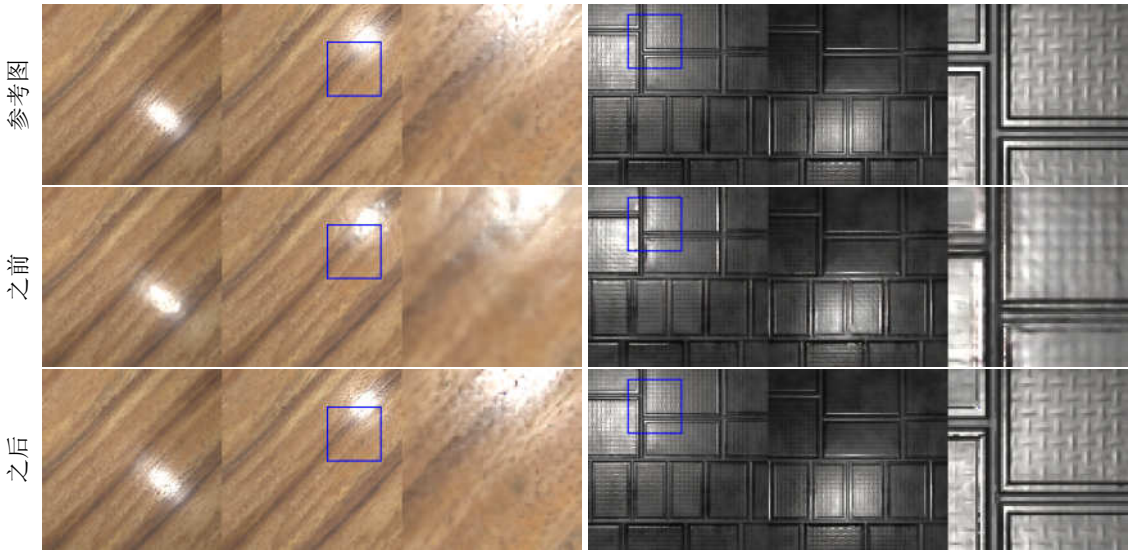


图 3.8　应用细节增强策略前后的结果对比，本实验中输入图片数量为 5。

本章提出了细节增强策略来解决上述细节丢失问题。细节增强是在深度逆渲染优化的基础上采用类似于经典逆渲染方法的策略（其优化目标参见式（3.1））来直接优化材质贴图，以进一步减小渲染误差。如之前所述，欠约束的逆渲染优化在初值不够好的情况下会倾向于得到有瑕疵的结果。本章提出的细节增强策略可以避免该问题的原因在于：首先，深度逆渲染优化得到的材质贴图本身已经是足够好的初值；其次，由于细节增强仅用于添加丢失的细节信息，因此只需要少量的迭代优化步骤即可收敛到理想的结果，保证了细节增强前后的结果不会偏差太远。图 3.8 展示了有无细节增强策略的材质重建结果对比，从图中可以看出，细节增强之前的可视化结果（第二行）相较于参考图（第一行）而言存在细节丢失和过度模糊的情况，而细节增强之后的可视化结果可以在保持整体特征不变的情况下拥有更多的高光反射细节和表面纹理细节。

讨论：细节增强思路在人脸纹理估计中的应用

将本章方法的主体思路应用于人脸高频纹理估计任务当中，可以实现基于单张人脸图片的高频人脸纹理估计。如之前所述，本章提出的基于逆渲染和数据驱动的表观建模方法中，第一步通过数据驱动的逆渲染优化过程实现初步的表观估计，第二步则以第一步的结果为起点，仅利用输入图片本身信息来进行进一步优化表观属性以实现细节增强。类似的思路可以应用于人脸纹理估计任务中：第一步通过数据驱动的人脸参数化模型[150]实现人脸低频的几何和纹理的初步估计，第二步则以第一步恢复的低频纹理为起点和正则约束，通过基于单张输入图片的进一步纹理优化来将原输入图片中的高频信息引入到纹理贴图当中。

具体而言，第一步基于 Blanz 等人[150]提出的三维可变形人脸模型（3D mor-

phable model, 3DMM）进行低频的人脸几何和纹理的恢复。3DMM 是通过对数百张人脸数据进行主成分分析而得到的一种人脸参数化表达模型，其中人脸的几何和纹理分别使用两组高维的参数向量来表达。基于 3DMM 的人脸重建过程可以概括为：首先，在输入图片中进行二维人脸关键点检测，并利用二维关键点和 3DMM 中三维关键点之间的语义对应关系来优化相机参数和 3DMM 的几何参数向量；其次，在假定场景光照可以由球面谐波函数表达而人脸材质满足 Lambertian 假设的情况下，通过最小化渲染误差来逆渲染优化 3DMM 的纹理参数向量。通过以上策略可以重建得到初步的人脸几何和纹理，但重建的纹理（图 3.9 (b)）中缺失很多重要的细节（例如毛孔、雀斑和胡子等）。第二步人脸细节增强的过程如下：首先，我们注意到，光照信息可以表达为高频人脸图片（即观测图片）和待恢复的高频纹理的比值，同时也可以表达为低频人脸图片（基于初步重建结果的渲染图片）和已恢复的低频纹理的比值。根据该不变性关系，容易得到待恢复的高频纹理图片（图 3.9 (c)）；其次，我们观察到，该高频纹理虽然保留了丰富的纹理细节但存在一定的光照残留，而低频纹理没有光照残留的问题但存在细节缺失，因此可以将低频纹理作为正则约束来实现高频纹理的高光去除，即最终的纹理图片（图 3.9 (d)）在整体上和低频纹理保持相似而同时其细节（通过梯度表达）和高频纹理接近；最后，在重光照和人脸编辑等任务中，需要将人脸区域的编辑操作传播到背景区域中以实现平滑过渡。以上方法可以实现基于单张图片的高频人脸纹理估计并成功应用于人脸重建、人脸重光照和人脸编辑等任务当中。

(a) 输入图片　(b) 低频纹理　(c) 高频纹理　(d) 最终纹理　(e) 重光照渲染

图 3.9　基于单张人脸图片的高频纹理估计结果

上述思路在复杂材质表观建模和人脸纹理估计两个任务上的成功应用可以体现出：

1. 数据驱动的逆渲染优化是一种有效的轻量化建模方式。逆渲染优化的优势主要体现在其可以充分利用观测数据中包含的有用信息和潜在的多视角一致性约束，同时，大数据先验可以为高度欠约束的逆渲染优化问题提供合适的正则化约束，因此二者结合后可以实现更加鲁棒的轻量化建模。

2. 细节增强需要更多的依赖于观测数据本身。数据驱动的逆渲染优化方法虽然可以得到合理的结果，但同时也倾向于模糊细节。细节缺失的原因在于：首先，

基于大数据的模型或表达均存在一定程度的信息压缩，会导致高频细节丢失；其次，数据驱动的逆渲染优化过程过分依赖于大数据约束而导致无法从观测图片中获得有效的高频信号。因此，细节增强的关键在于充分利用观测图片中蕴含的高频细节。

3. 数据驱动的逆渲染优化方法所得到的结果对于细节增强而言是足够好的初值。虽然仅基于观测图片进行表观建模可以有效恢复高频细节，但其存在优化无法收敛等潜在风险，导致其最终重建的表观无法实现表观、几何和光源的解耦。我们认为，数据驱动的逆渲染优化可以为后续的细节增强提供足够好的初始值和正则约束，使得细节增强中只需要在局部的光滑区域内进行迭代优化，避免了无法收敛的问题。

因此，以上思路的关键在于：1. 合理利用大数据先验来获得足够好的合理结果；2. 基于观测图片进行细节增强优化时需要保证尽量在初始点附近的局部光滑范围内进行。

3.4.3　高分辨率支持

得益于自编码器网络的全卷积设计，本章方法可以自然地扩展到高分辨率输入的情况。由于在提高输入材质贴图分辨率的情况下，自编码器网络编码后得到的隐空间变量的维度也随之扩展，进而在解码后可以得到和输入材质贴图相同分辨率的重建结果，因此在自编码器隐空间进行逆渲染优化后可以得到高分辨率的材质贴图。例如，对于训练中所使用的分辨率为 256×256 的 SVBRDF 贴图，其编码后的隐空间变量的维度是 $8\times8\times512$，而对于分辨率扩大到 16 倍（1024×1024）的 SVBRDF 贴图，编码后的隐空间变量维度也随之扩大 16 倍（$32\times32\times512$）。

3.4.4　基于多分辨率优化的加速策略

由于逆渲染优化的时间开销随图片分辨率增长而成比例增加，因此为了提高本章方法在高分辨率图片下的运行效率，我们提出一种基于多分辨率优化的加速策略。首先，我们将顶视角的输入图片降采样到 256×256 分辨率，并根据降采样后的图片来完成初始化。然后，我们在该分辨率下执行 $1k$ 步的深度逆渲染优化。接着，我们将优化得到的材质贴图通过双线性插值上采样到双倍分辨率，并将其作为该分辨率下的优化初始值，并在该分辨率下也执行 $1k$ 步的深度逆渲染优化。重复以上过程直至材质贴图达到目标分辨率，在目标分辨率下的深度逆渲染优化则执行 $2k$ 步以确保其收敛，之后再进行 200 步的细节增强优化。以上基于多分辨率优化的加速策略可以节省约一半的时间开销并保证其生成结果和直接在目标分辨率下运行本章方法所得到的结果相似。本章所展示的所有高分辨率下的材质重

建结果均采用该加速策略生成。

3.5　实验结果和分析

本节将首先介绍本章方法的具体实现细节，然后在合成数据集和真实采集数据上均进行测试并验证了本章方法可以处理任意数量输入图片，接着通过消融实验来验证各个组件的必要性，最后对本章方法和传统逆渲染方法进行对比分析。

3.5.1　实现细节

本章方法中的自编码器神经网络是基于 TensorFlow 深度学习框架[151]来实现，渲染损失函数中的可微渲染层也同样是基于 TensorFlow 中内置操作符进行组合来实现。自编码器网络的训练使用 Adam 优化器[152]，其参数设置如下：学习率为 10^{-4}，β_1 设为 0.5，其他参数遵守 TensorFlow 提供的默认值。网络参数采用均值为零、方差为 0.02 的正态分布进行随机初始化。自编码器采用端到端的训练策略共训练 $100k$ 个迭代步，批大小设置为 64。训练数据集为 Deschaintre 等人[25]所提供的 SVBRDF 数据集。在配有两块 NVIDIA GTX 1080Ti 显卡的机器上，自编码器网络的训练耗时约 16 个小时。

深度逆渲染优化和细节增强优化同样也是基于 TensorFlow 框架和 Adam 优化器来实现。Adam 优化器的学习率设置为 10^{-3}，而其他参数同上述网络训练过程中一致。深度逆渲染优化一共执行 $4k$ 步迭代，细节增强优化共执行 200 步迭代。

3.5.2　合成数据结果

我们首先在公开的 SVBRDF 数据上进行实验并与前人方法进行对比分析。由于在合成数据中已知的是 SVBRDF 贴图，因此输入图片的采集过程需要通过渲染来模拟。具体而言，每张输入图片对应的相机位置（也即共位点光源位置）通过在距物体平面固定距离的平面上随机采样生成，其相机视角方向则设置为指向被拍摄样本的中心。

图 3.10 展示了前人提出的基于单张图片的材质重建方法的结果[25-26]以及本章方法基于 1、2、5 和 20 张输入图片的材质重建结果，在本实验中本章方法均采用 Deschaintre 等人[25]方法的重建结果作为逆渲染优化的初始点。在给定单张输入图片的情况下，本章方法（图 3.10 第四行）可以取得比前人提出的基于单张图片的材质重建方法（图 3.10 第二行、第三行）更加准确的结果：本章方法重建的材质贴图和参考图（图 3.10 第一行）更为接近，其全新视角下可视化结果的瑕疵也更少且视觉上更加可信。图 3.11 展示了在更多材质样本上本章方法和其他基于

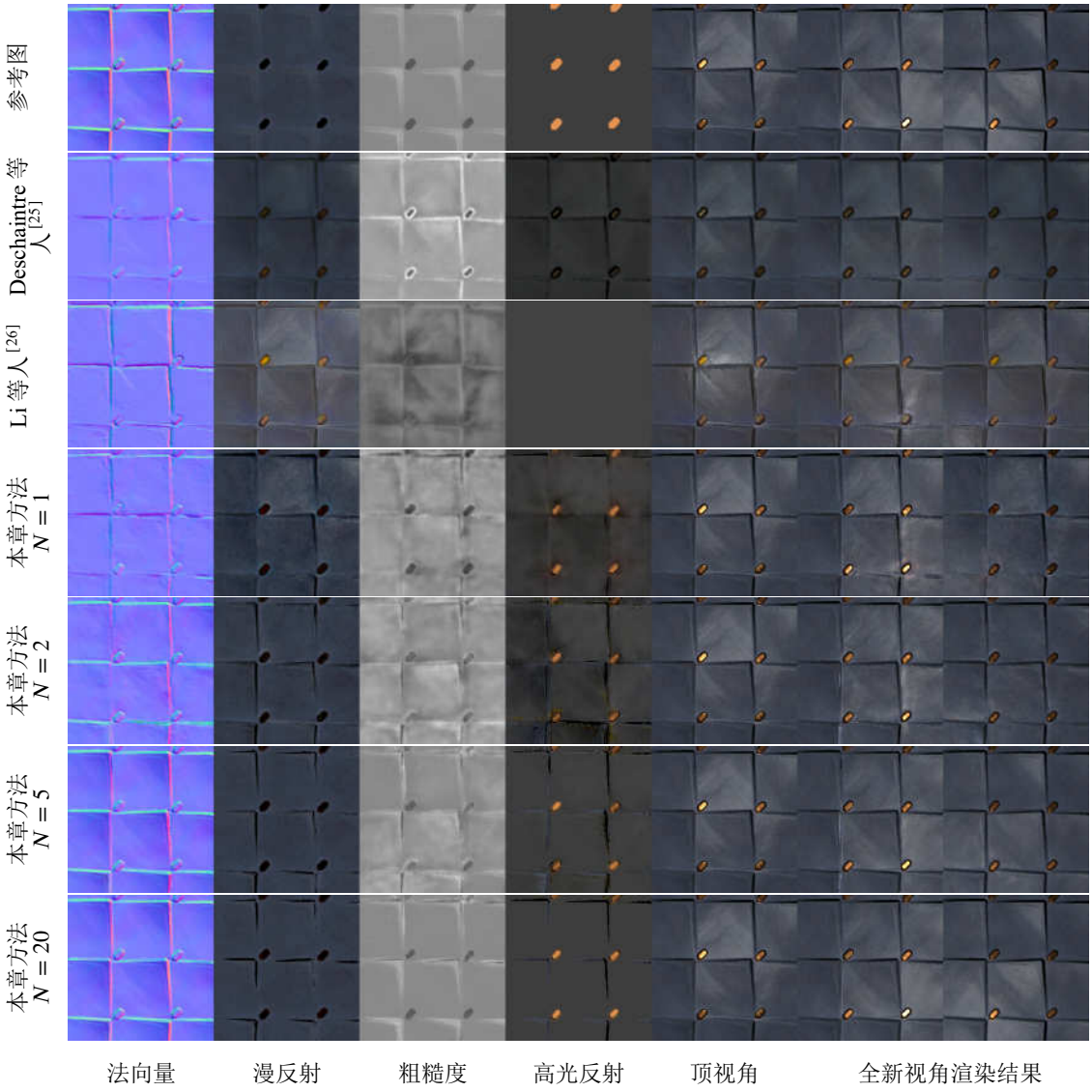


图 3.10　在合成数据上的材质重建结果

单张图片的材质重建方法的可视化对比，结果表明本章方法在仅有单张图片输入的情况下仍然可以取得比前人工作更好的结果。随着输入图片数量不断增加，本章方法的结果（图 3.10 第四行-第七行）的准确性也不断提高。需要说明的是，即使是 20 张输入图片，相比于整个可能的视角、光源方向空间而言仍然是非常稀疏的采样。

本章提出的基于逆渲染和数据驱动的表观建模方法可以支持多张图片作为输入，其好处除了可以提升重建材质贴图的质量外，还可以有效修正单张输入图片无法解决的二义性问题（即多个不同的 SVBRDF 对应于相同的渲染图片）。图 3.12 展示了本章方法针对带有二义性的材质样本的重建结果，该图中材质 A 和材质 B 具有相似的渲染结果，但其材质贴图完全不同。在仅给定单张输入图片的情况下（图 3.12 第二行），本章方法无法有效地区分两种材质，其结果中存在的差异主要

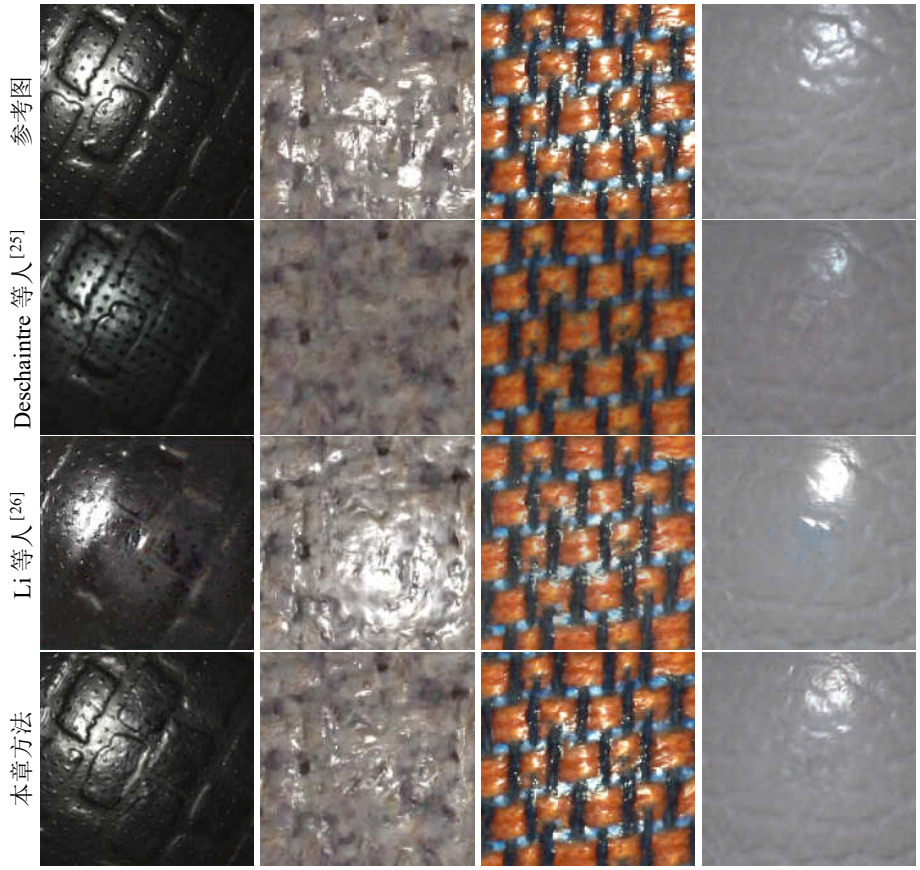


图 3.11　在合成数据上的基于单张图片的材质重建结果

来自于初始化网络[25]。然而，随着输入图片数量逐渐增多，两种材质间的二义性问题逐渐被解决。结果表明本章方法在仅给定两张输入图片（图 3.12 第三行）的情况下已经可以初步区分两种情况。当给定 20 张图片作为输入（图 3.12 最后一行）时，本章方法可以完全解决该二义性问题。

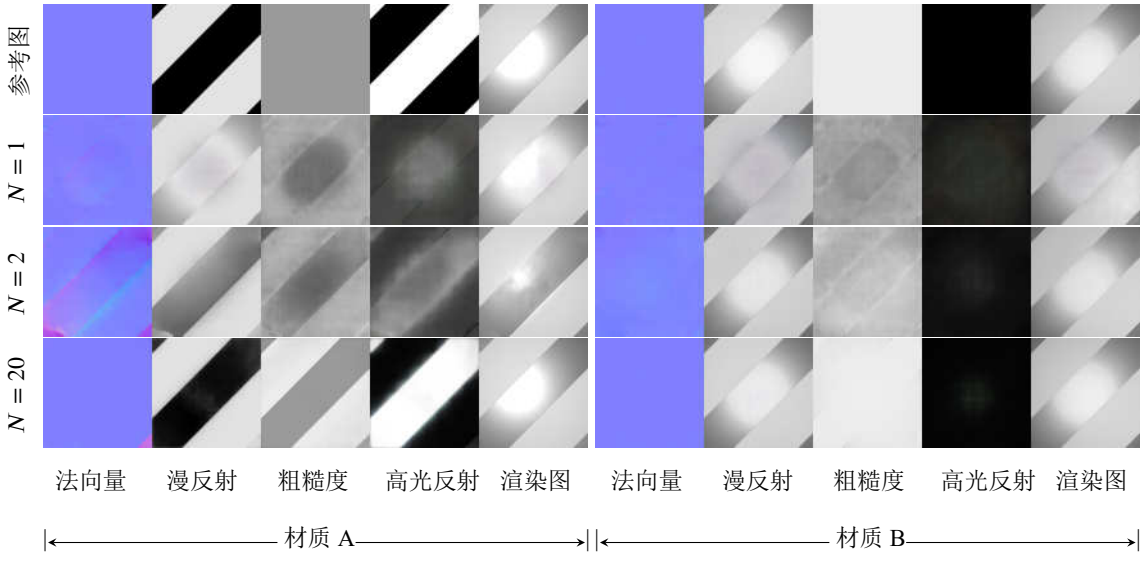


图 3.12　二义性材质样本的材质重建结果

表 3.1　三种优化策略的定量结果对比，最优值以粗体标注

N	漫反射	高光反射	粗糙度	法向量	材质贴图均值	渲染误差
			经典逆渲染			
1	0.016030	**0.02109**	0.08772	**0.003790**	0.03215	0.007594
2	0.009133	0.01818	0.07673	0.003293	0.02684	0.006141
5	0.006182	0.01485	0.06686	0.001340	0.02231	0.002163
20	0.003658	0.00724	0.05278	0.001198	0.01622	0.001092
			深度逆渲染			
1	**0.014400**	0.02123	0.07209	0.004214	0.02798	0.007443
2	0.006029	0.01725	0.06269	0.002284	0.02206	0.004717
5	0.003349	0.009078	**0.05515**	0.001042	0.01716	0.002503
20	0.002098	0.006228	0.04249	0.000715	0.01288	0.001637
			深度逆渲染 + 细节增强			
1	0.014440	0.02121	**0.07145**	0.004235	**0.02783**	**0.007394**
2	**0.005919**	**0.01722**	**0.06235**	**0.002079**	**0.02189**	**0.004369**
5	**0.002854**	**0.00817**	0.05548	**0.000489**	**0.01675**	**0.001437**
20	**0.000850**	**0.00447**	**0.04130**	**0.000273**	**0.01172**	**0.000413**

此外，我们在更大的测试数据集上对本章方法进行了定量分析。测试数据集由多种不同来源的共 42 个 SVBRDF 样本所构成：其中 20 个来自 Deschaintre 等人[25]提供的测试集，15 个来自 Aittala 等人[19]提供的数据，4 个来自 Free PBR 材质网站①，另外 3 个是由我们手工创建得到。测试数据集中的每个 SVBRDF 数据均统一裁剪到 256×256 分辨率，并基于随机采样的视角/光源方向来分别渲染 1、2、5 和 20 张图片作为输入。表 3.1 最下方一栏展示了本章方法的测试误差，我们所采用的误差度量包括：SVBRDF 贴图各个分量的 L_2 误差，SVBRDF 贴图的平均 L_2 误差以及 L_2 渲染误差，其中渲染误差是在随机采样的 100 个视角/光源组合对应的渲染图片上计算得到。

3.5.3　真实数据结果

我们在多个真实拍摄的材质样本上进行实验并验证了本章方法可以在真实样本上重建出高分辨率 (1024×1024) 的材质贴图。正如之前介绍，我们使用带有闪光灯的手机相机来完成低动态范围（Low dynamic range，LDR）照片的拍摄和采

① https://freepbr.com

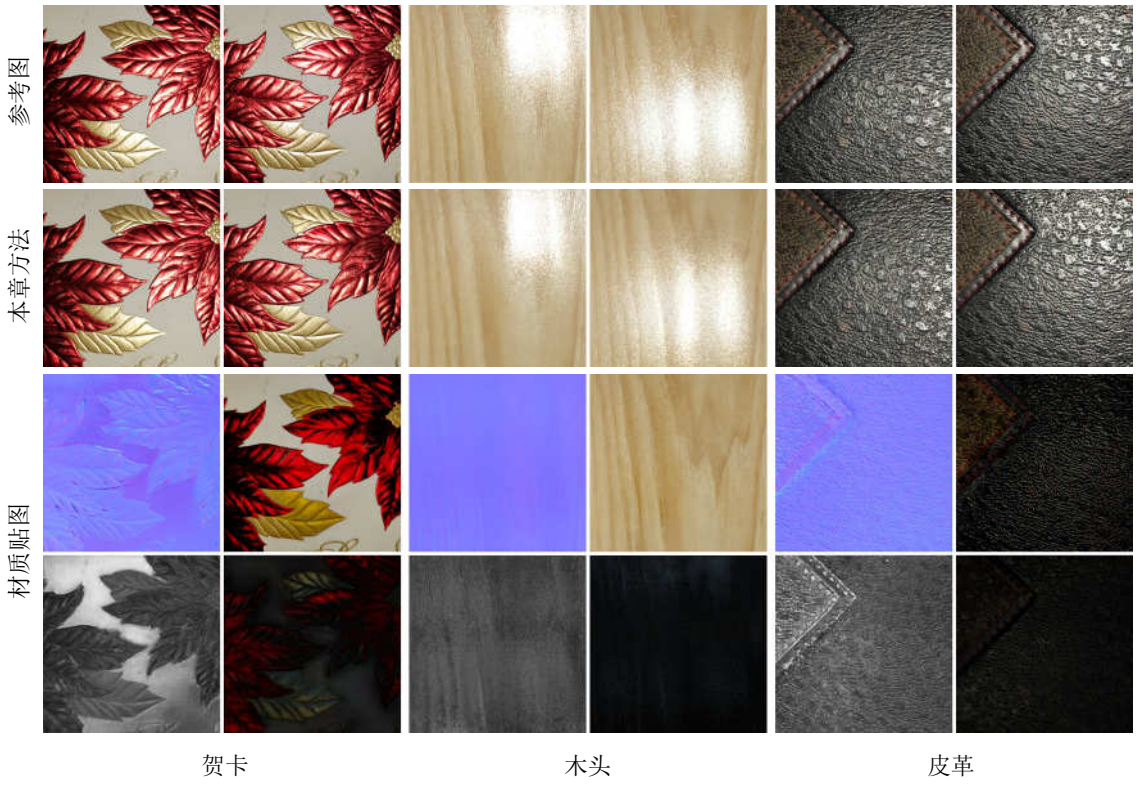


图 3.13　在真实拍摄数据上的材质重建结果

集。需要说明的是，拍摄图片中需要包含一张用于初始化的顶视角图片。在输入到本章方法前，每张拍摄图片需要经过逆伽马矫正（伽马值假定为 2.2）变换到线性空间。相机位置（也就是共位光源的位置）通过经典的相机标定方法完成估计，在拍摄时需要在样本周围放置若干棋盘格图案来辅助相机标定。

图 3.13 展示了本章方法在真实数据上重建得到的高分辨率材质贴图和其在全新视角下的可视化结果，其中贺卡、木头和皮革材质的输入图片数量分别为 20、5 和 2。实验表明本章方法针对多种不同类型的真实材质均可以得到高质量的材质估计。

3.5.4　消融实验

本节将从输入图片动态范围、训练损失函数、初始化准确性和初始化中的顶视角假设以及光照鲁棒性等几个方面对本章方法进行验证分析。

输入图片动态范围　本章方法既可以支持高动态范围输入图片（HDR）也可以支持支持低动态范围输入图片（LDR）。针对真实拍摄的低动态范围输入图片，我们仅需要在逆渲染优化的前向渲染中将渲染图片的像素值裁剪到 $[0, 1]$ 范围内，以保持渲染图片具有输入观测图片相同的数值范围。针对合成数据，我们在将低动态范围输入图片进行裁剪后还需对其进行量化操作（即从浮点数转为整数存储）。

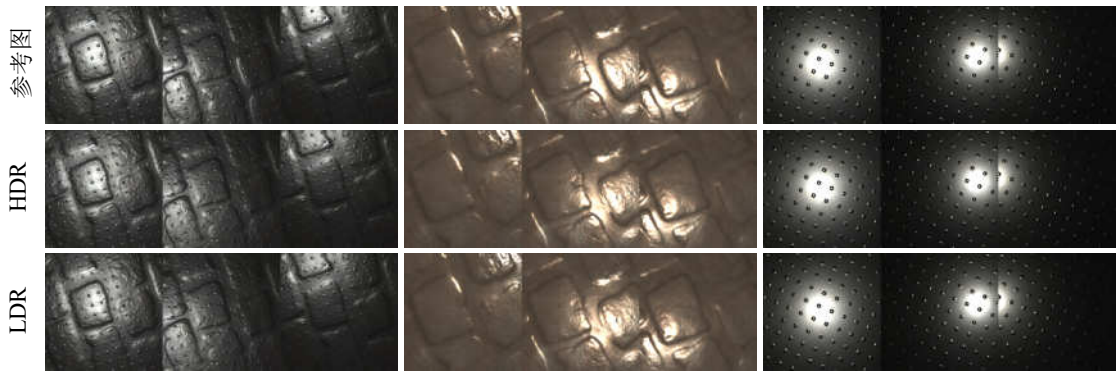


图 3.14　基于 HDR 输入图片和 LDR 输入图片的材质重建结果对比

图 3.14 展示了本章方法在三种选定材质上基于 LDR 和 HDR 输入图片的材质重建结果对比，该图中从左到右的三个样本的输入图片数量分别为 2、5 和 20。整体上，基于 LDR 和 HDR 输入图片的重建结果非常相似。即使 LDR 输入图片中高光区域的过曝像素会被裁剪而导致信息丢失，本章方法仍然可以基于 LDR 输入图片取得准确的材质恢复结果。本章方法基于 LDR 输入图片的重建结果仅在某些严重欠曝区域（即其像素值量化后为 0 的区域）存在一定瑕疵。

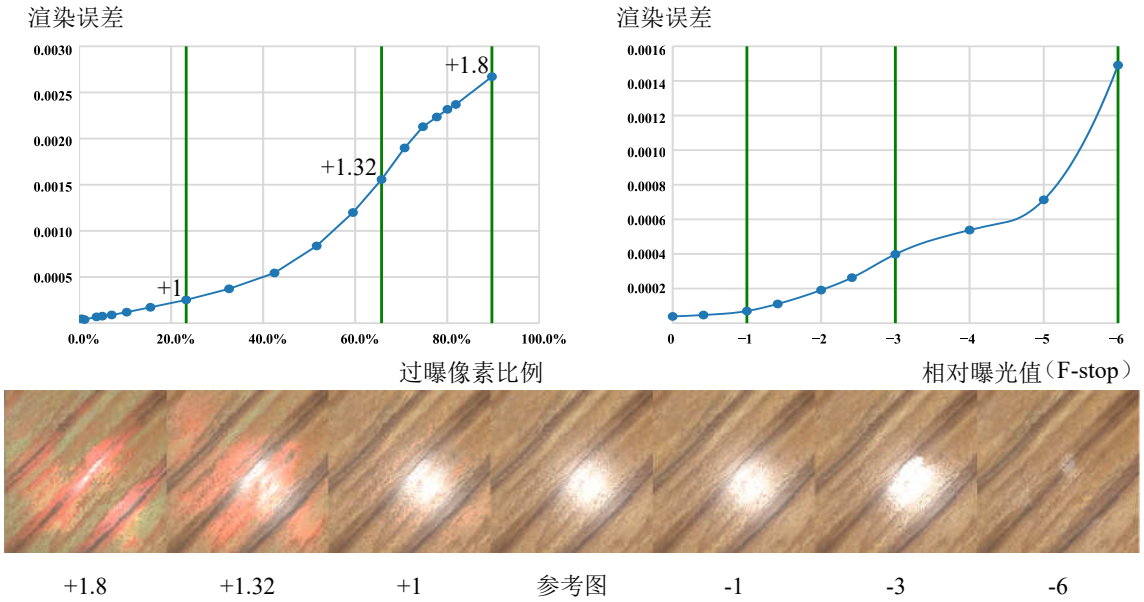


图 3.15　不同曝光度的 LDR 输入图片的材质重建结果

为了进一步研究本章方法在不同曝光度的观测图片下的鲁棒性，我们进行了定量分析。图 3.15 左上角展示了不同过曝比例下本章方法重建结果的渲染误差曲线。结合图 3.15 下方的可视化结果，本章方法在输入图片中过曝像素占比为 20% 时仍然可以取得高质量的重建结果，而随着过曝比例进一步提高，重建结果中会逐渐出现瑕疵。图 3.15 右上角展示了不同欠曝程度下本章方法重建结果的渲染误差曲线。整体上，本章方法基于欠曝输入图片的重建结果中瑕疵较少，只有在极

端欠曝（焦比 (F-stop) 为-6）的情况下会存在高光区域丢失的瑕疵。

训练损失函数　前人提出的基于单张图片的回归方法中，Li 等人[23]采用材质贴图损失函数，Deschaintre 等人[25]采用渲染损失函数，Li 等人[26]综合使用了材质贴图损失函数和渲染损失函数。本章方法中神经网络的训练目标与之前方法不同，本章方法的训练目标是构造适合于优化的隐空间而非直接学习从输入图片到材质贴图的映射。我们通过针对训练损失函数的消融实验来验证不同损失函数对重建材质贴图的影响。图 3.16 对比了本章方法使用三种训练损失函数（分别是材质贴图损失函数 $\mathcal{L}_{map}$，渲染损失函数 $\mathcal{L}_{render}$ 和二者综合使用 $\mathcal{L}_{map}+\frac{1}{9}\mathcal{L}_{render}$）下的材质重建结果。从图中可以看出，仅使用材质贴图损失函数得到的重建材质贴图更加锐利，但其渲染后的可视化结果存在明显瑕疵；仅使用渲染损失函数得到的重建材质贴图的渲染可视化结果更加合理，但其材质贴图存在过度模糊的问题（尤其是高光反射和粗糙度贴图）；二者的综合使用则可以在保持贴图细节和提升渲染结果质量之间取得更好的平衡。

光滑性约束使得自编码器隐空间更加连续并且更加适宜在其中执行优化。从图 3.17 可以看出，加上光滑性约束可以使得重建的材质贴图中保留更多的纹理细节且可以提升其渲染可视化结果的质量。

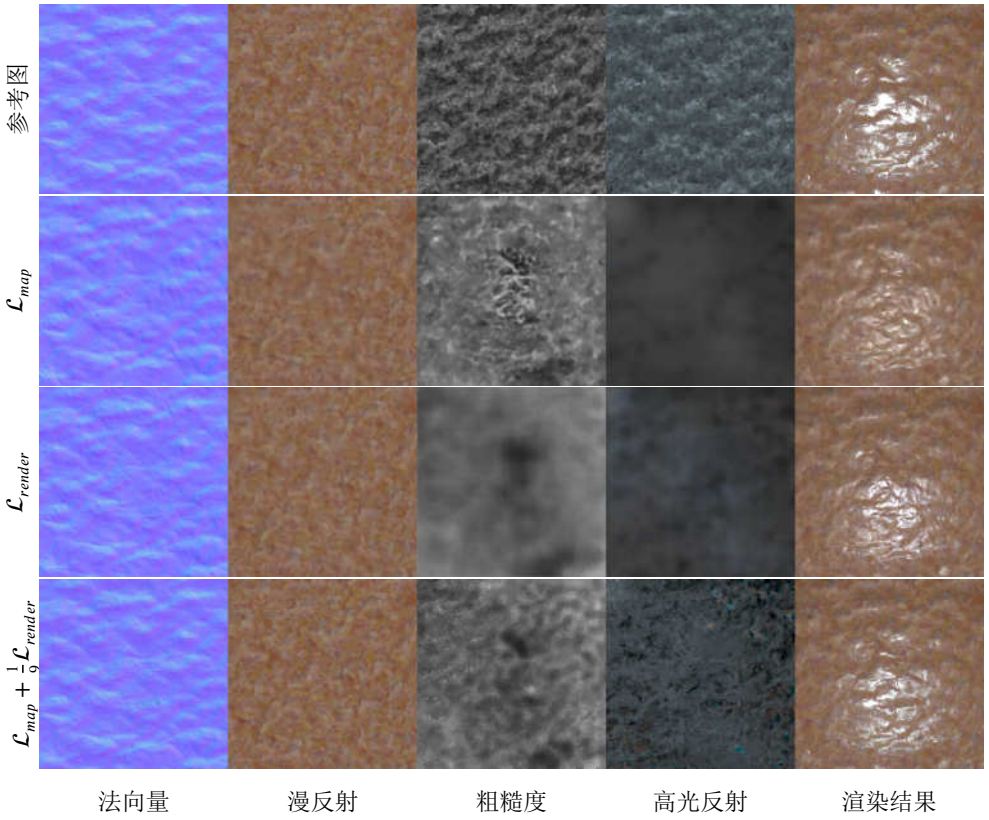


图 3.16　针对训练损失函数的消融实验结果，本实验中输入图片数量为 2。

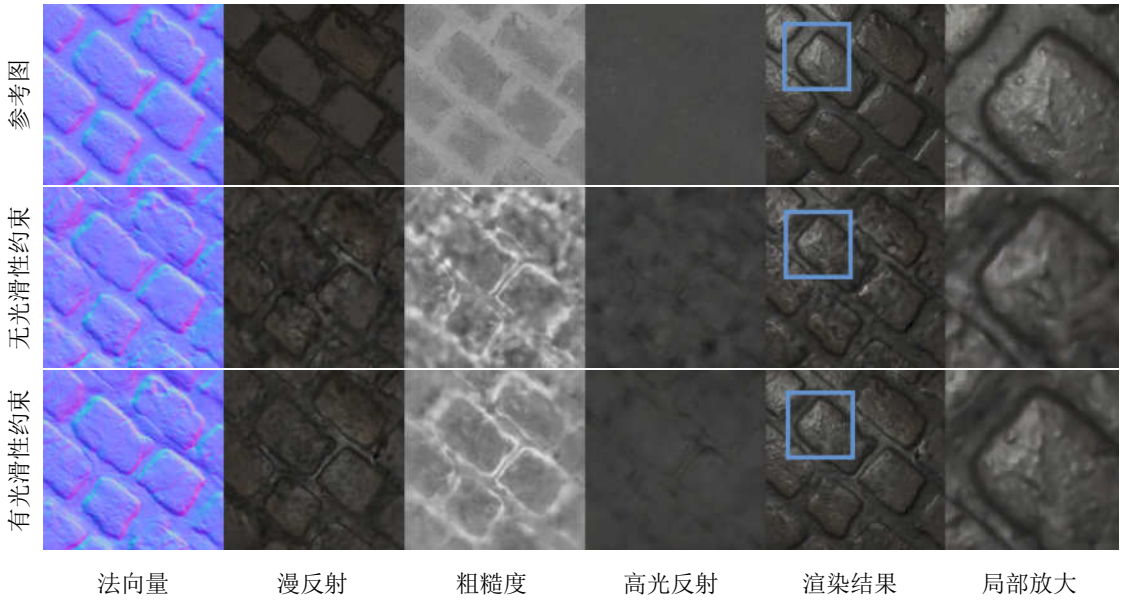


图 3.17　针对光滑性约束的消融实验结果，本实验中输入图片数量为 1。

初始化准确性　在深度逆渲染优化过程中，我们假定作为初始化的材质贴图是合理的，而当初始化结果不满足该假设时，优化会陷入局部最优。图 3.18 展示了一个由于低质量的初始化导致优化无法收敛的例子。图 3.18 第二行展示了若以参考值作为优化的初始点进行深度逆渲染优化，本章方法可以得到准确的材质估计，该结果证明材质贴图本身是可以被自编码器网络的隐空间准确表达（即隐空间中存在该材质贴图对应的隐变量）。因此，图 3.18 第三行所示的优化失败的原因在于初始化质量低，进而导致在逆渲染优化中一直无法跳出局部最优区域，使得最终无法在隐空间中找到被证实存在的准确解。

本章方法中的核心步骤（即深度逆渲染优化）本身是一个非线性优化过程，因此也会面临和其他任何非线性优化过程一样的问题：次优的初始点可能会导致其只能收敛到局部最优。图 3.18 展示的例子代表了表观建模中一种典型的因陷入局部最优而导致优化失败的情况，即初始化材质贴图的粗糙度比真实值大很多而其高光反射比真实值小很多，在该情况下，包含深度逆渲染优化在内的逆渲染优化策略会容易陷入初始点附近的局部最优区域中，导致无法找到全局最优解。

本章方法目前依赖于基于单张图片的材质估计方法[25-26]进行初始化。当用于初始化的单张输入图片没有充分展示出材质贴图中某个分量的特性或者存在二义性的特征时，初始化网络重建出的材质贴图可能是次优的。需要说明的是，本章方法并不局限于某种特定的算法进行初始化，任何可以提供合理 SVBRDF 的方法均可以为本章方法提供初始化。和未来可能出现的可处理多张输入图片的深度材质估计方法相比，本章方法仍然会具有两点明显优势：其一，本章方法采用逆渲染优化策略而非基于训练数据集的平均损失来进行回归，因而可以更加高效地利用

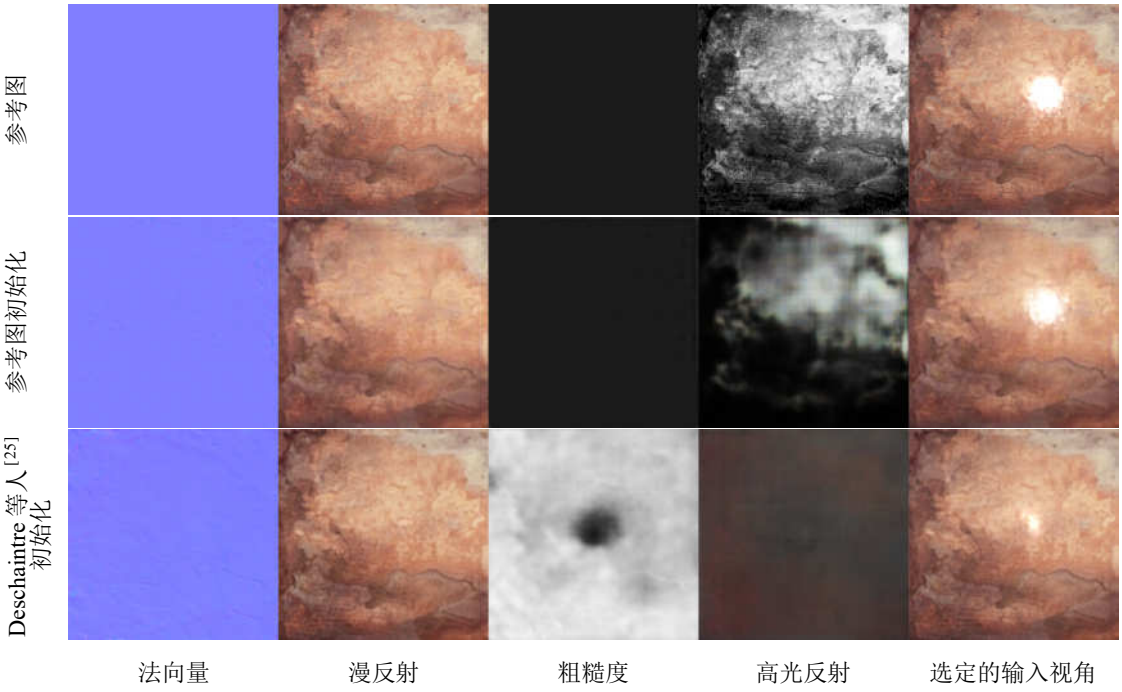


图 3.18　次优初始化导致本章方法无法收敛的示例

输入图片信息；其二，本章方法可以自然地支持高分辨率下的表观建模，无需额外的重新训练过程。

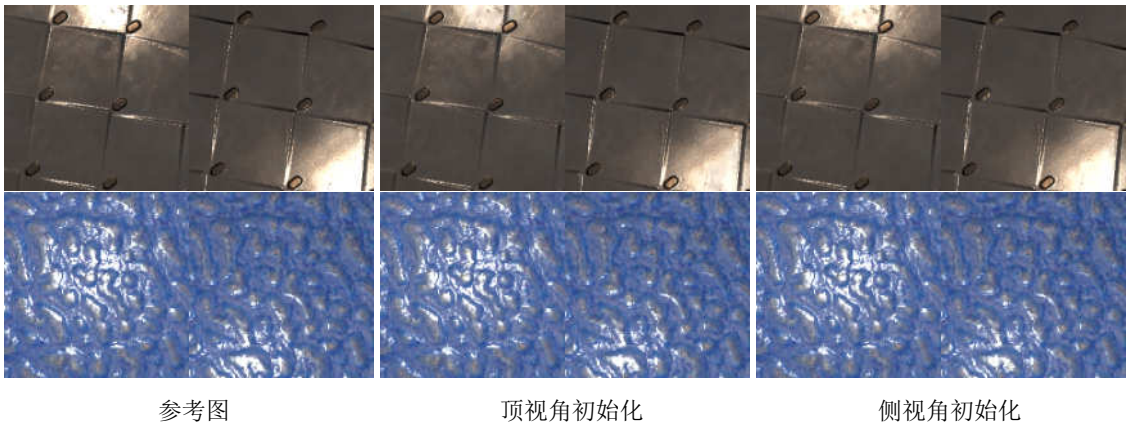


图 3.19　顶视角初始化和侧视角初始化的材质重建结果对比

顶视角图片假设　我们假定输入图片中包含一张顶视角图片的原因在于初始化步骤中前人工作往往依赖于顶视角图片，实际上本章方法对视角方向并无此限制。如果初始化方法（例如 Deschaintre 等人[25]）并不局限于顶视角图片，那么本章方法可以放松顶视角图片假设。图 3.19 展示了本章方法基于侧 45 度视角图片进行初始化的材质重建结果（以侧视角图片进行初始化时，整个输入图片集中均不包含顶视角图片），该实验中第一行所示材质样本的输入图片数量为 5，第二行材质样本的输入图片数量为 20。结果表明本章方法基于顶视角初始化和侧视角初始化所得

到的结果几乎一致。

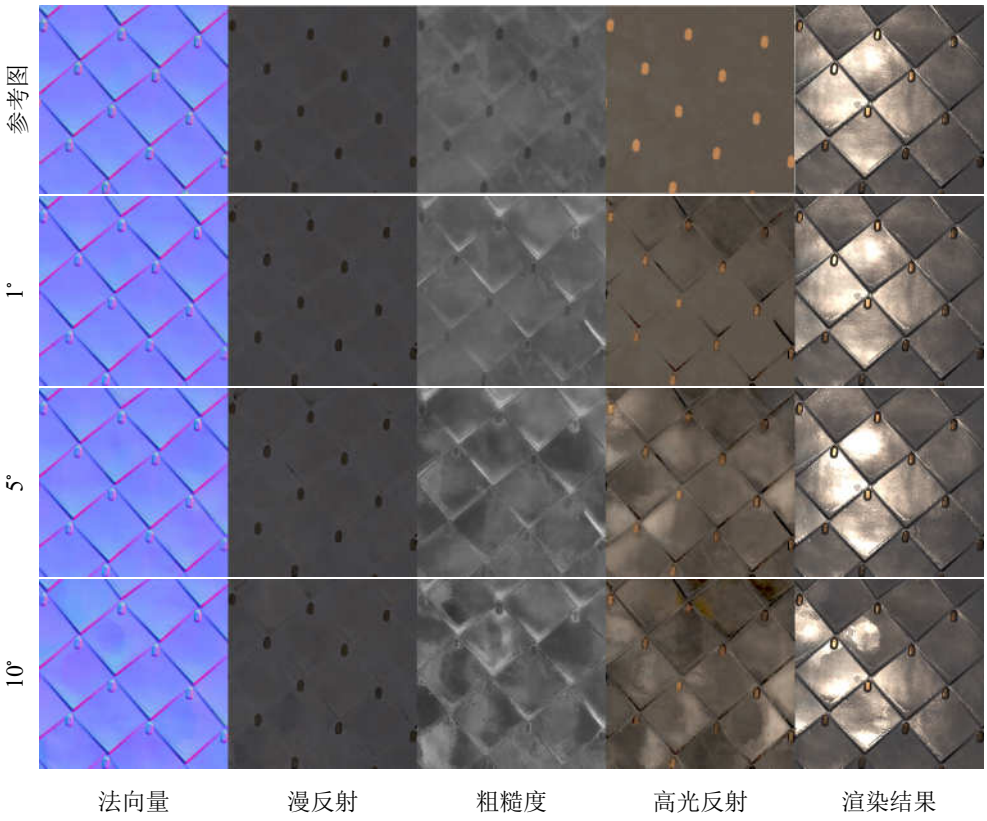


图 3.20　针对光源位置噪声的光源鲁棒性实验结果

光照鲁棒性　本章方法假定在数据采集时场景中仅包含位置已知且和相机共位的点光源，而不存在其他光源。我们通过两个实验来进一步说明在以上假设不完全满足的情况下本章方法的鲁棒性。

在第一个实验中，我们对每张输入图片所对应的光源位置添加不同程度的随机噪声扰动（随机扰动限制在以原光源方向为中心的圆锥形范围内，圆锥角度越大代表噪声越明显）。图 3.20 展示了本章方法在不同程度的光源位置噪声下得到的材质重建结果。结果表明光源位置的误差会导致重建得到的法向量、高光反射以及粗糙度等贴图中出现偏差。表 3.2 前 4 行列举了不同强度的光源位置噪声下的定量分析结果。图 3.20 和表 3.2 均表明本章方法对于 10 度以内的光源位置误差保持鲁棒。

在第二个实验中，我们在点光源基础上增加了不同亮度的环境光照（实验中使用经典的 Uffizi Gallery 环境光照贴图进行模拟），以验证本章方法对于采集环境非绝对暗室时的鲁棒性。图 3.21 展示了环境光照会影响重建结果中高光反射和粗糙度贴图的质量。表 3.2 后四行列举了不同亮度的环境光照下的定量结果，结果表明本章方法对于亮度在点光源亮度的 10% 以内的未知环境光照保持鲁棒。

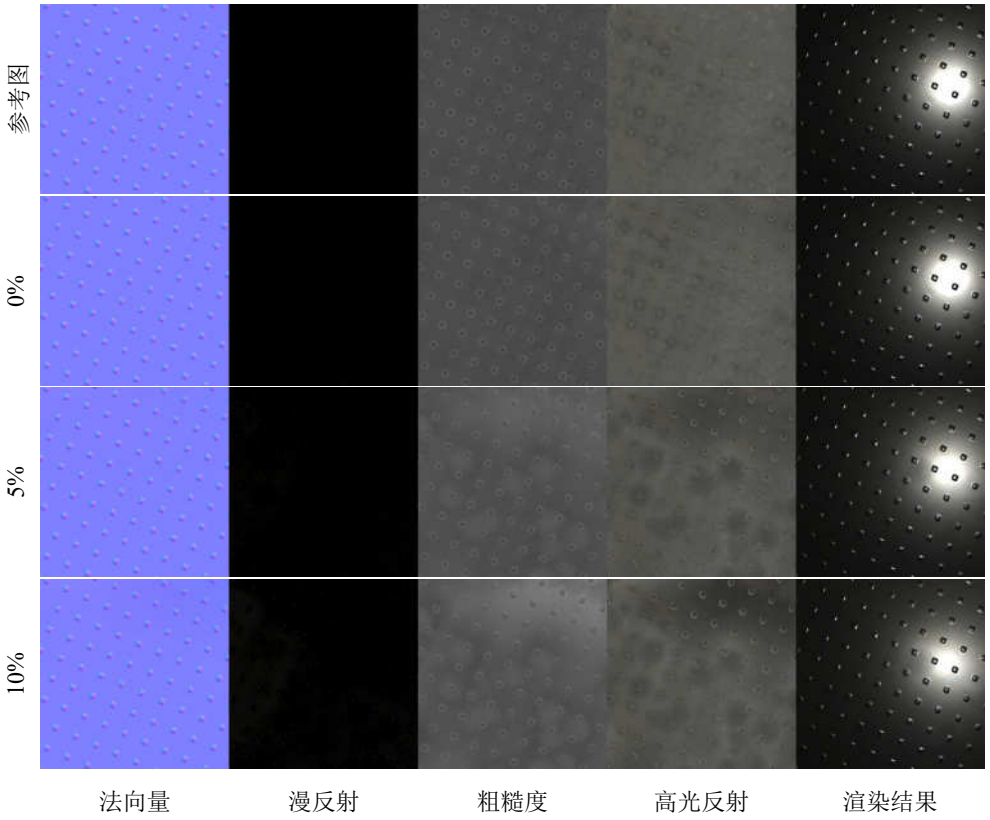


图 3.21　针对环境光照强度的光源鲁棒性实验结果

表 3.2　针对光源鲁棒性的定量结果分析

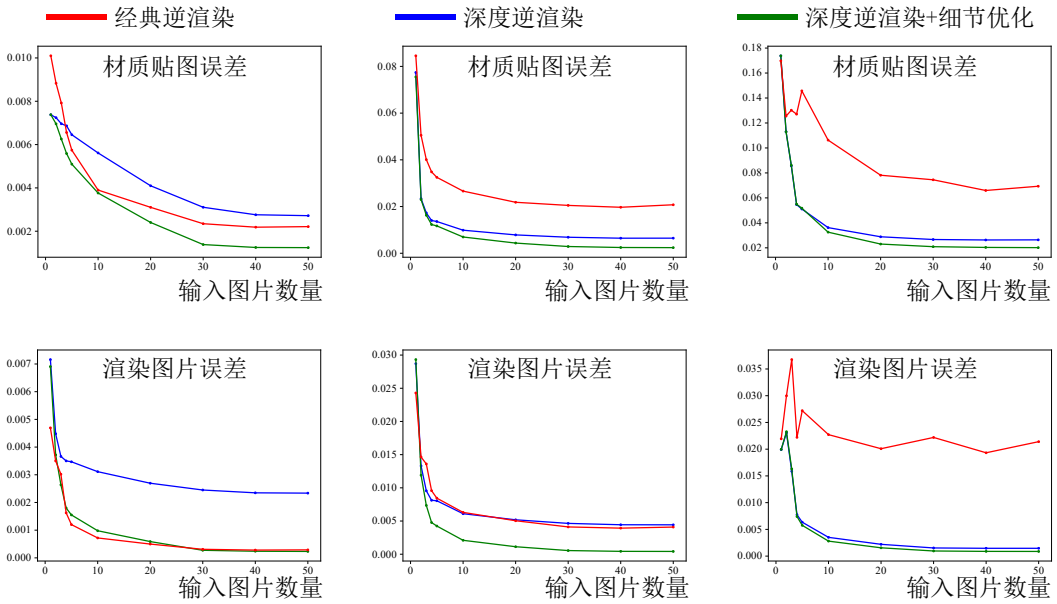
	漫反射	高光反射	粗糙度	法向量	贴图平均	渲染误差
1°	0.000901	0.004756	0.04143	0.000281	0.01184	0.000496
5°	0.001578	0.005981	0.04492	0.000365	0.01321	0.001144
10°	0.002427	0.007834	0.05208	0.000642	0.01575	0.002654
0°/ 0%	0.000850	0.004470	0.04130	0.000273	0.01172	0.000413
1%	0.000809	0.005403	0.04073	0.000280	0.01181	0.000481
5%	0.001663	0.005546	0.04033	0.000306	0.01196	0.000929
10%	0.003577	0.008345	0.04162	0.000374	0.01348	0.002270

3.5.5　讨论和分析

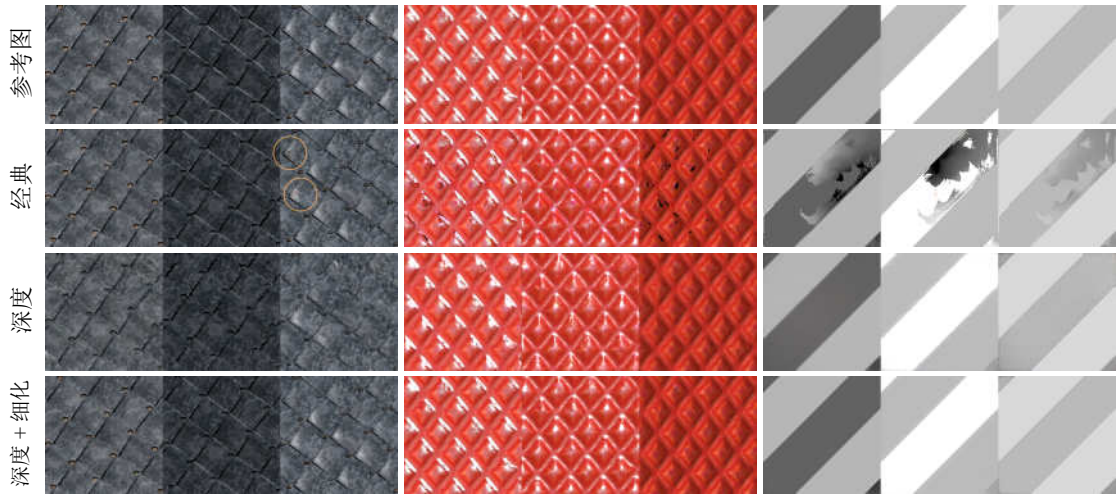
本章方法和经典逆渲染方法对比　本文 3.4.2 节中提到，作为后处理步骤的细节增强策略可以将自编码器网络丢失的细节重新补充回来。细节增强的优化过程和经典逆渲染方法十分类似，区别之处在于细节增强所需的迭代步骤更少且无需额外的人工设计的启发式正则约束。细节增强策略的优化过程无需额外正则约束的原因在于深度逆渲染优化后的材质重建结果通常而言已经足够准确，因此细节增强

优化可以在其局部范围内进行。一个自然的疑问是基于单张图片的深度表观建模方法的材质重建结果是否足够准确，换言之，是否可以直接基于这些方法所重建的 SVBRDF 进行类似细节增强的逆渲染优化。我们通过实验证明，前人提出的基于单张图片的深度表观建模方法[25-26]的结果作为逆渲染优化的起点是不够准确的，也就是说在没有自编码器网络的隐空间所提供的数据约束的情况下，直接在 SVBRDF 空间进行逆渲染优化会导致无法收敛。

图 3.22 比较了三种优化策略在三种选定的材质样本上的材质重建结果，三种优化策略包括：经典逆渲染、深度逆渲染以及深度逆渲染 + 细节增强。实验中，三



(a) 三种优化策略随输入图片数量变化的误差曲线



(b) 材质重建结果的可视化对比

图 3.22　三种优化策略在不同输入图片数量下的材质重建结果对比

种优化策略所对应的优化初始点和迭代总步骤均保持一致，且在该迭代总步数下三种优化策略均已收敛（即迭代的相邻步骤间的损失函数相对变化小于 1%）。

整体上，仅使用深度逆渲染策略可以预测出可信的结果但存在一定的细节丢失；经典逆渲染方法在某些情况下可以收敛到准确的结果，但在其他情况下其收敛后的结果存在明显的视觉瑕疵；本章方法（即深度逆渲染策略 + 细节增强策略）的结果既保留了丰富的纹理和材质细节，同时也不存在视觉瑕疵。此外，即使是对于本章方法和经典逆渲染方法均可以收敛的材质样本上，本章方法相较于经典逆渲染方法而言仍然存在优势，主要体现在本章方法需要更少的输入图片即可收敛。

图 3.22 (a) 中列举了三种优化策略在不同数量的输入图片下的 L_2 材质贴图误差曲线和 L_2 渲染误差曲线。我们使用两种误差度量的原因在于从材质贴图到渲染图片的渲染过程是非线性映射，因此材质贴图误差和渲染误差并不总是保持一致。例如，对于高光反射明显的材质而言，法线贴图中的微小误差（意味着较小的材质贴图误差）可能会导致渲染结果的视觉差异很大（意味着较大的渲染误差）；反过来，对于高光反射较小的材质而言，粗糙度贴图对于渲染结果的贡献并不明显，也就是说粗糙度贴图中的较大误差可能对应于较小的渲染误差。

从结果中我们发现，经典逆渲染方法通常倾向于生成带有明显噪声的材质贴图，因而其材质贴图误差一般较大（参见图 3.22 (a) 中红色曲线）。尽管在某些情况下，经典逆渲染方法的数值误差很小，但其结果在视觉上仍然存在明显的瑕疵。例如在图 3.22 (b) 中左侧所示的例子，三种优化策略中经典逆渲染方法的渲染误差最小，但从视觉上来看，经典逆渲染方法重建的材质贴图的可视化结果中存在明显的瑕疵（参见图中橙色圆圈区域内），本章方法的重建结果中则没有类似的瑕疵。仅使用深度逆渲染策略可以得到高质量且无瑕疵的重建结果，但存在细节丢失的问题。根据材质贴图本身所包含的高频信号的多少，深度逆渲染策略所造成的细节丢失对最终误差所造成的影响也不同。在多数情况下，深度逆渲染策略的结果（参见图 3.22 (a) 中蓝色曲线）比经典逆渲染方法更好。在深度逆渲染策略的基础上加上细节增强后可以有效解决细节丢失的问题，其结果（参见图 3.22 (a) 中绿色曲线）是三种优化策略中最优的。

此外，我们在之前介绍的包含 42 个 SVBRDF 数据的测试数据集上对三种优化策略进行了定量对比。表 3.1 展示了三种优化策略的数值误差。结果表明，本章方法（深度逆渲染策略 + 细节增强策略）的渲染误差在不同输入图片数量下均是最优的，而其材质贴图误差在大多数情况下也是最优的，只有个别情况下材质贴图误差的最优结果来自于其他两种策略。该结论也与图 3.22 的观察结果一致，即本章方法重建的材质贴图在数值上有时并非是最优的，但其视觉质量整体上是最

好的。

与后续工作的关联　本章提出的在深度学习构造的材质数据隐空间中进行逆渲染优化的表观建模思路被后续工作所沿用。Guo 等人[153]提出一种基于 StyleGAN2 的材质建模方法，其和本章工作的主要区别在于将自编码器神经网络替换为对抗生成网络，其核心思路是在对抗生成网络的隐空间中进行逆渲染优化。

此外，正如之前所介绍，本章结果默认采用 Deschaintre 等人[25]的方法进行初始化，但本章方法的初始化过程并不局限于某种特定算法。后续的基于单张输入图片的深度回归方法[154]和基于多张输入图片的深度回归方法[27]均可以为本章方法提供初始化。

3.6　本章小结

本章提出了一种可处理任意数量输入图片的统一材质建模框架，并支持高分辨率材质建模。本章方法估计的材质质量随输入图片数量的增加而提高：当输入图片数量较少时本章方法可以得到合理的预测结果，而当输入图片数量足够多时可以收敛到准确的结果。本框架直接在深度自编码器网络构造的隐空间中执行逆渲染优化，无需任何手工设计的启发式正则约束。本章还提出了多种增强策略来约束深度学习构造的隐空间以使其更加适合于优化。通过合成数据和真实数据上的一系列实验证明了本章方法可以从任意数量输入图片中估计高分辨率的材质贴图；此外，实验还表明本章方法即使在单张输入图片的情况下相较于之前工作仍然取得质量提升。

未来工作方面，我们将进一步提升深度逆渲染优化过程的初始化策略。目前本章方法依赖于已有工作进行初始化，已有方法训练的目标是提供视觉上合理的材质预测而并非专为逆渲染优化的初始化设计，因此探索初始化网络和逆渲染优化协同训练是未来待研究的方向。

第 4 章　基于深度场景表达的重光照

本文第 3 章中介绍了一种基于逆渲染和数据驱动的表观建模方法，该方法可以基于轻量化的采集条件拍摄的任意数量输入图片实现高质量的材质重建，从而可以为计算机图形学中的诸多应用提供大量高质量的材质数据。然而，面对真实世界复杂场景的数字化任务（即数字化重现真实场景在新视角或新光照条件下的高逼真结果），现有的几何重建和材质建模方法仍然存在困难：一方面，传统的几何表达（例如常见的三角形网格、体素及点云等）和材质模型（例如 BRDF 等）表达力有限，难以准确表达真实世界复杂物体（例如具有复杂散射性质的半透明物体、具有精细几何结构的毛发等）。另一方面，真实世界的复杂场景通常包含复杂的光传输效果，在采集和建模过程中往往难以将几何、材质、光源、全局光照等完全解耦。因此，针对真实世界复杂场景的数字化问题，我们迫切地需要研究一种全新的场景表达和相应的采集、渲染流程以满足实际应用的需求。

(a) 毛绒猫　(b) 貔貅　(c) 糖果碗　(d) 小红帽　(e) 金鱼　(f) 组合场景

图 4.1　基于深度场景表达的重光照方法的渲染结果

真实世界复杂场景的数字化重渲染是计算机图形学和视觉领域中极具挑战但同时有着广泛应用场景的研究方向。相关工作可以大体分为两类：基于建模的方法和基于图像的方法。其中，基于建模的方法[46-47,155]的核心思路是首先对场景中的各个组分（例如几何、材质、光源等）进行显式建模，而在得到场景的数字化表达后，新视角和新光源下的重渲染可以通过经典的渲染管线来得到。正如之前所述，由于实际采集条件限制以及模型表达力不足等因素，场景中的各个组分无法被精确重建，而重建的误差会影响最终重渲染的质量。例如，在几何重建时，拍摄过程很难保证可以无遮挡地采集到物体的每个部分，而几何重建的误差会直接影响重渲染时的光传输模拟过程。另一类基于图像的方法[72-73,103]则无需显式地进行场景重建，而是直接利用采集图片中的信息合成目标视角或光源下的结果，因

此其重渲染结果往往可以更好地保留原场景中的光传输细节。然而，现有的基于图像的方法依赖于密集的采样和专业的采集设备，限制了其使用范围。此外，现有的很多基于图像的方法仅支持动态视角重渲染而不支持重光照渲染。

4.1　本章引言

针对三维世界数字化问题，本章提出一种基于深度场景表达的重光照方法，以轻量化手持设备所拍摄的无结构化图片作为输入，支持 360 度自由视点下的重光照渲染（结果示例参见图 4.1 ）。本章方法的核心思路是将定义在粗糙几何上的可学习神经纹理和辐射亮度信息 (Radiance cues) 共同输入到场景相关的神经渲染网络来完成重渲染。一方面，由于本章方法属于基于图像的方法，因此本章方法无需显式地对场景中的光传输过程进行精细建模，而是利用基于图像的神经绘制管线来得到重渲染结果；另一方面，本章方法也从基于建模的方法中借鉴了三维几何信息，并将其融入到本章所提出的深度场景表达和神经渲染管线当中。需要说明的是，本章方法并不像传统基于建模的方法一样依赖于高精度的重建几何，而是仅将粗糙的场景几何作为提高多视角间一致性的约束，由于本章方法的最终渲染结果仍然是通过神经渲染网络预测得出，因此可以避免几何不准确所带来的渲染瑕疵。而本章方法与传统的基于图像的方法相比，几何信息的融入又可以大大降低其采集复杂性。

本章方法将经典的延迟光照渲染算法[156]和神经纹理[55]表达进行了结合，本章方法也可以叫做延迟神经光照方法。本章方法的核心步骤包括：第一步，将可学习的神经纹理通过粗糙几何的参数化 UV 展开投影到粗糙几何上；第二步则类似于经典延迟光照算法中的光照步骤，该步主要负责计算作为场景光源表达的辐射亮度信息，即粗糙几何在目标视角、光照和一系列预定义的同质材质下的渲染结果；最后一步，将辐射亮度信息和投影神经纹理结合起来，共同输入到场景相关的可学习神经渲染网络中来输出最终的重光照结果。粗糙几何和辐射亮度信息共同构成全新的深度场景表达，可以帮助神经渲染网络更好地完成全新视角生成和重光照渲染，其中，辐射亮度信息以一种图片友好的方式自然地编码了输入光照的信息，在扩展神经渲染网络以适应全新视角或光照下的表观变化方面发挥关键作用。

本章方法基于目标场景的无结构化图片数据集实现端到端的训练，无需要任何预训练步骤。当粗糙几何和真实几何相差过大或场景中的光传输过程过于复杂时，我们需要扩展神经纹理的通道数以及使用更大的神经渲染网络来提升神经网络表达能力，以支持整个场景 360 度自由视点的重光照渲染。然而在目前的 GPU

硬件下，增大神经网络容量会面临内存瓶颈，为此本章提出使用一种高效的视角划分策略来解决扩大神经网络容量时面临的显存限制。由于本章方法可以基于消费级手持设备（例如消费级相机或者手机等）完成数据的拍摄和采集，因此有更广泛的受众和使用场景。我们在多个包含复杂材质、多种光传输效果和复杂几何的合成场景和真实场景中测试并验证了我们方法的有效性。此外，本章还提出了光照增强策略，通过利用光传输过程的线性性质有效扩展了本章方法所支持的光源类型，使得本章方法不仅支持训练数据中的点光源，也可以支持环境光照等更为复杂的光源。

总之，本章方法的主要贡献包括：

- 针对带有复杂材质和光传输效果的真实场景，可基于轻量化手持设备拍摄的无结构化图片，实现 360 度自由视点重光照渲染；
- 提出基于神经纹理和辐射亮度信息的深度场景表达，使得整个流程不依赖于精确的场景建模；
- 提出全新的端到端的深度渲染管线，支持多种不同类型的光源和动态视角；
- 提出一种光照增强策略，可以扩展神经渲染网络适用的光源类型。

4.2　方法概览

本章方法的输入 $\{\mathbf{C}_k, \mathbf{p}_k, \mathbf{l}_k, \mathbf{M}_k\}_{k=1}^{N}$ 包括 N 张给定场景的输入图片 $\mathbf{C}_k$、每张输入图片对应的前景遮罩 $\mathbf{M}_k$、相机内参和外参 $\mathbf{p}_k$ 和输入光照 $\mathbf{l}_k$。本章方法并不假定相机视角和光源方向按照某种结构化方式分布，用户可以手持拍摄设备在自由移动相机过程中完成数据采集。此外，本章方法的输入还包括基于输入图片重建得到的场景粗糙几何 $\mathbf{G}$ 以及 M 个预定义的基材质 $\{b_i\}_{i=1}^{M}$。

和 Thies 等人[55]方法类似，我们通过一个 S-通道的可学习神经纹理 $\{\mathbf{T}_t\}_{t=1}^{S}$ 来编码视角相关的表观信息。神经纹理定义在粗糙几何 G 的 UV 纹理空间上。神经纹理和计算机图形学中广泛使用的纹理贴图类似，其区别在于每个纹素中不再存储法向量、表观、凹凸等具有明确物理含义的内容，而是在每个神经纹素中存储长度为 S 的可学习特征向量，这些特征向量会用于指导神经渲染网络如何计算最终的像素颜色。对于给定的某个视角 $\mathbf{p}$，我们首先计算定义在粗糙几何上的神经纹理在该视角下的投影神经纹理 $\mathbf{T}^p = \bar{\mathcal{P}}(\mathbf{T}, \mathbf{p}; \mathbf{G})$。和 Thies 等人[55]方法不同的是，我们并不直接将投影神经纹理输入到神经渲染网络 $\mathcal{R}$ 中，而是将投影神经纹理 $\mathbf{T}_t^P$ 和辐射亮度信息 $\mathbf{B}_i$ 的逐像素乘积输入到渲染神经网络 $\mathcal{R}(\mathbf{T}_t^p \odot \mathbf{B}_i)$ 当中，最后通过神经网络完成重渲染结果生成。辐射亮度信息是本章提出的一种全新光照表达形式，其定义为基材质在目标光照下的可视化渲染结果：$\mathbf{B}_i = \bar{\mathcal{R}}(b_i, \mathbf{p}_i, \mathbf{l}_i; \mathbf{G})$。

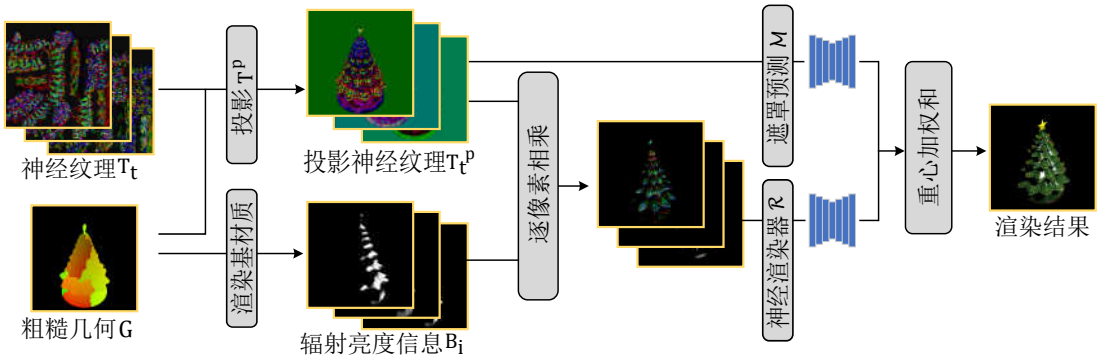


图 4.2　基于深度场景表达的重光照方法的流程示意图

此外，我们将投影神经纹理输入到遮罩预测网络 $\mathcal{M}$ 中完成前景遮罩 $\mathcal{M}(\mathbf{T}^p)$ 的预测，本章方法的最终渲染结果是神经渲染网络的输出结果和遮罩预测网络的输出结果的乘积。图 4.2 概述了本章提出的基于深度场景表达的重光照方法的流程。

神经纹理可以指导神经渲染网络如何计算输出图片中像素的颜色，反过来，神经渲染网络定义了神经纹理中特征向量的具体含义，对于每个场景而言，神经纹理和神经渲染网络需要协同进行训练，训练过程可以形式化地定义为：

$$\mathbf{T}^*, \mathcal{R}^*, \mathcal{M}^* = \underset{\mathbf{T},\mathcal{R},\mathcal{M}}{\operatorname{argmax}} \sum_{i}^{N} \mathcal{L}(\mathbf{C}_i, \mathbf{p}_i, \mathbf{l}_i, \mathbf{M}_i | \mathbf{T}, \mathcal{R}, \mathcal{M}), \tag{4.1}$$

其中，$\mathcal{L}(\cdot,\cdot)$ 是训练损失函数。关于神经网络训练相关的更多细节将在 4.5.3 节中介绍。

4.3　深度场景表达和神经渲染管线

本章方法的深度场景表达包括神经纹理和辐射亮度信息，将二者结合后输入到神经渲染网络中完成重光照渲染。本章将展开介绍其中各个模块的具体内容和实现细节。

4.3.1　神经纹理

本章方法中的神经纹理和 Thies 等人[55]方法中的神经纹理存在三点主要区别：其一，本章方法使用神经纹理的方式与 Thies 等人[55]方法不同，本章方法将神经纹理和辐射光照信息结合后输入到神经渲染网络，而 Thies 等人[55]是直接将神经纹理输入到神经渲染网络中。神经纹理的使用方式不同使得其编码的表观信息也不同。其二，本章方法与 Thies 等人[55]方法采用不同的神经纹理采样方法。Thies 等人[55]直接将各个层级内的 Mipmap 采样结果进行简单的平均作为最终多层次神经纹理的采样结果，而本章方法则使用传统计算机图形学中多层次纹理贴图的标

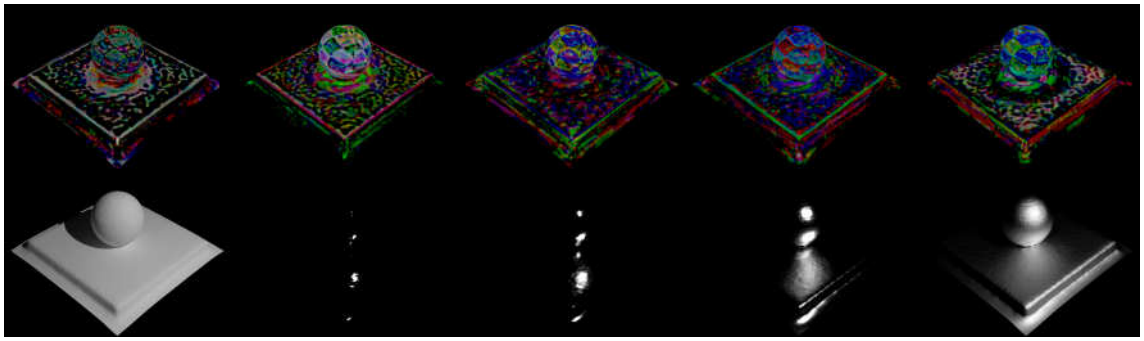


图 4.3　球体场景中投影神经纹理和辐射亮度信息的可视化

准采样流程对多层次神经纹理进行采样。此外，本章方法中的神经纹理只有原始分辨率层次是可学习变量，其他层次均是由上一个层次经过平均池化操作降采样而得到。其三，本章方法中的神经纹理的每个神经纹素中存储了更多的特征通道（30 个），并且要求其中从 $i \times 6$ 到 $(i+1) \times 6$ 的特征通道对应第 i 个基材质。

4.3.2　辐射亮度信息

经典延迟光照算法中光照信息通过漫反射光照图和高光反射光照图进行编码。在本章提出的延迟神经光照方法中，我们将经典延迟光照算法中的两个光照图泛化到更多的“光照图”，即利用 M 个辐射亮度信息图片（每个图片对应一个预定义的同质基材质）来将输入光照信息传递给神经渲染网络。利用一系列基材质来表达物体表观信息的思路和 Ren 等人[157]方法有相似之处，但其区别在于：Ren 等人[157]方法中将物体表观简单地表示为基材质表观的线性加权和，而本章方法首先将基材质表观和神经纹理相乘，然后通过神经渲染渲染网络将二者乘积非线性地映射为最终的物体表观。

在实现中，我们采用 $M = 5$ 个基材质 $\{b_i\}_{i=1}^{M}$，其中一个基材质使用 Lambertian BRDF 建模，而其他四个则采用粗糙度为 $\{0.02, 0.05, 0.13, 0.34\}$ 的 Cook-Torrance BRDF[158]建模。在辐射亮度信息 $\mathbf{B}_i = \bar{\mathcal{R}}(b_i, \mathbf{p}_i, \mathbf{l}_i; \mathbf{G})$ 的渲染过程中，本章方法通过基于 GPU 的路径跟踪渲染器来生成粗糙几何 $\mathbf{G}$ 在目标光照和给定基材质 b_i 下的带有全局光照的渲染图片。由于本章方法中的辐射亮度信息并不包含可学习参数，因此该基于 GPU 的路径跟踪渲染器并不需要是可微分的，大大提升了训练过程的效率并简化了方法的实现难度。需要说明的是，辐射亮度信息是一种适用于任意类型输入光照的光源表达，再结合 4.4 节介绍的光源增强策略，本章方法可以支持包含方向光源、环境光照在内的多种类型的光源；此外，虽然辐射亮度信息对应的基材质均为单色材质，但是为支持带有颜色的输入光照，辐射亮度信息使用三通道的 RGB 图片来表达。图 4.3 展示了合成的球体场景中投影神经纹理（以 RGB 图片编码）和对应的辐射亮度信息的可视化结果。

4.3.3　神经渲染网络

神经渲染网络的输入是投影神经纹理和辐射亮度信息的逐像素乘积 $\mathbf{T}_t^p \odot \mathbf{B}_i$。每个辐射亮度信息图片对应于神经纹理中的 6 个通道，由于辐射亮度信息本身是 RGB 三通道图片，因此需要辐射亮度信息拷贝后组成 6 个通道，再和投影神经纹理相乘。

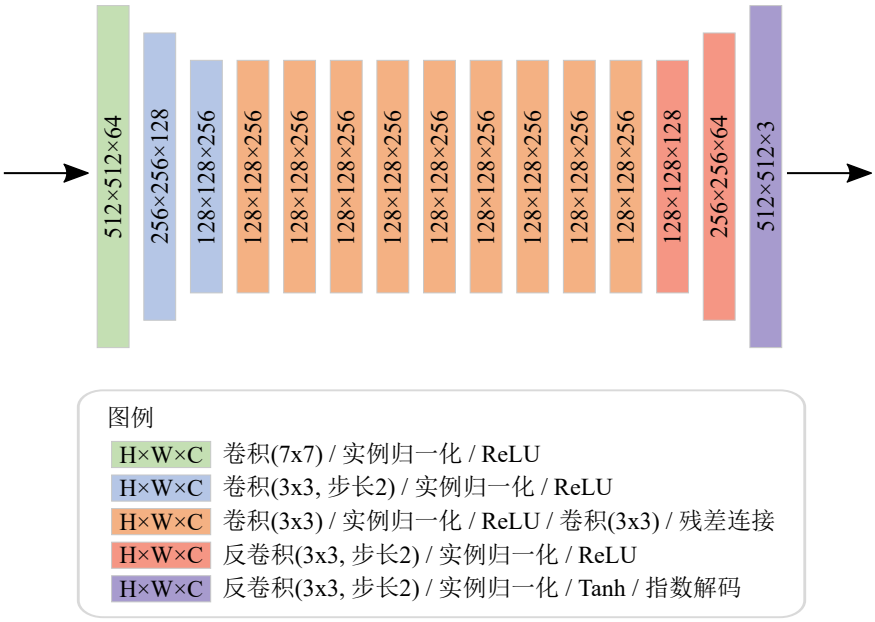


图 4.4　神经渲染网络架构

神经渲染网络采用视觉领域中的一种经典生成器网络架构[159-160]，该网络架构的详细细节参见图 4.4 。除最后一层采用 tanh 作为激活函数外，其他层均采用 ReLU 作为激活函数。实验表明该带有残差连接的网络架构相较于 Thies 等人[55]使用的 U-net 网络架构可以更好地表达场景的表观信息。

为了支持更大的动态范围，我们在投影神经纹理和辐射亮度信息的乘积后添加对数编码：

$$\frac{\log(x+\epsilon)}{\log(s)} + o, \tag{4.2}$$

其中，$\epsilon = \frac{1}{e}$ (e 是自然底数)，s 和 o 是进行曝光亮度归一化的缩放因子和偏置因子（其取值取决于输入图片 x 的动态范围）。在对数编码后，我们将其取值范围归一化到 $[-1,+1]$ 区间内再输入到神经渲染网络中。神经网络的原始输出也定义在对数空间，为了在对数空间中完成损失函数的计算，在训练过程中我们需要将以上对数编码和归一化过程同样应用到参考图片上。为了得到线性空间的渲染图片，我们需要对神经网络原始输出结果应用指数解码来逆变换式（4.2）中的对数编码。

神经渲染网络在本方法中同时发挥两个作用：其一是将投影神经纹理和辐射亮度信息的乘积转换为最终的像素颜色，即完成渲染任务；其二是负责修正粗糙几何带来的误差。

4.3.4　空间划分方案

当粗糙几何和真实几何相差过大或者场景中包含非常复杂的光传输效果时，神经渲染网络需要负责修复更大的误差，因而依赖于更大容量的神经网络。然而扩大网络规模会面临超出目前显卡显存限制的实际困难，因此本章提出一种基于视角的空间划分方案来解决该问题。相较于在单个神经纹理和神经渲染网络的基础上扩增神经纹理通道数的简单策略，我们将大量神经纹理通道进行划分，并为每个划分内的神经纹理通道单独训练一个对应的神经渲染网络。由于划分之间是相互独立的，因此不同划分内的神经纹理和神经渲染网络可以独立地进行并行训练。

实现中，我们使用 13 个独立的划分，其中每个划分内有 30 个神经纹理通道和一个对应的神经渲染网络。每个划分内的训练数据是所有训练数据的一个子集，划分内的训练数据所对应的视角方向均落在该划分范围内，而其光源方向则可能是任意方向。具体而言，我们首先将 13 个顶点均匀排布在视角方向的半球面上，其中每个顶点大约覆盖 60 度范围，其次基于 13 个顶点进行三角化操作来形成若干彼此相邻的三角形区域。每个顶点定义了一个划分，其覆盖范围包括该顶点的所有相邻三角形区域。在运行时，某个给定的视角方向一定属于某个三角形区域内，也就是硕，同时隶属于三角形三个顶点所对应的三个划分。本章方法会分别在三个划分内预测该给定视角下的重渲染图片，并通过三角形重心加权的方法生成最终的渲染图片。

4.3.5　神经遮罩

为了支持将重渲染结果融入到新的背景当中，我们需要估计出每个视角下的前景遮罩。我们提出使用神经网络来预测前景遮罩，该网络的输入是投影神经纹理。神经遮罩网络的网络架构和神经渲染网络类似，区别在于其中间层的通道数减半且输出是经过 sigmoid 激活函数的单通道遮罩。

4.4　光照增强机制

本章提出的辐射亮度信息表达的一大优势在于其可以兼容任意类型的光照，我们仅需要在目标光照下渲染对应的辐射亮度信息并将其输入到本章方法中即可

(a) 参考图　　(b) 有光照增强　　(c) 无光照增强

图 4.5　有无光照增强策略的重光照结果对比

得到目标光照下的渲染结果。然而，由于神经渲染网络在仅包含白色点光源的训练数据上进行训练，其对于带有面积或者带有颜色的光源没有任何概念，因此直接将环境光照下的辐射亮度信息输入到神经渲染网络中并能得到合理的渲染结果（参见图 4.5 (c)）。

为了解决该光源泛化性问题，本章提出一种光照增强机制来进一步优化神经渲染网络。光照增强机制的核心思路是利用经典的基于图像的重光照方法[103]来合成一个环境光照下的训练数据集，并在该增强数据集上进一步优化神经渲染网络。基于图像的重光照方法所需的方向光源下的基图片则可以通过本章方法得到：首先渲染方向光源下的辐射亮度信息，然后将其输入到在点光源下训练好的神经渲染网络中生成方向光源下的基图片。

然而，上述增强策略依赖于大量视角下的大量基图片，例如一个典型的场景需要 1500 个视角方向，如果环境光照贴图使用 $32 \times 32 \times 6$ 分辨率的立方体贴图存储，那么每个视角下需要 6144 个基图片，所有视角下共需要超过九百万个基图片。为此，本章提出使用重要性采样策略对环境光照进行采样，每个环境光照贴图使用 100 个光源采样（远少于原 6144 个）来近似。在我们的实现中，每个场景共有 7500 个增强训练数据：视角数量为 1500 个，每个视角下有 5 张在不同环境光照下的训练数据。每个视角下的 5 个环境光照贴图是从包含 90 个环境光照贴图的数据集中随机采样得到的。需要补充说明的是，针对高光反射非常强的场景，光源重要性采样得到的 100 个光源采样并不足以很好地近似原环境光照，因此神经渲染网络无法从合成的增强数据中学习到如何正确处理低频的环境光照。在这些场景（例如合成场景中的球体场景）中，我们转而使用所有的光源方向来生成增强数据。构造好增强数据后，我们将原始采集数据（点光源）和增强数据（环境光照）混合后一起用于最终的优化训练中。实验表明，上述光照增强策略可以很好地扩展本章方法，使之可以处理全新环境光照（不在环境光照数据集中）下的重光照渲染。

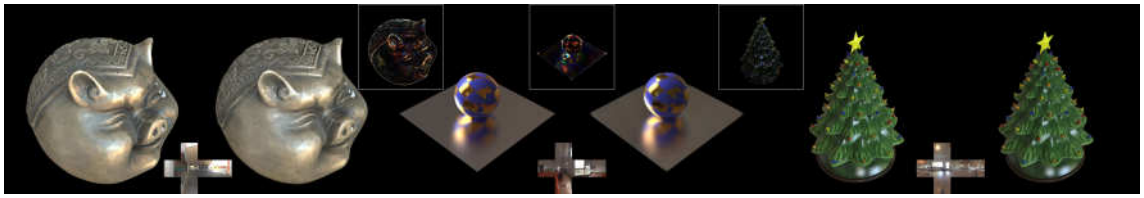


图 4.6　合成场景中环境光照下的重光照结果

图 4.5 对比了应用光照增强策略前后的可视化结果并验证了光照增强策略的有效性。图 4.6 展示了本章方法在三个合成场景中环境光照下的重光照结果，图中结果所对应的视角和环境光照均不包含于训练数据当中，图中白框区域展示了每个场景对应的误差图 (×5)。此外，图 4.1 (d-f) 展示了三个真实场景在环境光照下的重光照结果。在时间开销方面，每个划分需要约 5 个小时进行数据准备（基辐射亮度渲染和基图片生成）、1.5 个小时完成基于图像的重光照渲染以及额外 20 个小时来进行优化训练（以前一阶段训练结果为初始点）。

4.5　数据采集和训练细节

本节首先介绍真实场景的采集和数据预处理过程，然后介绍合成场景的模拟生成过程，最后介绍本章方法的训练细节。

4.5.1　真实场景采集和处理

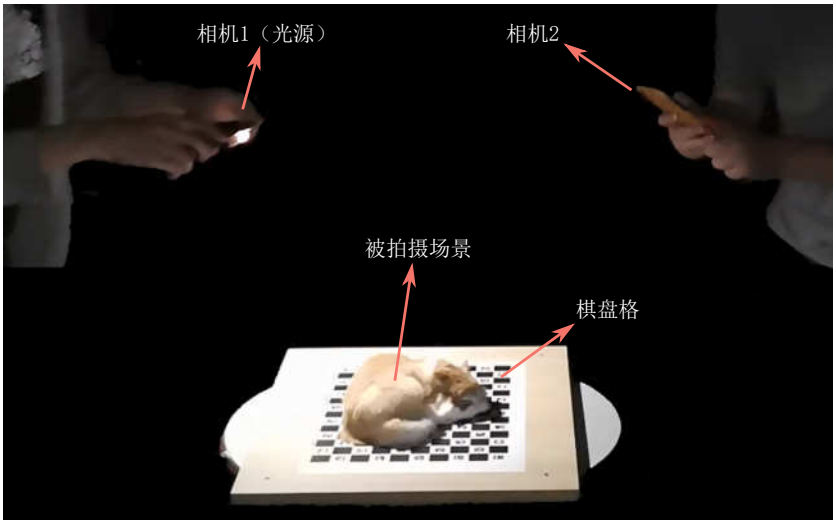


图 4.7　真实场景采集过程示意图

数据采集　本章提出了一种基于两个手持拍摄设备（手机或者 DSLR 相机）的轻量化采集方法（参见图 4.7 ）。在数据采集时，两个相机均开启视频拍摄模式并围绕

场景独立地进行移动，其中一个相机开启共位的闪光灯来照亮拍摄场景（假定场景中没有其他光源）。两个相机在采集中所起的作用并不相同：开闪光灯的相机是场景中的光源（以点光源建模），而不开闪光灯的相机负责拍摄输入图片。两个相机均由手持的方式进行移动，无需遵守任何预定义的轨迹，但移动的目标是获得关于视角方向和光源方向的良好覆盖。以开闪光灯的相机作为光源的原因在于基于多视角立体匹配技术可以方便地对相机外参进行估计，从而得到共位的光源位置的估计。在实际采集过程中，我们在被拍摄物体周围放置 ChArUco 棋盘格[161]以提升相机参数估计的鲁棒性和质量。在输入到本章方法前，每张拍摄图片需要经过逆伽马矫正（伽马值假定为 2.2）变换到线性空间。尽管理论上两个相机所拍摄的图片均可以用在训练当中，然而我们实验中发现仅使用无闪光灯相机所拍摄的数据已经足够完成整个流程的训练。

数据预处理　本章方法所依赖的粗糙几何是通过 COLMAP 方法[34]基于自然光照下拍摄的若干图片重建得到。尽管基于点光源下的拍摄图片也可以重建场景几何，但其中包含的强烈高光和阴影等着色效果会严重影响重建质量，因此我们使用自然光照下预拍摄的图片进行几何重建（该拍摄数据仅用于几何重建）。图 4.8 展示了本章所有场景对应的粗糙几何可视化结果，我们使用每个点处的法向量方向作为其可视化颜色。我们在 COLMAP 方法重建得到的粗糙几何的基础上进行了最小化的手工清洗，主要包括两类：其一，通过指定包围盒或者选择最大连续几何等方式来剔除背景部分的无关几何（参见图 4.8 右下角）；或者当 COLMAP 方法在某些局部区域失败时，手工对该区域进行修剪（该情况仅发生在球体和金鱼的例子当中，参见图 4.8 右上角）。

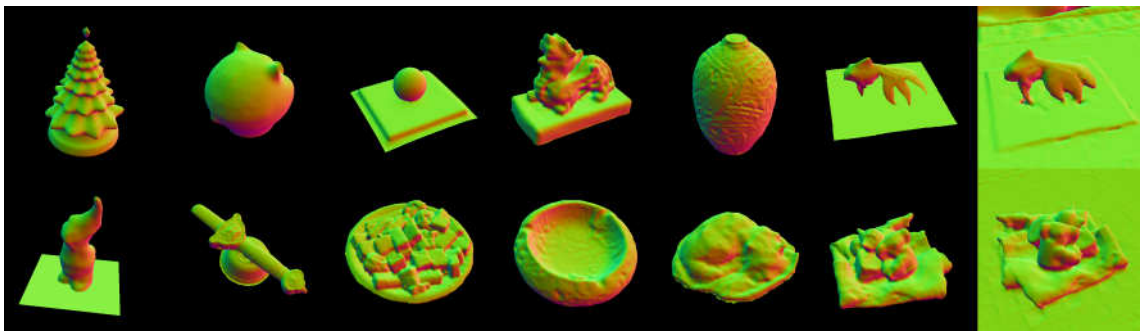


图 4.8　本章测试场景的粗糙几何可视化

每张图片对应的前景遮罩的生成方法如下：首先通过将粗糙几何的投影边缘进行收缩或膨胀得到三值图（trimap），然后通过闭式抠图算法[162]从 trimap 中生成前景遮罩的参考图。

真实场景介绍　我们一共采集了 9 个真实场景，包括：小红帽（带有与高光反射板之间的明显的高光互反射）；铜质花瓶（带有粗糙的高光反射）；糖果碗（带有粗糙几何未捕捉到的强烈互遮挡和阴影）；碗具（带有大范围遮挡）；宝剑（带有精细的纹理细节）；金鱼（带有高光互反射和精细纹理细节）；貔貅（带有半透明效果）；毛绒猫（带有复杂的毛发几何和材质）；组合场景（带有复杂的阴影、各向异性反射以及物体间互反射）。其中，金鱼、貔貅、毛绒猫和组合场景是基于 DSLR 相机拍摄，其他场景使用手机相机拍摄。每个真实场景输入图片数量从 6000 到 20000 不等（具体参见表 4.1 ）。需要说明的是，尽管本章方法的输入图片明显多于前人提出的基于建模的方法[47]，但是本章方法可以更好地处理带有复杂材质和光传输效果的复杂场景。本章方法属于基于图像的重光照方法，通常而言基于图像的重光照方法均需要更多的输入图片。假想我们将全新视角重光照渲染分解为两个独立步骤，即首先进行固定光照的全新视角合成，然后再进行固定视角的重光照渲染，那么 10000 张输入图片大体上相当于 100 个视角方向采样和 100 个光源方向采样，因此实际上该采样数量并不密集，对很多基于图像的渲染方法而言，100 个视角或光源采样并不足以实现高质量的视角插值或者重光照合成。此外，由于输入图片均是从连续的视频序列中提取而来，因此其在采集时间和代价方面相较于离散的照片拍摄方式有明显优势。在理想情况下，以 30 FPS 模式进行视频采集，则整体数据采集时间不超过 6 分钟。在实际采集中，由于场景中仅由单个闪光灯照亮，因此场景整体亮度偏暗，为避免运动模糊等问题，我们需要减慢相机移动的速度（最终输入图片是视频帧的子样本），实际需要的拍摄时间要比理论估计时间略长（大约每个场景需要 15-30 分钟）。未来，采用亮度更大的闪光灯或者视频帧率更高的相机均可以进一步降低采集时间开销。

4.5.2　合成场景

本章方法使用了 3 个合成场景，其相机标定和粗糙几何重建过程尽可能模拟真实采集的过程。三个合成场景包括：圣诞树 (包含复杂的几何、纹理以及着色、阴影效果)，猪（其背部包含细致的几何细节），球体（带有纹理的高光反射球放置于粗糙高光反射的平面上，展示出富有挑战的表观属性以及强烈的互反射）。每个合成场景包含 10000 个训练数据，其视角和方向通过在半球面内随机采样得到。

4.5.3　训练细节

针对每个场景，我们以拍摄的无结构化图片为训练集进行端到端的训练。训练损失函数包括渲染损失函数和遮罩损失函数两部分（二部分权重相等）：其中渲染损失函数是神经渲染网络输出的对数空间的预测图片和转换到对数空间的参考图

表 4.1　测试场景中的定量结果[a]

场景	输入数量	平均绝对误差	平均 LPIPS 误差	最大绝对误差	最大 LPIPS 误差
猪	10,000	0.0030	0.061	0.0055	0.160
球体（高光反射）	10,000	0.0007	0.003	0.0012	0.013
球体（漫反射）	10,000	0.0006	0.017	0.0016	0.039
球体（混合）	10,000	0.0014	0.035	0.0059	0.072
圣诞树	10,000	0.0017	0.043	0.0040	0.099
糖果碗	16,729	0.0089	0.051	0.0160	0.084
铜质花瓶	17,024	0.0017	0.037	0.0059	0.092
小红帽	14,132	0.0034	0.032	0.0110	0.056
碗具	19,682	0.0034	0.046	0.0094	0.066
宝剑	13,537	0.0024	0.015	0.0052	0.029
金鱼	13,032	0.0039	0.121	0.0170	0.180
毛绒猫	6,389	0.0019	0.018	0.0037	0.029
貔貅	13,928	0.0036	0.066	0.0061	0.130
组合场景	16,720	0.0040	0.066	0.0092	0.089

[a] 真实场景的误差值在包含超过 1000 个测试数据的测试集上计算得到；合成数据的误差值则在包含 1384 个测试数据的测试集上计算得到。

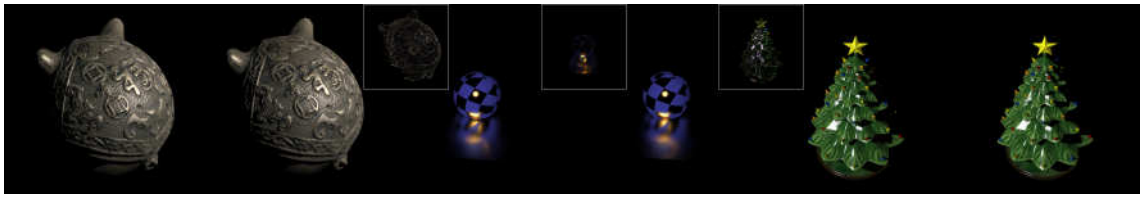
片之间的 L_1 损失函数，遮罩损失函数是网络预测遮罩和参考遮罩间的交叉熵损失函数。整个流程采用 TensorFlow 框架[151]实现，训练中采用 Adam 优化器[152]，其参数设置如下：学习率为 0.0002，β_1 为 0.9，β_2 为 0.999，其他超参数采用 TensorFlow 提供的默认值。网络训练中批大小设置为 1。在单个 NVIDIA P100 GPU 上，每个划分内的训练需要大约 20 个小时。本章提出的划分策略的额外优势在于不同划分内的训练过程完全独立，可以在多个 GPU 上并行进行。

4.6　实验结果和分析

本节首先在真实场景和合成场景中验证本章方法有效性，接着通过和已有方法进行对比分析来进一步说明本章方法的优势，然后通过消融实验验证每个组分和设计的必要性，最后分析本章方法的局限性。

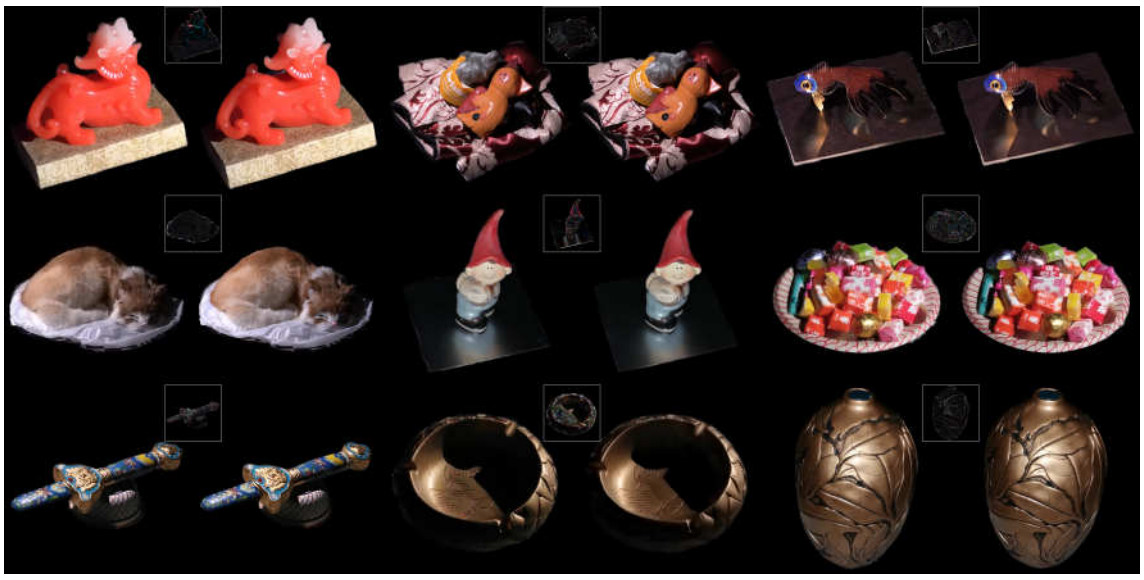
4.6.1　验证实验

图 4.9 和图 4.10 分别在合成场景和真实场景中初步验证了本章方法可以得到视觉上高质量的重光照结果，图中白框区域展示了各个场景所对应的误差图 (×5)。图 4.1 (a-c) 展示了方向光下的重光照结果，(d-f) 展示了环境光照下的重光照结果。我们在构建的测试数据集上进行了定量分析并进一步说明了本章方法的有效性。合成场景的测试数据集包含了随机采样而得到的 1384 个视角/光源组合，真实场景的测试数据集则包含了从采集数据中随机采样得到的超过 1000 个数据（和训练数据集没有重叠）。在定量实验中我们使用了四种误差度量，分别是平均绝对误差、最大绝对误差、平均 LPIPS 误差和最大 LPIPS 误差，其中 LPIPS[2] 是一种通过深度学习构造的基于感知的误差度量函数。



参考图　本章方法　参考图　本章方法　参考图　本章方法

图 4.9　合成场景在方向光下的重光照结果



参考图　本章方法　参考图　本章方法　参考图　本章方法

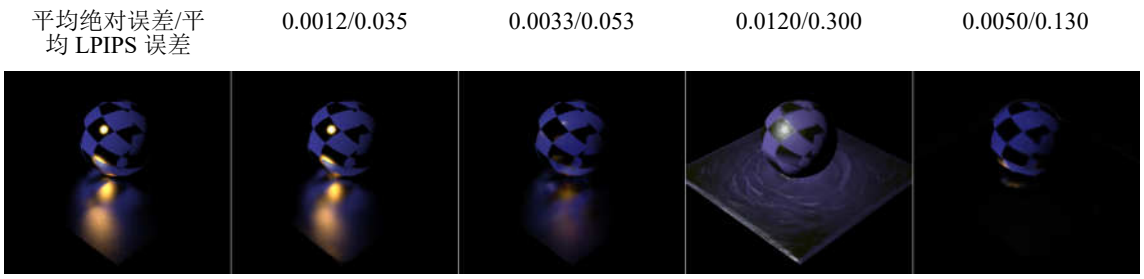
图 4.10　真实场景在方向光下的重光照结果

表 4.1 展示了本章方法在测试场景中的定量结果，结果表明本章方法可以在合成场景和真实场景中均得到高准确性的重渲染结果。其中，糖果碗的绝对误差数值偏大是由于手机相机拍摄图片中存在诸如运动模糊、失焦模糊、传感器噪声等

瑕疵；金鱼的 LPIPS 误差偏高是由于参考图中金鱼下方的高光板上存在闪烁高光（glints）纹理，该细节难以被准确捕捉和重建。此外，图 4.18 展示了本章方法在更多视角和光源下的重光照结果。

4.6.2　对比实验

前人提出的重光照方法中，尚没有工作可以针对复杂场景（即包含复杂材质属性和光传输效果的真实场景）以无结构化图片作为输入同时支持 360 度自由视点重光照渲染。我们整理了一些求解类似问题的已有工作进行比较。



(a) 参考图　(b) 本章方法　(c) Xu 等人[116]　(d) Nam 等人[47]　(e) Thies 等人[55]扩展

图 4.11　本章方法和已有工作的结果对比

Xu 等人[116]提出一种固定视角的重光照方法，该工作与其后续提出的多视角生成工作[54]相结合可以实现多视角的重光照渲染。虽然该方法可以泛化到训练场景外的一般场景中，但该方法局限于半球面内的重光照并且无法处理复杂的长距离全局光照效果（例如图 4.11 (c) 中球体和下方平面间的互反射）。该方法和本章方法的更多对比结果参见图 4.12 。

Nam 等人[47]提出一种基于手持设备所拍摄的后向散射数据进行物体几何和材质协同建模的方法。然而，该方法不考虑互反射效果，无法正确处理包含强烈互反射的场景（参见图 4.11 (d)）。实验中，该方法使用和本章方法相同的粗糙几何，材质重建结果和渲染可视化结果均由 Nam 等人友情提供。

Thies 等人[55]提出基于神经纹理的全新视角生成方法，并不支持重光照渲染。该方法提出将视角方向的三阶球面谐波系数和神经纹理的其中 9 个通道相乘后再输入到后续的渲染网络当中。考虑到球面谐波函数在预计算辐射亮度传输和逆渲染等领域被广泛用于光照的编码和表达，因此将光源方向也采用和视角方向类似的方法进行编码（即光源方向也采用三阶球面谐波系数表达并与 9 个神经纹理通道相乘）是 Thies 等人[55]方法的一种直接扩展。尽管该扩展理论上具有重光照的能力，但是结果表明（参见图 4.11 (e)）该扩展无法很好地学习到从输入光照信息（光源方向的球面谐波表达）到最终表观变化间的复杂映射。

参考图　　本章方法　　Xu 等人[116]

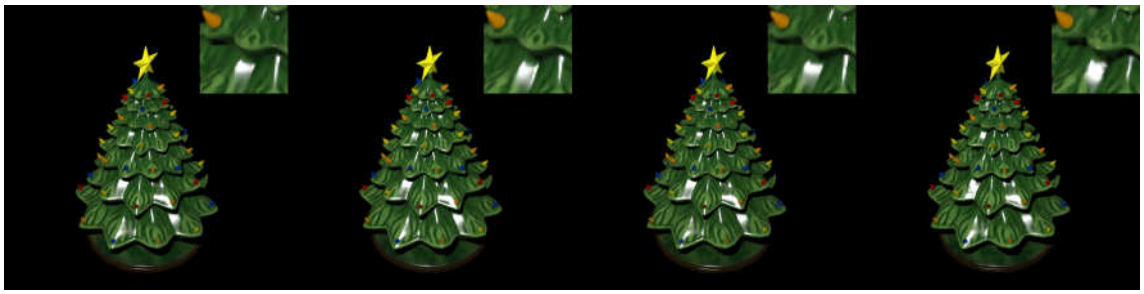
图 4.12　本章方法和 Xu 等人方法在其他合成场景中的结果对比

4.6.3　消融实验

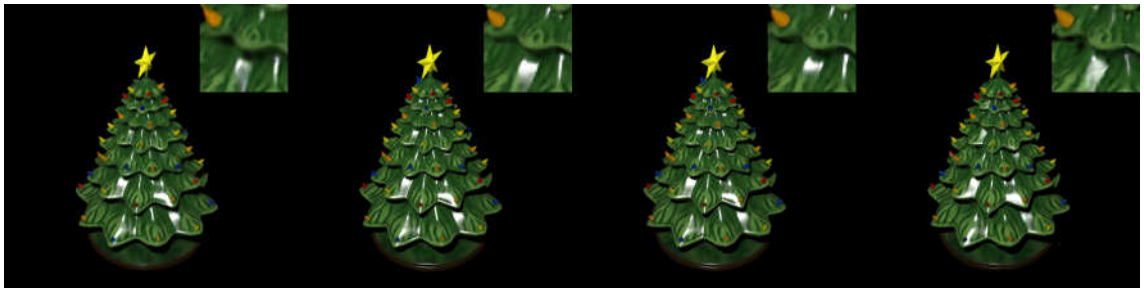
网络架构　Thies 等人[55]提出神经纹理表达并将其应用于全新视角生成。然而，我们实验发现 Thies 等人[55]所使用的神经渲染网络架构并不适合于本章的重光照渲染。Thies 等人[55]所使用的 U-net 网络架构无法准确预测高光细节等表观信息（图 4.13 (c)）并会导致在移动视角的测试视频序列中出现明显的抖动。本章所采用的带有残差结构的生成器网络架构则可以更好地复原表观信息（图 4.13 (b)）且不存在明显的视频抖动。

辐射亮度信息中的基材质数量　在经典的延迟光照算法中，由于已知物体的材质属性，因此仅需要漫反射和高光反射两个“光照图”就足以表达输入光照信息。由于本章方法并没有物体的真实材质信息，因此辐射亮度信息依赖多个不同的基材质来提供物体表观的粗糙估计。图 4.13 (d-f) 探索了不同数量的基材质对重光照的影响。需要注意的是，为公平地进行比较，我们在改变基材质数量的时候保证了每个基材质所对应的神经纹理通道数恒定。图中结果表明，和使用 5 个基材质（图 4.13 (b)）相比，增加基材质数量（图 4.13 (f)）带来的效果提升有限，而减少基材质数量（图 4.13 (d, e)）则会显著降低重光照质量（体现在高光细节准确性下降）。由于增加基材质数量会增加辐射亮度信息的渲染时间开销，因此我们最终选择 5 个基材质以取得质量和效率上的平衡。表 4.2 的定量比较结果也证实了以上观察。

(a) 参考图　(d) 1 个基材质　(g) 3 通道每基材质　(j) 2500 张输入图片



(b) 本章方法　(e) 2 个基材质　(h) 9 通道每基材质　(k) 1000 张输入图片



(c) Thies 等人[55]网络架构　(f) 9 个基材质　(i) 15 通道每基材质　(l) 500 张输入图片

图 4.13　消融实验可视化结果

输入图片数量　图 4.13 (j-l) 展示了本章方法在给定不同数量的输入图片（分别是 2500、1000 和 500 张，均是从完整的 10000 张中采样得到）时的重光照结果。本章方法默认采用 10000 张输入图片，其结果参见图 4.13 (b)。从图中可以看出，2500 张输入图片下，本章方法已经可以得到合理的重光照结果，而进一步增加输入图片数量可以继续提升重光照结果的准确性和增加更多细节。在移动视角的连续测试序列中，增加输入图片数量可以显著降低视频的抖动和帧间不连续性。一个重要的观察是，在变化光源的测试中，即使是在给定较少的输入图片数量的情况下，本章方法也并不存在明显的抖动和帧间不连续性。该观察提示我们提升输入图片数量的主要目的是修正粗糙几何的误差从而支持自由视点重渲染。表 4.2 的定量结果也进一步说明了增加输入图片数量可以降低误差。

神经纹理通道数量　每个辐射亮度信息（也就是每个基材质）分配多少与之对应的神经纹理通道是本章方法的一个重要超参数。图 4.13 (g-i) 表明增加神经纹理通

表 4.2　消融实验定量结果[a]

变种	平均绝对误差	平均 LPIPS 误差
Thies 等人[55]网络架构	0.0042	0.073
小划分范围 (30°)	0.0039	0.045
中等划分范围 (60°)	**0.0029**	**0.038**
大划分范围 (90°)	0.0042	0.049
1 个基材质	0.0057	0.049
2 个基材质	0.0038	0.044
5 个基材质	**0.0029**	**0.038**
9 个基材质	0.0029	0.038
3 通道每基材质	0.0031	0.038
6 通道每基材质	**0.0029**	**0.038**
9 通道每基材质	0.0030	0.038
15 通道每基材质	0.0028	0.037
500 张输入图片	0.0049	0.053
1000 张输入图片	0.0042	0.045
2500 张输入图片	0.0032	0.040
10000 张输入图片	**0.0029**	**0.038**

[a] 表中所示误差值是在所有合成场景的测试集上计算得到的平均误差，其中每个合成场景包含 1384 个测试数据。本章方法的默认配置以粗体标注。

道数量可以带来一定的视觉质量提升。从表 4.2 的定量结果中看，提升神经纹理通道数量带来的结果质量提升并不显著。值得一提的是，虽然每个基材质仅对应 3 个神经纹理通道的变种与 Thies 等人[55]的神经纹理通道数相同，但本章方法变种提供了额外的高质量重光照渲染的能力。本章方法选择每个基材质对应 6 个神经纹理通道的配置，该配置可以在准确性、训练时间和推理效率等方面取得更好的平衡。

划分方案　本文 4.3.4 节中提出了一种基于视角方向的神经纹理通道划分方案，可以有效解决扩大神经网络所带来的显存限制问题。该方案根据视角划分神经纹理通道，而每个划分均带有一个独立的神经渲染网络。为了更好地理解该划分方案的必要性和优势，我们对比了以下四种可选的方案：

1. 单个大网络方案，该方案不带有任何空间划分，仅包含单个神经纹理和单个

神经渲染网络，其神经纹理的通道数等同于本章提出的划分方案中所有划分中神经纹理通道之和；

2. 联合训练方案，该方案的空间划分方法和我们方案一致（即根据视角划分神经纹理通道，每个划分内包含一个独立的神经渲染网络）。区别主要体现在训练机制上，我们的方案中各个划分内独立进行训练，而该方案在训练时将所有划分内的神经纹理和神经渲染网络联合训练。具体来说，对于某个给定视角而言，其渲染结果是该视角所属三角形区域的三个顶点处的预测结果的重心加权和，训练时通过计算加权后的渲染结果和参考图间的误差进行反向传播；
3. 共享神经渲染网络方案，该方案中各个划分有独立的神经纹理，但所有划分共享同一个神经渲染网络；
4. 共享神经纹理方案，该方案中所有划分共享同一个神经纹理，但各个划分有独立的神经渲染网络。

(a) 参考图　(b) 本章方法　(c) 联合训练方案　(d) 共享神经渲染网络方案

(e) 共享神经纹理方案　(f) 单个大网络方案　(g) 30° 划分范围　(h) 90° 划分范围

图 4.14　针对划分方案和划分数量的消融实验结果

需要说明的是，完整半球面需要 13 个划分来覆盖，以上几种候选方案均会超过 GPU 的显存限制。在本消融实验中，我们仅在单个三角形区域内（包含 3 个顶点，对应 3 个划分）进行各方案的训练和对比。图 4.14 展示了在合成场景圣诞树中不同划分方案的可视化对比。从图中可以看出，本章提出的划分方案（图 4.14 (b)）可以生成和参考结果（图 4.14 (a)）视觉上相似的渲染结果并且保持高光细节准确。联合训练方案预测结果的高光部分存在明显偏差。共享神经渲染网络方案

表 4.3　不同划分方案的定量结果对比[a]，最优值以粗体标注

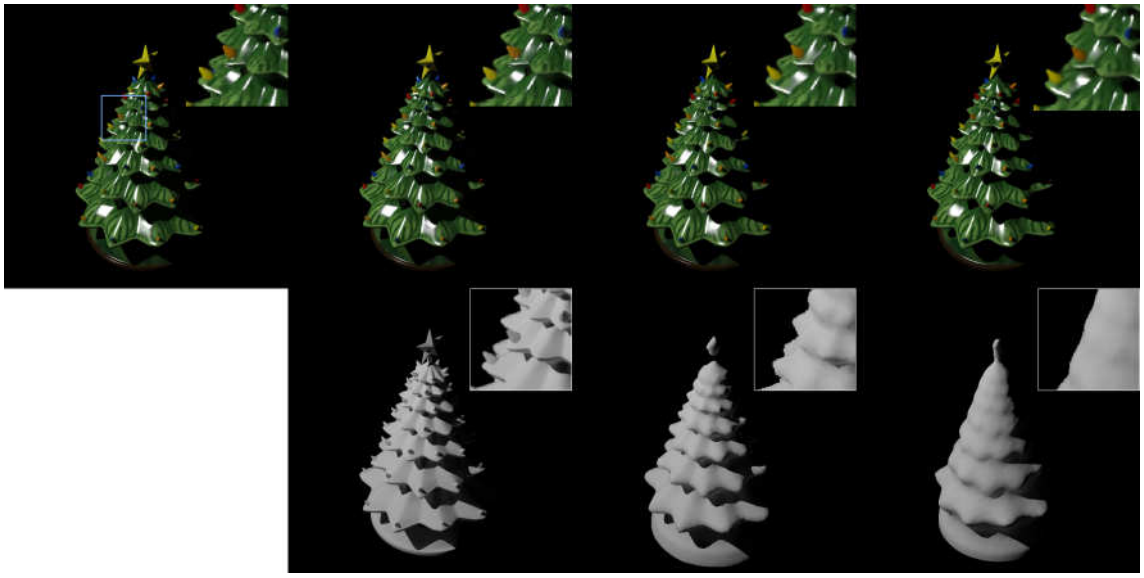
划分方案	平均绝对误差	平均 LPIPS 误差
单个大网络方案	**0.0026**	0.059
联合训练方案	0.0036	0.067
共享神经渲染网络方案	0.0029	0.063
共享神经纹理方案	0.0040	0.071
本章方案	0.0027	**0.056**

[a] 表中所示误差值在包含 180 个测试数据的测试集上计算得到的平均误差。

的结果比*共享神经纹理方案*的结果更加准确，该观察提示我们针对神经纹理进行划分是更加关键和必要的。虽然*单个大网络方案*的结果和本章方法相似，然而本章方法不受限于显存限制。表 4.3 的定量比较也证实了可视化结果中的观察。

划分数量　本章提出的划分方案中将半球面划分成 13 个区域，相邻顶点间相距约 60°。在神经纹理的通道总和保持恒定以及训练图片总数不变的情况下，划分的数量取决于两个相互独立的因素：其一，每个划分内的神经纹理通道是否足够准确表达该划分范围内的表观变化；其二，每个划分内的训练图片数量是否足够丰富以避免过拟合。实验中我们发现过大的划分覆盖范围和过小的划分覆盖范围均会造成重光照渲染结果的质量下降：一方面，神经纹理通道总数决定了每个划分所覆盖范围的上限，如果划分覆盖范围过大，会导致拟合质量下降。图 4.14 (h) 中展示了 90° 划分范围的结果，相较于 60° 划分范围而言，其生成质量更低且其时间抖动更加明显。另一方面，减小划分覆盖范围会增加划分数量，同时导致每个划分内的训练数据减少（每个视角范围内训练数据中的光源多样性也随之降低），使得最终的渲染质量（图 4.14 (g)）也同样出现下降。表 4.2 的定量比较结果也进一步证实了该结论。

几何准确性　本章方法对粗糙几何的误差保持鲁棒。图 4.15 展示了本章方法基于不同准确度粗糙几何的重光照结果对比。图 4.15 (c) 所示粗糙几何的质量和真实采集并重建得到的粗糙几何质量类似，在该粗糙度下，本章方法仅需要 2500 张输入图片进行训练即可实现高质量的重光照渲染。对于更加粗糙的几何（图 4.15 (d)），在给定更多的输入图片（10000 张）的情况下，本章方法仍然可以得到很好的重渲染结果。此外，更加精确的几何（图 4.15 (b)）可以显著地降低训练所需的输入图片数量（1000 张），并且有助于细小高光细节的准确预测。综上，粗糙几何的准确



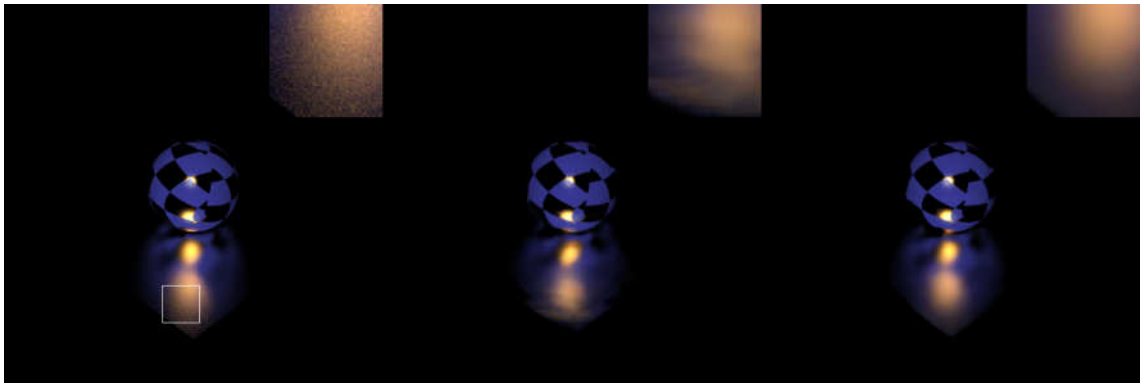
(a) 参考图　(b) 准确几何　(c) 粗糙几何　(d) 极粗糙几何

图 4.15　针对粗糙几何准确度的消融实验结果

度和训练所需的输入视角数量紧密关联，更加准确的几何先验有助于本章方法实现较大距离的视角间的高质量插值。此外，粗糙几何的几何准确度和本章方法所需视角数量间的关系还取决于物体本身的几何和材质复杂性。

辐射亮度信息中的全局光照　如之前所述，本章方法中的辐射亮度信息是使用路径跟踪算法渲染得到的带有全局光照的图片。在辐射亮度信息中包含全局光照信息的初衷是期望神经渲染网络可以有效利用辐射亮度信息中的全局光照提示信息并在此基础上合成正确的全局光照。然而，由于辐射亮度信息的渲染过程基于粗糙几何和预定义的同质基材质，因此其包含的全局光照信息并不准确，错误的全局光照信息严重依赖于神经渲染网络进行修正。一个自然的疑问是辐射亮度信息中的全局光照信息是否是必要的，也就是说，利用神经渲染网络来修正输入中并不准确的全局光照是否比从头合成全局光照更加容易。图 4.16 在带有强烈间接光照效果的球体场景中展示了针对辐射亮度信息是否包含全局光照的消融实验结果。结果表明，即使辐射亮度信息不包含全局光照信息，本章方法仍然可以预测出合理的全局光照效果，然而辐射亮度信息中包含全局光照信息可以进一步提升本章方法的全局光照预测质量。

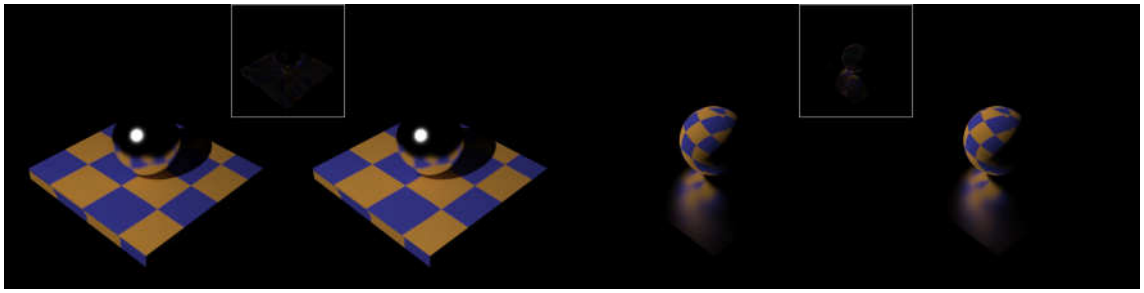
图 4.17 在球体场景的两个变种场景（变种场景包含和原始场景不同的材质组合，左侧为高光反射球在漫反射平面上，右侧为漫反射球在高光反射平面上）中进一步展示了本章方法可以有效生成复杂的全局光照效果（该例子中主要体现为不同材质间的复杂互反射），图中白框区域展示了每个场景的误差图 (×5)。虽然辐射



(a) 参考图　　(b) 无全局光照　　(c) 有全局光照

图 4.16　辐射亮度信息是否包含全局光照的结果对比

亮度信息对应的每个基材质均是同质的不透明材质，但本章方法仍然可以有效处理空间变化的复杂材质甚至是半透明材质。实际上，神经渲染网络的主要作用正是将不准确的输入信息非线性地映射到最终的准确渲染结果，其典型例子除本节讨论的基于输入中不准确的全局光照提示信息生成正确的全局光照外，还包括增加输入中不包含的几何细节、修正由于法向量存在错误或者几何存在缺失而导致的表观误差等。此外，消融实验也表明，物体的粗糙几何和辐射亮度信息越精确，则神经渲染网络越容易得到更加准确的结果。探索辐射亮度信息的其他编码方式（例如将直接光照和间接光照分开编码）也是未来有趣的研究方向。



参考图　　本章方法　　参考图　　本章方法

图 4.17　不同材质间复杂互反射效果的验证实验结果

辐射亮度信息和神经纹理的结合方式　本章方法中辐射亮度和神经纹理的结合方式是逐像素相乘，我们也对拼接等其他结合方式进行了实验。然而实验表明，采用拼接方式后整个训练过程更加不稳定且其最终结果质量也更低。经过分析我们认为，逐像素相乘可以显式地将辐射亮度信息与固定的神经纹理通道进行耦合，其约束比采用拼接方式更强。探索其他可能的结合方式也是未来有价值的研究方向。

4.6.4　局限性分析

本小节主要讨论本章方法的局限性。

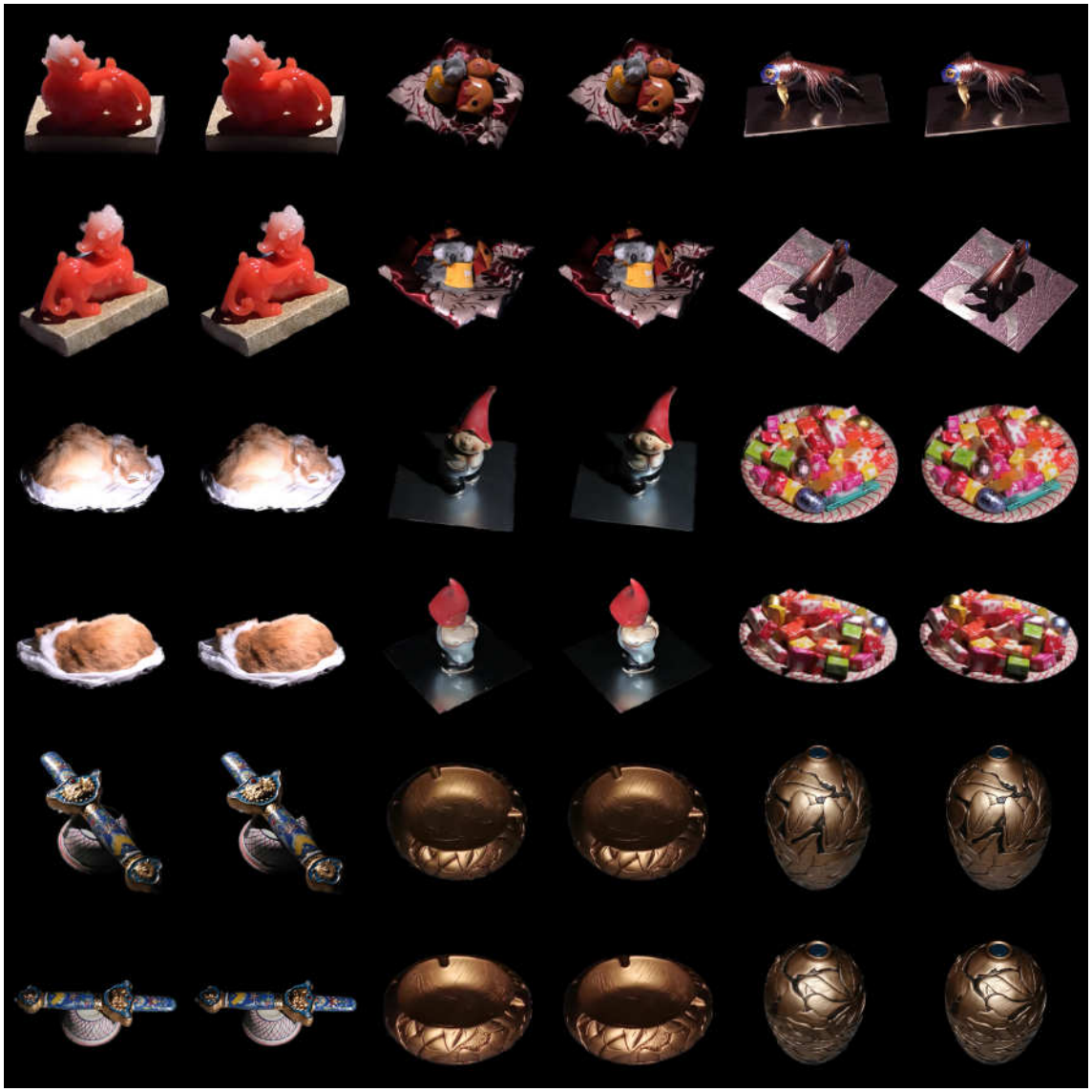


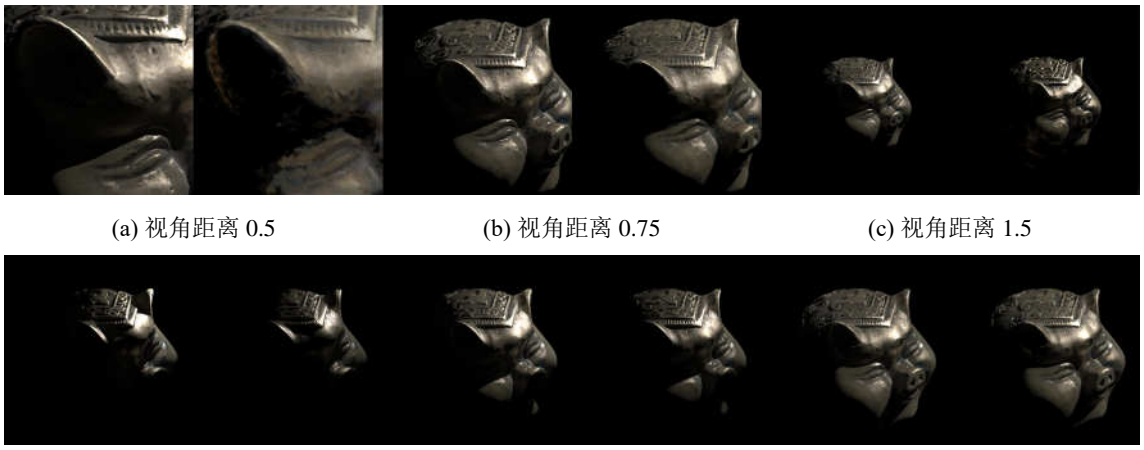
图 4.18　真实场景的更多重光照结果

某些情况下，本章方法无法准确预测物体表面的细小高光。我们分析其原因可能是：首先，由于细小的高光丢失只会在损失函数中引入较小的局部误差，因此在训练过程中很难准确拟合；其次，我们观察到粗糙几何的准确性会极大地影响细小高光的预测，粗糙几何的误差越小则本章方法预测的细小高光也越准确（参见图 4.15 ）。另一个与以上原因一致的观察是：对于那些缺失细小高光的预测图片，其输入的辐射亮度信息中通常也不包含与这些细小高光相关的信息。

本章方法对相机标定的误差敏感。虽然本章方法中的神经渲染网络可以修正粗糙几何和观测图片间的不一致性，但无法修正错误的相机标定带来的误差。当相机标定误差过大时，本章方法生成的连续视角移动的测试序列会出现“反复摇摆”的瑕疵。此外，虽然本章方法对于不同准确度的粗糙几何具有很高的鲁棒性，但无法有效处理存在大面积缺失的粗糙几何。

光照增强机制的准确性受限于本章方法在点光源数据集下训练的版本所合成的基图片的精确性。最重要的一种误差来源是训练数据存在无法覆盖完整动态范围的问题，该问题在包含强烈高光反射的场景中易于出现。在这种情况下，本章方法在点光源数据集下训练的版本也无法有效覆盖完整的动态范围，导致经过基于图像的重光照方法得到的增强数据存在明显的错误。此外，为了降低合成增强数据的时间开销，我们提出使用重要性采样策略得到的 100 个光源采样来近似环境光照。对于 100 个光源采样无法准确近似环境光照的场景，则必须退化为使用完整的光源采样，大大提高了时间开销。

在理论上，本章提出的辐射亮度信息可以在任何距离的视角和光源下渲染。然而，实际上，神经渲染网络只能在训练数据所覆盖的距离范围内生成合理的重光照结果。我们在合成场景猪中尝试了不同视角和光照距离下的重光照渲染，其结果参见图 4.19 。该图中所标注距离均为相对值，以输入数据所对应的视角和光源距离为 1.0。实验中发现，本章方法对于拉远的视角和光源（参见图 4.19 (c,f)）可以得到视觉上合理的重光照结果。本章方法对于在有限范围内拉近的视角和光源（参见图 4.19 (b,e)）也可以得到鲁棒的结果，但对于距离物体非常近的视角或光源（参见图 4.19 (a,d)）则会生成带有瑕疵的渲染结果。另外，我们发现应用光照增强机制后的神经渲染网络对于光源距离的改变更加鲁棒。



(a) 视角距离 0.5　(b) 视角距离 0.75　(c) 视角距离 1.5

(d) 光源距离 0.25　(e) 光源距离 0.42　(f) 光源距离 2.5

图 4.19　不同视角和光照距离下的重光照结果

最后，本章方法可以基于无结构化的输入图片实现高质量的重光照渲染，但仍然对采集方式存在一定偏好。在数据采集时，我们期望可以获得对视角方向和光源方向的 4D 外积空间的良好覆盖，需要尽量避免视角方向和光源方向存在强关联的采集方式。例如，相机和光源共位是一种常见的采集方法，但该方法实际上只能获得 4D 空间的一个 2D 采样，其覆盖范围严重不足，因而不适用于本章任务。在实际采集过程中，我们控制其中一个作为光源的相机的移动速度大约是另一个

相机的三倍，以避免二者之间存在强耦合。

4.7　本章小结

本章提出了一种可基于无结构化输入图片实现 360 度自由视点重光照的方法。无结构化的输入图片是由本章提出的包含两个手持移动相机所构成的轻量化采集设备完成采集。本章方法无需显式对复杂场景进行精确地几何或材质重建，可以大大降低采集过程的工作量。本章方法是通过融合深度场景表达和经典延迟光照方法的优点而构成的一种神经渲染管线，支持视点和光源的动态变化。此外，本章提出了光照增强机制来利用光传输的线性性质扩展本章方法所适用的光源类型。

未来工作方面，我们将进一步探索降低输入图片数量的方法，例如引入更高效的视角插值方法或者更好地利用场景中不同区域间的表观相似性。

第 5 章　基于深度绘制管线的全局光照绘制

本文第 4 章中介绍了一种基于深度场景表达的重光照方法，该方法针对真实世界复杂场景难以精确建模的挑战，提出使用基于粗糙几何先验的深度场景表达（包括定义在粗糙几何的神经纹理和负责编码输入光照的辐射亮度信息），再通过神经渲染管线来修正粗糙几何的误差以完成最终的高质量渲染。该方法属于基于图像的渲染方法，从概念上讲，基于图像的渲染方法可视为基于低维观测（输入图片均是二维数据）进行高维场景（几何和材质均定义在高维空间）的隐式推理并最终输出低维预测（渲染图片同样是二维数据）的过程。然而，高维场景信息在很多虚拟场景的渲染任务中是已知信息，例如在游戏制作或者电影特效制作当中，场景信息可以通过专业美术人员设计得到或者通过诸如本文第 3 章中介绍的材质建模及其他几何建模方法重建得到，因此并不需要基于低维观测来隐式推理高维场景。基于已知场景信息进行全局光照渲染具有形式化的明确定义和大量已有方法，该问题的主要难点和挑战在于如何快速地实现高质量渲染以满足众多实际应用的效率和质量需求。

(a) 厨房　(b) 客厅　(c) 浴室 2　(d) 康奈尔盒

图 5.1　基于深度绘制管线的全局光照绘制方法的渲染结果

全局光照的快速渲染是计算机图形学中一个非常重要且富有挑战的研究问题。全局光照可以为渲染结果提供诸多可以极大提升其真实感的视觉效果。相关工作可以分为离线渲染方法和实时渲染方法。其中，经典的离线渲染方法诸如路径跟踪[10,163-164]和光子映射[165-167]均可以实现照片级真实的渲染质量。然而，离线方法通常依赖于大量的采样和高昂的时间开销（单张无噪声图片的渲染需要耗费几分钟甚至几小时），因此并不适用于交互式的应用当中。实时或者交互级的全局光照渲染方法则可以大致分成三个类别：第一类是近似渲染方法，例如屏幕空间渲染方法[123-126]和基于体的渲染方法[168-170]等，该类方法通常受限于其算法假设或输入信息不足等问题，仅可以处理低频的全局光照效果；第二类方法是基于滤波或

者降噪的方法，该类方法通过对低采样数的有明显噪声的路径跟踪渲染图片进行降噪来实现全局光照渲染，然而由于光线跟踪的求交过程需要耗费大量时间，因此该类方法在带有复杂全局光照效果的场景中的运行效率仍然有限；第三类方法是基于预计算的方法，典型的工作包括光照贴图方法和预计算辐射传输方法。由于现有的基于预计算的方法通常仅适用于某些特定情况，例如静态光源[171-172]或点光源[61,137]，因此无法扩展到更一般的渲染场景中。总之，现有的离线渲染方法和实时渲染方法无法在渲染质量和运行效率两方面均达到实际应用要求。

5.1　本章引言

本章提出一种全新的基于深度绘制管线的全局光照绘制方法，可以实现动态面光源下静态场景的全频率全局光照的快速渲染（结果示例参见图 5.1 ）。本章方法的核心思路是使用深度神经网络来学习从场景输入信息（包括着色点属性、视角以及输入光照信息）到全局光照的复杂映射。本章方法可以视为一种基于预计算的方法，神经网络的训练过程充当本章方法中的预计算步骤。和经典的基于预计算的方法（例如光照贴图方法和光照探针方法等）相比，本章方法可以支持更加复杂的场景和全局光照效果（例如高光反射物体和动态面光源下的全局光照渲染），但同时本章方法和经典的基于预计算的方法拥有类似的设置，因此可以在相关应用中替换传统方法实现质量提升。

本章方法使用深度全连接网络（也被称作多层感知机，简写为 MLP）来对全局光照进行建模，由于该全连接网络在本章方法中用于生成渲染图片，因此也被称作神经渲染网络。从概念上讲，训练数据的渲染和神经渲染网络的训练过程是对于场景辐射场的采样和拟合过程，而训练好的神经网络的权重是一种紧凑的深度场景表达，最终运行时的渲染可视为基于该深度场景表达的插值过程。由于全局光照本身是高维且高度非线性的，因此如何使用紧凑的神经网络来高效地拟合高频全局光照是极富挑战的问题。本章提出三个策略来解决该挑战：其一，通过位置编码技术[173-174]将低维的输入向量转换到高维，相关研究表明位置编码技术可以帮助全连接网络更好地拟合高频细节。其二，使用组合式光照表达来对输入光照建模。如何将输入的动态面光源信息以神经渲染网络友好的方式进行建模尤为关键，本章提出的组合式光照表达可以充分结合多种不同表达的优势，使得神经渲染网络可以高效学习输入光照和最终表观变化间的复杂映射。其三，通过卷积神经网络（CNN）在不同着色点之间共享信息，该过程可以扩大全连接网络的感受野，从而提升本章方法的鲁棒性并降低神经网络的训练难度。此外，本章提出一种可选的基于材质划分的运行时加速方案，一方面可以在不影响渲染质量的

情况下显著降低整体的计算量从而提升渲染效率，另一方面还可以大大降低本章方法的存储代价。

本章方法可以在带有动态面光源的静态场景中实现交互级的全频率全局光照渲染。本章方法可以支持丰富的全局光照效果，例如：高光互反射（图 5.1 (a, b, d)）、焦散（图 5.1 (d)）、镜面反射（图 5.1 (b, c)）和色溢（图 5.1 (c, d)）等。我们在 8 个室内场景中进行测试并验证了本章方法的有效性。此外，由于屏幕空间 CNN 网络的全卷积结构可以自然地支持任意分辨率，而全连接的深度渲染网络是对各个着色点独立进行计算，因此本章方法无需额外训练即可支持高分辨率的全局光照预测。得益于本章提出的高效的深度场景表达，本章方法仅占用不超过 56MB（应用推理加速方案后仅需不到 5MB）的存储空间，因此适合于在多种实际应用中使用。

总之，本章方法的主要贡献包括：

- 针对包含复杂的光传输过程的室内场景，可以实现其在动态面光源下全频率全局光照的快速渲染；
- 提出一种端到端的深度绘制管线，其输入仅包括直接光照和其他屏幕空间贴图，可方便地集成到任何现有的实时渲染器当中；
- 提出一种组合式输入光照表达，以适合于神经渲染的方式对动态面光源进行建模；
- 提出一种屏幕空间神经贴图，可通过屏幕空间信息共享来提升神经渲染网络的学习效率；
- 提出一种基于材质划分的加速方案，实现运行时渲染效率的提升。

5.2　方法概览

正如本文 1.1.1 节所述，在不考虑参与性介质的情况下，物体表面某个着色点处的出射辐射亮度可以使用渲染方程[10]建模。考虑到本章叙述内容的完整性，我们将渲染方程的具体形式列举如下，公式中符号含义参考式（1.2）处的介绍：

$$L(p,\omega_v) = L_e(p,\omega_v) + \int_{S^2} f_p(\omega_v,\omega_i) L_i(p,\omega_i) |n_p \cdot \omega_i| \,\mathrm{d}\omega_i, \qquad (5.1)$$

根据入射辐射亮度来自光源或其他非发光物体，可以将式（5.1）分解为直接光照 $L_d(p,\omega_v)$ 和间接光照 $L_*(p,\omega_v)$ 两项。其中，面光源照射下的直接光照可以采用基于线性变换余弦函数的渲染方法[175-176]实时计算得到，而面光源下的间接光照的渲染则更具挑战性且通常需要耗费更多的时间。

本章方法的核心观察是静态场景中某个着色点处的间接光照 $L_*(p,\omega_v)$ 可以被

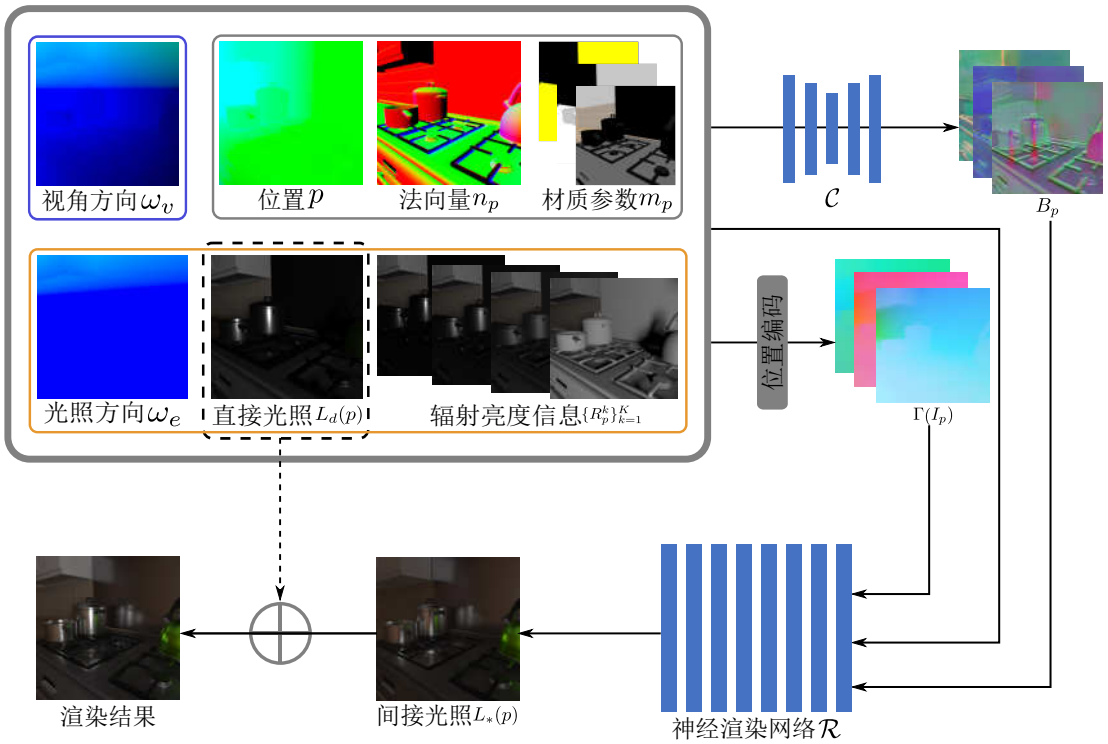


图 5.2　基于深度绘制管线的全局光照绘制方法的流程示意图

着色点位置、视角和光源所决定，换言之，着色点位置、视角和光源固定后其间接光照也是不变的。基于以上观察，我们可以将静态场景的间接光照重写为：

$$L_*(p,\omega_v) = \mathcal{F}(g(p),\omega_v,L_e(p)), \tag{5.2}$$

其中，$g(p)$ 代表着色点处的属性向量，包括位置 p，法向量 n_p 和材质参数 m_p；$L_e(p)$ 代表场景中的面光源；$\mathcal{F}$ 则代表从输入信息到间接光照的高度非线性的复杂映射关系。需要说明的是，实际上，在静态场景中 $g(p)$ 可以被位置 p 完全决定， 5.3.1 节将展开介绍在 $g(p)$ 中引入其他属性的原因。

本章方法的核心思路是利用深度神经网络来表达上述复杂映射 $\mathcal{F}$。本章方法可以分为预处理阶段和渲染阶段两个步骤。在预处理阶段中，本章方法使用离线渲染算法（路径跟踪）来生成带有全局光照效果的训练数据，并基于该训练数据集对深度神经网络进行端到端训练。训练好的神经网络可视为关于整个场景全局光照信息的一种紧凑的深度表达，借助该表达本章方法可以在无需任何复杂空间数据结构和昂贵存储代价的情况下，实现高频全局光照的插值和预测。在渲染阶段，直接光照和其他作为输入的屏幕空间贴图可以通过经典实时渲染管线得到，间接光照则是将以上信息输入到训练好的神经渲染网络中来通过网络前向预测得到。

图 5.2 概述了本章提出的基于深度绘制管线的全局光照绘制方法的流程，概括来讲：首先，本章方法的输入 I_p 包括三部分：着色点信息 $g(p)=\{p,n_p,m_p\}$、视角

信息 ω_v 和组合式光源信息 $L_e(p) = \{L_d(p), \{R_p^k\}_{k=1}^K, \omega_e\}$；接着，我们将以上输入 I_p 通过位置编码映射到高维空间得到特征 $\Gamma(I_p)$；同时，将输入通过卷积神经网络 C 得到屏幕空间神经贴图 $B_p = C(I_p)$；最后，将 I_p、$\Gamma(I_p)$ 和 B_p 拼接后输入到神经渲染网络 $\mathcal{R}$ 中得到间接光照预测 $L_*(p)$，加上直接光照 $L_d(p)$ 后即可得到最终渲染结果。以上流程中的各个组件的设计思路和具体内容将在本章后续进行详细介绍。

5.3　深度绘制管线

尽管深度神经网络适合于拟合高维空间的函数，但是基于紧凑的神经网络来学习从低维输入到高频全局光照的复杂映射仍然存在挑战。本节将首先介绍如何以一种神经网络友好的方式对输入信息进行编码，接着介绍基于全连接网络的神经渲染网络的细节，最后介绍关于本方法的训练和渲染流程。

5.3.1　神经网络友好的输入表达

考虑到从低维输入到全局光照的映射极为复杂，为了更高效地使用深度神经网络对全局光照进行建模，我们需要仔细设计输入信息的表达。如之前所述，静态场景中某个着色点处的间接光照取决于着色点位置、视角和光源。本节具体介绍每个部分的具体表达方式以及位置编码、屏幕空间神经贴图等内容。

着色点信息　理论上，静态场景中着色点位置 p 足以完整表达该着色点（其他辅助属性均是位置的函数）。然而，每个着色点仅给定位置信息的话，神经网络不仅需要预测每个点处的全局光照信息，还需要隐式地推测着色点处的其他和表观紧密相关的辅助信息（如空间变化的法向量和其他材质信息等），更多的学习任务意味着需要更大规模的网络。我们提出将着色点位置信息和其他辅助信息拼接后作为该着色点的信息表达，可以在保持预测质量不受影响的情况下显著降低神经网络的规模，从而保证了交互级的运行效率和较小的存储开销。在本章方法实现中，每个着色点处的信息可以表示为输入向量 $g(p)$，包括位置 p、法向量 n_p 和材质参数 m_p（材质参数包括漫反射颜色、高光反射颜色和粗糙度）。需要说明的是，对于镜面反射材质，本章方法在着色点处追加一条反射光线找到光路中的下一个交点，以该交点处的输入向量作为着色点处的输入向量。

视角信息　在不考虑相机镜片大小的情况下（即假定相机可以使用针孔相机模型建模），视角信息可以直接通过每个着色点处的视角方向 ω_v 来表达，视角方向即是每个着色点的位置和相机位置的差值向量。

输入光照信息　输入面光源 $L_e(p)$ 的表达则更加困难。尽管理论上可以简单地使用面光源的所有顶点位置和光源亮度来描述面光源，但是在该简单的光源表达下，神经网络难以有效地完成复杂且高频映射关系的高质量学习。本章提出一种神经网络友好的动态面光源表达，有效解决了光源信息难以传递到神经渲染网络的难题。光照表达中的冗余性使得神经网络可以在不同情况下利用光照表达中最有效的部分信息，从而更加高效地感知输出表观和输入光照间的关系。例如，在合成邻近漫反射物体间的色溢效果时需要更多地依赖于屏幕空间的光照信息，而合成离屏物体的高光反射效果时则需要利用全局光照信息。

具体而言，本章所使用的组合式光照表达包括屏幕空间光照信息（由直接光照和辐射亮度信息两部分构成）以及全局光照信息。

直接光照　直接光照 L_d 是着色点处的 BRDF 和光源的入射光照二者乘积的积分。直接光照作为一种基于物理的表达同时考虑了物体材质和光源信息，可以为神经渲染网络提供每个着色点处输入光照的有效信息。针对面光源下直接光照的快速渲染问题，本章方法首先采用线性变换余弦方法[175]对 BRDF 进行近似，然后以解析地方式快速计算不考虑可见性的直接光照积分，最后使用 Heitz 等人[176]提出的比率估计方法完成直接光照中的软阴影渲染。

辐射亮度信息　对于高光泽度或者镜面反射的物体，由于其 BRDF 是狄拉克分布（几乎处处取值为 0），其直接光照信息几乎是全黑的，因此难以为神经网络提供关于输入光照的有效信息。本章提出采用辐射亮度信息作为额外的输入光照表达，该表达可以在直接光照失效的情况下为神经网络提供所需的输入光照信息。本章中的辐射亮度信息 $R_p^k = L_d(p, \omega_v | b^k)$ 是一组定义在精确几何和 K 个预定义基材质 $\{b^k\}_{k=1}^K$ 上的直接光照渲染图片。本章中的辐射亮度信息和第 4 章中的辐射亮度信息主要有三点不同之处：其一，本章中的辐射亮度信息定义在精确几何而非粗糙几何上；其二，本章中的辐射亮度信息在渲染中只考虑直接光照而不考虑间接光照；其三，本章仅针对非漫反射材质的物体使用预定义的基材质来渲染辐射亮度信息，对于漫反射材质则在渲染辐射亮度过程中保持原材质。本章方法中使用辐射亮度信息的主要目的是为高光泽度或镜面反射物体提供补充的光照信息。

在本章实现中，基材质数量设置为 $K = 4$ 来覆盖不同的材质频率。具体地，基材质包括一个 Lambertian 漫反射材质和三个粗糙度分别为 0.05、0.13、0.34 的 Cook-Torrance 材质。由于辐射亮度信息并不包含间接光照，因此可以将辐射亮度信息 R_p^k 和直接光照 L_d 一同在延迟渲染管线中进行渲染，通过共享 G-Buffer 中的信息和阴影计算中光线求交的中间结果实现整体效率

的提升。

全局光照方向贴图　除屏幕空间的输入光照表达外，本章方法还提出通过输入光照方向 ω_e 作为全局的光照表达。对于每个着色点而言，其光照方向为着色点位置和面光源中心位置的差值。全局光照方向贴图可以将全局的光照位置信息扩散到每个着色点处，有助于神经网络学习远距离的全局光照效果并且可以帮助解决屏幕空间表达无法解决的二义性问题。例如，两种不同的输入光照可能会有类似的屏幕空间渲染结果，导致神经网络无法有效地进行区分，而全局光照方向贴图可以更直接地描述光源信息。

综合以上三点，本章使用的组合式光照表达可形式化地表示为：$L_e(p) = \{L_d(p), \{R_p^k\}_{k=1}^K, \omega_e\}$.

位置编码技术　实验中发现，如果直接将着色点属性 $g(p)$、视角信息 ω_v 和输入光照 $L_e(p)$ 输入到神经渲染网络中，神经网络会倾向于输出过度模糊的结果，无法准确地捕捉高频的全局光照效果。该观察和其他领域的前人工作[59,174,177]所得结论一致。

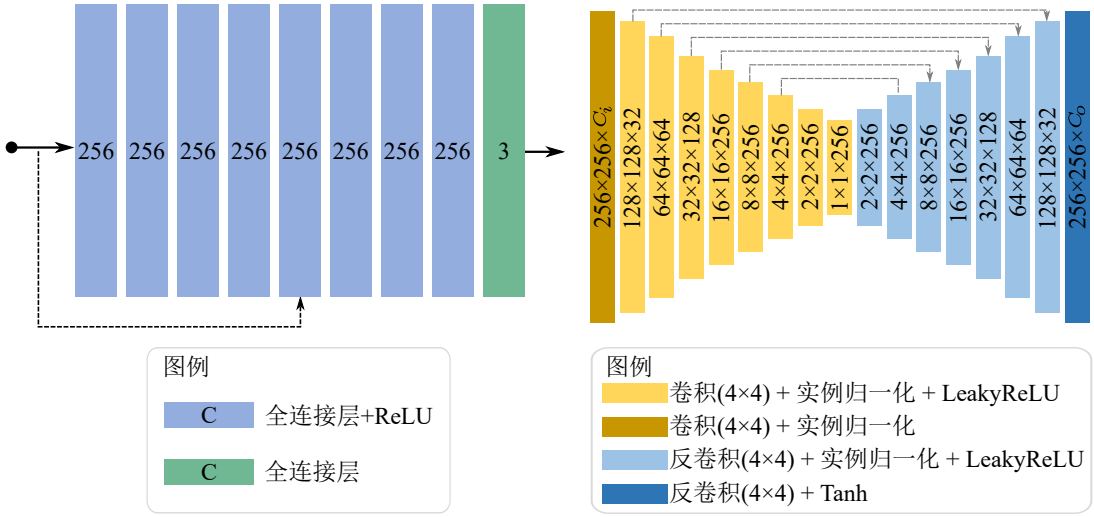
本章方法采用与 Mildenhall 等人[59]类似的位置编码技术，将输入信息通过预定义的傅里叶特征映射来变换到高维空间：

$$\begin{aligned} \gamma(x) &= \{\gamma_0(x), \cdots, \gamma_L(x)\}, \\ \text{where } \gamma_l(x) &= \{\sin(2^{l-1}\pi x), \cos(2^{l-1}\pi x)\}, l \in [0, L] \end{aligned} \tag{5.3}$$

实现中，本章方法对原始输入 $I_p = \{g(p), \omega_v, L_e(p)\}$ 的各个组分独立地应用上述映射 $\gamma(\cdot)$，参数 L 取值为 9。前人工作中[59]仅对位置和视角应用该变换，考虑到全局光照生成问题中视角、光源的移动以及材质参数的改变均会导致高频的表观变化，因此本章方法不仅对位置和视角应用该映射，而且还对其他着色点信息和输入光源信息也应用该映射。

应用位置编码可以使得神经渲染网络更加容易捕捉到高频的全局光照细节。编码后的高维空间特征 $\Gamma(I_p)$ 和原始输入 I_p 拼接后一同输入到神经渲染网络中。本文 5.5.4 节中对位置编码技术和其他可选的输入编码方法进行对比分析。

屏幕空间神经贴图　本节之前介绍的输入信息表达均是针对单个着色点，并未考虑邻近着色点间的信息共享。实际上，对很多全局光照效果的生成而言，仅从单个着色点出发来预测是非常低效的。邻近着色点信息进行共享可以使得某些全局光照效果的生成更加容易。例如：在合成两个邻近高光物体间的高光互反射或两个邻近漫反射平面间的色溢等全局光照效果时，邻近着色点的信息（例如其材质



(a) 全连接神经渲染网络 $\mathcal{R}$ 网络架构　　(b) 屏幕空间神经贴图提取网络 $\mathcal{C}$ 网络架构

图 5.3　本章方法中的神经网络架构

颜色）是非常有用的辅助信息。

本章提出使用卷积神经网络 $\mathcal{C}$ 来从输入 $\boldsymbol{I}_p$ 中提取屏幕空间神经贴图 $\boldsymbol{B}$ 以更好地利用屏幕空间内不同物体之间的上下文信息。在实现中，本章方法使用全卷积的 U-Net 网络结构[178]（参见图 5.3 (b)），其中编码器网络负责将逐像素的输入向量编码为紧凑的隐空间向量，而解码器网络则将全局的隐空间向量扩散回各个像素中。屏幕空间神经贴图 $\boldsymbol{B}$ 的空间分辨率和输入图片相同，其每个每个像素处存储高维的神经特征向量 $\boldsymbol{B}_p$。

本章提出的卷积神经网络模块有如下优势：

- 屏幕空间神经贴图扩宽了全连接神经渲染网络的感受野并将屏幕空间的全局信息扩散到各个着色点上；
- 全卷积的网络结构可以在无需任何重新训练的情况下自然地支持不同分辨率的输入。

本节小结　每个着色点处的神经网络友好的输入表达 $\boldsymbol{I}_p^+ = \{\boldsymbol{I}_p, \Gamma(\boldsymbol{I}_p), \boldsymbol{B}_p\}$ 包括：

1. 由着色点属性向量、视角信息和组合式光照表达构成的原始输入向量 $\boldsymbol{I}_p$；
2. $\boldsymbol{I}_p$ 经过位置编码技术转换后的高维空间特征 $\Gamma(\boldsymbol{I}_p)$；
3. 编码了屏幕空间全局信息的神经特征向量 $\boldsymbol{B}_p$。

5.3.2　基于全连接网络的神经渲染网络

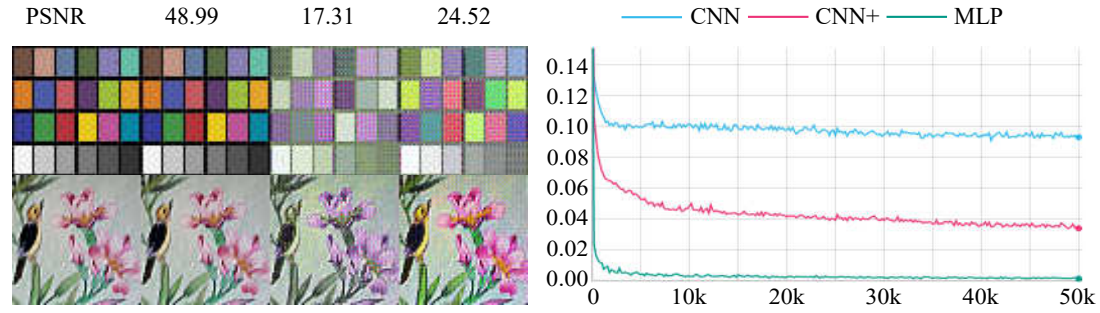
神经渲染网络 $\mathcal{R}$ 以 $\boldsymbol{I}_p^+$ 为输入来预测间接光照 $L_*(p, \omega_v)$。神经渲染网络采用和 Park 等人[56]方法中类似的全连接网络架构（参见图 5.3 (a)）。全连接渲染网络

和提取屏幕空间神经贴图的全卷积网络以端到端的方式共同训练：屏幕空间神经贴图可以指导全连接渲染网络如何更有效地计算全局光照结果，而全连接渲染网络可以帮助卷积神经网络提取到更加有效的屏幕空间特征。

神经渲染网络采用全连接网络架构而非经典的卷积神经网络架构的原因在于：全连接网络可以独立地学习各个着色点处的输入信息到其全局光照的映射，而卷积神经网络则不可避免地会受到邻居像素的干扰。换言之，全连接网络可以更好地利用每个着色点处的多数据间一致性，并有效学习到各个输入属性到对应输出结果间的解耦映射关系。举例而言，考虑某个着色点处的两个不同批的输入数据，二者除输入光照信息外的其他输入信息均保持一致，那么全连接网络会忽略两个批之间的共享输入信息（如视角、位置、法向量和材质等）并有效利用输入中唯一不同的光照信息来生成二者不同的表观结果。然而，卷积神经网络在执行屏幕空间的卷积操作后，该着色点处的共享输入信息会被邻域像素的信息所“污染”，导致卷积神经网络更加难以学习到不同输入属性到输出结果间的解耦映射。在典型的渲染合成数据集中，每个着色点会包含大量符合上述特性的数据采样，全连接网络可以更加有效地利用多数据一致性来学习到解耦的映射关系。

玩具示例：卷积神经网络和全连接网络对比

我们基于一个玩具示例来进一步验证以上观察。该玩具示例的任务是：通过神经网络（MLP 或 CNN）来学习从 **HSV** 色彩空间图片到 **RGB** 色彩空间图片的映射关系。该映射关系实际上拥有解析形式并且在不同像素间不存在任何依赖性。该玩具示例任务和本章的全局光照生成任务存在以下共性：其一，两个任务的目标都是学习从低维输入向量到低维输出向量间的复杂映射且输入向量中的不同属性影响输出的不同方面，例如玩具示例中输入的 **H** 颜色通道影响结果的色度而全局光照生成任务中材质参数影响渲染结果的局部表观；其二，从输入到输出的映射完全取决于单个数据点（如单个像素或单个着色点）本身，而与邻域信息无关。



(a) 参考图　(b) MLP　(c) CNN　(d) CNN+　(e) 训练收敛曲线

图 5.4　玩具示例的实验结果

该玩具示例的实验设置如下：首先，在 **HSV** 色彩空间中随机采样 1000 个分辨率为 64×64 的图片，**HSV** 的每个通道的值均通过在 $[0, 1]$ 范围的随机采样来得到。其次，利用解析的定义关系得到其对应的 **RGB** 空间的参考图片。在该训练数据集上分别完成卷积神经网络和全连接网络的训练。图 5.4 展示了本玩具示例实验的结果，图中显示，卷积神经网络（图 5.4 (c)）无法生成高质量的色彩空间转换结果，而可学习参数量与之相似的全连接网络（图 5.4 (b)）则可以准确地复原色彩空间转换后的结果。即使换用可学习参数更大（约 15 倍）的卷积神经网络（图 5.4 (d)），其重建结果仍然不够准确。定量误差数据（图 5.4 (a-d) 图片上方）和训练收敛曲线（图 5.4 (e)）进一步验证了以上观察。综上，该玩具示例实验表明全连接网络更加适合于学习满足类似性质的复杂映射关系。

5.3.3　训练和渲染

如之前所述，针对每个场景，全连接渲染网络 $\mathcal{R}$ 和卷积神经网络 C 需要一同进行训练，训练过程可以形式化地定义为：

$$\mathcal{R}^*, C^* = \underset{\mathcal{R},C}{\operatorname{argmax}} \sum_{i}^{N} \mathcal{L}(I_p^+, L_p | \mathcal{R}, C), \tag{5.4}$$

其中，$\mathcal{L}(\cdot,\cdot)$ 代表训练损失函数，L_p 是着色点 p 处的全局光照参考值，N 是训练数据个数。

训练损失函数包含像素损失函数 $\mathcal{L}_{pix}(\cdot,\cdot)$ 和感知损失函数 $\mathcal{L}_{per}(\cdot,\cdot)$ 两项：

$$\mathcal{L}(a, b) = \mathcal{L}_{pix}(a, b) + \lambda \mathcal{L}_{per}(a, b), \tag{5.5}$$

其中，λ 代表两项间的权重因子，实现中设置为 1.0。像素损失函数定义为像素值经过对数编码后的 L_1 距离：

$$\mathcal{L}_{pix}(a, b) = \| \log(a + \epsilon) - \log(b + \epsilon) \|_1, \tag{5.6}$$

其中，ϵ 取值为 $1.0/e$。感知损失函数 L_{per} 是 Zhang 等人[2]提出的基于人类视觉感知的图片距离度量，具体定义为图片通过预训练神经网络提取的特征图间的距离。按照 Zhang 等人[2]的建议，本章方法中使用 VGG16 网络的前 5 个卷积层特征来计算感知损失函数。为了与后文在测试时使用的感知误差 (LPIPS_{Alex}) 进行区分，训练中的感知损失函数可以记作 LPIPS_{VGG16}，二者的主要区别是用于提取深度特征的网络模型不同。

训练细节　本章方法基于 TensorFlow 框架[151]和 Adam 优化器[152]来实现，其中 Adam 优化器的超参数设置如下：学习率为 10^{-4}，$\beta_1 = 0.9$，$\beta_2 = 0.999$。训练中批

大小为 1，训练共有 1000k 个迭代步。训练单个场景在单张 NVIDIA RTX 2080Ti 显卡上大约需要 22 个小时。

训练数据生成　每个场景的训练数据包含约 5000 个视角，各个场景的具体训练图片数量参见表 5.1 ）。训练数据中每张图片对应的视角和光照按照如下流程得到：对于视角方向，在给定的包围盒中分别随机采样相机位置和相机看向的目标位置；对于面光源，在给定包围盒范围内随机采样光源中心位置，并在给定范围内随机采样面光源的大小。训练数据通过基于 GPU 的路径跟踪渲染器（Pharr 等人[179]提出的 PBRT-v4）完成渲染，训练数据分辨率为 256×256，每个像素采样数为 1024。为了尽可能充分地捕捉全局光照效果，路径跟踪中采样路径的最大长度设置为 16。我们使用 OptiX 降噪器[138]对渲染后的图片进行进一步的降噪处理。每个场景的训练数据生成过程在两张 RTX 2080 显卡上需要大约 20-26 个小时。

渲染流程　神经渲染网络训练结束后，本章方法可以据此来生成动态视角和动态面光源下的全局光照渲染结果：首先，使用经典的实时渲染管线生成 G-Buffer（包含位置贴图、视角方向贴图、光源方向贴图、法线贴图和材质贴图）、直接光照和辐射亮度信息；其次，将原始输入分别通过位置编码和卷积神经网络来得到高维编码输入和屏幕空间特征；然后，将以上内容拼接后输入到全连接神经渲染网络中得到间接光照的预测结果；最后，将直接光照和预测得到的间接光照相加即可得到全局光照渲染结果。

得益于光传输过程的线性性质，本章方法可以自然地支持多光源的情况。在渲染时，先通过本章方法得到每个面光源单独照射下的子渲染结果，再将所有子渲染结果求和即可得到最终的多光源照射下的渲染结果。理论上，多光源下本章方法的计算复杂性随光源数量增加而线性增长。在实际应用中，由于多个光源间的计算彼此独立，因此可以并行地计算每个光源下的渲染结果以降低整体的时间开销。图 5.12 最后一行的结果验证了本章方法可以生成多光源下高质量的全局光照渲染结果。

5.4　基于材质划分的加速方案

本节将介绍一种可选的基于材质划分的运行时加速方案。该加速方案的整体目标是在不降低生成质量的前提下实现运行时计算开销下降，其核心思路是通过将大网络替换为小网络来降低计算量从而达到加速的目的。然而，直接减小神经网络容量会导致其表达力下降，进而难以准确表达高频全局光照信息。本章提出基

(a) 厨房　(b) 浴室 2　(c) 康奈尔盒　(d) 浴室

(e) 客厅 2　(f) 客厅　(g) 卧室　(h) 楼梯间

图 5.5　本章测试场景示意图

于划分的策略来将单个大网络替换为多个小网络，其中每个小网络只负责拟合场景中部分着色点的全局光照信息。划分的标准是着色点的材质属性，具体而言，漫反射材质的着色点使用一个规模较小的网络而高光反射材质的着色点使用一个规模相对较大的网络。该加速方案的显著优势在于用户容易根据应用场景和使用需求来进行效率和质量间的平衡，例如在效率优先级非常高的应用中可以更加“激进地”减小子网络规模。此外，本章方法可以根据材质属性来分开调整两个子网络的规模，从而有效避免单个大网络规模减小时容易出现的各种视觉瑕疵。

本章提出的加速方案的具体实现流程如下：

- 根据每个着色点的材质属性生成划分索引。输入图片中包含的着色点输入向量需要根据着色点的材质属性进行重排，以将输入图片分解为两个子网络的输入向量。划分索引是着色点在原图片空间中的坐标到其在输入向量中的坐标间的双向映射关系。
- 基于划分索引，生成漫反射网络的输入向量和高光反射网络的输入向量。
- 在每个划分中基于子网络进行独立地前向查询，以得到各个划分的输出向量。
- 根据划分索引，将输出向量反向映射回原图片空间，得到最终的渲染结果。

5.5　实验结果

本节将首先介绍本章所使用的测试场景，然后通过可视化对比和定量分析验证本章方法有效性，接着介绍和其他已有工作的对比实验，最后介绍消融实验的

设计及结果。

5.5.1　测试场景简介

本章所使用的 8 个测试场景（如图 5.5 所示）均展示出复杂的全局光照效果。表 5.1 展示了每个场景中包含的三角形网格顶点个数以表明各个场景的几何复杂度。八个测试场景包括：*厨房*，带有多个物体间的强烈高光互反射效果；*浴室*，带有复杂的光传输过程；*浴室 2*，带有强烈的间接光照和镜面反射；*康奈尔盒*，带有色溢、高光反射和焦散；*客厅*，带有高光互反射和镜面反射；*客厅 2*，带有丰富的纹理细节、色溢和镜面反射；*卧室*，带有强烈间接光照和镜面反射；*楼梯间*，带有高光互反射效果。

5.5.2　结果验证

图 5.6 和图 5.12 展示了本章方法在多个场景中的渲染结果，验证了本章方法可以生成高质量的全局光照结果。图 5.13 展示了作为输入的直接光照和本章方法预测的间接光照。本章方法在不同的光源条件下均保持鲁棒：图 5.12 最后一行展示了本章方法在包含多个面光源的场景中可以生成高质量的渲染结果；图 5.7 展示了本章方法在不同大小的面光源下的渲染结果，从图中可以看出，对于如图 5.7

参考图　本章方法　误差图 ×5　参考图　本章方法　误差图 ×5

图 5.6　本章方法的全局光照渲染结果

表 5.1　测试场景中的定量结果[a]

场景	输入数量	顶点数	MAE	MSE	SSIM	PSNR	LPIPS_{Alex}
厨房	4,900	443*k*	0.0064	0.00034	0.984	37.82	0.0270
浴室	5,300	258*k*	0.0062	0.00041	0.985	36.85	0.0249
浴室 *2*	4,915	696*k*	0.0074	0.00020	0.983	39.96	0.0206
康奈尔盒	5,250	1,127*k*	0.0053	0.00017	0.992	40.99	0.0133
客厅	5,786	117*k*	0.0018	0.00004	0.996	47.10	0.0105
客厅 *2*	5,217	3,596*k*	0.0088	0.00028	0.988	36.74	0.0146
卧室	6,250	1,052*k*	0.0083	0.00050	0.985	36.67	0.0233
楼梯间	6,000	101*k*	0.0054	0.00011	0.991	42.78	0.0169

[a] 表中每个场景的误差值是在包含 100 个全新视角/光源的测试集上计算得到的平均误差。表中，SSIM 指结构相似性，PSNR 指峰值信噪比，LPIPS_{Alex} 是基于 AlexNet 网络的感知误差。

(a) 所示的极小面光源甚至是点光源本章方法可以生成带有硬阴影边界的高质量渲染结果，对于如图 5.7 (b) 所示的更大的面光源本章方法可以得到带有软阴影的渲染结果，而对于如图 5.7 (c) 所示的非常大的面光源，本章方法可以得到符合预期的几乎没有任何阴影的渲染结果。

(a) 极小面光源　(b) 中等面光源　(c) 极大面光源

图 5.7　不同尺寸面光源下的渲染结果

表 5.1 展示了本章方法在各个测试场景中的定量结果。在定量对比中，本章使用平均绝对误差 (MAE)、均方误差 (MSE)、结构相似性 (SSIM)、峰值信噪比 (PSNR) 和感知误差 (LPIPS_{Alex}) 等误差度量来对本章方法和参考结果进行详细评估。

本章方法针对每个场景仅需要约 55.9 MB 的存储空间。本章方法主要包括 G-Buffer 生成、直接光照渲染和网络推理三部分，在 256×256 分辨率下，每部分的时间开销分别为：G-Buffer 生成平均需要 0.7 ms；直接光照渲染（包括直接光照图

片和 4 个辐射亮度信息图片的渲染）中，不考虑可见性的直接光照渲染需要约 6.0 ms，随机阴影计算约 10 ms；网络推理所需时间仅与网络大小有关而与具体场景无关，该步所需时间约为 28.3 ms。总之，本章方法可以实现 22 FPS 的运行效率。需要说明的是，以上空间占用和网络推理时间均为不使用基于材质划分的加速方案的结果，而在应用加速方案后，空间占用小于 5 MB，网络推理时间可进一步缩短 25% 到 49%。

5.5.3　对比实验

前人工作中尚没有方法可以实现动态视角和动态面光源下全局光照的快速渲染。我们整理了如下几个求解类似问题的相关工作进行比较（可视化对比结果参见图 5.8 ）：

(a) 参考图　(b) 本章方法　(c) RRF　(d) BCN　(e) CNNR　(f) RTRT　(g) LPV

图 5.8　本章方法和已有工作的结果对比

Ren 等人[61]提出的 RRF 方法支持生成静态场景在动态点光源下的全局光照，该方法假定场景中仅包含点光源且该方法中的光源表达无法轻易扩展到面光源的情况。为更加公平地对比，我们将该方法的神经网络大小扩展到同本章方法相同。如图 5.8 (c) 所示，RRF 方法可以得到合理的渲染结果但无法准确捕捉高光互反射效果，该结果表明 RRF 方法中的光源表达并不足以处理面光源等复杂光照。

Xin 等人[126]提出的 BCN 方法是一种基于屏幕空间的全局光照渲染方法，该方法可以基于屏幕空间贴图来生成漫反射物体间包含单次散射的间接光照。尽管该工作可以在不重新训练的情况下处理全新场景，但是为了使得比较更加公平，我们仍然对其进行了重新训练。如图 5.8 (d) 所示，BCN 方法倾向于得到过度模糊的渲染结果且无法有效预测高光反射相关的全局光照效果，该结果说明基于屏幕空间的策略并不适用于包含多次高光反射的复杂全局光照渲染任务。

Granskog 等人[60]提出的 CNNR 方法是一种基于深度场景表达的绘制方法，该方法提出了一种可以将场景几何、材质和光照解耦的深度场景表达并将其应用于间接光照预测任务中。在对比实验中，我们对该方法进行了简化，使之仅考虑动态光源而保持几何和材质恒定。需要说明的是，该方法在生成动态光源下的渲染结果时，首先需要 3 张在相同光源但不同视角下的全局光照渲染图片作为输入，然后将这 3 张渲染图片和 G-Buffer 一同输入到神经网络中，来得到全局光照预测结果。由于准备 3 张输入图片的开销已经远大于直接渲染目标图片的代价，因此该方法在动态光源任务中并不实用。如图 5.8 (e) 所示，CNNR 方法预测的全局光照渲染结果没有本章方法的结果准确（例如图 5.8 所示第一个例子中铁锅在墙面的高光反射和第二个例子中的软阴影等）。此外，该方法无法准确预测高光反射物体的表观和带有复杂纹理的材质细节（例如桌子和地面的木制纹理）。

RTRT 方法是通过结合低采样数光线跟踪和后处理降噪算法来实现全局光照渲染，实现中采用基于 OptiX 的实时路径跟踪渲染器和 OptiX 降噪器。如图 5.8 (f) 所示，RTRT 方法可以捕捉大致的全局光照效果但是由于输入噪声过大仍然无法得到高质量且无瑕疵的渲染结果。

LPV[168-169]方法使用多个虚拟点光源来近似场景中的间接光照并将其存储在三维体纹理中。针对漫反射场景，该方法可以实现实时全局光照预测，因而被广泛应用于实时游戏渲染当中。在对比试验中，我们手动将测试场景导入虚幻引擎并使用虚幻引擎中的 LPV 算法完成渲染。如图 5.8 (g) 所示，LPV 方法无法有效预测与高光反射物体相关的全局光照效果。需要说明的是，虚幻引擎中的材质模型和光照模型和其他对比方法存在差异，因此 LPV 方法的渲染结果和其他方法相比存在一些细微的表观差异。

表 5.2　本章方法和已有工作的定量结果对比[a]，最优值以粗体标注

方法	MAE	MSE	SSIM	PSNR	LPIPS_{Alex}
RRF[61]	0.0222	0.00197	0.918	30.06	0.1172
BCN[126]	0.0190	0.00231	0.917	30.41	0.1579
CNNR[60]	0.0215	0.00606	0.949	28.33	0.0925
RTRT	0.0117	0.00106	0.952	32.86	0.1263
本章方法	**0.0064**	**0.00034**	**0.984**	**37.82**	**0.0270**

[a] 表中所示误差值是在*厨房*场景中包含 100 个全新视角/光照的测试集上计算得到的平均误差。

表 5.2 的定量对比结果进一步说明了本章方法的优势，表中所示误差数据是在包含 100 个测试数据的测试集上计算得到，测试数据的光源和视角方向通过随机

表 5.3　消融实验定量结果[a]，最优值以粗体标注，次优值以下划线标注

变种	MAE	MSE	SSIM	PSNR	LPIPS_{Alex}
Zhu 等人[159]网络结构	0.0078	0.00050	0.977	36.15	0.0399
宽度 128	0.0082	0.00055	0.976	35.86	0.0403
深度 4	0.0077	0.00048	0.977	36.52	0.0370
无屏幕空间神经贴图	0.0099	0.00054	0.968	35.70	0.0412
无感知误差	**0.0051**	**0.00025**	**0.987**	**39.71**	0.0334
无位置编码	0.0071	0.00042	0.981	37.15	0.0335
SIREN 编码	0.0112	0.00108	0.963	33.28	0.0647
SIREN* 编码	0.0106	0.00099	0.967	33.60	0.0570
1000 张输入图片	0.0137	0.00130	0.951	32.34	0.0734
2500 张输入图片	0.0108	0.00092	0.964	33.76	0.0567
光源位置	0.0207	0.00152	0.926	31.73	0.0604
光源位置 + 其他属性	0.0109	0.00157	0.969	32.60	0.0466
光源位置 + 直接光照	0.0079	0.00040	0.981	37.49	0.0279
本章方法	0.0064	0.00034	0.984	37.82	**0.0270**

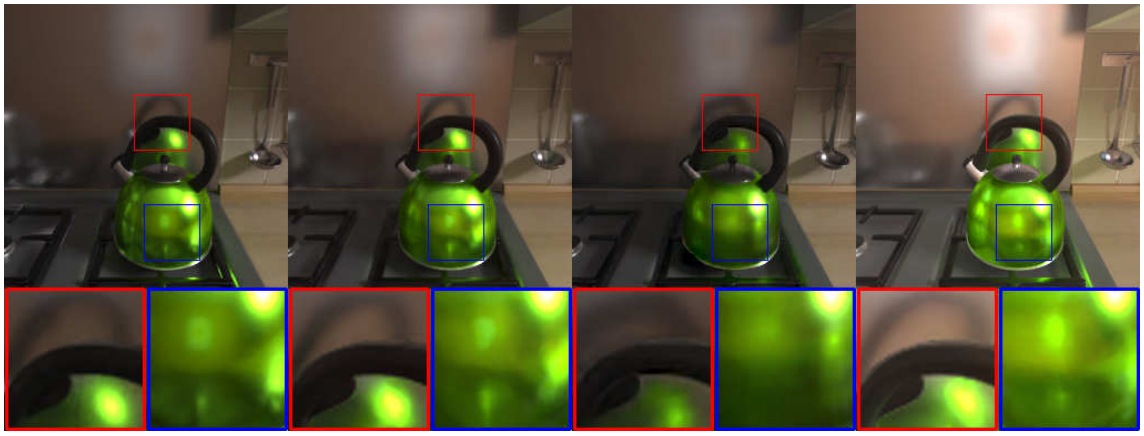
[a] 表中所示误差值是在厨房场景中包含 100 个全新视角/光照的测试集上计算得到的平均误差。

采样生成。从表中可以看出，本章方法在所有误差度量上均一致性地优于其他相关工作。

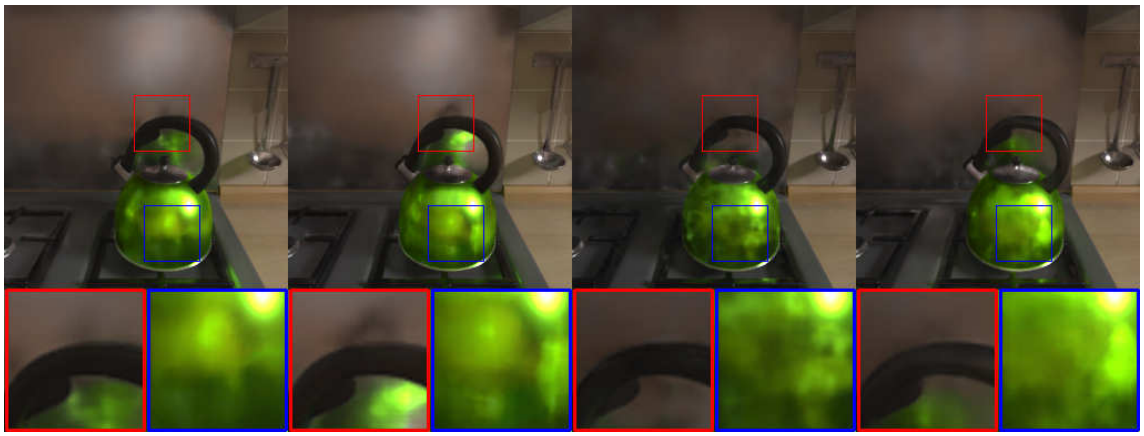
5.5.4　消融实验

神经渲染网络架构　本文 5.3.2 节提到全连接网络比卷积神经网络更加适合于学习从低维输入到全局光照的复杂映射。我们将本章方法使用的全连接网络和 Zhu 等人[159]提出的基于卷积神经网络的生成器网络进行了对比。如图 5.9 所示，本章使用的全连接网络（图 5.9 (b)）的全局光照渲染结果比经典的卷积神经网络 (图 5.9 (c)) 的渲染结果更准确。表 5.3 的定量比较结果也与可视化对比的结论一致。

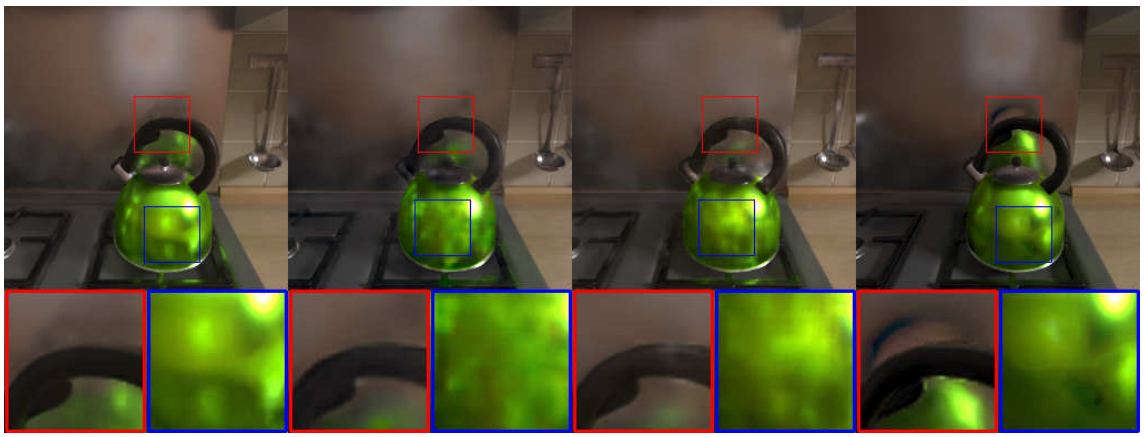
神经渲染网络大小　本章默认使用的全连接神经网络的深度为 8，宽度为 256。图 5.9 (e, f) 分别展示了减小网络宽度和深度后的渲染结果。从图中可以看出，减小网络规模会导致明显的结果质量下降。此外，减小网络宽度造成的质量下降比减小网络深度造成的质量下降更加显著，这表明对于本章任务而言网络宽度对于高质量结果的生成更加关键。



(a) 参考图　(b) 本章方法　(c) Zhu 等人[159]网络结构　(d) 无屏幕空间神经贴图



(e) 宽度 128　(f) 深度 4　(g) 1000 张输入图片　(h) 2500 张输入图片



(i) 无位置编码　(j) SIREN 编码　(k) SIREN* 编码　(l) 无感知损失函数

图 5.9　消融实验可视化结果

屏幕空间神经贴图　虽然在理论上静态场景中某个着色点处的全局光照可以由着色点位置、视角和光源完全决定，然而由于全局光照是由光线在多个物体间散射而形成，因此实际上不同物体间的信息共享可以有效提升神经渲染网络预测全局光照的质量。本章提出使用屏幕空间神经贴图来将屏幕空间内的物体的层次信息扩散到每个着色点上。如图 5.9 (b, d) 所示，屏幕空间神经贴图有助于消除生成结

果中的视觉瑕疵并提升结果准确性。表 5.3 的定量比较结果也说明屏幕空间神经贴图可以显著提升结果质量。

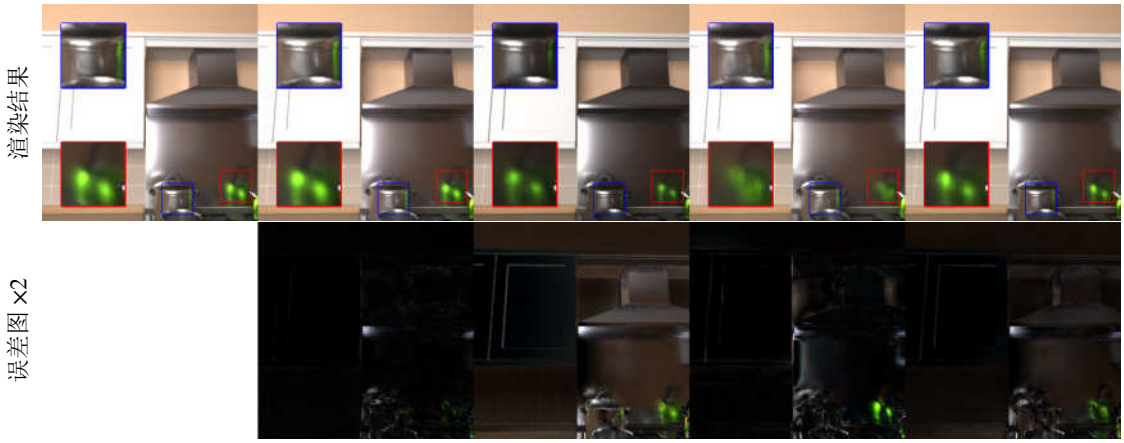
输入编码　位置编码技术可以帮助全连接网络更好地拟合高频函数。由于全局光照中包含丰富的高频细节，因此位置编码技术有助于本章提出的神经渲染网络捕捉更多的表观细节。图 5.9 (b, i) 展示了本章方法有无位置编码的渲染结果。为保证对比公平，我们在无位置编码的变种中，通过重复输入信息来保持神经网络输入的总通道数恒定。从图中可以看出，即使没有位置编码技术本章方法仍然可以得到合理的渲染结果。然而通过位置编码技术将低维输入向量映射到高维后，可以进一步提升结果质量。此外，我们也与近年来在全新视角生成领域获得大量应用的 SIREN[180]（SInusoidal REpresentation Networks）方法进行了比较：图 5.9 (j) 展示了使用 SIREN 方法的渲染结果，而图 5.9 (k) 是使用 SIREN* 方法的渲染结果，SIREN* 是指使用 SIREN 技术的同时通过重复输入信息来保持输入总通道数和本章方法一致。图中结果显示 SIREN 和 SIREN* 的结果均明显差于无位置编码的变种和本章方法的结果，该结果说明 ReLU 激活函数和位置编码技术在本章方法中均发挥重要作用。

训练损失函数　本章方法综合使用像素损失函数和感知损失函数 (LPIPS_{VGG}) 进行训练。在消融实验中，尽管加入感知损失函数后会导致测试集中传统误差度量均出现一定程度的增大（见表 5.3 ），但感知损失函数可以帮助神经网络聚焦于视觉上重要区域的学习而非噪声、过曝等不重要区域的学习，从而提升渲染结果的整体视觉质量（参见表 5.3 中 LPIPS_{Alex} 以及图 5.9 (b, l)）。

输入图片数量　本章针对输入图片数量（影响对视角/光照组合的覆盖）对全局光照预测结果的影响进行了实验。图 5.9 (g, h) 展示了从包含 5000 张图片的原始训练集中随机采样 1000 张和 2500 张图片进行训练的渲染结果。从图中看出，减少输入图片数量后，渲染结果更加模糊且无法很好地捕捉高频的间接光照。以上结果表明，训练数据能否覆盖足够密的视角/光照组合对于本章方法进行全局光照预测至关重要。此外，图片中的视觉瑕疵还会导致视频测试序列中时间上的抖动。我们在各个场景中使用了 5000 到 6000 张输入图片以在覆盖面和时间开销之间取得平衡。

输入光照表达　本章方法针对动态光源提出全局光源位置和屏幕空间光源表达（直接光照和辐射亮度信息）相结合的组合式光照表达。我们针对组合式光照表达中各个组分对最终渲染结果的影响进行了消融实验，该实验的可视化对比参见

图 5.10 。仅使用光源中心位置作为光照表达会忽略面光源的大小和朝向，无法完整地描述动态面光源。图 5.10 (c) 和表 5.3 也证明了基于该表达无法生成高质量的全局光照渲染结果。从误差图来看，该方案预测的间接光照存在整体亮度偏暗的问题。在理论上，使用面光源位置加上法向量朝向和所有顶点的坐标作为光照表达可以完整描述面光源信息，然而实验表明，该表达（图 5.10 (d)）可以在一定程度上缓解间接光照亮度偏暗的问题但仍然无法准确捕捉高频反射细节，说明神经网络友好的输入光照表达至关重要。通过结合光源位置和直接光照信息（图 5.10 (e)），本章方法可以得到合理的全局光照预测，而加上辐射亮度信息（图 5.10 (b)）可以进一步提升高光反射区域的间接光照预测准确性。



(a) 参考图　(b) 本章方法　(c) 光源位置　(d) 光源位置 + 其他属性　(e) 光源位置 + 直接光照

图 5.10　针对输入光照表达的消融实验结果

加速方案　本文 5.4 节介绍的基于材质划分的加速方案可以实现运行时渲染效率的提升和存储代价的降低。通过设定不同的子网络规模可以在运行效率和生成质量之间进行权衡。本章中，加速方案变种命名为 Dx/y-Sm/n-Cl，其中，Dx/y 指漫反射子网络的深度为 x 而宽度为 y，同理，Sm/n 指高光反射子网络的深度为 m 而宽度为 n，Cl 指卷积神经网络深度为 l。图 5.11 展示了不同加速方案变种的可视化对比结果。从图中可以看出，不同加速方案变种均可以得到合理的全局光照渲染结果。表 5.4 列举了不同加速方案变种的定量对比结果。综合可视化对比和定量对比的结果，D4/32-S6/256-C3 方案在取得约 25% 加速的同时可以得到最优的全局光照渲染结果，D4/32-S4/256-C3 方案的生成结果质量与基线方案几乎一致但可以实现 40% 左右的加速，D3/32-S4/128-C2 方案则可以实现 50% 左右的加速且其结果不包含明显的视觉瑕疵。在实际应用中可根据任务需求来选择合适的加速方案。

参考图　　基线方法　　D2/32-S4/128-C2　　D4/32-S4/256-C3　　D4/32-S6/256-C3

图 5.11　不同加速方案的可视化结果对比

表 5.4　不同加速方案的定量结果对比，最优值以粗体标注，次优值以下划线标注

方案	MAE	MSE	SSIM	PSNR	LPIPS_{Alex}	加速比	空间占用
基线方法	<u>0.0065</u>	<u>0.000353</u>	0.9838	<u>37.79</u>	<u>0.0253</u>	-	55.8 MB
D2/32-S4/128-C2	0.0088	0.000591	0.9747	35.65	0.0420	49.21%	1.39 MB
D4/32-S4/256-C3	0.0067	0.000358	<u>0.9838</u>	37.76	0.0293	40.02%	3.53 MB
D4/32-S6/256-C3	**0.0064**	**0.000311**	**0.9859**	**38.37**	0.0251	25.28%	4.79 MB

5.6　讨论分析

本节首先讨论本章方法和传统的基于学习的屏幕空间方法之间的不同之处并说明本章方法的优势，接着介绍本章方法的两种简单扩展以支持快速高分辨率渲染和材质编辑，最后分析本章方法的局限性。

5.6.1　与基于学习的屏幕空间方法对比分析

尽管本章方法的输入信息是屏幕空间中各个着色点处的信息，但是本章方法仍然与基于学习的屏幕空间方法[125-126]存在根本性的区别。基于学习的屏幕空间方法以屏幕空间贴图为输入，通过深度神经网络来预测动态场景中的全局光照。该类方法中神经网络的作用是从任意给定场景的屏幕空间贴图中预测全局光照，神

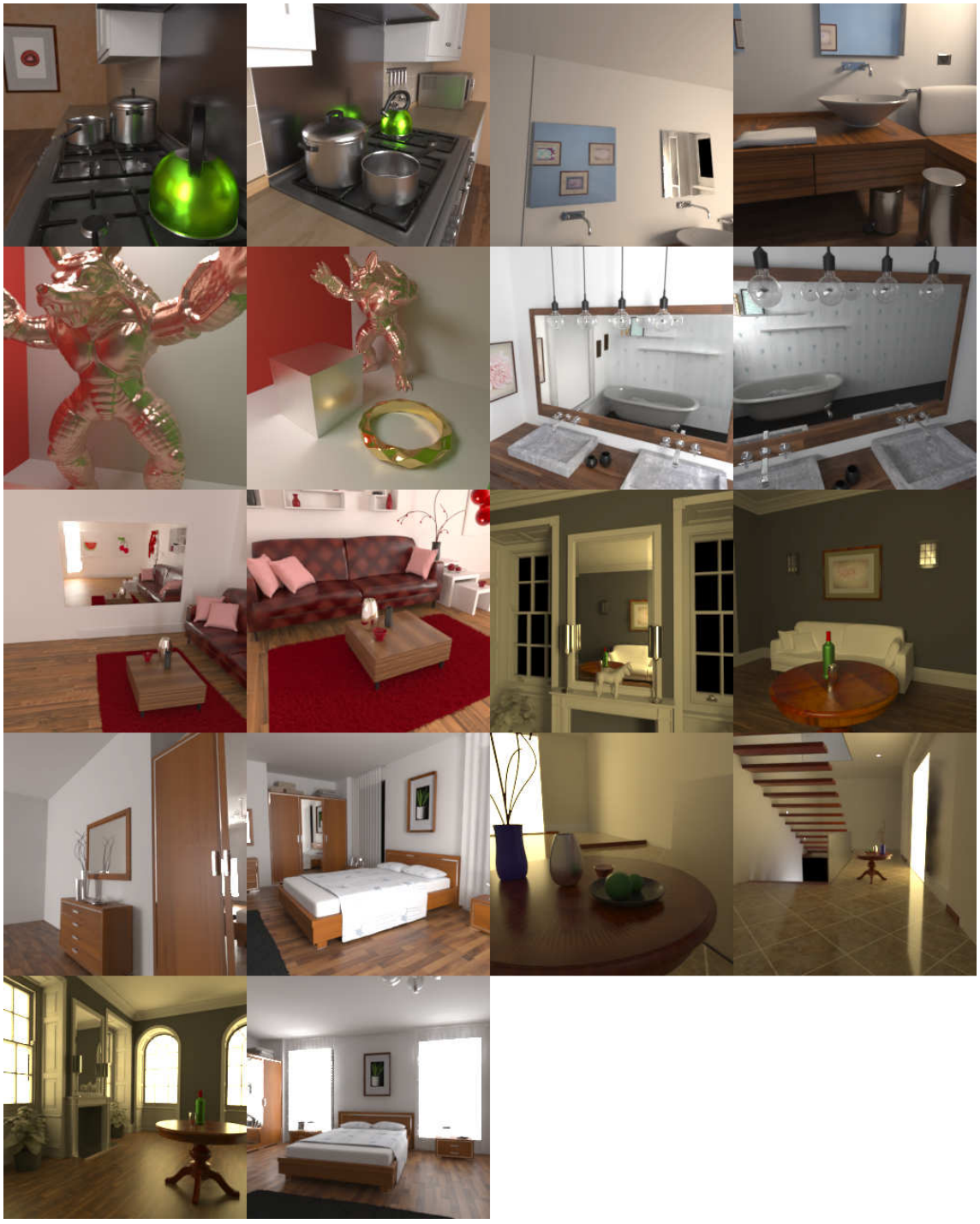


图 5.12　本章方法的更多全局光照渲染结果

经网络的参数中并不存储任何场景特定的信息。与之相反，本章方法假定静态场景，神经网络负责生成特定场景在动态视角和动态光源下的全局光照结果。训练数据生成和神经网络训练可以视为对特定场景的辐射场的采样和拟合；训练后神经网络的参数权重是该场景辐射场的一种紧凑表达；在运行时，给定某个着色点处的输入，本章方法通过查询场景辐射场紧凑表达来得到最终的全局光照渲染结

(a) 直接光照　(b) 间接光照　(c) 直接光照　(d) 间接光照

图 5.13　直接光照和间接光照可视化

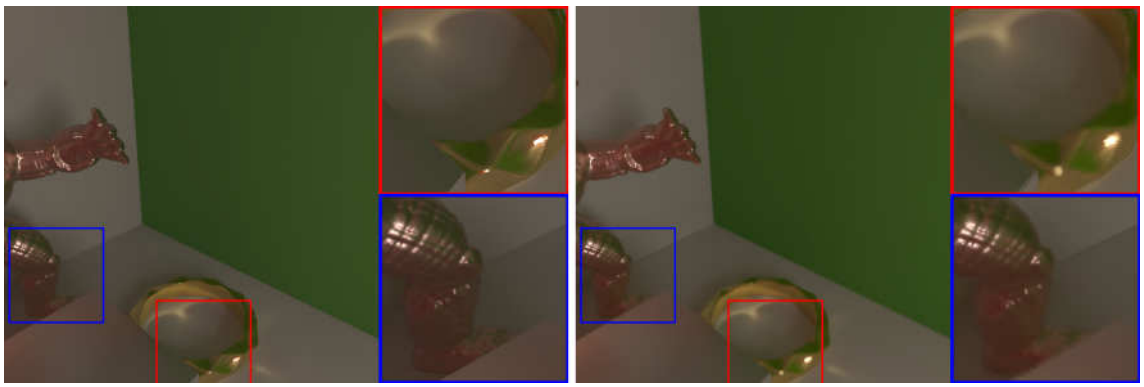
果。

理论上，显式的全局场景表达（例如点云或体素表达等）有助于增强神经网络对场景的理解。然而，为了使得本章方法尽可能轻量化，本章方法仅使用屏幕空间输入而不依赖于复杂的数据结构和预采样等繁琐过程，使得本章方法容易与现有的实时渲染管线相结合。为了使得从屏幕空间输入到全局光照的映射更加容易学习，本章提出采用神经网络友好的输入表达对原始输入进行编码。

5.6.2　基于联合双边上采样的高分辨率渲染

如之前所述，本章方法不需要任何重新训练即可支持更高分辨率下的全局光照渲染。在实践中，为实现运行时交互级的运行效率，可使用联合双边上采样方

法对本章方法预测的低分辨率间接光照结果进行上采样（以 G-Buffer 中的法线贴图或位置贴图作为指导图），最终的全局光照渲染结果是上采样后的间接光照和高分辨率的直接光照之和。图 5.14 展示了康奈尔盒场景中通过高分辨率管线预测得到的全局光照结果和通过以上双边上采样流程得到的全局光照结果的对比，结果显示基于联合双边上采样的高分辨率渲染流程可以生成合理的高分辨率全局光照结果并保持交互级的运行效率。



(a) 高分辨率管线　　(b) 联合双边上采样管线

图 5.14　不同渲染管线的高分辨率渲染结果对比

5.6.3　材质编辑

由于本章方法中神经渲染网络成功学习到了从输入材质参数到最终表观结果间的复杂映射，因此本章方法可以提供一定的材质编辑能力。实验发现，本章方法不需要任何重新训练或者数据增强即可支持场景中漫反射纹理的编辑。图 5.15 左侧两列展示了本章方法对漫反射纹理进行编辑的结果，其中第一行展示了训练数据集中原本的墙壁贴画，第二行则是运行时对其进行编辑后的结果，该结果说明本章方法可以得到鲁棒的漫反射纹理编辑结果。

然而，对于一般性的材质编辑任务（例如离屏物体造成的高光反射），本章方法仍然存在困难。为进一步扩展本章方法以支持更广泛的材质编辑，本章提出将全局材质特征向量作为每个着色点的补充输入信息。材质特征向量是场景中所有材质可变的物体（如图 5.15 右侧两列中的水壶和铁锅）的材质参数集合。将全局材质特征向量作为每个着色点输入信息的一部分，可以使得每个着色点均感知到全局的材质编辑信号，进而可以支持更一般化的材质编辑操作。为实现上述扩展方案，训练数据需要做额外的数据增广。具体而言，首先需要收集场景中所有材质可变的物体，其次在每个训练样本生成时，通过随机采样的方式为其生成可变材质参数（包括高光反射颜色和粗糙度），最后完成数据集中所有训练样本的渲染。如图 5.15 右侧两列所示，应用上述扩展后本章方法可以支持多次散射的高光互反

射等更加复杂的材质编辑操作。

图 5.15　材质编辑可视化

5.6.4　局限性分析

本章方法主要包含三点局限性。

首先，尽管在理论上本章方法可以支持任意视点和面光源，然而实际上本章方法和其他任何数据驱动的方法一样受限于训练数据覆盖范围。本章方法仅能针对训练数据所覆盖的视角、光照空间内的视角和光照预测出高质量的全局光照预测。图 5.16 (a) 展示了一个本章方法失败的例子，该例子的视角和光照组合并不在训练集覆盖范围内。

(a) 参考图　(b) 本章方法　(c) 参考图　(d) 本章方法

图 5.16　本章方法失败情况示例

其次，本章方法支持多种材质（例如漫反射、高光反射、镜面反射材质等），也可以合理地生成浴室 *2* 场景中的玻璃材质的灯泡。然而，本章方法所使用的材质模型（仅考虑非透明物体的表面散射）和光照表达均不足以准确描述一般性的折射物体。图 5.16 (b) 展示了康奈尔盒变种场景（将高光反射的立方体替换为玻璃材质的球体）的渲染结果，结果表明本章方法无法准确地捕捉包含折射的光传输效

果。

最后，本章方法无法支持复杂场景中任意的材质编辑。本文 5.6.3 节介绍的扩展策略可以提升本章方法材质编辑能力但仍然存在局限性。如果场景中具有大量材质可变物体，那么材质特征向量会迅速膨胀以至于神经渲染网络难以准确捕捉输入材质参数和最终的表观变化之间的联系。

5.7　本章小结

本章提出一种支持静态场景在动态视角和动态面光源下全频率全局光照快速渲染的神经渲染方法。本章使用深度全连接网络对每个着色点到其全局光照的复杂映射关系进行建模。本章提出的神经网络友好的输入表达在降低神经网络规模方面发挥重要的作用，而紧凑的神经网络表达一方面可以提升运行时的渲染效率，另一方面可以降低存储开销。位置编码技术和屏幕空间神经贴图则有效提升了本章方法捕捉高频且复杂的全局光照效果的能力。本章方法仅以直接光照和若干屏幕空间贴图为输入而无需任何复杂数据结构或预计算过程，使其非常容易集成到现有的实时渲染管线当中。此外，本章提出一种基于材质划分的加速方案，以多个小网络替换单个大网络，有效降低了神经网络的整体参数量，从而实现了运行时加速和存储代价降低两方面的作用。

未来工作方面，我们将进一步探索在训练数据生成过程中如何自适应且高效地进行视角和光源采样。此外，扩展本章方法以支持如透明物体、半透明物体等复杂材质或完全动态的场景也是非常有价值的研究方向。

第 6 章　总结与展望

本文围绕神经渲染中采集和建模、存储和表达以及绘制和可视化这一主线进行研究，提出了基于逆渲染和数据驱动的表观建模、基于深度场景表达的重光照以及基于深度绘制管线的全局光照绘制等方法，通过将深度学习引入到传统真实感渲染问题中实现了更加轻量化、高质量且高效率的算法方案。本章首先概括总结本文提出的多个神经渲染方法，随后介绍并分析神经渲染领域未来的重要研究方向。

6.1　研究工作总结

在各行各业对三维内容的需求日益增长的今天，真实感渲染作为计算机图形学中一直以来的核心研究领域也迎来了新的发展契机和方向。神经渲染，即基于深度学习的真实感渲染，不仅可以有效解决传统真实感渲染方法存在的问题，也可以拓宽其使用场景。神经渲染的研究方向同真实感渲染一样可以分为采集和建模、存储和表达以及绘制和可视化三个领域，本文针对这些领域所存在的问题和挑战展开研究并提出一系列算法。

采集和建模是计算机图形学方法的基础，其中针对复杂表观属性的采集和建模在真实感渲染当中尤为关键。传统的轻量化表观建模方法均存在明显不足：经典逆渲染方法依赖于大量输入图片或对材质的强假设，而基于单张图片的深度学习方法则存在重建质量不够高和无法轻易扩展到更多图片等缺陷。本文第 3 章提出一种支持从任意数量输入图片中进行平面物体表观建模的统一框架。该方法的核心思路是在基于深度学习构建的表观数据隐空间中进行逆渲染优化，表观数据隐空间可以充分利用表观数据中蕴含的先验信息来为优化过程提供先验，使得整个优化过程中无需任何手工设计的启发式正则约束。为构建适合于逆渲染优化的表观数据隐空间，本文第 3 章提出了解码器网络中去掉批归一化层、综合考虑贴图损失函数和渲染损失函数、训练过程增加隐空间光滑性约束以及利用基于单张图片的前人工作进行初始化等一系列策略。此外，该方法所使用的结合深度逆渲染策略和细节增强策略的思路也可以应用于基于单张图片的人脸高频纹理重建中。实验表明，本文第 3 章所述方法在给定单张或少量输入图片的情况下，可以给出可信且优于前人方法的重建结果，而随着输入图片数量的增加，其重建质量不断提升最终可以收敛到准确的结果。此外，该方法支持任意分辨率的输入图片，在无需任何重新训练的情况下可以实现高分辨率下的高质量表观建模。

当面对真实世界中非常复杂的场景时，采集和建模往往难以成功或其结果存在明显瑕疵，针对这种场景，基于图像的绘制方法是另一种常用的数字化重渲染方法。本文第 4 章提出一种以轻量化采集设备所拍摄的无结构输入图片作为输入支持 360 度自由视点重渲染的方法。该方法属于基于图像的绘制方法，因此不依赖于精确的场景重建，可以直接利用输入图片中蕴含的表观细节来合成全新视角和光照下的渲染结果。本文第 4 章核心贡献是提出了深度场景表达和神经渲染管线，其中深度场景表达包括定义在粗糙几何上的神经纹理和辐射亮度信息，二者相乘后输入到神经渲染网络中完成最终渲染图片的生成。该方法提出了一种基于视角的空间划分策略以在扩大网络规模的同时保证不超出显存限制。此外，该方法提出一种光照增强策略以扩展其所支持的光源类型。在多个复杂场景中的实验表明本文第 4 章所述方法可以实现高质量的 360 度自由视点重光照渲染。

基于已知三维场景信息的真实感渲染是另一个重要的研究问题，在该问题中不需要像基于图像的绘制方法一样利用低维观测来隐式推理高维场景，而是关注于给定高维场景信息后如何快速且高质量的实现全局光照渲染。本文第 5 章提出一种支持动态面光源下全频率全局光照快速绘制的深度绘制管线。该方法的核心思路是使用深度全连接网络对着色点到其全局光照的复杂映射关系进行建模，训练数据的渲染和神经网络的训练过程可视为场景反射场的采样和存储过程，运行时神经网络的推理可视为对预计算好的场景反射场进行查询和插值。为在保证网络规模紧凑的前提下实现高质量的学习，该方法提出使用神经网络友好的输入表达对输入信息建模、使用位置编码技术进行信息编码以及使用屏幕空间神经贴图进行屏幕内信息共享等策略。此外，该方法提出一种基于材质划分的加速方案，可降低其存储代价并提升运行时渲染效率。实验表明，本文第 5 章所述方法可以支持包括高光互反射、焦散、色溢和镜面反射在内的多种复杂全局光照效果的快速渲染。

总之，本文在采集和建模、存储和表达以及绘制和可视化等三个领域展开研究并提出多种神经渲染算法，通过将深度表达和神经网络技术以合适的方式同真实感渲染问题相结合，实现了降低采集代价、提升结果质量、提升运行效率以及扩展应用场景等目标。

6.2　未来工作展望

近年来，神经渲染方法被广泛应用于各类计算机图形学的任务当中，展示出极大的发展潜力。尽管深度学习是一种通用的数据驱动技术并在计算机视觉等其他领域大获成功，然而将其与计算机图形学问题进行结合时仍然依赖于研究人员

对于计算机图形学领域知识和具体问题的深入了解和思考。综合考虑技术发展和业界需求，本文认为神经渲染领域有如下几个重要且富有挑战的研究方向：

轻量化全场景建模　轻量化的几何和表观协同建模是未来的重要研究方向，其研究意义体现在：一方面，轻量化采集条件可以大大降低相关数据的采集成本；另一方面，自动化采集流程有助于构建大规模且高质量的三维数字资产库，可进一步应用于计算机视觉领域或虚拟现实、元宇宙等场景中。几何和表观协同建模是极富挑战的研究问题，基于二维图片观测进行高维几何和表观数据重建的关键思路或技术可能包含：1. 近年来蓬勃发展的基于物理的可微分渲染技术，可微分渲染可以建立图片空间梯度到任意场景三维参数梯度之间的桥梁，并且基于物理的可微分渲染考虑了完整的光传输过程使得正向渲染本身的表达能力大大提高，从而有助于逆渲染过程的解耦。2. 结合深度学习提供的数据先验有助于缓解该欠约束优化问题所存在的二义性难题。本文第 3 章提出在自编码器网络的隐空间执行逆渲染优化的思路可以提升表观建模过程的质量和鲁棒性，类似的思路和方法可以应用于几何和表观的协同建模当中。

统一可编辑材质表达　真实感渲染中表观数据通常采用 BRDF 进行建模，常见的 BRDF 模型可以分为解析 BRDF 模型和测量 BRDF 模型两类。其中解析 BRDF 模型具有参数意义明确、易于编辑、计算代价低廉等优势，然而受限于模型本身表达力限制无法表达很多真实的复杂材质；测量 BRDF 模型则具有更强的表达能力，但具有无法编辑且依赖于高昂的存储和计算代价等不足。探索一种可以兼具二者优势的统一材质表达是很有价值的研究方向，一方面，该表达需要可以涵盖真实世界中许多复杂材质；另一方面，该表达需要提供易于直接编辑的材质参数或其他修改方式。通过深度学习技术将解析 BRDF 模型和测量 BRDF 投影到公共隐空间并借助解析 BRDF 的编辑性来扩充测量 BRDF 的编辑性是一种富有潜力的研究方向。

实时全局光照渲染　实时全局光照渲染是真实感渲染领域一直以来的重要研究方向。长久以来，真实感渲染方法或是着眼于运行效率或是追求高真实感，难以在质量和效率之间取得平衡。近年来，随着基于 GPU 光线跟踪效率的不断提升，实时全局光照渲染逐渐成为热点。一些富有价值的研究方向包括：1. 综合考虑光线跟踪算法中蒙特卡洛采样步骤和后处理降噪步骤，当前的大多数降噪算法是一种独立的后处理步骤，实际上采样和降噪息息相关，某些难以有效降噪的区域可以通过提高采样数来提升质量，反过来，某些依赖于大采样数的噪声区域实际上可能

易于被降噪算法消除，因此综合考虑采样过程和降噪算法是提升效率和质量的重要方向；2. 结合场景知识的降噪算法研究，针对一般性场景的通用降噪算法一般存在明显的质量瓶颈，难以对各类几何和材质均实现高质量且鲁棒的降噪。因此，结合具体场景知识的降噪算法在特定领域和任务中具有重要作用，例如针对高逼真数字人的渲染任务，由于人脸本身具有很强的数据先验，因此可以借助大量人脸数据中蕴含的一致性特征来探索更加适合于人脸的降噪算法；3. 利用神经网络代替传统方法中的预计算步骤或信息缓存表达。本文第 5 章采用了神经网络作为光传输过程预计算表达，可以同时实现高质量渲染和低存储开销。实际上，基于预计算或基于复杂数据结构进行信息缓存的渲染方法在计算机图形学中广泛存在，而神经网络作为一种更加紧凑的数据驱动表达可以有效地结合到该类方法当中。

动态场景数字化渲染　基于图像的绘制方法一直是数字化重渲染领域的重点。目前针对静态场景在动态视角和动态光源下的重渲染是研究的重点方向，例如本文第 4 章所提出的同时支持自由视角和动态光源的神经渲染管线即是典型代表。展望未来，数字化渲染有如下几点前沿探索方向：1. 轻量化采集，本文第 4 章虽然仅依赖于两个手持相机这样轻量化的采集设备，但仍然需要大量观测图片作为输入，探索更加轻量化的数字化渲染方法是可行且有价值的研究方向。2. 可编辑数字化渲染，基于图像的绘制方法往往难以编辑和调整，限制了该类方法的应用场景，因此未来可以继续探索具有材质编辑和几何编辑（包括几何变换、删除或添加等操作）能力的数字化渲染方法。支持完全动态场景的数字化渲染方法则是更加富有挑战但同时也更加重要的研究问题。3. 神经渲染用于多模态内容理解和生成，深度学习方法是文字和语音以及图片和视频等领域的主流方法，神经渲染方法由于与这些方法具有相似的架构和底层模块，因而适合于与之结合进行多模态信息的整合和处理。

参考文献

[1] Heitz E. Sampling the GGX distribution of visible normals[J/OL]. Journal of Computer Graphics Techniques (JCGT), 2018, 7(4): 1-13[2022-02-16]. http://jcgt.org/published/0007/04/01/.

[2] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 586-595[2022-02-16]. https://doi.org/10.1109/CVPR.2018.00068.

[3] MPC Moving Picture Company. The Lion King[EB/OL]. 2019[2022-02-16]. https://www.mpcfilm.com/film/the-lion-king.

[4] 游科互动科技有限公司. 黑神话：悟空[EB/OL]. 2021[2022-02-16]. https://heishenhua.com/img/screenshot/blackmyth_wukong_screenshot_031.jpg.

[5] Mitra N J, Pauly M. Shadow art[J/OL]. ACM Transactions on Graphics (TOG), 2009, 28(5): 1-7[2022-02-16]. https://doi.org/10.1145/1618452.1618502.

[6] Meta. Introducing horizon workrooms: Remote collaboration reimagined[N/OL]. 2021-08-19[2022-02-16]. https://about.fb.com/news/2021/08/introducing-horizon-workrooms-remote-collaboration-reimagined/.

[7] Musée du Louvre. The Mona Lisa in virtual reality in your own home[EB/OL]. (2021-02-23)[2022-02-16]. https://www.louvre.fr/en/what-s-on/life-at-the-museum/the-mona-lisa-in-virtual-reality-in-your-own-home.

[8] ZAMAK design. Developing innovative design concepts faster with SOLIDWORKS® Visualize[EB/OL]. (2017-02-22)[2022-02-16]. https://blogs.solidworks.com/solidworksblog/wp-content/uploads/sites/2/2017/02/chambre.jpg.

[9] Chandrasekhar S. Radiative transfer[M]. Dover Publications, 1960.

[10] Kajiya J T. The rendering equation[C/OL]//Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 1986: 143-150[2022-02-16]. https://doi.org/10.1145/15922.15902.

[11] Tewari A, Fried O, Thies J, et al. State of the art on neural rendering[J/OL]. Computer Graphics Forum (CGF), 2020, 39(2): 701-727[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14022.

[12] 赵烨梓, 王璐, 徐延宁, 等. 基于机器学习的三维场景高度真实感绘制方法综述[J/OL]. 软件学报, 2022, 33(01): 356-376. https://doi.org/10.13328/j.cnki.jos.006334.

[13] Weinmann M, Klein R. Advances in geometry and reflectance acquisition (course notes)[C/OL]//SIGGRAPH Asia 2015 Courses. 2015: 1:1-1:71[2022-02-16]. https://doi.org/10.1145/2818143.2818165.

[14] Dorsey J, Rushmeier H, Sillion F. Digital modeling of material appearance[M]. Morgan Kaufmann, 2007.

[15] Palma G, Callieri M, Dellepiane M, et al. A statistical method for SVBRDF approximation from video sequences in general lighting conditions[J/OL]. Computer Graphics Forum (CGF), 2012, 31(4): 1491-1500[2022-02-16]. 10.1111/j.1467-8659.2012.03145.x.

[16] Dong Y, Chen G, Peers P, et al. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting[J/OL]. ACM Transactions on Graphics (TOG), 2014, 33 (6): 193:1-193:12[2022-02-16]. https://doi.org/10.1145/2661229.2661283.

[17] Riviere J, Peers P, Ghosh A. Mobile surface reflectometry[J/OL]. Computer Graphics Forum (CGF), 2016, 35(1): 191-202[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12719.

[18] Hui Z, Sunkavalli K, Lee J Y, et al. Reflectance capture using univariate sampling of BRDFs [C/OL]//2017 IEEE International Conference on Computer Vision (ICCV). 2017: 5372-5380 [2022-02-16]. https://doi.org/10.1109/ICCV.2017.573.

[19] Aittala M, Weyrich T, Lehtinen J. Two-shot SVBRDF capture for stationary materials[J/OL]. ACM Transactions on Graphics (TOG), 2015, 34(4): 110:1-110:13[2022-02-16]. https://doi.org/10.1145/2766967.

[20] 冯洁, 李博, 周秉锋. 基于像素聚类的空间变化表面材质建模[J]. 图学学报, 2021, 42(01): 94-100.

[21] Xu Z, Nielsen J B, Yu J, et al. Minimal BRDF sampling for two-shot near-field reflectance acquisition[J/OL]. ACM Transactions on Graphics (TOG), 2016, 35(6): 188:1-188:12[2022-02-16]. https://doi.org/10.1145/2980179.2982396.

[22] Zhou Z, Chen G, Dong Y, et al. Sparse-as-possible SVBRDF acquisition[J/OL]. ACM Transactions on Graphics (TOG), 2016, 35(6): 189:1-189:12[2022-02-16]. https://doi.org/10.1145/2980179.2980247.

[23] Li X, Dong Y, Peers P, et al. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks[J/OL]. ACM Transactions on Graphics (TOG), 2017, 36(4): 45:1-45:11[2022-02-16]. https://doi.org/10.1145/3072959.3073641.

[24] Ye W, Li X, Dong Y, et al. Single image surface appearance modeling with self-augmented CNNs and inexact supervision[J/OL]. Computer Graphics Forum (CGF), 2018, 37(7): 201-211 [2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13560.

[25] Deschaintre V, Aittala M, Durand F, et al. Single-image SVBRDF capture with a rendering-aware deep network[J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(4): 128:1-128:15 [2022-02-16]. https://doi.org/10.1145/3197517.3201378.

[26] Li Z, Sunkavalli K, Chandraker M. Materials for masses: SVBRDF acquisition with a single mobile phone image[C/OL]//Proceedings of the 15th European Conference on Computer Vision (ECCV). 2018: 74-90[2022-02-16]. https://doi.org/10.1007/978-3-030-01219-9_5.

[27] Deschaintre V, Aittala M, Durand F, et al. Flexible SVBRDF capture with a multi-image deep network[J/OL]. Computer Graphics Forum (CGF), 2019, 38(4): 1-13[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13765.

[28] Woodham R J. Photometric method for determining surface orientation from multiple images [J/OL]. Optical Engineering, 1980, 19(1): 139-144[2022-02-16]. https://doi.org/10.1117/12.7972479.

[29] Seitz S M, Curless B, Diebel J, et al. A comparison and evaluation of multi-view stereo reconstruction algorithms[C/OL]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). 2006: 519-528[2022-02-16]. https://doi.org/10.1109/CVPR.2006.19.

[30] Witkin A P. Shape from contour[R]. MIT Computer Science & Artificial Intelligence Laboratory, 1980.

[31] Brady M, Yuille A. An extremum principle for shape from contour[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 1984, PAMI-6(3): 288-301[2022-02-16]. https://doi.org/10.1109/TPAMI.1984.4767521.

[32] Laurentini A. The visual hull for understanding shapes from contours: a survey[C/OL]//Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings. 2003: 25-28[2022-02-16]. https://doi.org/10.1109/ISSPA.2003.1224631.

[33] Zhang R, Tsai P S, Cryer J E, et al. Shape-from-shading: a survey[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690-706[2022-02-16]. https://doi.org/10.1109/34.784284.

[34] Schönberger J L, Frahm J M. Structure-from-motion revisited[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 4104-4113[2022-02-16]. https://doi.org/10.1109/CVPR.2016.445.

[35] Chen K, Lai Y K, Wu Y X, et al. Automatic semantic modeling of indoor scenes from low-quality RGB-D data using contextual information[J/OL]. ACM Transactions on Graphics (TOG), 2014, 33(6): 208:1-208:12[2022-02-16]. https://doi.org/10.1145/2661229.2661239.

[36] Xu K, Chen K, Fu H, et al. Sketch2Scene: Sketch-based co-retrieval and co-placement of 3D models[J/OL]. ACM Transactions on Graphics (TOG), 2013, 32(4): 123:1-123:15[2022-02-16]. https://doi.org/10.1145/2461912.2461968.

[37] Wu Z, Song S, Khosla A, et al. 3D ShapeNets: A deep representation for volumetric shapes [C/OL]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 1912-1920[2022-02-16]. https://doi.org/10.1109/CVPR.2015.7298801.

[38] Chang A X, Funkhouser T, Guibas L, et al. ShapeNet: An information-rich 3D model repository [J/OL]. CoRR, 2015, abs/1512.03012[2022-02-16]. http://arxiv.org/abs/1512.03012.

[39] Fan H, Su H, Guibas L. A point set generation network for 3D object reconstruction from a single image[C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2463-2471[2022-02-16]. https://doi.org/10.1109/CVPR.2017.264.

[40] Groueix T, Fisher M, Kim V G, et al. A papier-mâché approach to learning 3D surface generation [C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 216-224[2022-02-16]. https://doi.org/10.1109/CVPR.2018.00030.

[41] Choy C B, Xu D, Gwak J, et al. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction[C/OL]//Proceedings of the 14th European Conference on Computer Vision (ECCV). 2016: 628-644[2022-02-16]. https://doi.org/10.1007/978-3-319-46484-8_38.

[42] Tulsiani S, Zhou T, Efros A A, et al. Multi-view supervision for single-view reconstruction via differentiable ray consistency[C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 209-217[2022-02-16]. https://doi.org/10.1109/CVPR.2017.30.

[43] Wu J, Wang Y, Xue T, et al. MarrNet: 3D shape reconstruction via 2.5D sketches[C/OL]// Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017: 540-550[2022-02-16]. https://proceedings.neurips.cc/paper/2017/file/ad972f10e0800b49d76fed33a21f6698-Paper.pdf.

[44] Kanazawa A, Tulsiani S, Efros A A, et al. Learning category-specific mesh reconstruction from image collections[C/OL]//Proceedings of the 15th European Conference on Computer Vision (ECCV). 2018: 386-402[2022-02-16]. https://doi.org/10.1007/978-3-030-01267-0_23.

[45] Holroyd M, Lawrence J, Zickler T. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance[J/OL]. ACM Transactions on Graphics (TOG), 2010, 29 (4): 99:1-99:12[2022-02-16]. https://doi.org/10.1145/1778765.1778836.

[46] Xia R, Dong Y, Peers P, et al. Recovering shape and spatially-varying surface reflectance under unknown illumination[J/OL]. ACM Transactions on Graphics (TOG), 2016, 35(6): 187:1-187:12[2022-02-16]. https://doi.org/10.1145/2980179.2980248.

[47] Nam G, Lee J H, Gutierrez D, et al. Practical SVBRDF acquisition of 3D objects with unstructured flash photography[J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(6): 267:1-267:12[2022-02-16]. https://doi.org/10.1145/3272127.3275017.

[48] Li Z, Xu Z, Ramamoorthi R, et al. Learning to reconstruct shape and spatially-varying reflectance from a single image[J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(6): 269:1-269:11[2022-02-16]. https://doi.org/10.1145/3272127.3275055.

[49] Kang K, Xie C, He C, et al. Learning efficient illumination multiplexing for joint capture of reflectance and shape[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(6): 165:1-165:12[2022-02-16]. https://doi.org/10.1145/3355089.3356492.

[50] Bi S, Xu Z, Sunkavalli K, et al. Deep 3D capture: Geometry and reflectance from sparse multi-view images[C/OL]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 5959-5968[2022-02-16]. https://doi.org/10.1109/CVPR42600.2020.00600.

[51] Jimenez Rezende D, Eslami S M A, Mohamed S, et al. Unsupervised learning of 3D structure from images[C/OL]//Advances in Neural Information Processing Systems 29 (NIPS 2016). 2016: 5003-5011[2022-02-16]. https://proceedings.neurips.cc/paper/2016/file/1d94108e907bb8311d8802b48fd54b4a-Paper.pdf.

[52] Wang P S, Liu Y, Guo Y X, et al. O-CNN: Octree-based convolutional neural networks for 3D shape analysis[J/OL]. ACM Transactions on Graphics (TOG), 2017, 36(4): 72:1-72:11 [2022-02-16]. https://doi.org/10.1145/3072959.3073608.

[53] Henzler P, Mitra N, Ritschel T. Escaping plato's cave: 3D shape from adversarial rendering [C/OL]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 9983-9992[2022-02-16]. https://doi.org/10.1109/ICCV.2019.01008.

[54] Xu Z, Bi S, Sunkavalli K, et al. Deep view synthesis from sparse photometric images[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 76:1-76:13[2022-02-16]. https://doi.org/10.1145/3306346.3323007.

[55] Thies J, Zollhöfer M, Nießner M. Deferred neural rendering: Image synthesis using neural textures[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 66:1-66:12[2022-02-16]. https://doi.org/10.1145/3306346.3323035.

[56] Park J J, Florence P, Straub J, et al. DeepSDF: Learning continuous signed distance functions for shape representation[C/OL]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 165-174[2022-02-16]. https://doi.org/10.1109/CVPR.2019.00025.

[57] Mescheder L, Oechsle M, Niemeyer M, et al. Occupancy networks: Learning 3D reconstruction in function space[C/OL]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 4455-4465[2022-02-16]. https://doi.org/10.1109/CVPR.2019.00459.

[58] Sitzmann V, Zollhöfer M, Wetzstein G. Scene representation networks: Continuous 3D-structure-aware neural scene representations[C/OL]//Advances in Neural Information Processing Systems 32 (NeurIPS 2019). 2019: 1121-1132[2022-02-16]. https://proceedings.neurips.cc/paper/2019/file/b5dc4e5d9b495d0196f61d45b26ef33e-Paper.pdf.

[59] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: Representing scenes as neural radiance fields for view synthesis[C/OL]//Proceedings of the 16th European Conference on Computer Vision (ECCV). 2020: 405-421[2022-02-16]. https://doi.org/10.1007/978-3-030-58452-8_24.

[60] Granskog J, Rousselle F, Papas M, et al. Compositional neural scene representations for shading inference[J/OL]. ACM Transactions on Graphics (TOG), 2020, 39(4): 135:1-135:13[2022-02-16]. https://doi.org/10.1145/3386569.3392475.

[61] Ren P, Wang J, Gong M, et al. Global illumination with radiance regression functions[J/OL]. ACM Transactions on Graphics (TOG), 2013, 32(4): 130:1-130:12[2022-02-16]. https://doi.org/10.1145/2461912.2462009.

[62] Sun T, Barron J T, Tsai Y T, et al. Single image portrait relighting[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 79:1-79:12[2022-02-16]. https://doi.org/10.1145/3306346.3323008.

[63] Rainer G, Jakob W, Ghosh A, et al. Neural BTF compression and interpolation[J/OL]. Computer Graphics Forum (CGF), 2019, 38(2): 235-244[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13633.

[64] Rainer G, Ghosh A, Jakob W, et al. Unified neural encoding of BTFs[J/OL]. Computer Graphics Forum (CGF), 2020, 39(2): 167-178[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13921.

[65] Kuznetsov A, Mullia K, Xu Z, et al. NeuMIP: Multi-resolution neural materials[J/OL]. ACM Transactions on Graphics (TOG), 2021, 40(4): 175:1-175:13[2022-02-16]. https://doi.org/10.1145/3450626.3459795.

[66] Matusik W. A data-driven reflectance model[D]. Massachusetts Institute of Technology, 2003.

[67] Filip J, Vávra R. Template-based sampling of anisotropic BRDFs[J/OL]. Computer Graphics Forum (CGF), 2014, 33(7): 91-99[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12477.

[68] Dupuy J, Jakob W. An adaptive parameterization for efficient material acquisition and rendering [J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(6): 274:1-274:14[2022-02-16]. https://doi.org/10.1145/3272127.3275059.

[69] Sun T, Jensen H W, Ramamoorthi R. Connecting measured BRDFs to analytic BRDFs by data-driven diffuse-specular separation[J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(6): 273:1-273:15[2022-02-16]. https://doi.org/10.1145/3272127.3275026.

[70] Hu B, Guo J, Chen Y, et al. DeepBRDF: A deep representation for manipulating measured BRDF[J/OL]. Computer Graphics Forum (CGF), 2020, 39(2): 157-166[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13920.

[71] Zheng C, Zheng R, Wang R, et al. A compact representation of measured BRDFs using neural processes[J/OL]. ACM Transactions on Graphics (TOG), 2021, 41(2): 14:1-14:15[2022-02-16]. https://doi.org/10.1145/3490385.

[72] Levoy M, Hanrahan P. Light field rendering[C/OL]//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 1996: 31-42[2022-02-16]. https://doi.org/10.1145/237170.237199.

[73] Gortler S J, Grzeszczuk R, Szeliski R, et al. The lumigraph[C/OL]//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 1996: 43-54[2022-02-16]. https://doi.org/10.1145/237170.237200.

[74] Buehler C, Bosse M, McMillan L, et al. Unstructured lumigraph rendering[C/OL]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 2001: 425-432[2022-02-16]. https://doi.org/10.1145/383259.383309.

[75] Chaurasia G, Duchene S, Sorkine-Hornung O, et al. Depth synthesis and local warps for plausible image-based navigation[J/OL]. ACM Transactions on Graphics (TOG), 2013, 32(3): 30:1-30:12[2022-02-16]. https://doi.org/10.1145/2487228.2487238.

[76] Hedman P, Ritschel T, Drettakis G, et al. Scalable inside-out image-based rendering[J/OL]. ACM Transactions on Graphics (TOG), 2016, 35(6): 231:1-231:11[2022-02-16]. https://doi.org/10.1145/2980179.2982420.

[77] Penner E, Zhang L. Soft 3D reconstruction for view synthesis[J/OL]. ACM Transactions on Graphics (TOG), 2017, 36(6): 235:1-235:11[2022-02-16]. https://doi.org/10.1145/3130800.3130855.

[78] Hedman P, Philip J, Price T, et al. Deep blending for free-viewpoint image-based rendering [J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(6): 257:1-257:15[2022-02-16]. https://doi.org/10.1145/3272127.3275084.

[79] Wood D N, Azuma D I, Aldinger K, et al. Surface light fields for 3D photography[C/OL]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 2000: 287-296[2022-02-16]. https://doi.org/10.1145/344779.344925.

[80] Chen A, Wu M, Zhang Y, et al. Deep surface light fields[J/OL]. Proceedings of the ACM on Computer Graphics and Interactive Techniques, 2018, 1(1): 14:1-14:17[2022-02-16]. https://doi.org/10.1145/3203192.

[81] Sitzmann V, Thies J, Heide F, et al. DeepVoxels: Learning persistent 3D feature embeddings [C/OL]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 2432-2441[2022-02-16]. https://doi.org/10.1109/CVPR.2019.00254.

[82] Zhu J Y, Zhang Z, Zhang C, et al. Visual object networks: Image generation with disentangled 3D representations[C/OL]//Advances in Neural Information Processing Systems 31 (NeurIPS 2018). 2018: 118-129[2022-02-16]. https://proceedings.neurips.cc/paper/2018/file/92cc227532d17e56e07902b254dfad10-Paper.pdf.

[83] Nguyen-Phuoc T H, Li C, Balaban S, et al. RenderNet: A deep convolutional network for differentiable rendering from 3D shapes[C/OL]//Advances in Neural Information Processing Systems 31 (NeurIPS 2018). 2018: 7902-7912[2022-02-16]. https://proceedings.neurips.cc/paper/2018/file/68d3743587f71fbaa5062152985aff40-Paper.pdf.

[84] Nguyen-Phuoc T, Li C, Theis L, et al. HoloGAN: Unsupervised learning of 3D representations from natural images[C/OL]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 7587-7596[2022-02-16]. https://doi.org/10.1109/ICCV.2019.00768.

[85] Lombardi S, Simon T, Saragih J, et al. Neural volumes: Learning dynamic renderable volumes from images[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 65:1-65:14[2022-02-16]. https://doi.org/10.1145/3306346.3323020.

[86] Zhou T, Tucker R, Flynn J, et al. Stereo magnification: Learning view synthesis using multiplane images[J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(4): 65:1-65:12[2022-02-16]. https://doi.org/10.1145/3197517.3201323.

[87] Mildenhall B, Srinivasan P P, Ortiz-Cayon R, et al. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 29:1-29:14[2022-02-16]. https://doi.org/10.1145/3306346.3322980.

[88] Flynn J, Broxton M, Debevec P, et al. DeepView: View synthesis with learned gradient descent [C/OL]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 2362-2371[2022-02-16]. https://doi.org/10.1109/CVPR.2019.00247.

[89] Saito S, Huang Z, Natsume R, et al. PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization[C/OL]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 2304-2314[2022-02-16]. https://doi.org/10.1109/ICCV.2019.00239.

[90] 常远, 盖孟. 基于神经辐射场的视点合成算法综述[J]. 图学学报, 2021, 39(03): 376-384.

[91] Zhou T, Tulsiani S, Sun W, et al. View synthesis by appearance flow[C/OL]//Proceedings of the 14th European Conference on Computer Vision (ECCV). 2016: 286-301[2022-02-16]. https://doi.org/10.1007/978-3-319-46493-0_18.

[92] Jin S, Liu R, Ji Y, et al. Learning to dodge a bullet: Concyclic view morphing via deep learning [C/OL]//Proceedings of the 15th European Conference on Computer Vision (ECCV). 2018: 230-246[2022-02-16]. https://doi.org/10.1007/978-3-030-01264-9_14.

[93] Park E, Yang J, Yumer E, et al. Transformation-grounded image generation network for novel 3D view synthesis[C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 702-711[2022-02-16]. https://doi.org/10.1109/CVPR.2017.82.

[94] Liu M, He X, Salzmann M. Geometry-aware deep network for single-image novel view synthesis[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 4616-4624[2022-02-16]. https://doi.org/10.1109/CVPR.2018.00485.

[95] Sun S H, Huh M, Liao Y H, et al. Multi-view to novel view: Synthesizing novel views with self-learned confidence[C/OL]//Proceedings of the 15th European Conference on Computer Vision (ECCV). 2018: 162-178[2022-02-16]. https://doi.org/10.1007/978-3-030-01219-9_10.

[96] Kalantari N K, Wang T C, Ramamoorthi R. Learning-based view synthesis for light field cameras[J/OL]. ACM Transactions on Graphics (TOG), 2016, 35(6): 193:1-193:10[2022-02-16]. https://doi.org/10.1145/2980179.2980251.

[97] Srinivasan P P, Wang T, Sreelal A, et al. Learning to synthesize a 4D RGBD light field from a single image[C/OL]//2017 IEEE International Conference on Computer Vision (ICCV). 2017: 2262-2270[2022-02-16]. https://doi.org/10.1109/ICCV.2017.246.

[98] Srinivasan P P, Tucker R, Barron J T, et al. Pushing the boundaries of view extrapolation with multiplane images[C/OL]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 175-184[2022-02-16]. https://doi.org/10.1109/CVPR.2019.00026.

[99] Yang J, Reed S E, Yang M H, et al. Weakly-supervised disentangling with recurrent transformations for 3D view synthesis[C/OL]//Advances in Neural Information Processing Systems 28 (NIPS 2015). 2015: 1099-1107[2022-02-16]. https://proceedings.neurips.cc/paper/2015/file/109a0ca3bc27f3e96597370d5c8cf03d-Paper.pdf.

[100] Yan X, Yang J, Yumer E, et al. Perspective transformer nets: Learning single-view 3D object reconstruction without 3D supervision[C/OL]//Advances in Neural Information Processing Systems 29 (NIPS 2016). 2016: 1704-1712[2022-02-16]. https://proceedings.neurips.cc/paper/2016/file/e820a45f1dfc7b95282d10b6087e11c0-Paper.pdf.

[101] Ji D, Kwon J, McFarland M, et al. Deep view morphing[C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 7092-7100[2022-02-16]. https://doi.org/10.1109/CVPR.2017.750.

[102] Olszewski K, Tulyakov S, Woodford O, et al. Transformable bottleneck networks[C/OL]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 7647-7656[2022-02-16]. https://doi.org/10.1109/ICCV.2019.00774.

[103] Debevec P, Hawkins T, Tchou C, et al. Acquiring the reflectance field of a human face[C/OL]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 2000: 145-156[2022-02-16]. https://doi.org/10.1145/344779.344855.

[104] Furukawa R, Kawasaki H, Ikeuchi K, et al. Appearance based object modeling using texture database: Acquisition, compression and rendering[C/OL]//Proceedings of the 13th Eurographics Workshop on Rendering. 2002: 257-266[2022-02-16]. https://doi.org/10.5555/581896.581929.

[105] Peers P, Mahajan D K, Lamond B, et al. Compressive light transport sensing[J/OL]. ACM Transactions on Graphics (TOG), 2009, 28(1): 3:1-3:18[2022-02-16]. https://doi.org/10.1145/1477926.1477929.

[106] Malzbender T, Gelb D, Wolters H. Polynomial texture maps[C/OL]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 2001: 519-528[2022-02-16]. https://doi.org/10.1145/383259.383320.

[107] Li G, Wu C, Stoll C, et al. Capturing relightable human performances under general uncontrolled illumination[J/OL]. Computer Graphics Forum (CGF), 2013, 32(2pt3): 275-284[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12047.

[108] Guo K, Lincoln P, Davidson P, et al. The relightables: Volumetric performance capture of humans with realistic relighting[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(6): 217:1-217:19[2022-02-16]. https://doi.org/10.1145/3355089.3356571.

[109] Haber T, Fuchs C, Bekaer P, et al. Relighting objects from image collections[C/OL]//2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2009: 627-634[2022-02-16]. https://doi.org/10.1109/CVPR.2009.5206753.

[110] Hauagge D, Wehrwein S, Upchurch P, et al. Reasoning about photo collections using models of outdoor illumination[C/OL]//Proceedings of the British Machine Vision Conference (BMVC). 2014: 1-12[2022-02-16]. https://doi.org/10.5244/C.28.78.

[111] Imber J, Guillemaut J Y, Hilton A. Intrinsic textures for relightable free-viewpoint video [C/OL]//Proceedings of the 13th European Conference on Computer Vision (ECCV). 2014: 392-407[2022-02-16]. https://doi.org/10.1007/978-3-319-10605-2_26.

[112] Meshry M, Goldman D B, Khamis S, et al. Neural rerendering in the wild[C/OL]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 6871-6880[2022-02-16]. https://doi.org/10.1109/CVPR.2019.00704.

[113] Chen Z, Chen A, Zhang G, et al. A neural rendering framework for free-viewpoint relighting [C/OL]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 5598-5609[2022-02-16]. https://doi.org/10.1109/CVPR42600.2020.00564.

[114] Philip J, Gharbi M, Zhou T, et al. Multi-view relighting using a geometry-aware network[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 78:1-78:14[2022-02-16]. https://doi.org/10.1145/3306346.3323013.

[115] Kanamori Y, Endo Y. Relighting humans: Occlusion-aware inverse rendering for full-body human images[J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(6): 270:1-270:11[2022-02-16]. https://doi.org/10.1145/3272127.3275104.

[116] Xu Z, Sunkavalli K, Hadap S, et al. Deep image-based relighting from optimal sparse samples [J/OL]. ACM Transactions on Graphics (TOG), 2018, 37(4): 126:1-126:13[2022-02-16]. https://doi.org/10.1145/3197517.3201313.

[117] Ren P, Dong Y, Lin S, et al. Image based relighting using neural networks[J/OL]. ACM Transactions on Graphics (TOG), 2015, 34(4): 111:1-111:12[2022-02-16]. https://doi.org/10.1145/2766899.

[118] Meka A, Häne C, Pandey R, et al. Deep reflectance fields: High-quality facial reflectance field inference from color gradient illumination[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 77:1-77:12[2022-02-16]. https://doi.org/10.1145/3306346.3323027.

[119] Bi S, Xu Z, Sunkavalli K, et al. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images[C/OL]//Proceedings of the 16th European Conference on Computer Vision (ECCV). 2020: 294-311[2022-02-16]. https://doi.org/10.1007/978-3-030-58580-8_18.

[120] Bi S, Xu Z, Srinivasan P, et al. Neural reflectance fields for appearance acquisition[J/OL]. CoRR, 2020, abs/2008.03824[2022-02-16]. https://arxiv.org/abs/2008.03824.

[121] Srinivasan P P, Deng B, Zhang X, et al. NeRV: Neural reflectance and visibility fields for relighting and view synthesis[C/OL]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 7491-7500[2022-02-16]. https://doi.org/10.1109/CVPR46437.2021.00741.

[122] Dachsbacher C, Stamminger M. Reflective shadow maps[C/OL]//Proceedings of the 2005 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D). 2005: 203-231 [2022-02-16]. https://doi.org/10.1145/1053427.1053460.

[123] Ritschel T, Grosch T, Seidel H P. Approximating dynamic global illumination in image space [C/OL]//Proceedings of the 2009 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D). 2009: 75-82[2022-02-16]. https://doi.org/10.1145/1507149.1507161.

[124] Robison A, Shirley P. Image space gathering[C/OL]//Proceedings of the Conference on High Performance Graphics 2009. 2009: 91-98[2022-02-16]. https://doi.org/10.1145/1572769.1572784.

[125] Nalbach O, Arabadzhiyska E, Mehta D, et al. Deep shading: Convolutional neural networks for screen space shading[J/OL]. Computer Graphics Forum (CGF), 2017, 36(4): 65-78[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13225.

[126] Xin H, Zheng S, Xu K, et al. Lightweight bilateral convolutional neural networks for interactive single-bounce diffuse indirect illumination[J/OL]. IEEE Transactions on Visualization and Computer Graphics, 2020: 1-1[2022-02-16]. https://doi.org/10.1109/TVCG.2020.3023129.

[127] Sloan P P, Kautz J, Snyder J. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments[C/OL]//Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 2002: 527-536[2022-02-16]. https://doi.org/10.1145/566570.566612.

[128] Sloan P P, Hall J, Hart J, et al. Clustered principal components for precomputed radiance transfer [J/OL]. ACM Transactions on Graphics (TOG), 2003, 22(3): 382-391[2022-02-16]. https://doi.org/10.1145/882262.882281.

[129] Ng R, Ramamoorthi R, Hanrahan P. Triple product wavelet integrals for all-frequency relighting [J/OL]. ACM Transactions on Graphics (TOG), 2004, 23(3): 477-487[2022-02-16]. https://doi.org/10.1145/1015706.1015749.

[130] Ramamoorthi R. Precomputation-based rendering[M]. NOW Publishers, 2009.

[131] Wang R, Tran J, Luebke D. All-frequency relighting of glossy objects[J/OL]. ACM Transactions on Graphics (TOG), 2006, 25(2): 293-318[2022-02-16]. https://doi.org/10.1145/1138450.1138456.

[132] Green P, Kautz J, Matusik W, et al. View-dependent precomputed light transport using nonlinear Gaussian function approximations[C/OL]//Proceedings of the 2006 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D). 2006: 7-14[2022-02-16]. https://doi.org/10.1145/1111411.1111413.

[133] Cohen M, Wallace J. Radiosity and realistic image synthesis[M]. Morgan Kaufmann, 1993.

[134] Greger G, Shirley P, Hubbard P M, et al. The irradiance volume[J/OL]. IEEE Computer Graphics and Applications, 1998, 18(2): 32-43[2022-02-16]. https://doi.org/10.1109/38.656788.

[135] McGuire M, Mara M, Nowrouzezahrai D, et al. Real-time global illumination using precomputed light field probes[C/OL]//Proceedings of the 2017 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D). 2017: 2:1-2:11[2022-02-16]. https://doi.org/10.1145/3023368.3023378.

[136] Rodriguez S, Leimkühler T, Prakash S, et al. Glossy probe reprojection for interactive global illumination[J/OL]. ACM Transactions on Graphics (TOG), 2020, 39(6): 237:1-237:16[2022-02-16]. https://doi.org/10.1145/3414685.3417823.

[137] 刘晓芸, 姚承宗, 曾晓勤. 基于 RBF 神经网络的全局光照实时绘制[J]. 计算机仿真, 2021, 39(09): 424-428.

[138] Chaitanya C R A, Kaplanyan A S, Schied C, et al. Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder[J/OL]. ACM Transactions on Graphics (TOG), 2017, 36(4): 98:1-98:12[2022-02-16]. https://doi.org/10.1145/3072959.3073601.

[139] Bako S, Vogels T, Mcwilliams B, et al. Kernel-predicting convolutional networks for denoising Monte Carlo renderings[J/OL]. ACM Transactions on Graphics (TOG), 2017, 36(4): 97:1-97:14[2022-02-16]. https://doi.org/10.1145/3072959.3073708.

[140] Xu B, Zhang J, Wang R, et al. Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(6): 224:1-224:12[2022-02-16]. https://doi.org/10.1145/3355089.3356547.

[141] 曾峥. 蒙特卡洛渲染算法的高效降噪方法[D/OL]. 山东大学, 2021. https://doi.org/10.27272/d.cnki.gshdu.2021.002596.

[142] Müller T, Mcwilliams B, Rousselle F, et al. Neural importance sampling[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(5): 145:1-145:19[2022-02-16]. https://doi.org/10.1145/3341156.

[143] Zheng Q, Zwicker M. Learning to importance sample in primary sample space[J/OL]. Computer Graphics Forum (CGF), 2019, 38(2): 169-179[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13628.

[144] Bitterli B, Wyman C, Pharr M, et al. Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting[J/OL]. ACM Transactions on Graphics (TOG), 2020, 39(4): 148:1-148:17[2022-02-16]. https://doi.org/10.1145/3386569.3392481.

[145] Müller T, Rousselle F, Novák J, et al. Real-time neural radiance caching for path tracing[J/OL]. ACM Transactions on Graphics (TOG), 2021, 40(4): 36:1-36:16[2022-02-16]. https://doi.org/10.1145/3450626.3459812.

[146] Hasselgren J, Munkberg J, Salvi M, et al. Neural temporal adaptive sampling and denoising[J/OL]. Computer Graphics Forum (CGF), 2020, 39(2): 147-155[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13919.

[147] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J/OL]. Science, 2006, 313(5786): 504-507[2022-02-16]. https://www.science.org/doi/abs/10.1126/science.1127647.

[148] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C/OL]//Proceedings of the 32nd International Conference on International Conference on Machine Learning (ICML). 2015: 448-456[2022-02-16]. https://doi.org/10.5555/3045118.3045167.

[149] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J/OL]. Journal of Machine Learning Research, 2008, 9(86): 2579-2605[2022-02-16]. http://jmlr.org/papers/v9/vandermaaten08a.html.

[150] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[C/OL]//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 1999: 187-194[2022-02-16]. https://doi.org/10.1145/311535.311556.

[151] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems[EB/OL]. 2015[2022-02-16]. https://www.tensorflow.org/.

[152] Kingma D P, Ba J. Adam: A method for stochastic optimization[C/OL]//Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). 2015: 1-13[2022-02-16]. http://arxiv.org/abs/1412.6980.

[153] Guo Y, Smith C, Hašan M, et al. MaterialGAN: Reflectance capture using a generative SVBRDF model[J/OL]. ACM Transactions on Graphics (TOG), 2020, 39(6): 254:1-254:13[2022-02-16]. https://doi.org/10.1145/3414685.3417779.

[154] Guo J, Lai S, Tao C, et al. Highlight-aware two-stream network for single-image SVBRDF acquisition[J/OL]. ACM Transactions on Graphics (TOG), 2021, 40(4): 123:1-123:14[2022-02-16]. https://doi.org/10.1145/3450626.3459854.

[155] Lensch H P A, Kautz J, Goesele M, et al. Image-based reconstruction of spatial appearance and geometric detail[J/OL]. ACM Transactions on Graphics (TOG), 2003, 22(2): 267:1-267:24 [2022-02-16]. https://doi.org/10.1145/636886.636891.

[156] Geldreich R, Pritchard M, Brooks J. Deferred lighting and shading[C]//Game Developers Conference (GDC) 2004 Presentation. 2004.

[157] Ren P, Wang J, Snyder J, et al. Pocket reflectometry[J/OL]. ACM Transactions on Graphics (TOG), 2011, 30(4): 45:1-45:10[2022-02-16]. https://doi.org/10.1145/2010324.1964940.

[158] Cook R L, Torrance K E. A reflectance model for computer graphics[J/OL]. ACM Transactions on Graphics (TOG), 1982, 1(1): 7-24[2022-02-16]. https://doi.org/10.1145/357290.357293.

[159] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C/OL]//2017 IEEE International Conference on Computer Vision (ICCV). 2017: 2242-2251[2022-02-16]. https://doi.org/10.1109/ICCV.2017.244.

[160] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution [C/OL]//Proceedings of the 14th European Conference on Computer Vision (ECCV). 2016: 694-711[2022-02-16]. https://doi.org/10.1007/978-3-319-46475-6_43.

[161] OpenCV. Detection of charuco boards[EB/OL]. 2021[2022-02-16]. https://docs.opencv.org/3.4/df/d4a/tutorial_charuco_detection.html.

[162] Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2008, 30(2): 228-242[2022-02-16]. https://doi.org/10.1109/TPAMI.2007.1177.

[163] Lafortune E P, Willems Y D. Rendering participating media with bidirectional path tracing [C/OL]//Proceedings of the Eurographics Workshop on Rendering Techniques '96. 1996: 91-100[2022-02-16]. https://doi.org/10.5555/275458.275468.

[164] Veach E, Guibas L J. Optimally combining sampling techniques for Monte Carlo rendering [C/OL]//Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 1995: 419-428[2022-02-16]. https://doi.org/10.1145/218380.218498.

[165] Jensen H W. Global illumination using photon maps[C/OL]//Rendering Techniques '96. 1996: 21-30[2022-02-16]. https://doi.org/10.1007/978-3-7091-7484-5_3.

[166] Hachisuka T, Ogaki S, Jensen H W. Progressive photon mapping[J/OL]. ACM Transactions on Graphics (TOG), 2008, 27(5): 130:1-130:8[2022-02-16]. https://doi.org/10.1145/1409060.1409083.

[167] Hachisuka T, Jensen H W. Stochastic progressive photon mapping[J/OL]. ACM Transactions on Graphics (TOG), 2009, 28(5): 1-8[2022-02-16]. https://doi.org/10.1145/1618452.1618487.

[168] Kaplanyan A. Light propagation volumes in CryEngine 3[C/OL]//ACM SIGGRAPH 2009 Courses. 2009: 1[2022-02-16]. https://doi.org/10.1145/1667239.1667243.

[169] Kaplanyan A, Dachsbacher C. Cascaded light propagation volumes for real-time indirect illumination[C/OL]//Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D). 2010: 99-107[2022-02-16]. https://doi.org/10.1145/1730804.1730821.

[170] Crassin C, Neyret F, Sainz M, et al. Interactive indirect illumination using voxel cone tracing [J/OL]. Computer Graphics Forum (CGF), 2011, 30(7): 1921-1930[2022-02-16]. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2011.02063.x.

[171] id Software. Quake III Arena[EB/OL]. 1999[2022-02-16]. https://github.com/id-Software/Quake-III-Arena.

[172] Sun X, Zhou K, Chen Y, et al. Interactive relighting with dynamic BRDFs[J/OL]. ACM Transactions on Graphics (TOG), 2007, 26(3): 27:1-27:10[2022-02-16]. https://doi.org/10.1145/1276377.1276411.

[173] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C/OL]//Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017: 6000-6010[2022-02-16]. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[174] Tancik M, Srinivasan P, Mildenhall B, et al. Fourier features let networks learn high frequency functions in low dimensional domains[C/OL]//Advances in Neural Information Processing Systems 33 (NeurIPS 2020). 2020: 7537-7547[2022-02-16]. https://proceedings.neurips.cc/paper/2020/file/55053683268957697aa39fba6f231c68-Paper.pdf.

[175] Heitz E, Dupuy J, Hill S, et al. Real-time polygonal-light shading with linearly transformed cosines[J/OL]. ACM Transactions on Graphics (TOG), 2016, 35(4): 41:1-41:8[2022-02-16]. https://doi.org/10.1145/2897824.2925895.

[176] Heitz E, Hill S, McGuire M. Combining analytic direct illumination and stochastic shadows [C/OL]//Proceedings of the 2018 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D). 2018: 2:1-2:11[2022-02-16]. https://doi.org/10.1145/3190834.3190852.

[177] Rahaman N, Baratin A, Arpit D, et al. On the spectral bias of neural networks[C/OL]//Proceedings of the 36th International Conference on Machine Learning (ICML). 2019: 5301-5310[2022-02-16]. https://proceedings.mlr.press/v97/rahaman19a.html.

[178] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks [C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 5967-5976[2022-02-16]. https://doi.org/10.1109/CVPR.2017.632.

[179] Pharr M, Jakob W, Humphreys G. Physically based rendering: From theory to implementation [M]. 3rd ed. Morgan Kaufmann, 2016.

[180] Sitzmann V, Martel J, Bergman A, et al. Implicit neural representations with periodic activation functions[C/OL]//Advances in Neural Information Processing Systems 33 (NeurIPS 2020). 2020: 7462-7473[2022-02-16]. https://proceedings.neurips.cc/paper/2020/file/53c04118df112c13a8c34b38343b9c10-Paper.pdf.

致 谢

博士生活虽然即将结束，但在攻读博士期间的收获却是我一生受用的宝贵财富。五年的时间里，我遇到诸多良师益友，不仅在学术和专业方便帮助我快速成长，也给严肃的学术生活带来一抹温馨。

首先非常感谢我的博士生导师徐昆副教授，徐老师在我攻读博士期间的言传身教使我受益终身。徐老师为我提供了宽松自由的科研氛围，允许我自主地进行科研探索并同时给予充分的指导和帮助，使得我可以快速入门计算机图形学研究并不断获得成长。徐老师充分了解并尊重个体发展差异，根据我们自身不同的人生目标和特长给予个性化的培养方式和资源，鼓励并推荐我利用暑假和其他空余时间前往业界进行实习，使得我不仅可以在科研上有所进展也可以提前积累关于未来职业生涯的认识和经验。徐老师在图形学领域具有广泛的积累和深入的洞见，也对我们提出成为所在领域专家的期许。

在微软亚洲研究院网络图形组进行科研实习期间，有幸受到童欣老师和董悦老师的悉心指导，我逐渐找到了科研方向并做出一些有意义、有价值的研究工作。童欣老师对待科研工作既严谨求实又充满热情，在讨论交流中童欣老师对于计算机图形学领域高屋建瓴的认知和对于具体问题打破砂锅问到底的态度令我受益匪浅，为我们树立了优秀科研人的标杆。董悦老师在实习期间给予了我诸多耐心指导和帮助，感谢董老师不厌其烦的答疑解惑，不仅让我收获了具体的回答和知识，更是帮助我建立大胆表露未知和承认不足的勇气和品格。此外，感谢陈国军博士和李潇博士在科研合作的整个过程中的耐心帮助，从中我受益良多。感谢威廉玛丽学院的 Pieter Peers 教授的指导。感谢叶文杰博士、郭雨潇博士、王鹏帅博士在服务器使用和实验方面提供的帮助。感谢网络图形组的其他同事和同学，每日的组会讨论中使我收益颇丰。

在其他实习和平日学习过程中我也遇到了多位良师益友，在此感谢南京大学过洁老师带我走进图形学的大门，感谢马里千博士、黄浩智博士在实习中的照顾和建议，感谢实验室所有同学在平日的帮助。

最后，我要感谢我的妈妈，你给予了我最为无私的爱和永远的鼓励与理解，同时也为我树立了善良、勇敢且坚强的人生榜样。感谢我的女友谢忠赟从大学入学到博士毕业一路与我携手同行，感谢你带来的每一份爱护与快乐。感谢我的哥哥、姥姥、姥爷和其他亲人的帮助。你们是我永远的港湾和后盾。

个人简历、在学期间完成的相关学术成果

个人简历

1995 年 3 月 6 日出生于山西省吕梁市。

2012 年 9 月考入南京大学生命科学专业, 2014 年 9 月转入计算机科学与技术专业, 2017 年 7 月本科毕业并获得理学学士学位。

2017 年 9 月免试进入清华大学计算机科学与技术系攻读博士至今。

在学期间完成的相关学术成果

学术论文：

[1] **高端**, 徐昆. 基于单张图片的人脸高频纹理估计 [C]. 第二十二届中国计算机辅助设计与图形学大会（CAD/CG 2019）, 2019.

[2] **Gao D**, Li X, Dong Y, Peers P, Xu K, Tong X. Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images[J/OL]. ACM Transactions on Graphics (TOG), 2019, 38(4): 134:1-134:15. https://doi.org/10.1145/3306346.3323042.

[3] **Gao D**, Chen G, Dong Y, Peers P, Xu K, Tong X. Deferred neural lighting: free-viewpoint relighting from unstructured photographs[J/OL]. ACM Transactions on Graphics (TOG), 2020, 39(6): 258:1-258:15. https://doi.org/10.1145/3414685.3417767.

[4] **Gao D**, Mu H, Xu K. Deep global illumination: interactive indirect illumination prediction under dynamic area lights[J]. (已投稿 IEEE Transactions on Visualization and Computer Graphics).

国家发明专利：

[5] **高端**, 周峙龙, 凌永根, 黄浩智, 胡事民, 徐昆, 刘威. 三维人脸模型构建方法、装置、计算机设备及存储介质: 中国, CN110675413B[P/OL]. 2020-11-13[2022-02-16]. http://epub.cnipa.gov.cn/patent/CN110675413B.

[6] 马里千, **高端**. 人脸图像数据采集方法及人脸图像数据采集装置: 中国, CN108876891B[P/OL]. 2021.12.28[2022-02-16]. http://epub.cnipa.gov.cn/patent/CN108876891B.

[7] 马里千, **高端**. 人脸图像合成方法和装置: 中国, CN107644455B[P/OL]. 2022.02.22[2022-03-01]. http://epub.cnipa.gov.cn/patent/CN107644455B.

[8] **高端**, 徐昆, 顾澄宇, 廖晶堂. 渲染方法和装置: 中国, 202111543782.X (专利申

请号)[P]. 2021.12.16.