谨以此文献给我的家人、老师和朋友们!

---------- 举雅琨

非朗伯光度立体的深度学习模型

学位论文完成日期：_____

指导教师签字：_____

答辩委员会成员签字：_____

_____

_____

_____

_____

_____

_____

独 创 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含未获（注：如没有其他需要特别声明的，本栏可空）或其他教育机构的学位或证书使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：　　签字日期：　　年　　月　　日

学位论文版权使用授权书

本学位论文作者完全了解国家有关保留、使用学位论文的法律、法规和学校有关规定，并同意以下事项：

1、学校有权保留并向国家有关部门或机构送交本学位论文的复印件和磁盘，允许论文被查阅和借阅；

2、学校可以将本学位论文的全部或部分内容编入学校学位论文数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文；

3、学校可以基于教学及科研需要合理使用本学位论文。

需保密的学位论文在解密后适用本授权书。

学位论文作者签名：　　导师签字：

签字日期：　　年　　月　　日　　签字日期：　　年　　月　　日

非朗伯光度立体的深度学习模型

摘　要

三维重建是计算机视觉中的一个基础且重要的问题，准确的三维信息可以使人脸识别和语义分割等高层视觉任务的精确度大幅提高。光度立体是一种单视角的三维重建方法，其利用不同光照方向下的图像（三张及以上）提供的灰度变化线索恢复物体表面法向。不同于双目视觉和运动恢复结构等方法利用变化的视角测量稀疏的点云，光度立体在弱纹理的条件下也可以重建出逐像素的稠密表面法向。因此，光度立体视觉在许多高精度要求的三维重建任务中有着不可替代的地位，例如文物重建、海床测绘、月表地貌重建和工业缺陷检测等。

传统的最小二乘法光度立体算法基于朗伯表面（漫反射表面）的假设，但现实世界的物体几乎不具有理想朗伯表面的特性，例如陶瓷和金属，这限制了其实际应用。为了解决这个限制，后续的方法更多地关注具有更灵活反射函数的非朗伯表面，例如采用双向反射分布函数来拟合一般反射。然而这些方法仅能处理部分表面材质的物体，并且存在优化不稳定的问题。近年来，部分学者利用数据驱动的方法来拟合非朗勃物体的表面法向，取得了比非数据驱动的传统方法更高的精确度。得益于深度学习技术的快速发展，光度立体的深度学习模型参数逐渐增多，神经网络架构变得越来越复杂，且对硬件的要求也越来越高。但是，近期研究表明，持续增加网络参数并不能显著提升表面法向的重建精度。本文认为，阻碍基于深度学习的光度立体算法提升精度的典型原因有以下几点：第一，现有方法面临细节缺失的挑战，这对形状变化迅速的高频表面区域影响很大，例如皱纹、边缘和其他复杂结构，在这些区域，重建的表面法向模糊，误差较大；第二，现有方法仅采用单一的法向角度余弦损失函数，缺乏额外的监督信息，难以提高重建精度；第三，先前的方法未能结合物理先验的信息，仅使用光度立体图像作为输入来学习从图像域到法向域的映射，很难获得更好的重建结果。

针对上述问题，本文对光度立体的深度学习模型进行了多方面的研究。论文的创新点和主要贡献概括如下：

（1）本文提出了一种自适应注意力光度立体模型。该模型利用注意力加权的法向重建损失，在自监督框架下为高频表达的区域施加更高权重的细节保护损失，避免了先前单一的基于欧氏距离损失带来的采样模糊，从而提升了物体表面

法向的重建精度。

（2）本文在自适应注意力光度立体模型的基础上，进行了进一步实验研究，发现光度立体图像的高频表达主要受物体表面复杂结构和物体表面变化的材质影响，进而提出了一种归一化的高频区域增强光度立体模型。该模型在物体表面材质剧烈变化的区域上也取得了高精度的重建结果，实验表明所提出的模型获得了最佳的表面法向重建精度。

（3）本文提出了一种重光照-光度立体双重监督模型。该模型利用输入的光度立体图像重建物体表面法向，并进一步从表面法向中回归重建的光度立体图像，形成了闭环并提供额外图像重建损失的监督。相对于先前单一余弦损失的方法，该模型提升了物体表面法向的重建精度并且可以生成任意指定的重光照光度立体图像。

（4）在重光照-光度立体双重监督模型的基础上，本文进一步提出了重渲染-光度立体三重监督模型。该模型利用编码的材质信息，通过重渲染网络生成任意表面材质和光照下的重渲染光度立体图像，并为模型提供了余弦损失、图像重建损失和图像变化损失三种监督。该模型可以同时重建高精度的物体表面法向和重渲染光度立体图像，解决了光度立体数据集样本扩充的问题。

（5）本文提出了一种融合物理先验的光度立体模型。相比现有仅使用光度立体图像的跨域映射框架，本文将物体模型中最小二乘法得到的初始法向与光度立体图像融合，使得模型在法向域内学习映射。物理先验的初始法向可以减少求解表面法向的函数空间，使其从图像域到法向域的映射转变为法向域内映射。此外，在融合物体先验的框架之上，本文提出了局部亲和力特征模块以更好地重建高精度的表面法向。

关键词：**光度立体**，**表面法向**，**非朗伯材质**，**三维重建**，**深度学习**

Deep Learning Models for Non-Lambertian Photometric Stereo

Abstract

Three-dimensional (3D) reconstruction is a basic and pivotal problem in computer vision, accurate 3D information can greatly improve the performance of high-level vision tasks, such as face recognition and semantic segmentation. Photometric stereo recovers the surface normal of an object, in a single view, from various shading cues under multiple images (three or more images) with different illuminations. Unlike binocular and multi-view stereo that use different scenes from viewpoints to triangulate sparse 3D points, photometric stereo can recover per-pixel dense surface normal, and prevails in reconstructing weak texture objects. Photometric stereo, therefore, plays an irreplaceable role in many high-precision 3D reconstruction tasks, such as cultural relic reconstruction, seabed mapping, moon surface reconstruction, and industrial defect detection.

The conventional least squares based photometric stereo algorithm assumes the Lambertian reflectance (diffuse surface) model. However, ideal Lambertian surfaces barely exist in the real world, which limits the application of the algorithm in general real-world materials, such as ceramics and metals. To deal with this limitation, subsequent methods focus more on non-Lambertian surfaces with more flexible reflectance functions, such as adopting the bidirectional reflectance distribution functions to model the general reflectance. However, these methods are accurate for limited categories of materials and suffer from unstable optimization. Recently, some researchers have investigated data-driven methods to approximate the surface normals of non-Lambertian objects, and achieved improved accuracy than the conventional non-data-driven photometric stereo methods. Due to the explosive development of deep learning, the parameters of the deep learning models of photometric stereo methods have gradually increased, the neural network architectures have become increasingly complex, associated with the higher requirements of hardware. However, recent studies have shown that continuously increasing datasets and network parameters cannot significantly improve the accuracy of surface normals. This thesis argues that the typical reasons hindering the

improvement of deep learning based photometric stereo algorithms are as follows: first, the existing methods face the challenge of missing details, which are greatly limited in high-frequency surface regions with rapid shape variations, such as crinkles, edges, and other complex structures. In these regions, the estimated surface normal is blurred with large error; second, existing methods only employ the Cosine loss function of normals, which lack additional supervision and are difficult to improve the accuracy; third, the previous methods fail to incorporate the priors information, which learns the mapping from the image domain to the normal domain. It is difficult to obtain better surface normals using only the photometric stereo images as the inputs.

To overcome the aforementioned problems, this thesis studies the deep learning models of the photometric stereo method in multiple aspects, as follows:

(1) This thesis proposes an adaptive attention photometric stereo model, which utilizes the attention-weight normal loss to impose a higher-weight detail-preserving loss on the high-frequency region, in a self-supervised framework. It avoids the sampling blur due to previous single Euclidean distance loss, and improves the accuracy of the reconstructed surface normals.

(2) Based on the adaptive attention photometric stereo model, experimental research is conducted and finds that the high-frequency expression of photometric stereo images is mainly affected by the complex structure and the changing material of the surface. This thesis further introduces a normalized high-frequency region enhanced photometric stereo model. It achieves high-precision reconstruction results on the surface with steep changing materials. The proposed model outperforms the state-of-the-art photometric stereo methods.

(3) This thesis reports a re-illumination & photometric stereo dual-supervised model. The model reconstructs the surface normals from the input photometric stereo images, and further regresses the reconstructed photometric stereo images back, which forms a closed-loop structure and therefore provides additional supervision of image reconstruction loss. Compared with the previous single Cosine loss, the model improves the accuracy of the reconstructed surface normals, and can generate arbitrarily re-illuminated photometric stereo images.

(4) Based on the re-illumination & photometric stereo dual-supervised model, this thesis further introduces a re-rendering & photometric stereo triple-supervised model. The model utilizes the encoded material information to generate re-rendered photometric stereo images with arbitrary surface materials and illumination directions, through the re-rendering network. According to that, three types of supervision are provided, as the Cosine loss, the image reconstruction loss, and the image transform loss. This model can simultaneously reconstruct high-precision surface normals and re-rendered photometric stereo images, which solves the problem of the expansion in photometric stereo datasets.

(5) This thesis proposes a physical prior incorporated photometric stereo model. Compared with the existing cross-domain mapping framework that only uses photometric stereo images, this thesis fuses the least square initial surface normals under physical model with photometric stereo images, to learn mapping in the normal-domain. The initial surface normals under physical prior can reduce the space of learning functions for the surface normals, which transforms the mapping from the image domain to the normal domain into the mapping in the same normal domain. Furthermore, under the framework of the physical prior incorporated model, this thesis adopts a local affinity feature module to reconstruct the surface normal better.

Key Words: Photometric Stereo, Surface Normal, Non-Lambertian reflectance, 3D Reconstruction, Deep Learning

目　录

1 绪论

1.1 研究背景及意义

计算机视觉是一门利用相机和计算机代替人眼和大脑对客观世界的场景进行识别、感知和理解的科学[1]。三维重建是计算机视觉的重要领域，其目标是利用计算机获得客观世界物体的三维模型。获得图像的三维结构是许多计算机视觉应用的关键，因为其可以提高对图像本身的精确理解和感知[2,3]。总的来说，主流的三维重建方法可以概括为几何法和光度法两类。所谓几何法，即多视角几何[4,5,6,7]，该类方法利用不同视角下的相机对同一场景进行拍摄，通过不同视角下的特征匹配和相机的几何成像关系恢复出三维坐标。然而，多视角几何的难点在于特征的匹配，往往匹配中的误差会影响到局部重建的精细度。此外，该类方法在弱纹理区域难以产生准确的匹配，从而产生成重建的三维点云稀疏等问题。

以光度立体（photometric stereo）为代表的光度法[8,9,10,11] 则利用光度信息来恢复物体的几何形状。光度立体法通过相机拍摄得到的三张及以上的图像像素点的明暗程度，计算光照反射模型并得到物体表面的法向信息[9]。与多视角几何相比，光度立体只受光照条件的影响而不受物体轮廓、特征点及纹理信息的影响。光度立体技术可以扩展到具有镜面反射的非朗伯表面三维重建[12,13,14]，在这类物体上，即使采用更复杂繁琐的激光扫描法[15] 和结构光法[16] 也难以取得令人满意的结果。同时，光度立体可以计算出每个像素点的稠密表面法向，进行精密的三维重建。因此，光度立体法在三维重建中有重要的地位，如图 1-1所示，常被用于精细重建，例如文物三维重建[17]、海床测绘[18]、表面缺陷检测[19] 和月球地貌重建[20] 等众多任务。此外，随着社会科技的进步，光度立体技术也被逐渐应用于虚拟现实[21]、天气预报[22] 及服饰绘制[23] 等新领域。

图 1-1 光度立体常被用于精细重建稠密的表面结构

1.2 研究目标

随着深度学习技术的蓬勃发展，基于深度学习的光度立体模型[24,25,26,27,28]也被国内外学者广泛提出，这些模型的网络结构越来越复杂，需要学习的模型参数越来越多。然而近期的研究表明，持续增加训练数据量和网络的复杂度并不能显著提高基于深度学习的光度立体的重建精准度。为了进一步挖掘深度学习光度立体算法的表面法向重建精度，本文的主要研究目标如下：

（1）研究高频增强的光度立体模型，以提高方法在高频区域表面法向重建的准确度和清晰度。现有的光度立体方法在物体表面的褶皱、边缘等高频区域存在模糊和重建表面法向错误的问题，而这些高频区域正是人们最关注、需要准确重建的部分。究其原因，先前基于深度学习的光度立体方法仅采用基于欧几里得距离的损失函数，例如 L1 损失，L2 损失和余弦损失，而这些基于欧氏距离的损失函数由于平均采样，导致预测总是倾向于回归总体的平均值[29,30]，因此很难约束高频信息。

（2）研究如何光度立体任务中构建额外的监督信息。现有的深度学习光度立体方法仅采用重建法向和真值法向之间的余弦角度损失或者 L2 距离损失进行网络的训练优化，而缺乏额外的监督信息。盲目地增加模型的复杂度很难进一步提高非朗伯表面材质下表面法向重建的准确性，尤其是与阴影、镜面反射和非凸结构相关的区域。

（3）研究如何对有限真值法向的光度立体数据进行数据扩充。数据是深度学习算法的核心，训练数据集的充足与否对深度神经网络的效果至关重要。而基于深度学习的光度立体数据集，由于客观世界里真实的物体难以获得并配准其表面法向真值[31]，因而面临着训练数据不足的问题。

（4）研究利用光度立体图像自身的先验信息来辅助跨域的映射，以提升重建效果。对于基于深度学习的光度立体任务而言，现有的方法都遵循学习从图像域到法向域的映射这一框架，即将光度立体图像作为神经网络的输入，将预测的表面法向作为神经网络的输出。然而，这种跨域的映射并不能取得十分准确的结果。

1.3 研究内容及创新点

针对 1.2节提出的四个目标展开研究，本文的研究内容和创新点可以归纳为：

（1）为解决现有深度学习光度立体预测法向的物体表面褶皱、边缘等高频区

域重建的表面法向存在模糊、误差大的问题，本文首先提出了一种自适应注意力光度立体模型。该模型利用自监督学习得到的注意力图为注意力加权的法向重建损失提供权重，从而对高频的区域施加高比例的细节保护损失惩罚。因此该方法可以有效改善基于欧式距离损失带来的高频区域重建的表面法向模糊的问题。

（2）本文进一步研究了光度立体图像中出现高频的原因。其一是由于物体表面的复杂结构，这在上述提出的自适应注意力光度立体模型中已经被解决。另一种高频表达的原因则是由于物体表面剧烈变化的材质，在这种情况下该区域的表面法向仍是平滑的，而在此处使用上述模型会由于施加了较大的细节保护损失而导致重建结果变差。因此，本文在研究内容（1）的基础上进一步提出了归一化的高频区域增强光度立体模型。该方法同时解决了物体复杂结构和变化的材质两种原因导致的表面法向重建误差问题。

（3）为解决现有光度立体方法单一余弦损失监督的问题，本文提出了一种重光照-光度立体双重监督模型。该方法在传统的表面法向生成网路的基础上提出了双重回归网络，以生成重光照的光度立体图像。因此，除常规的表面法向提供的监督（余弦损失或 L2 损失）外，利用重光照图像与输入的光度立体图像形成了额外的图像重建损失，使模型形成闭环结构，提升表面法向重建精度。

（4）为缓解训练数据不足的问题，本文在研究内容（3）的基础上提出了重渲染-光度立体三重监督模型。该方法在重渲染网络引入了编码的材质信息，从而使其可以重渲染光度立体图像而不局限于与输入图像相同的表面材质。重渲染-光度立体三重监督模型可以从输入的光度立体图像渲染出任意表面材质，任意光照方向的图像，实现数据集的扩充。

（5）本文提出了一种融合物理先验的光度立体模型。该方法通过物理模型下最小二乘法得到的朗伯先验初始法向与光度立体图像融合，从而将基于学习的光度立体任务从现有的图像域至法向域映射框架变换至法向域内映射问题。在融合物理先验的框架之上，我们又提出了局部亲和力特征模块以通过显式挖掘相邻特征的关系来学习丰富的结构表示，更好地重建高精度的表面法向。

1.4 本文的组织结构

本文围绕基于深度学习的光度立体算法展开研究，图 1-2展示了本文的组织机构和各章节的关系与逻辑。

主要内容及结构安排如下：

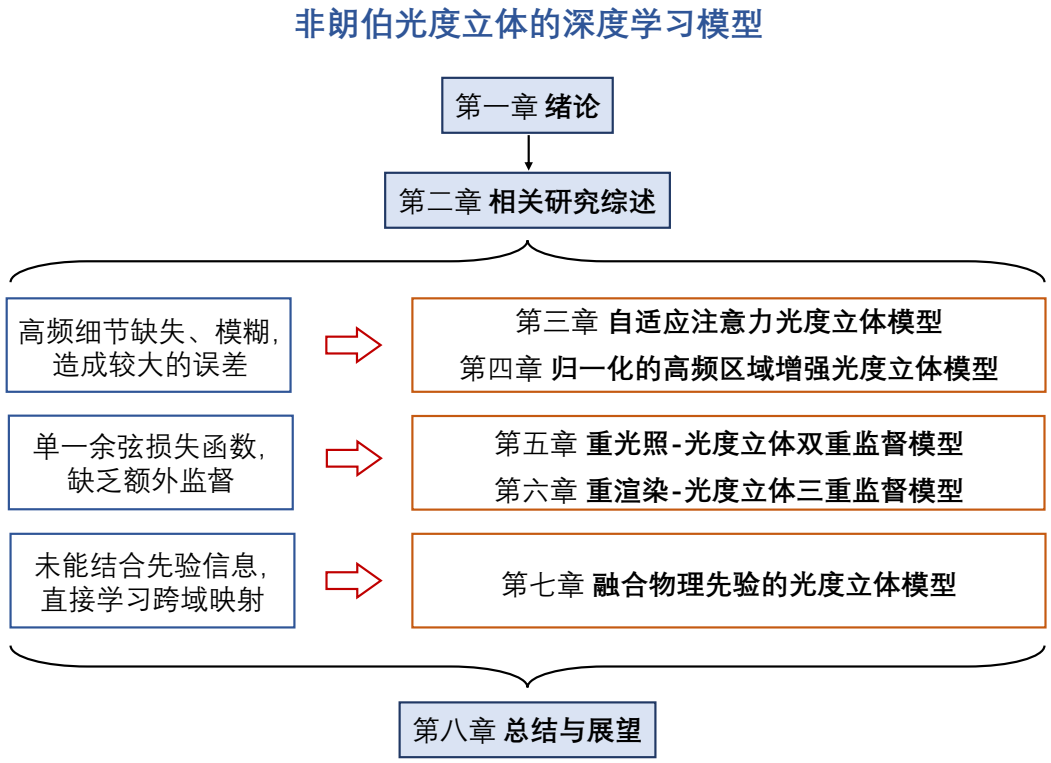


图 1-2 本文组织结构和各个章节关系

第一章为绪论，介绍了本文的研究背景、主要研究工作、创新点以及本文的组织结构。

第二章为相关背景知识和研究综述，对本文涉及到的基本知识和相关领域的工作展开了综述。重点介绍了非朗伯光度立体中各种类型的传统方法和基于深度学习的方法，并分析了其优缺点。

第三章提出了自适应注意力光度立体模型，其引入注意力加权的法向重建损失，为复杂结构区域施加高权重的细节保护损失，在复杂区域能重建更清晰的表面法向。

第四章则在第三章的基础上，提出了归一化的高频区域增强光度立体模型。解决了物体表面复杂结构区域和物体表面材质变化区域的重建误差，并取得了最佳的表面法向重建精度。

第五章提出了重光照-光度立体双重监督模型。该模型能够利用输入光度立体图像重建物体表面法向，并从表面法向中回归出重建光度立体图像，使模型形成闭环，提供额外图像重建损失的监督。

第六章在第五章的基础上，进一步提出了重渲染-光度立体三重监督模型。利用编码的材质信息，通过重渲染网络生成任意表面材质、光照下的重渲染光度

立体图像，实现对少量样本的光度立体数据集的扩充。

第七章提出了融合物理先验的光度立体模型将物理模型下的初始法向与光度立体图像融合，以学习差分特征，将图像域到法向域的跨域映射转变为学习法向空间的域内映射任务。基于该框架之上，本章进一步提出了局部亲和力特征模块，实现高精度的表面法向重建。

第八章为总结与展望。对本文的研究内容进行了总结，并对未来工作进行展望，指出今后进一步的研究方向。

2 相关背景知识和研究综述

本章对论文中涉及到的传统和基于深度学习的光度立体算法的国内外研究现状进行了综述并分析其中的特点及缺陷。在综述光度立体算法之前，本章先介绍了双向反射分布函数和光照成像模型等基础知识。在本文中，除特殊声明外，以普通小写字母或小写希腊字母表示标量，以粗体小写字母或粗体希腊字母表示矢量，以粗体大写字母或粗体大写希腊字母表示矩阵。

2.1 双向反射分布函数

双向反射分布函数（bidirectional reflectance distribution function，BRDF）[32]描述了物体表面入射光和反射光的关系。对于一个方向的入射光，物体表面会将光反射到表面上半球的各个方向，而不同方向反射的比例是不同的。因此，我们用 BRDF 来表示在该表面上指定方向的反射光（reflected light）和入射光（incident light）的比例关系。通常，使用球坐标系来表示表面上半球的入射光方向 (θ_i, ϕ_i) 和反射光方向 (θ_r, ϕ_r)，其中 θ 为仰角，ϕ 为方位角。那么，BRDF 则可以表示为一个四元函数 f_{BRDF}：

$$f_{BRDF}(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{L_r(\theta_r, \phi_r)}{E_i(\theta_i, \phi_i)}, \tag{2-1}$$

其中 $L_r(\theta_r, \phi_r)$ 表示反射光的辐亮度（单位投影面积和单位立体角上的辐射通量），$E_i(\theta_i, \phi_i)$ 表示入射光的辐照度（单位面积接收到的辐射通量）。

2.2 光照成像模型

假设一个具有线性辐射响应的正交投影相机，有来自上半球的定向光照明，且观察方向 (θ_r, ϕ_r) 平行于指向世界坐标系原点的 z 轴，图像坐标与世界坐标系 xy 坐标对齐。依照成像模型[33,26,34]，那么拍摄的图像中某一点像素的亮度 $\tilde{o}$ 与接收到的反射光的辐亮度 $L_r(\theta_r, \phi_r)$ 的关系可以表示为：

$$\tilde{o} = cL_r(\theta_r, \phi_r), \tag{2-2}$$

其中 c 为相机自身的参数，它不随外界环境的变化而变化，可被看作为常数。根据 2.1节中介绍的 BRDF，结合式 (2-1) 和式 (2-2)，可以得到相机观测的亮度与入射光源的辐照度关系

$$\tilde{o} = cf_{BRDF}(\theta_i, \phi_i, \theta_r, \phi_r)E_i(\theta_i, \phi_i) + \epsilon, \tag{2-3}$$

其中 ϵ 表示全局光照信息（例如互反射[35] 等）和噪声。式 (2-3) 中，入射光源辐照度 E_i 与正投影方向的辐照度的关系可以写为 $E_i = Ecos\theta$，其中 $cos\theta$ 为光源方向与照射表面法向的夹角，且有 $cos\theta = \boldsymbol{n}^\top \boldsymbol{l}$，其中 $\boldsymbol{n}$、$\boldsymbol{l}$ 为归一化后的表面法向和入射光源方向。在入射光源标定好的情况下（即方向 $\boldsymbol{l}$ 和辐照度 E 已知），则可以定义归一化的像素亮度 $o = \tilde{o}/E$。因此式 (2-3) 亦可以写作：

$$o = \boldsymbol{\rho}(\theta_i, \phi_i, \theta_r, \phi_r) \max(\boldsymbol{n}^\top \boldsymbol{l}, 0) + \epsilon, \tag{2-4}$$

其中 $\boldsymbol{\rho}(\theta_i, \phi_i, \theta_r, \phi_r) = cf_{BRDF}(\theta_i, \phi_i, \theta_r, \phi_r)$ 以表示广义的反射率，$\max(\boldsymbol{n}^\top \boldsymbol{l}, 0)$ 用来保证附加阴影处的像素亮度不为负值。式 (2-4) 即为最终的光照成像模型，其建立了基本的二维图像与三维结构之间的联系。

2.3 朗伯光度立体技术

光度立体所要解决的问题，可以看作是式 (2-4) 的逆问题，即给定已知的拍摄图像的像素亮度 o，求解得到该像素位置上物体的表面法向 $\boldsymbol{n}$。1970 年 Horn[8] 最早提出了从阴影恢复形状技术（shape from shading，SFS），从单张图像，利用其亮度变化恢复物体表面法向。然而由于需要求解的表面法向的自由度大于方程个数，是一个欠约束的问题[36]，这导致实际三维重建效果较差。因此，Woodham[9] 在 1980 年受此启发提出了从三张及以上固定视角图像恢复物体表面法向的光度立体方法（图 2-1）。该方法需要假设待重建物体表面具有理想漫反射表面，即符合朗伯模型假设[37]：观测到的表面强度不随观察方向（即反射光 (θ_r, ϕ_r)）的变化而变化，换句话说，入射光在朗伯表面上向各个方向以相同的强度发散。在这种假设下，即认为 BRDF 函数 ρ 是一个常数，反映该表面本身的反射率。在不考虑投射阴影和全局光照的情况下，式 (2-4) 可以被简化为：

$$o = \rho\, \boldsymbol{l}^\top \boldsymbol{n}. \tag{2-5}$$

考虑有 j 个光源照射的光度立体系统，即有光照方向矩阵 $\boldsymbol{L} = \{\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_j\} \in \mathbb{R}^{3\times j}$，拍摄的图像中某一点像素强度向量 $\boldsymbol{o} = \{o_1, o_2, \cdots, o_j\}^\top \in \mathbb{R}^j$。则式 (2-5) 的矩阵形式可以表示为：

$$\boldsymbol{o} = \rho\, \boldsymbol{L}^\top \boldsymbol{n}, \tag{2-6}$$

其中 $\boldsymbol{n} = (n_x, n_y, n_z)$ 为归一化的表面法向，因此只含有两个未知数，此外，亦有 ρ 未知。因此，光度立体算法[9] 需要至少三张不同光照方向下的拍摄图像通过最小

图 2-1 光度立体示意图

二乘法（least square，LS）找到最小误差的未归一化表面法向 $\hat{\boldsymbol{n}}$：$\min_{\hat{\boldsymbol{n}}}\|\boldsymbol{o}-\boldsymbol{L}^{\top}\hat{\boldsymbol{n}}\|^2$。其求解过程如下：

$$\|\boldsymbol{o}-\boldsymbol{L}^{\top}\hat{\boldsymbol{n}}\|^2=\boldsymbol{o}^{\top}\boldsymbol{o}+\hat{\boldsymbol{n}}^{\top}\boldsymbol{L}\boldsymbol{L}^{\top}\hat{\boldsymbol{n}}-2\hat{\boldsymbol{n}}^{\top}\boldsymbol{L}\boldsymbol{o}, \tag{2-7}$$

此时，对 $\hat{\boldsymbol{n}}$ 求导，可得：

$$2\boldsymbol{L}\boldsymbol{L}^{\top}\hat{\boldsymbol{n}}-2\boldsymbol{L}\boldsymbol{o}=0, \tag{2-8}$$

即有：

$$\hat{\boldsymbol{n}}=(\boldsymbol{L}\boldsymbol{L}^{\top})^{-1}\boldsymbol{L}\boldsymbol{o}, \tag{2-9}$$

至此可以求得未经归一化的表面法向 $\hat{\boldsymbol{n}}$，且其模长 $\|\hat{\boldsymbol{n}}\|=\rho$。因此，最终归一化表面法向 $\boldsymbol{n}$ 可以得知：

$$\boldsymbol{n}=\frac{\hat{\boldsymbol{n}}}{\sqrt{\hat{\boldsymbol{n}}^{\top}\hat{\boldsymbol{n}}}}. \tag{2-10}$$

然而上述的基于最小二乘法光度立体算法局限于朗伯表面的假设。在现实世界中，几乎不存在理想朗伯表面的物体，常见的材料如陶瓷、金属、塑料等都带有非朗伯的表面属性，这严重限制了朗伯光度立体方法在实际场景中的应用。

2.4 非朗伯光度立体技术

真实世界的物体几乎没有朗伯材质，难以避免有非朗伯特性。如图 2-2所示，非朗伯表面材质存在镜面反射、投射阴影等影响朗伯假设下线性模型的因素，导致最小二乘法[9] 难以求解准确的表面法向。为了满足现实世界一般反射的需要，国内外学者提出了不同的策略。通常参考[31] 的分类法，本文将非朗伯光度立体技术简要分为四类：异常值剔除法、建模复杂反射模型法、基于 BRDF 性质法和基于深度学习法。更全面的光度立体综述亦可以在文献[38,39,40] 等中找到。此外，光度立体也衍生出很多相关问题，例如：非标定光度立体、近场点光源光度立体、彩色光度立体等，这些问题将在 2.5中进行介绍。

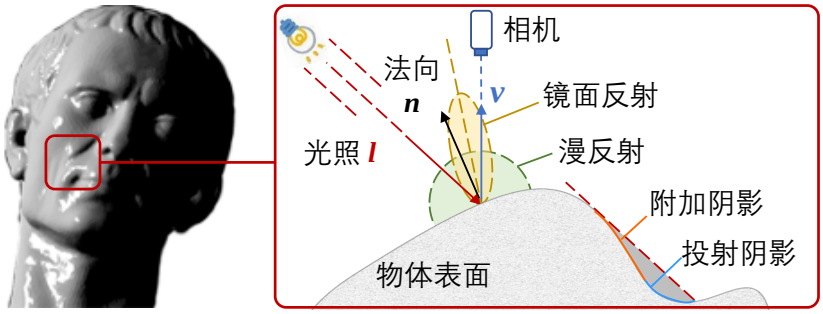


图 2-2 非朗伯表面材质

2.4.1 基于异常值剔除的方法

基于异常值剔除的方法通常假设非朗伯表面（例如高光）是局部的、稀疏的，因此这些非朗伯部分可以利用数理统计方法被当作异常值剔除。此时保留的物体表面的镜面反射项可以忽略不计而只剩下漫反射部分，因此可以利用朗伯光度立体的方法重建物体表面法向。早期的一些相关工作仅是从多张光度立体图像中设定标准阈值，将超过该阈值的部分认为是异常值（高光）并剔除，用朗伯光度立体计算物体的表面法向[41]。此外，Chandraker 等人[42] 利用图割法分离出异常值和朗伯部分并求解表面法向；Wu 等人[43] 利用马尔可夫随机场提取朗伯反射图像；Mukaigawa 等人[44] 利用随机抽样一致方法（RANSAC）对光度立体图像进行线性分解从而剔除了阴影和高光等异常值并求解表面法向；Verbiest 等人[45] 利用最大似然估计将阴影等异常值信息去除。最近，一些鲁棒统计学方法也被用于异常值剔除中，这类方法通过分解低秩矩阵（代表朗伯分项）和稀疏矩阵（代表异常值分项）来解决非朗伯光度立体问题。Wu 等人[46] 利用鲁棒主成分分析算法将光度立体问题转化为恢复具有缺失条目和损坏条目的最小化矩阵的秩的问题，以处理存在阴影和镜面反射的表面材质；Ikehata 等人[47] 利用改进的秩等于 3 分解代替秩最小化分解，从而获得更好的表面法向重建效果。

尽管异常值剔除法能在一些镜面反射表面中准确分离漫反射分项而使得光度立体算法可以处理非朗伯表面，但是异常值剔除法不能处理广泛而柔和的镜面反射，因为这些区域与朗伯部分没有明显的边界，因此很难被剔除。此外，近些年有着较好效果的低秩矩阵方法[46,47] 则需要较多的输入图片来剔除异常值，且迭代轮数多、计算量大。因此，在稀疏光照下的非朗伯光度立体依然是一个具有挑战性的问题。

2.4.2 基于建模复杂反射模型的方法

不同于将非朗伯表面当作异常值并剔除掉，建模复杂反射模型法利用高光、阴影等非朗伯区域的全部数据来建模并拟合复杂的物体表面反射系数，因此这类方法具有利用所有可用数据的优势。这些反射模型通常采用复杂的多项式函数来近似真实世界的材料，例如 Georghiades 等人[48] 利用简化的 Torrance-Sparrow 光照模型拟合非朗伯光度立体中物体表面的反射特性；Tozza 等人[49] 利用基于 Blinn-Phong 的光照模型求解带有高光的物体的表面法向和表面材质；Yeung 等人[50] 利用镜面尖峰模型同时求解物体表面法向和表面材质特性；Chen 等人[12] 利用微表面的椭球法向分布函数拟合求解物体的表面法向和表面反射特性；Chung 等人[51] 利用基于 Ward 的光照模型由物体的投射阴影轮廓求解物体表面反射特性。进一步地，Goldman 等人[52] 和 Ackermann 等人[53] 认为物体的表面材质可以用几种不同的 Ward 模型线性组合表达，并对不同的表面反射属性采用不同的加权权重以计算物体表面法向和反射特性。

尽管建模复杂反射模型法可以精确的建模非朗伯表面特性与表面法向，但是不同材质的表面反射特性变化很大，其往往只能对有限种类的物体表面有比较好的重建结果。此外由于这类方法使用了复杂的反射模型，导致高度非线性优化等问题，使其计算变的复杂并容易导致优化不稳定。

2.4.3 基于 BRDF 性质的方法

由于建模复杂反射模型法往往只能在有限种类的表面材质下取得理想的表面法向重建结果，因此近年来，一些学者致力于寻找适用于更多表面材质的特性，利用这些一般特性，光度立体有可能处理更广泛类型的材料，本文称这类方法为基于 BRDF 性质法。这些 BRDF 的一般属性包括各向同性（isotropic）、互易性（reciprocity）和单调性（monotonicity）等。BRDF 的各向同性的特点是如果法向 $\boldsymbol{n}$ 在观察方向 $\boldsymbol{v}$ 和光照方向 $\boldsymbol{l}$ 所跨越的平面上对称，则可以观察到相等的反射率值；互易性则是指当观察方向 $\boldsymbol{v}$ 和光照方向 $\boldsymbol{l}$ 互换时，亦可以观察到相等的反射率值。Hui 等人[54] 表明各向同性的 BRDF 表面可以从材质数据库中构建字典来表示正则化表面法线并同时求解表面反射率。相似地，由于大多数的各向同性 BRDF 可以通过二元函数拟合，因此 Alldrin 等人[55] 采用 2D 的离散表来表示 BRDF 性质，进而迭代计算物体的表面法向。进一步地，通过假设这个二元函数在一个维度上是单调的，Shi 等人[13] 搜索维持 BRDF 单调性的候选值来

估计物体表面法向的仰角；而 Li 等人[56] 亦利用球面线性内插法测量各向同性的 BRDF 来计算物体表面法向的方位角和仰角。此外，Higo 等人[57] 通过进一步结合 BRDF 的各向同性、单调性和可见性等性质的约束，以处理由单个高光瓣组成的一般材料表面。最近，一些工作[58,59] 提出了基于双变量的 BRDF 表达式来求解物体表面法向，这些方法先采用使用两个阈值分别排除阴影和镜面反射，而剩下变化缓慢的表面反射率。其中，文献[58] 使用双多项式表示对这种低频反射进行建模，而文献[59] 将其建模为中心方向未知的高光瓣的总和。

相比建模复杂反射模型法，基于 BRDF 性质法可以对更多种类的物体表面进行准确的表面法向预测。然而，这类方法需要更为苛刻的光度立体系统，例如需要均匀分布的光照方向等。

2.4.4 基于深度学习的方法

最早使用神经网络的光度立体方法甚至可以追溯至上世纪：1993 年，Iwahori 等人[60] 首次采用一层隐藏层的神经网络从三幅光度立体图像中恢复表面法向。之后的十余年亦有部分学者利用逐渐增多层数的神经网络拟合表面法向[61,62]。然而这些早期的基于人工神经网络的工作，或仅能处理朗伯反射表面[61,62]，或需要使用与目标对象相同材料的参考球体进行预训练[60]，从某种层面来说，这类似 Hertzmann 等人提出的基于标定物的光度立体方法[63]。然而由于需要对每种材料进行预训练或限制朗伯反射率，这些早期的基于神经网络的方法难以用于真实应用场景。

近年来现代深度神经网络（深度学习）在许多视觉领域都取得了极大的成功，例如目标检测[64]、语义分割[65]、深度估计[66] 和超分辨率[67] 等。受深度神经网络强大拟合能力的启发，最近国内外学者亦将深度学习方法引入非朗伯光度立体问题。2017 年，Santo 等人[24] 第一次将现代深度神经网络用于光度立体问题，利用 drop-out 层来模拟阴影，在非朗伯材质上取得了优于传统方法的结果。随后 Santo 等人[68] 将 2017 年的工作进行了扩展，将表面法向的预测扩展至表面材质的预测。然而上述方法使用的是全连接网络，因而缺乏邻域信息的约束，并且其方法只能处理预先设定好光照方向的固定张数的图像。

为了使深度学习方法能实际应用于光度立体重建任务中，后续有更多的工作集中于基于卷积神经网络架构的标定光度立体网络。Chen 等人[25,69] 提出了基于全卷积的光度立体网络（PS-FCN），该方法采用全卷积网络（FCN）[65] 提取特

征和回归表面法向，并利用最大池化层特征聚合处理不定数目的输入特征。以最大池化层聚合特征的优点是可以自然地忽略那些没有被激活的特征，而保留最显著的特征，这在一定程度上可以避免阴影等区域的影响。沿着这一方向，一些以全卷积结构为基础的工作被相继提出，例如 Wang 等人[70] 提出了融合并置光的光度立体网络，采用全连接网络与最大池化特征融操作结合的方式学习表面法向，文献[71] 提出高分辨提取网络和条件卷积的光度立体深度模型，并在不同分辨率的特征上进行了多尺度最大池化特征融合。文献[72] 提出了深浅层和全局局部信息融合的深度学习方法。Taniai 和 Maehara[26] 提出了无监督的基于学习的光度立体方法，利用渲染模型逆渲染光度立体图像，并采用重建损失来最小化逆向渲染图像与原图的差异，从而获得的表面法向。然而，这一类方法较少的关注图像间光照的变化，并丢弃大量非最大值激活特征，对图像信息的利用率不高。此外，一些方法也试图替换最大池化层特征聚合这一提取方式，以获得更高的特征融合利用率，例如文献[73] 利用流形学习中等距特征映射方法[74] 对高维光度立体特征进行融合，但由于流形学习方法会截断模型的反向传播过程，因此需要先以最大池化特征聚合对模型预训练，比较繁琐。

不同于上述文献对整个图像的特征进行融合，另一类方法则利用观察图的方法逐像素的对不定数目的特征进行处理。Ikehata[75] 首先提出了观察图这一概念，其根据光照的方向将像素的强度值投射在二维的观察图上，以克服需要固定数量输入的问题，并将此观察图作为输入特征输入网络进行学习。随后 Zheng 等人[27] 和 Li 等人[76] 在此基础上利用对称性和插值回归将观察图的方法应用到稀疏输入的场景，使其在较少的光度立体图像输入下依然能取得不错的重建精度。然而这一类方法本质上依然是逐像素处理的方法，因此割裂了相邻像素的约束关系，这在一定程度上影响了重建精确度。最近 Yao 等人[77] 利用图卷积网络，将不定数目的图像进行逐像素的信息融合，并采用卷积网络对整个 patch 信息进行处理，实现了较好的表面法向重建效果。此外，Logothetis 等人[78] 提出了一种可以兼顾学习全局表示的逐像素观察图方法，改进了文献[75] 中获取观察图缓慢的问题。

2.5 其它特殊场景下的光度立体

在 2.4节中重点探讨的经典场景光度立体算法，即遵循光源方向已知、平行光源和完全暗室等条件。此外，还有一些工作将重点放在了如何突破经典场景光

度立体算法的严格约束，例如非标定光度立体、近场点光源光度立体和彩色光度立体等任务。

2.5.1 非标定光度立体

标定的光度立体方法假设光源方向已知，但是其代价是繁琐的光源方向校准，需要在数据捕获期间将额外的校准对象（标定球，标定块等工具）[79,80,81] 放置在拍摄场景中。更重要的是，放置的具有镜面反射性质的标定球球经常会引起其他物体的相互反射，对表面法向重建产生影响。而未标定光度立体算法，由于不需要标定光照方向，仍有较大的实际应用价值。

大多数现有的未标定光度立体方法基于矩阵分解[82,83,84]。限制于朗伯表面材质的情况，其利用表面可积性约束专注于解决形状光模糊问题，例如广义浅浮雕模糊性[85]。为了进一步解决这种广义浅浮雕模糊性，许多方法利用了额外的线索，如相互反射[86]、镜面反射[87]、表面材质反照率先验[88]、各向同性反射对称性[89] 或朗伯漫反射最大值[90] 等。此外一些基于流形嵌入的非标定光度立体方法[91,92,93] 可以处理具有一般双向反射率分布函数的表面（即非朗伯材质）。但是这些方法通常假设光照方向均匀分布在物体四周。

除上述基于传统方法的非标定光度立体算法外，一些学者也利用深度学习技术处理光源未知的光度立体图像。Chen 等人[25] 采用一阶段的网络，可以同时处理标定和未标定情况下的光度立体任务。然而由于此方法没有显式地预测光源的方向和强度，因此在非标定情况下重建的表面法向有较大的误差。随后一些工作采用两阶段的网络，在第一阶段首先从光度立体图像中预测光源的方向和强度，再将预测的光源信息和光度立体图像一起输入第二阶段网络，以预测物体的表面法向[94,28]。由于在第一阶段估计的光源信息可以可视化，所以这类方法在表面法向重建上有着更高的精度并具有更好的可解释性。

2.5.2 近场点光源光度立体

在一个光度立体的拍摄系统中，若照明光源距离被摄物体或场景比较近时，理想化的平行光源假设将不再成立[95]。此时应该引入更为精细的近场点光源的光照模型。在近场点光源模型中，其光照方向随空间位置的变化而变化，并且其光照强度也随距离的增加而衰减[96]。Iwahori 等人[97] 首先利用照度平方反比定律拟合衰减的近点光照。近点光照模型会使得相机成像过程变的极其复杂，即使在朗伯表面的假设下，同一像素位置上的表面法向求解在不同光照方向下也

是非线性的。就求解表面法向而言，这种非线性可以看作是一个非凸优化问题，可以采用基于迭代优化的方法来解决[98,99,100]，这些方法基于成像模型，利用前一步的预测结果，交替估计表面法向和反照率。另一类方法则基于变分的方法，将非线性嵌入到偏微分方程中[101,102]。其中 Mecca 等人[101] 首先考虑两个图像的强度比，并将问题表述为准线性偏微分方程，随后的工作[102] 则进一步将基于朗伯反射模型的假设放松到 Blinn-Phong 镜面反射模型[103]。此外最近一些学者也利用深度学习技术重建近场点光源下的表面法向，取得了更准确的预测精度[104,105]。

2.5.3 彩色光度立体

传统的光度立体方法需要拍摄三张及以上不同光照方向的图像，以求解表面法向，同时在拍摄时亦需要保证视角和物体固定。然而在实际的拍摄时，会经常遇到动态或者非刚性物体。此时，则应该引入彩色光度立体的方法（也被称为多光谱光度立体）。基于彩色光度立体的方法通过将不同的颜色通道视为独立图像来简化数据捕获，也就是说理论上仅需拍摄一幅图像就可实现光度立体的表面法向重建。通常来说，这类方法采用红绿蓝三种不同颜色的灯光，从不同方向同时照射待重建物体[10,106,107,108]。彩色光度立体面临的挑战是光源频谱、表面反射率和相机响应的混杂，这导致表面法向求解的不确定性。因此通常需要额外提供的先验信息才能求解出准确的表面法向。例如 Hernandez 等人[109] 使用同样材质下的平面进行预校准，获得了织物的准确表面法向，Anderson 等人[110] 则采用 Kinect 相机或双目立体粗略估计的三维信息作为初始输入，并反复优化迭代以求解准确的表面法向。近年来亦有部分学者将深度学习引入彩色光度立体任务。Lu 等人[111] 提出了一种基于学习的初始深度预测网络，并将深度学习预测的初始深度输入彩色光度立体算法中求解。一些端到端的一阶段彩色光度立体深度模型也被提出[112,113]，并能取得更好的表面法向重建效果。

2.6 光度立体数据集

深度学习的关键是训练数据。基于深度学习的光度立体任务训练需要物体在不同光照方向下的图像和对应的表面法向真值。然而获取真实物体的表面法向真值是一项困难且耗时的任务。例如 Shi 等人[31] 花费了三年的时间才构建了一个 10 个物体的光度立体数据集，他们利用手持三维扫描仪获得真实的三维信

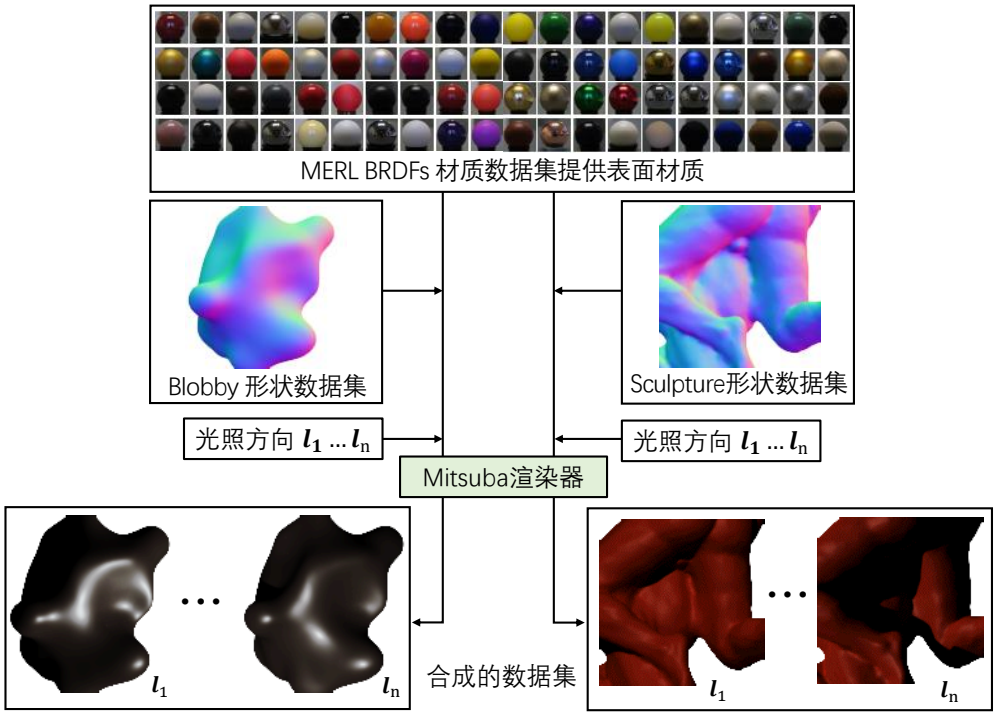


图 2-3 合成数据的渲染示意图

息，然后与拍摄图像进行再次配准。因此作为需要大量训练数据的光度立体深度模型，学者们通常采用合成的数据集进行训练，并在真实拍摄的数据集上进行测试，以验证提出的模型的有效性和鲁棒性。因此本节将光度立体数据集分为合成数据集和真实拍摄数据集两种。

2.6.1 合成数据集

合成的光度立体数据集本质上是对三维模型的渲染，因此需要三维模型、材质和指定的光照方向。例如在文献[25]中，作者利用两个三维形状数据集 Blobby 形状数据集[114] 和 Sculpture 形状数据集[115] 提供渲染的三维模型，利用 MERL BRDFs 数据集[116] 提供表面材质，并在物体上半球的空间内随机选择光照方向。如图 2-3所示，作者使用基于物理的光线追踪器 Mitsuba[117] 将三维模型、表面材质和光照渲染为所需的光度立体图像。

具体来说，MERL BRDFs 数据集[116] 包含了 100 种不同的真实世界材质 BRDF。Blobby 形状数据集[114] 包含了 10 个不同形状的物体模型。每一个物体模型，作者选择了 1296 个视角（36 个方位角 × 36 个仰角），并在 MERL BRDFs 数据集中从 100 种表面材质种随机抽取两种用来渲染模型，最终得到了 25920 个样本（$10 \times 1296 \times 2$）。对于每一个样本，作者从物体的上半球空间范围内随机采

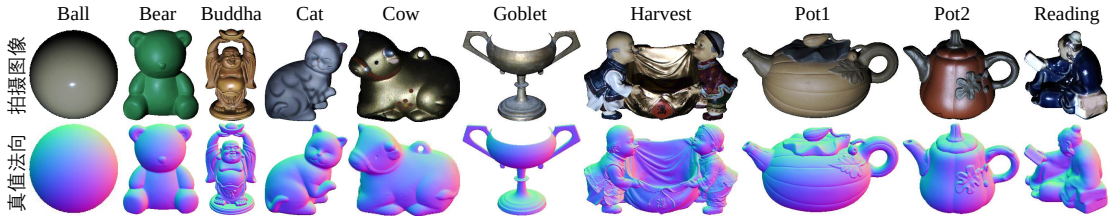


图 2-4 DiLiGenT 数据集，其包含 10 个带有真值表面法向的物体

样了 64 个不同光照方向，并渲染出 64 张分辨率为 128 × 128 的光度立体图像。此外 Sculpture 形状数据集[115] 包含了 8 个结构更为复杂的雕塑模型。Sculpture 形状数据集包含了 59292 个样本，每个样本有 64 张不同光照方向下的光度立体图像。

此外在文献[75] 中，CyclesPS 合成数据集也被提出。作者从互联网上下载了若干物体的三维模型并采用迪士尼 BSDF 材质数据集[118] 进行渲染。相比于 MERl BRDF 数据集[116]，迪士尼 BSDF 材质数据集集成了由 11 个参数控制的五类不同的 BRDF，可以更全面地表示真实世界中物体的表面材质。

2.6.2 真实拍摄数据集

DiLiGenT 数据集[31] 是被广泛使用的真实拍摄的光度立体数据集，该数据集内物体具有强非朗伯材质的表面和复杂的型状结构，如图 2-4所示，数据集包含 10 个带有法向真值的真实拍摄物体。对于每一个物体，都有 96 张来自不同光照方向下的光度立体图像。每张图像的分辨率为 512 × 612。

此外还有一些真实拍摄的光度立体数据集，例如 Gourd & Apple 数据集[55] 和 Light Stage Data Gallery 数据集[119]。其中 Gourd & Apple 数据集由三个物体组成，分别是 Gourd1、Gourd2 和 Apple，分别有 102、98 和 112 张不同光照方向下的光度立体图像。而 Light Stage Data Gallery 数据集由六个物体组成，每个物体提供了高达 253 张不同光照下的光度立体图像。然而上述这两个真实拍摄的数据集缺乏对应的表面法向真值，因此只能用于定性的实验。

2.7 本章小结

本章主要介绍了非朗伯光度立体涉及到的部分理论的基础知识和相关领域的国内外最新进展。首先本章在 2.1和 2.2中介绍了双向反射分布函数和光照成像模型，进而引出 2.3朗伯假设下的最小二乘法光度立体技术。其次本章重点介绍并分析了几种非朗伯光度立体方法的优缺点，包括基于异常值剔除的方法

（2.4.1）、基于建模复杂反射模型的方法（2.4.2）、基于 BRDF 性质的方法（2.4.3）和基于深度学习的方法（2.4.4）。随后本章还介绍了一些光度立体衍生出的相关领域（2.5）。最后在 2.6中，本章介绍了主流的光度立体数据集。在下一章中为了解决现有基于深度学习的光度立体方法在高频的表面结构（例如褶皱、边缘）中存在模糊和细节缺失的问题，本文首先提出了自适应注意力光度立体模型。

3 自适应注意力光度立体模型

3.1 研究背景

在上一章中，本文介绍了并分析了非朗伯光度立体的相关方法。传统的非朗伯光度立体方法，例如异常值剔除法[46,45,44]和建模反射模型法[53,49,12]仅能处理有限种类的表面材质且面临优化不稳定的问题[40]。近年来深度学习技术[120]亦被应用于光度立体三维重建任务中[24,25,75,71]，这些方法可以更灵活地处理非朗伯材质的物体，重建更准确的表面法向。然而现有的基于深度学习的光度立体方法在高频的表面结构（例如褶皱、边缘）中存在模糊和细节缺失的问题，这严重影响了表面法向的重建准确性。

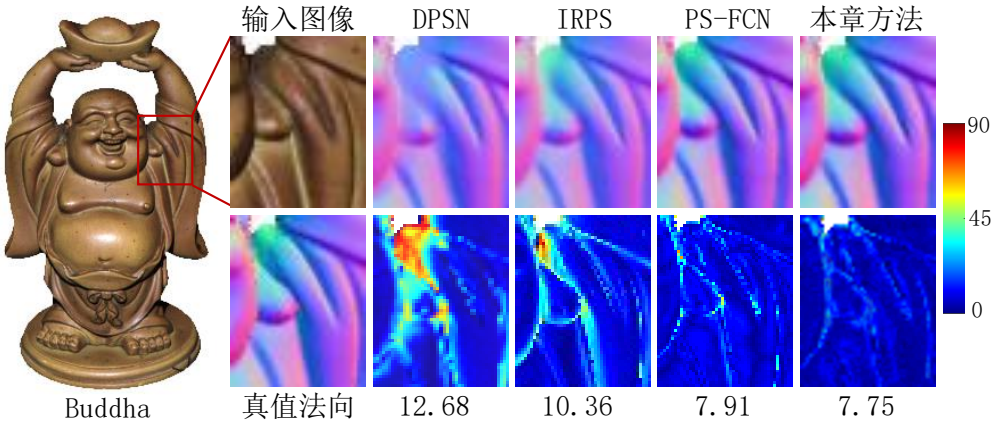


图 3-1 DiLiGenT 数据集[31]中物体 Buddha 的表面法向重建精度比较

如图 3-1所示，以 DiLiGenT 数据集[31]中物体 Buddha 为例，红色框中肩部衣服存在褶皱（具有复杂的表面结构），将本章提出的自适应注意力光度立体模型与若干基于深度学习的光度立体模型进行了比较，例如 DPSN[24]，IRPS[26]，GPS-Net[77]，PS-FCN[25]（图中第一行代表输入的图像和不同方法重建的表面法向，而第二行代表真值法向和对应方法预测的法向的角度误差图。误差图下方的数字代表预测的法向与真值法向的平均角度误差。角度误差越小，则重建精度越高），可以发现现有的基于学习的光度立体技术在褶皱、边缘等复杂形状处难以获得清晰准确的表面法向重建。这是因为先前基于深度学习的光度立体方法均采用单一的基于欧式距离的损失函数来优化网络，例如绝对值损失函数（L1 损失）、平方差损失函数（L2 损失）和余弦损失函数。而这些基于欧式距离的损失函数的优化趋向于回归总体的平均值[29]，难以约束预测图像中的高频变化，带

来图像模糊和过度平滑[121,122]。

为解决上述问题，本章提出了一种自适应注意力光度立体模型。该模型利用注意力加权的法向重建损失，在高频的复杂结构处施加合适的边缘保护损失来保证细节。通过这种方式，提出的模型显著减少了预测表面法向高频细节的模糊和误差，提高了表面法向的重建精度。

3.2 模型概述

本章提出了一种自适应注意力光度立体模型，以提高现有基于学习的方法在高频复杂结构区域（例如边缘和褶皱）重建的表面法向精度。本章的方法不是对所有法向图上的像素使用统一的损失约束，而是对每个像素采用自监督方式学习注意力权重损失，避免在这些高频区域中产生模糊的表面法向重建结果。如图 3-2所示，本章提出了自适应注意力光度立体模型。该模型分别利用表面法向生成网络（见 3.3）和注意力生成网络（见 3.4），从输入的光度立体图像和光照方向中分别生成预测的表面法向和注意力图。生成的注意力图上的像素值为注意力加权的法向重建损失提供了逐像素的权重，该权重决定了对高频细节保留更佳的梯度损失的比例（见 3.5）。详细的消融实验证明了提出的注意力加权的法向重建损失和生成网络的有效性。在多个公共真实数据集上进行的大量实验表明，与现有方法相比，本章提出的自适应注意力光度立体深度模型优于传统的光度立体算法和基于学习的方法。按照顺序，首先介绍表面法向生成网络和注意力生成网络，再介绍注意力加权的法向重建损失。

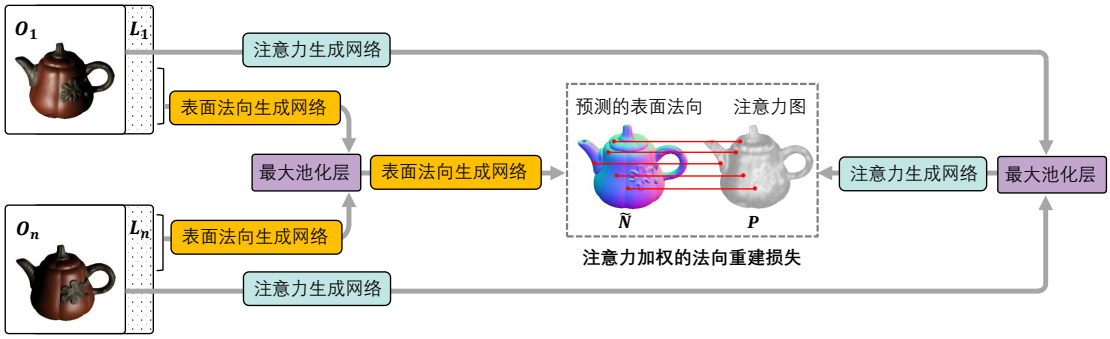


图 3-2 自适应注意力光度立体深度模型的总体结构

3.3 表面法向生成网络

表面法向生成网络旨在生成拍摄物体的表面法向图，其由特征提取器 f_{NE}、最大池化聚合层和特征回归器 f_{NR} 组成。

首先，作为标定光度立体（即已知光照方向）的表面法向预测任务，应首先将光度立体图像与光照方向信息进行融合。对于 n 张输入的光度立体图像 $\boldsymbol{O}_1$, $\boldsymbol{O}_2, \cdots, \boldsymbol{O}_n \in \mathbb{R}^{3\times H\times W}$ 和对应的光照方向 $\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n \in \mathbb{R}^3$，其中 $\mathbb{R}^{3\times H\times W}$ 代表具有 $H \times W$ 分辨率的 RGB 图像，$\mathbb{R}^3$ 代表三维笛卡尔坐标系下的方向向量。依照先前基于学习的光度立体模型的常规的做法[25,70,71,73]，将笛卡尔坐标系下每一副图像对应的光照方向 $\boldsymbol{l}_j \in \mathbb{R}^3$ 沿 H 和 W 的方向复制，扩展至与图像具有相同分辨率大小的张量 $\boldsymbol{L}_j \in \mathbb{R}^{3\times H\times W}$，其中 $j \in \{1, 2, \cdots, n\}$。随后，将图像 $\boldsymbol{O}_j$ 与扩展得到的对应光照 $\boldsymbol{L}_j$ 利用拼接操作沿第一维度拼接起来，记为张量 $\boldsymbol{\Phi}_j \in \mathbb{R}^{6\times H\times W}$。在第一维度中，前三维为光度立体图像的 RGB 通道，后三维为对应的光照方向在笛卡尔坐标系下的 x, y, z 分量。

表面法向生成网络中的 f_{NE} 可以看作是一个 n 路分支的共享权重特征提取器，首先从得到的 n 个拼接张量 $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_n \in \mathbb{R}^{6\times H\times W}$ 中生成 n 个提取的特征 $\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \cdots, \boldsymbol{\Psi}_n \in \mathbb{R}^{256\times \frac{1}{4}H\times \frac{1}{4}W}$，即：

$$\boldsymbol{\Psi}_j = f_{NE}(\boldsymbol{\Phi}_j; \theta_{NE}),\ j \in \{1, 2, \cdots, n\}, \tag{3-1}$$

θ_{NE} 表示特征提取器的可学习参数。利用设计的 f_{NE} 来提取输入张量 $\boldsymbol{\Phi}_j$ 中的特征信息。具体来说，特征提取器 f_{NE} 的主体由四个残差模块组成[123]，激活函数设置为 Leaky Relu。残差结构是一种广泛使用的特征提取结构，它可以有效地避免深度模型优化时梯度消失的问题。残差模块通过捷径连接（shortcut connections）的方式，将浅层的信息传入深层。因此深层在输入特征基础上学习到新的特征，从而拥有更好的性能。表 3-1展示了提出的特征提取器 f_{NE} 的具体结构。

在光度立体系统中，输入的光度立体图像的数量是不定的。为了保证提出的模型的实际应用价值，模型不能限定输入的光度立体图像数量以及光照方向的顺序。尽管卷积层在固定特征数目的融合中有着成功的应用，但是卷积层要求在训练和测试期间具有固定数量的输入特征通道，很难处理可变通道数的特征。对于上述可变的输入数量 n，共享权重特征提取器 f_{NE} 可用于从每个输入的 $\boldsymbol{\Phi}_j$ 中提取特征，得到 $\boldsymbol{\Psi}_j$。但是，模型需要额外的特征聚合层将 n 个特征聚合成具有固定数量的聚合特征，才能利用卷积网络对其进行特征回归，生成预测的表面法向。

为了解决光度立体深度模型存在不定数目的特征这一难题，本章采用了最

表 3-1 表面法向生成网络中特征提取器 f_{NE} 的网络结构

输入	操作	卷积核大小	步长	输出
$3 \times H \times W$	卷积层 1	3×3	1	$64 \times H \times W$
$64 \times H \times W$	卷积层 2	3×3	1	$64 \times H \times W$
$64 \times H \times W$	卷积层 3	3×3	1	$64 \times H \times W$
$64 \times H \times W$	捷径连接 1（卷积层 2、卷积层 3）			$64 \times H \times W$
$64 \times H \times W$	卷积层 4	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 5	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	捷径连接 2（卷积层 4、卷积层 5）			$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 6	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 7	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	捷径连接 3（卷积层 6、卷积层 7）			$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 8	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 9	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	捷径连接 4（卷积层 8、卷积层 9）			$256 \times \frac{1}{4}H \times \frac{1}{4}W$

大池化层这一操作，以将 n 个不固定的特征聚合为具有固定通道数的特征。池化层曾经在多个任务中被用来聚合不定数目的特征，例如 Wiles 和 Zisserman[115] 使用最大池化层来聚合来自不同视图的特征以进行三维体素的重建。Hartmann 等人[124] 则采用平均池化层来聚合特征，以学习特征间的相似性。Chen 等人也将最大池化层应用于基于学习的光度立体模型中[25,94,69]。根据文献[69] 所述，采用最大池化层作为特征聚合的方式，可以从所有特征中提取最显著的信息，而忽略掉未激活的特征（如阴影等）。而其它方法则有一定的缺陷，例如，循环神经网络（RNN）[125] 也可以处理不定数目的输入特征，但是 RNN 却对输入特征的顺序敏感，也就是说，RNN 会学习到特定的光照方向变化模式，从而使光度立体模型变的不通用；平均池化层也可以处理顺序和数量不定的特征，但是在光度立体任务中，平均池化聚合的特征就可能会被某些图像中的未激活的阴影区域所影响，而产生错误的结果。

图 3-3展示了最大池化层聚合不定数目特征的示意图，对于输入的任意 n 个特征，最大池化层保留对应位置的最大值（图 3-3中的特征数值越大则越红，反之，则以绿色表示）。在我们提出的表面法向生成网络中，特征聚合的过程可以表示为：

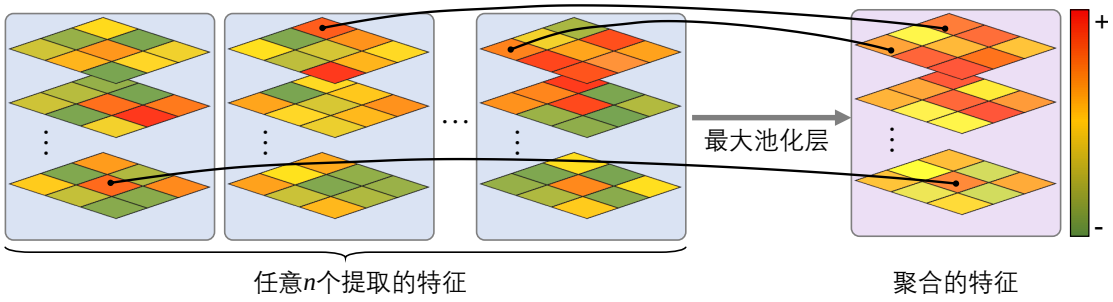


图 3-3 最大池化层聚合不定数目特征的示意图

$$\boldsymbol{\Psi}_{\max} = \bigcup_{i}^{\frac{1}{4}H\times\frac{1}{4}W} \max(\boldsymbol{\Psi}_{1,i}, \boldsymbol{\Psi}_{2,i}, \ldots, \boldsymbol{\Psi}_{n,i}), \tag{3-2}$$

其中 $\boldsymbol{\Psi}_{\max}$ 表示最大池化层聚合后的特征，下标 i 表示特征分辨率 $\frac{1}{4}H \times \frac{1}{4}W$ 中位置的索引。

在获得聚合的特征 $\boldsymbol{\Psi}_{\max}$ 后，本章提出了特征回归器 f_{NR}，以生成预测的表面法向 $\tilde{\boldsymbol{N}}$，记作：

$$\tilde{\boldsymbol{N}} = f_{NR}(\boldsymbol{\Psi}_{\max}; \theta_{NR}), \tag{3-3}$$

其中 θ_{NR} 表示特征回归器 f_{NR} 中可学习的参数。f_{NR} 由三层卷积层和两层转置卷积层组成，并在末端附加一个 L2 归一化层以生成归一化的表面法向（即逐像素点的法向向量模长为 1）。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。表 3-2展示了特征回归器 f_{NR} 的具体网络结构。

表 3-2 表面法向生成网络中特征回归器 f_{NR} 的网络结构

输入	操作	卷积核大小	步长	输出
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 1	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 1	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 2	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 2	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 3	3×3	1	$3 \times H \times W$
$3 \times H \times W$	L2 归一化层			$3 \times H \times W$

3.4 注意力生成网络

注意力生成网络旨在生成物体的注意力图，其由特征提取器 f_{AE}、最大池化聚合层和特征回归器 f_{AR} 组成。

与表面法向生成网络的特征提取器 f_{NE} 不同的是，f_{AE} 直接从图像 $\boldsymbol{O}_j$ 中提取特征，而不是从拼接的 $\boldsymbol{\Phi}_j$ 张量。这是因为注意力生成网络关注的是光度立体图像中高频信息的程度，而与光源的照射方向无关，因此特征提取器 f_{AE} 可以表示为：

$$\boldsymbol{\Gamma}_j = f_{AE}(\boldsymbol{O}_j, f_{Grad}(\boldsymbol{O}_j); \theta_{AE}),\ j \in \{1, 2, \cdots, n\}, \tag{3-4}$$

其中 θ_{AE} 表示特征提取器的可学习参数，$\boldsymbol{\Gamma}_j \in \mathbb{R}^{128\times\frac{1}{2}H\times\frac{1}{2}W}$ 是注意力生成网络中特征提取器 f_{AE} 提取的特征，$f_{Grad}(\boldsymbol{O}_j)$ 表示输入图像 $\boldsymbol{O}_j$ 的梯度信息，用于加强输入图像的高频信息。表 3-3展示了特征提取器 f_{AE} 的具体网络结构，可以看出，特征提取器有七个卷积层，其中特征图通过步长为 2 的卷积层被下采样两次，然后又通过转置卷积层上采样一次，最终的输出特征 $\boldsymbol{\Gamma}_j$ 的分辨率仅被下采样一次（$\frac{1}{2}H \times \frac{1}{2}W$）。这样操作的目的是增加网络的感受野，以提升性能。

表 3-3 注意力生成网络中特征提取器 f_{AE} 的网络结构

输入	操作	卷积核大小	步长	输出
$3\times H\times W$	卷积层 1	3×3	1	$64\times H\times W$
$3\times H\times W$	保边层 1			$3\times H\times W$
$64\times H\times W$	拼接（卷积层 1、保边层 1）			$67\times H\times W$
$67\times H\times W$	卷积层 2	3×3	1	$64\times H\times W$
$64\times H\times W$	卷积层 3	3×3	2	$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W$	卷积层 4	3×3	1	$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W$	卷积层 5	3×3	2	$256\times\frac{1}{4}H\times\frac{1}{4}W$
$256\times\frac{1}{4}H\times\frac{1}{4}W$	卷积层 6	3×3	1	$256\times\frac{1}{4}H\times\frac{1}{4}W$
$256\times\frac{1}{4}H\times\frac{1}{4}W$	转置卷积层 1	4×4	2	$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W$	卷积层 7	3×3	1	$128\times\frac{1}{2}H\times\frac{1}{2}W$

注意力生成网络同样利用最大池化层来聚合不定数量的 n 个 f_{AE} 提取的特征 $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \cdots, \boldsymbol{\Gamma}_n$，以获得聚合特征 $\boldsymbol{\Gamma}_{\max}$：

$$\mathbf{\Gamma}_{\max} = \bigcup_{i}^{\frac{1}{2}H\times\frac{1}{2}W} \max(\mathbf{\Gamma}_{1,i}, \mathbf{\Gamma}_{2,i}, \ldots, \mathbf{\Gamma}_{n,i}). \tag{3-5}$$

在获得聚合的特征 $\mathbf{\Gamma}_{\max}$ 之后，本节提出了特征回归器 f_{AR}，以生成注意图 $\boldsymbol{P}$:

$$\boldsymbol{P} = f_{AR}(\mathbf{\Gamma}_{\max}; \theta_{AR}), \tag{3-6}$$

其中 θ_{AR} 表示特征回归器 f_{AR} 中学习的参数。f_{AR} 由三层卷积层和一层转置卷积层组成,以生成注意力图 $\boldsymbol{P}$。除了最后一层的卷积层激活函数被设置为 Sigmoid 外，其余层的激活函数均为 Leaky ReLu。在最后一层卷积层使用 Sigmoid 作为激活函数的目的是 Sigmoid 可以生成 0 至 1 范围内的归一化特征。表 3-4展示了特征回归器 f_{AR} 的具体网络结构。

表 3-4 注意力生成网络中特征回归器 f_{AR} 的网络结构

输入	操作	卷积核大小	步长	输出
$128\times\frac{1}{2}H\times\frac{1}{2}W$	卷积层 1	3×3	1	$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W$	转置卷积层 1	4×4	2	$64\times H\times W$
$64\times H\times W$	卷积层 2	3×3	1	$64\times H\times W$
$64\times H\times W$	卷积层 3	3×3	1	$1\times H\times W$

3.5 注意力加权的法向重建损失

为了实现对物体表面高频的褶皱和低频的平坦区域施加不同损失约束网络的目的，本章提出了一种注意力加权的法向重建损失函数，以在不同的区域提供不同损失组合。如图 3-4所示，左图表示表面法向图，右图则表示注意力图。红色框表示高频的法向区域，这些区域有复杂的表面结构，而绿色框则表示低频的法向区域，这些区域的表面结构更加平滑。在表面法向图上，我们希望在红色框所示的复杂形状区域，模型提供更高权重的细节保护损失函数，而对于绿色框所示的平坦形状区域，模型就仅提供常用的欧式距离损失即可，例如 L2 损失和余弦损失。为了实现这一目的，本章因而引入了注意力图这一概念，利用注意力图的值提供不同损失组合的权重。

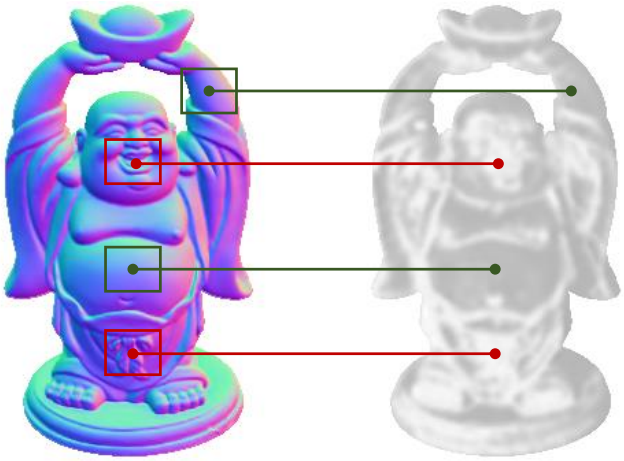


图 3-4 法向图与其对应的注意力图

注意力图是与预测的表面法向图有相同分辨率 $H \times W$ 的单通道图像，也就是说，注意力图上的每一个像素点都与重建的法向图对应。这里用下标 $_i$ 表示图像上像素位置的索引。首先，提出的注意力权重损失 $\mathcal{L}_{\text{attention}}$ 是一个逐像素的损失函数，通过最小化注意力加权的法向重建损失，可以优化上述 3.3节和 3.4节中的可学习参数 θ_{NE}、θ_{NR}、θ_{AE} 和 θ_{AR} ，它被表示为：

$$\mathcal{L}_{\text{Attention}} = \frac{1}{HW}\sum_{i}^{HW}\mathcal{L}_i, \tag{3-7}$$

其中 $\mathcal{L}_i$ 是在像素位置 i 上的损失函数，可以被表示为：

$$\mathcal{L}_i = P_i\mathcal{L}_{\text{Gradient}}(\boldsymbol{N}_i, \tilde{\boldsymbol{N}}_i) + \lambda(1-P_i)\mathcal{L}_{\text{Cosine}}(\boldsymbol{N}_i, \tilde{\boldsymbol{N}}_i), \tag{3-8}$$

其中 $\mathcal{L}_{\text{Gradient}}$ 代表梯度损失，$\mathcal{L}_{\text{Cosine}}$ 则代表余弦损失，λ 是一个超参数，用来平衡 $\mathcal{L}_{\text{Gradient}}$ 和 $\mathcal{L}_{\text{Cosine}}$ 两项损失函数，λ 的值被设置为 8。

在逐像素的损失 $\mathcal{L}_i$ 中，第一项梯度损失 $\mathcal{L}_{\text{Gradient}}$ 被定义为像素位置 i 上真值法向 $\boldsymbol{N}_i$ 和模型预测的法向 $\tilde{\boldsymbol{N}}_i$ 的梯度的绝对值误差，可以被表示为：

$$\mathcal{L}_{\text{Gradient}}(\boldsymbol{N}_i, \tilde{\boldsymbol{N}}_i) = |\triangle\boldsymbol{N}_i^x - \triangle\tilde{\boldsymbol{N}}_i^x| + |\triangle\boldsymbol{N}_i^y - \triangle\tilde{\boldsymbol{N}}_i^y|, \tag{3-9}$$

其中 $\triangle\boldsymbol{N}_i^x$，$\triangle\tilde{\boldsymbol{N}}_i^x$ 分别代表 i 像素位置上真值法向 $\boldsymbol{N}_i$ 和预测法向 $\tilde{\boldsymbol{N}}_i$ 在 x 方向上的梯度，相似地，$\triangle\boldsymbol{N}_i^y$，$\triangle\tilde{\boldsymbol{N}}_i^y$ 分别代表 i 像素位置上真值法向 $\boldsymbol{N}_i$ 和预测法向 $\tilde{\boldsymbol{N}}_i$ 在 y 方向上的梯度。梯度损失可以锐化不连续或高曲率的表面，并防止这

些高频区域被模糊[126]。因此，这里采用梯度损失作为保边损失来约束高频信息的完整性。

在逐像素的损失 $\mathcal{L}_i$ 中，第二项余弦角度损失 $\mathcal{L}_{\text{Cosine}}$ 被定义为像素位置 i 上真值法向 $\boldsymbol{N}_i$ 和模型预测的法向 $\tilde{\boldsymbol{N}}_i$ 的角度误差，可以被表示为：

$$\mathcal{L}_{\text{Cosine}}(\boldsymbol{N}_i, \tilde{\boldsymbol{N}}_i) = 1 - \boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i, \tag{3-10}$$

其中 $\cdot$ 操作代表点乘。可以看出，预测的表面法向 $\tilde{\boldsymbol{N}}_i$ 与真值法向 $\boldsymbol{N}_i$ 越相似，则其点乘 $\boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i$ 越接近 1，此时式 (3-10) 的值越接近 0。

然而需要注意的是，式 (3-9) 所示的梯度损失，只约束了相邻像素间法向的变化程度，而不能约束该像素法向本身的方向向量（即该像素的值的大小）。因此盲目地添加梯度损失 $\mathcal{L}_{\text{Gradient}}$ 其实会稀释余弦角度损失 $\mathcal{L}_{\text{Cosine}}$ 的对法向量角度直接的约束作用，导致更大的误差。事实上，仅采用梯度损失 $\mathcal{L}_{\text{Gradient}}$ 会导致光度立体深度模型的不收敛，这也是在逐像素的损失 $\mathcal{L}_i$ 中第二项余弦角度损失前的权重处添加一个超参数 λ，并经验地设置为 8 的原因，这在 3.6.2中的实验中有更详细的讨论。但是在物体表面的复杂结构区域，又必须添加一定的保边损失（梯度损失）来使得恢复的表面法向足够清晰。因此本章提出的注意力加权的法向重建损失可以依据所处理的表面复杂程度（频率高低）以施加不同的梯度损失权重，以获得更好的重建结果。

3.6 实验结果

本节对提出的自适应注意力光度立体模型进行了实验验证。首先对提出的方法进行了消融实验与分析，随后在 DiLiGenT 数据集[31] 上将提出的模型与先前的传统光度立体方法和基于深度学习的方法进行了比较，实验结果证明了提出方法的有效性。为了合理地评估预测的表面法向，采用了广泛使用的平均角度误差 (MAE) 来衡量重建的准确性，其计算公式为：

$$\text{MAE} = \frac{1}{T}\sum_{i}^{T} \cos^{-1}\left(\boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i\right), \tag{3-11}$$

其中 T 为图像掩模中物体表面所在位置的像素总数，其不包括背景位置上的像素。此外由于复杂结构区域的表面法线误差通常更大，我们还测量了法向图中角度误差小于 15° 和 30° 的像素占物体表面所有像素的百分比，分别表示为 $err_{<15^\circ}$

和 $err_{<30^\circ}$，其可以更好地反映出高频区域的误差。

3.6.1 实验设置

本章提出的自适应注意力光度立体模型使用默认的 Adam 优化器进行优化（β_1 = 0.9 and β_2 = 0.999），初始的学习率则被设为 0.002 且每 5 个 epoch 减半。训练的数据集为采用 MERL BRDFs 数据集[116] 渲染的 Blobby 形状数据集[114] 和 Sculpture 形状数据集[115]，总计 84360 个用于训练的样本。输入的样本的分辨率为 32 × 32 且每一个样本在训练时都有 32 张不同光照方向的 32 × 32 光度立体图像作为输入，即图 3-2中的 n 为 32。在单个 RTX 3080Ti 显卡上训练了 40 个 epoch 且采用的 batchsize 为 32。

3.6.2 消融实验与分析

本节使用 MAE、err_{15° 和 err_{30° 三个指标，对提出的自适应注意力光度立体模型进行了消融实验，实验数据来自 Blobby 形状数据集[114] 和 Sculpture 形状数据集[115] 的验证集中的 852 样本（每个样本采用 32 张不同光照的输入图像）。我们通过将注意力加权的法向重建损失与固定组合损失以及常规的余弦损失进行比较来评估注意力权重损失在表面法向重建中的有效性。实验结果如表 3-5所示。数字代表验证集中所有样本的平均 MAE，以度为单位（越低越好）。百分比表示角度误差小于 15° 或 30°（越高越好）的像素的比例。最佳效果以粗体显示。

表 3-5 不同损失函数的比较

方法	MAE	err_{15°	err_{30°
仅余弦损失 $\mathcal{L}_{\text{Cosine}}$	13.10	81.25%	92.32%
仅梯度损失 $\mathcal{L}_{\text{Gradient}}$	82.41	0.33%	2.93%
$\mathcal{L}_{\text{Cosine}}$+$\mathcal{L}_{\text{Gradient}}$	15.48	80.91%	92.80
注意力加权的法向重建损失 $\mathcal{L}_{\text{Attention}}$	**11.77**	**83.07%**	**93.49%**

如表 3-5所示，注意力加权的法向重建损失 $\mathcal{L}_{\text{Attention}}$ 在所有指标中的表现始终优于其他损失。较高的 err_{15° 和 err_{30° 意味着更少的像素位置上的重建法向有较大的误差。在这些结构复杂的区域中，注意力加权的法向重建损失通过保持高频信息的完整性而减少重建表面法向的模糊程度。同时，注意力加权的法向重建损失在 MAE 方面也表现最好。这是因为注意力权重损失在平坦（低频）区域，

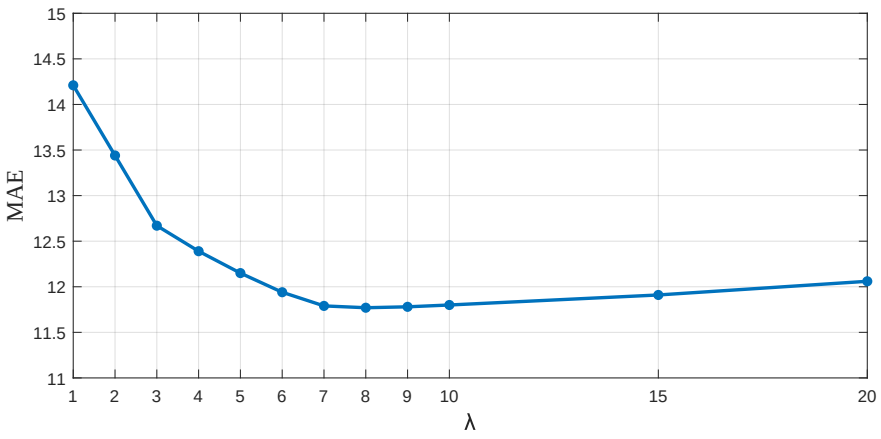


图 3-5 注意力加权的法向重建损失中使用不同的超参数 λ 对表面法向重建精度的影响

有更低的梯度损失权重，避免了对重建表面法向的不利影响。

有趣的是，在使用固定组合（1：1 比例）的余弦损失和梯度损失时，可以发现 MAE 变的更差，但与仅使用余弦损失 $\mathcal{L}_{\text{Cosine}}$ 相比，err_{30° 取得了更好的效果。由于当相邻像素之间存在较大的不连续差异时，就会激活梯度损失，而较大的角度误差主要存在于物体的边缘和结构复杂的区域。因此，添加固定比例的梯度损失必然可以对这些区域带来更好的约束。但是与仅使用余弦损失相比，在衡量图像中所有像素上的法向准确性上（即 MAE 指标）显示出更差的结果。

此外，也可以看出，仅使用梯度损失 $\mathcal{L}_{\text{Gradient}}$ 无法使我们的网络收敛。正如3.5所述，梯度损失只约束了相邻像素间法向的变化程度，而不能约束该像素法向本身的方向向量（即该像素的值的大小），因而无法优化网络。这也可以解释为什么在 $\mathcal{L}_{\text{Cosine}}$+$\mathcal{L}_{\text{Gradient}}$（固定组合损失）中 MAE 会变的更差，因为梯度损失稀释了余弦损失对预测表面法向直接的约束。

由于存在梯度损失无法在实质上优化表面法向这一问题，因此在提出的注意力加权损失 $\mathcal{L}_{\text{Attention}}$ 中，需要保证足够的余弦损失这一前提，如式 (3-8) 所示，在余弦损失项的权重中，模型额外添加了一个超参数 λ 作为保护性阈值。图3-5展示了注意力加权的法向重建损失中超参数 λ 对表面法向重建精度的影响。

如图 3-5所示，不同的超参数 λ 会显著影响提出方法的重建精度。具体来说，实验测试了 λ 取 1 至 20 的范围内的重建法向误差结果。当 λ 的值取 8 时，可以得到误差最小的表面法向。注意，当 λ 为 1 时，相当于提出的注意力加权的法向重建损失没有任何保护性约束，即注意力图的权重即为梯度损失的权重，而此

时提出的方法重建结果 MAE 比较差，甚至单独使用余弦损失 $\mathcal{L}_{\text{Cosine}}$ 时的精度。但是，其误差小于固定 1：1 组合的损失 $\mathcal{L}_{\text{Cosine}}+\mathcal{L}_{\text{Gradient}}$，这证明我们提出的注意力图提供权重方式优于固定的 1：1 权重的方式。图 3-5实验表明，在注意力加权的法向重建损失的框架下，较小的梯度损失 $\mathcal{L}_{\text{Gradient}}$ 占比（即较大的 λ 值）可以提高重建的精度，而增大梯度误差损失比例（即较小的 λ 值）则会显著的增大重建法向的误差。这是因为梯度误差损失可以保证预测表面法向的细节信息，提高清晰度。尽管如此，梯度误差损失却不能约束表面法向本身的值，过大的比例会稀释余弦损失对预测的表面法向的约束。因此，通过实验确定最合理超参数 λ 的值能才够获得最佳的表面法向重建效果。

3.6.3 DiLiGenT 数据集对比实验结果

为了显示自适应注意力光度立体模型的有效性，本章在多个真实拍摄的数据集上与其他方法做了对比实验。首先，本节在 DiLiGenT 数据集[31] 上进行了比较。如 2.6.2中所述，DiLiGenT 数据集包含十个真实的具有较强非朗伯特性和复杂结构的物体。本节将提出的自适应注意力光度立体模型与多个传统方法（以作者的姓氏的第一个字母 + 年份命名，LS 则代表最小二乘的基准方法[9]）和基于深度学习的方法（以网络简称命名）进行了广泛的比较。表 3-6展示了在 96 张输入的光度立体图像下各个方法对 DiLiGenT 数据集[31] 中十个物体的表面法向重建的 MAE 值。粗体的值代表最佳性能，而下划线的值代表次佳性能。图 3-6和 3-7进一步展示了可视化的结果。在图 3-6中，红色框是具有复杂结构（高频信息）的区域。

表 3-6 和图 3-6将提出的自适应注意力光度立体模型与传统的校准光度立体方法和基于深度学习的光度立体方法进行了比较。我们提出的方法在大多数对象上实现了最先进的结果，在使用 96 张光度立体图像作为输入时，平均的 MAE 为 7.92。在图 3-7中可以看出，输出的注意力图准确的体现了高频褶皱和边缘区域的位置，并对这些区域提供了更高的权重值。

图 3-6展示了一些带有红色框的例子，例如 Buddha 的腰带以及 Pot2 的花朵。可以看出，在这些复杂的区域注意力图被激活，产生较高的权值。因此与其他方法相比，本章提出的方法的误差图在这些区域中显示出较低的角度误差。在这些区域中，注意力权重损失中梯度损失的权重较大，说明提出的方法学习到了高频信息完整性的模式。提出的方法重建的表面法向保持清晰的边缘并减少模糊。

图 3-6 DiLiGenT 数据集中[31] 物体 Buddha、Bear、Pot2 和 Harvest 四个物体的可视化比较

表 3-6 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均使用 96 幅图像进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
LS[9]	4.10	8.39	14.92	8.41	25.60	18.50	30.62	8.89	14.65	19.80	15.39
ST12[13]	13.58	19.44	18.37	12.34	7.62	17.80	19.30	10.37	9.84	17.17	14.58
IW12[47]	2.54	7.32	11.11	7.21	25.70	16.25	29.26	7.74	14.09	16.17	13.74
WG10[46]	2.06	6.50	10.91	6.73	25.89	15.70	30.01	7.18	13.12	15.39	13.35
HM10[57]	3.55	11.48	13.05	8.40	14.95	14.89	21.79	10.85	16.37	16.82	13.22
AZ08[55]	2.71	5.96	12.54	6.53	21.48	13.93	30.50	7.23	11.03	14.17	12.61
GC10[52]	3.21	6.62	14.85	8.22	9.55	14.22	27.84	8.53	7.90	19.07	12.00
IA14[59]	3.34	7.11	10.47	6.74	13.05	9.71	25.95	6.64	8.77	14.19	10.60
ST14[58]	1.74	6.12	10.60	6.12	13.93	10.09	25.44	6.51	8.78	13.63	10.30
SPLINE-Net[27]	4.51	5.28	10.36	6.49	7.44	9.62	17.93	8.29	10.89	15.50	9.63
DPSN[24]	2.02	6.31	12.68	6.54	8.01	11.28	16.86	7.05	7.86	15.51	9.41
IRPS[26]	**1.47**	5.79	10.36	**5.44**	**6.32**	11.47	22.59	**6.09**	7.76	**11.03**	8.83
LMPS[76]	2.40	5.23	9.89	6.11	7.98	8.61	16.18	6.54	7.48	13.68	8.41
PS-FCN[25]	2.82	7.55	7.91	6.16	7.33	8.60	15.85	7.13	7.25	13.33	8.39
Manifold-PSN[73]	3.05	6.31	**7.39**	6.22	7.34	8.85	**15.01**	7.07	7.01	12.65	8.09
提出的方法	2.93	**4.86**	7.75	6.14	6.86	**8.42**	15.44	6.92	**6.97**	12.90	**7.92**

图 3-7 DiLiGenT 数据集中[31] 不同物体通过提出的自适应注意力光度立体模型预测的注意力图、重建的表面法向和对应的角度误差图

相比之下，IRPS[26]、PS-FCN[25] 和 DPSN[24]，仅使用单个角度损失，在具有高频的表面法向区域中表现不佳。这是因为传统余弦损失的采样平滑了高频信息的表达[29]。此外在图 3-7中，本节将 DiLiGenT 数据集[31] 生成的注意力图也进行

了可视化，可以看出，注意力图被激活的位置准确的出现在了高频的表面结构区域，这使得注意力权重损失在这部分区域提供更高的边缘保护损失，从而获得更清晰的表面重建结果。

此外可以看出，提出的方法在物体 Ball 上没有达到最佳性能，这是一个具有特别简单结构的物体。如图 3-8所示，黄色框是镜面反射的区域，具有非常平滑的表面结构，而提出的自适应注意力光度立体网络却生成了被激活的注意力图，其原因可能是镜面反射在非常简单的结构时误导了注意力网络，因为镜面反射是唯一的高频信息，也导致了较大的梯度变化。不过，从表 3-6中，也可以看出，几乎所有的基于深度学习光度立体方法，都难以在物体 Ball 上取得比传统方法更好的结果，这可能是因为基于深度学习方法大多采用基于卷积神经网络的架构，在非常简单的、没有互相反射的物体上，过大的感受野非但不能从邻域信息中收益，反而会带来模糊和误差，导致整体结果变差。

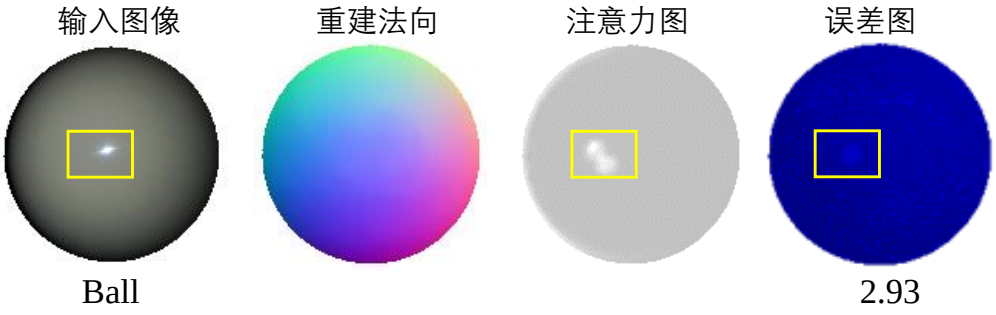


图 3-8 在物体 Ball 上的结果，黄色框代表存在镜面反射的区域

3.6.4 其他数据集实验结果

除了在 DiLiGenT 数据集上进行了比较，本节还在真实拍摄的 Light Stage Data Gallery 数据集[119] 上进行了实验。图 3-9显示了提出的模型的结果，以进一步证明其可迁移性。由于缺乏表面法向真值，本节定性地展示了模型的重建结果。Light Stage Data Gallery 数据集[119] 由六个物体组成，为每个对象提供 253 个图像和相应的光照方向以及强度。本实验选择 $n = 144$ 作为输入的光度立体图像数量。

如图 3-9所示，重建的表面法向准确的反映了物体的形状。红色框是具有高频信息的区域。例如物体 Knight standing 的裙子是由粗糙的布料材料制成的。可以看出，模型的结果也显示了该区域的粗糙表面法向，以及注意力图中较高的权重。同样，袖子区域也证明了本章方法的准确性和有效性。此外物体 Knight

图 3-9 Light Stage Data Gallery 数据集[119] 上的定性结果。

fighting 的表面法向和注意力图带有一些噪声。这可能是由于 Light Stage Data Gallery 数据集[119] 中的图像质量比较差。相机的 CCD 传感器无法抑制黑暗环境中的噪点。因此输入图像中存在的高频噪声有可能会激活注意力图，进而可能影响提出的模型。

3.7 本章小结

本章提出一种自适应注意力光度立体模型。消融实验表明，注意力加权的法向重建损失有利于更准确的重建，特别是在结构复杂的区域。对公共 DiLiGenT 数据集[31] 的大量实验表明，模型在校准的光度立体任务中取得了良好的表面法向重建结果。在 DiLiGenT 数据集中，模型的平均 MAE 为 7.92 度。可视化的比较表明提出的模型具有处理复杂结构区域的能力，模型可以在高频区域实现最佳的表面法向重建。此外，所提出的注意力加权的法向重建损失还可以为其他回归任务提供框架，例如深度估计和图像增强。在这些任务中，注意力权重损失可以学习自适应的惩罚，以减少模糊，输出清晰的估计结果。

尽管如此，自适应注意力光度立体模型也有一些缺陷。例如模型重建的表面法向精度比一些最新的基于学习的光度立体立体方法差，例如 PS-FCN (Norm.)[69]

和 CNN-PS[75]。这是因为本章的模型难以区分激活注意力图，例如物体的平坦表面上区域具有剧烈的材质变化，则依然可以导致较高的梯度损失权重，而这些区域的表面法向真值应该非常的平滑，因此会造成较大的误差。在一下章中，本文提出了归一化的高频增强光度立体模型，以解决上述问题，并进一步提高了表面法向重建精度。

4 归一化的高频区域增强光度立体模型

4.1 研究背景

上一章提出了一种自适应注意力的光度立体模型。该模型利用注意力加权的法向重建损失赋予不同频率的区域不同权重的损失组合，以获得清晰准确的表面法向重建结果。尽管上一章中提出的模型在复杂的表面结构区域可以取得更清晰的重建结果，但是该模型也遇到了相应的问题。如图 4-1所示，在 DiLiGenT 数据集[31] 的物体 Cat 上，背部的黄色框区域存在暗色的花纹，这是由于物体表面存在空间变化的材质造成的。可以看到该区域对应的注意力图上，也出现了较高的权重值。这是因为变化的材质在图像中也表现为梯度的变化（由于材质的突然变化导致相邻像素不连续），因此也可以激活注意力图，使其产生较高的权值。然而此处的表面法向并没有剧烈变化，即物体的形状结构是平坦的。但是被激活的注意力图却在此处依然提供了较高的梯度损失权重。较高的梯度损失稀疏余弦损失对法相方向直接的约束，导致更差的结果。

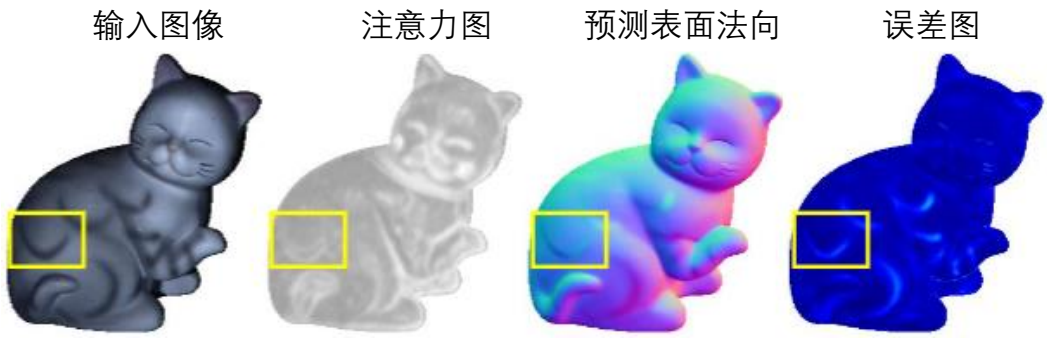


图 4-1 第 3章中提出的自适应注意力光度立体模型在物体 Cat 上的重建结果

为此，本章提出了一种归一化的高频区域增强光度立体模型，对复杂结构和材质变化的表面都能取得精确的表面法向重建结果。利用光度立体图像的归一化操作，消除剧烈变化的表面材质带来的图像高频表达。实验表明，本章提出的归一化的高频区域增强光度立体模型可以准确清晰地恢复出物体的复杂三维结构区域和材质剧烈变化区域的表面法向，并在多个广泛使用的数据集上达到最佳的表面法向重建精度。

4.2 模型概述

本章提出了一种归一化的高频区域增强光度立体模型，以改进标定光照的光度立体任务中表面法向重建，尤其是对于那些复杂结构的预测，其框架如图4-2所示。在第 3章中提出的注意力加权的法向重建损失基础上，本章进一步对观测的光度立体图像采用了归一化处理，明确区分高频表示是由表面复杂结构激活还是空间变化的表面材质激活，并消除空间变化的表面材质对注意力权重损失的影响。此外，本章采用并行高分辨率结构生成深度特征和多尺度的最大池化层聚合不同分辨率下学习到的特征，以提高表面法向的高分辨率细节。实验部分对提出方法的每个部分进行了详细的网络分析和消融分析。在公共基准数据集上进行的大量实验表明，本章提出的归一化的高频区域增强光度立体模型显著优于传统的校准光度立体算法和最先进的基于深度学习的方法。

本章将详细介绍各个提出的部分。由于注意力生成网络和注意力加权的法向重建损失已经在上一章中的 3.4和 3.5详细介绍，本章不再赘述。因此按照顺序，首先介绍观察图像归一化的方法，再介绍归一化的高频区域增强光度立体模型中的高分辨三维结构生成网络。

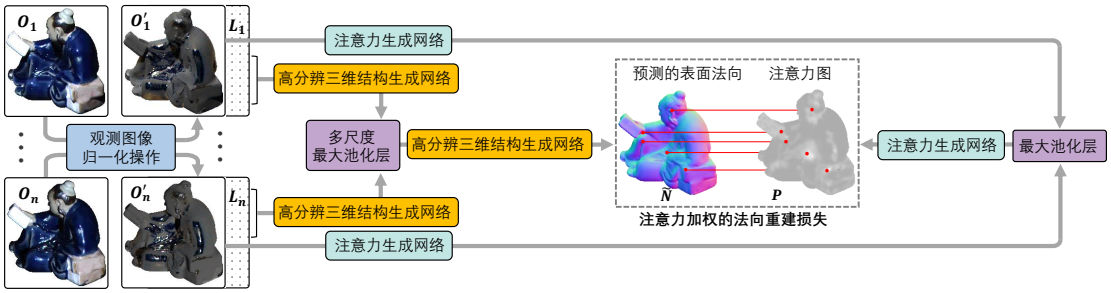


图 4-2 归一化的高频区域增强光度立体模型的总体结构

4.3 观察图像归一化操作

事实上，一个真实世界中的物体难免包含空间变化的材质，例如图 4-1中物体 Cat 背部的条纹，它们在第 3章中提出的自适应注意力光度立体模型中也被误认是高频率的复杂结构区域。然而，在这些空间变化的材质区域中相应的表面法向应该是平滑的，并且不应随着表面反射率的变化而改变。因此，这些区域对第3章中提出的模型有负面影响。

为了解决这个问题，本章提出的方法采用观察图像归一化方法来消除空间变化的材质的影响，其有效性已在先前的光度立体任务[69,93,91] 中得到证明。

给定观察图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n$ 中相同位置的像素 $o_1, o_2, \cdots, o_n$，对于第 j 张图像中的同一置上的像素 o_j 来说，其归一化操作如下式所示：

$$o'_j = \frac{o_j}{\sqrt{o_1^2 + o_2^2 + \cdots + o_n^2}},\ j \in \{1, 2, \cdots, n\}, \tag{4-1}$$

其中 o'_j 表示归一化后的像素强度。对图像中所有位置的像素做上述归一化操作，即可得到归一化后的图像 $\boldsymbol{O'}_1, \boldsymbol{O'}_2, \cdots, \boldsymbol{O'}_n$。注意，对像素 o_j 归一化时，应分别处理其 RGB 通道的值，即式 (4-1) 处理的是单通道的灰度值。

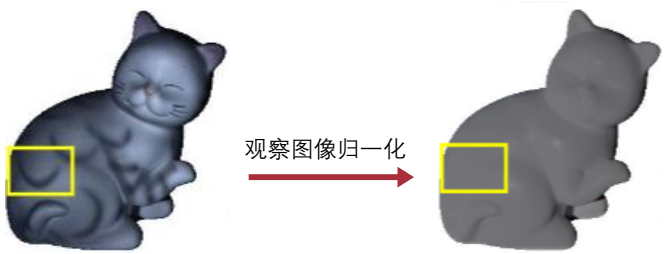
若物体表面材质具有朗伯反射特性，那么式 (2-4) 中的反射率 $\boldsymbol{\rho}(\theta_i, \phi_i, \theta_r, \phi_r)$ 会退化成一个常数 ρ。在不考虑全局光照的影响下，式 (2-4) 退化为：

$$o = \rho \max(\boldsymbol{n}^\top \boldsymbol{l}, 0), \tag{4-2}$$

将其带入归一化的式 (4-1) 中，即可得到：

$$o'_j = \frac{\max\left(\boldsymbol{n}^\top \boldsymbol{l}_j, 0\right)}{\sqrt{\max\left(\boldsymbol{n}^\top \boldsymbol{l}_1, 0\right)^2 + \max\left(\boldsymbol{n}^\top \boldsymbol{l}_2, 0\right)^2 + \cdots + \max\left(\boldsymbol{n}^\top \boldsymbol{l}_n, 0\right)^2}}, \tag{4-3}$$

可以看出，反射率 ρ（表面材质）的影响被消除，归一化后的像素点的值与其材质无关。图 4-3展示了对 DiLiGenT 数据集[31] 中物体 Cat 的归一化处理。经过归一化操作后，可以看出，黄色框中条纹的影响在归一化后的图中被全消除。



Cat

图 4-3 物体 Cat 的原始光度立体图像与经过图像归一化处理后的图像

然而，式 (4-2) 中所示的朗伯假设通常不存在。因此，本节采用的观察图像归一化方法推导的朗伯假设下的式 (4-3) 在非朗伯表面似乎不成立。但是实验证明，观察图像归一化在非朗伯表面的图像中依然取得非常好的效果。其原因有多个方面。第一，非朗伯表面中，镜面反射和投射阴影等非朗伯特性所在的区域是稀疏的，表面大部分的区域仍然可以被近似认为是朗伯反射[58]，因此式 (4-3)

依然有效。第二，对于在某些光照方向下具有镜面高光的区域，在其他光照方向下并不具有镜面反射（即物体表面的高光随着光照方向的改变而改变）。经过式(4-1) 的观察图像归一化操作后，在其他光照方向下没有镜面反射的像素点的值会被抑制（分母中由于某个 m_j 存在镜面反射会变大）[69]。在图 4-4中，物体 Ball 的例子直观地展示了这种影响，其中黄色框表示原图中存在镜面反射的区域，可以看出该光照方向下没有镜面反射的像素点的归一化后的值被抑制。然而，本章提出的模型采用了最大池化层的特征聚合方式，仅保留相同位置上的最大值特征。因此这些从观察图像归一化操作中被抑制的特征在后续的最大池化层中会被自然的丢弃，而不影响表面重建的精度。

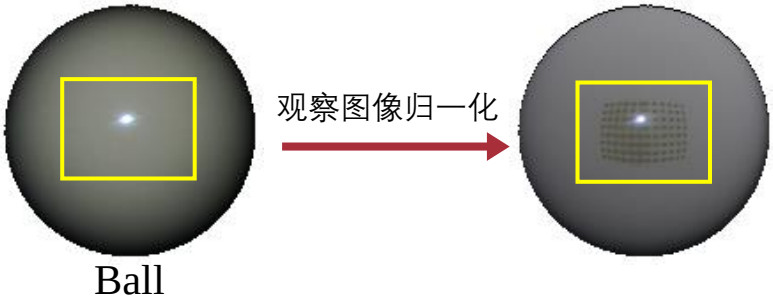


图 4-4 物体 Ball 的原始光度立体图像与经过图像归一化处理后的图像

4.4 高分辨三维结构生成网络

高分辨三维结构生成网络旨在从任意数量 n 的光度立体图像及对应的光照方向中重建出物体表面的法向 $\tilde{\boldsymbol{N}}$。高分辨三维结构生成网络由三部分组成，分别是高分辨特征提取器 f_{GE}，多尺度最大池化层特征聚合和特征回归器 f_{GR}，其具体的网络结构示意图如图 4-5所示。

作为标定的光度立体任务，模型首先将观察图像归一化的光度立体图像 $\boldsymbol{O'}_1, \boldsymbol{O'}_2, \cdots, \boldsymbol{O'}_n \in \mathbb{R}^{3\times H\times W}$ 与扩展后的光照方向 $\boldsymbol{L}_1, \boldsymbol{L}_2, \cdots, \boldsymbol{L}_n \in \mathbb{R}^{3\times H\times W}$ 利用拼接操作沿第一维度拼接起来，记为张量 $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_n \in \mathbb{R}^{6\times H\times W}$，其中第一维度中前三维为归一化的光度立体图像的 RGB 通道，后三维为光照方向的 x, y, z 分量。

高分辨三维结构生成网络的高分辨特征提取器 f_{GE} 可以看作是 n 路分支的共享权重特征提取器，可以表示为：

$$\boldsymbol{\Psi}_j^{fr}, \boldsymbol{\Psi}_j^{hr}, \boldsymbol{\Psi}_j^{qr} = f_{GE}(\boldsymbol{\Phi}_j; \theta_{GE}),\ j \in \{1, 2, \cdots, n\}, \tag{4-4}$$

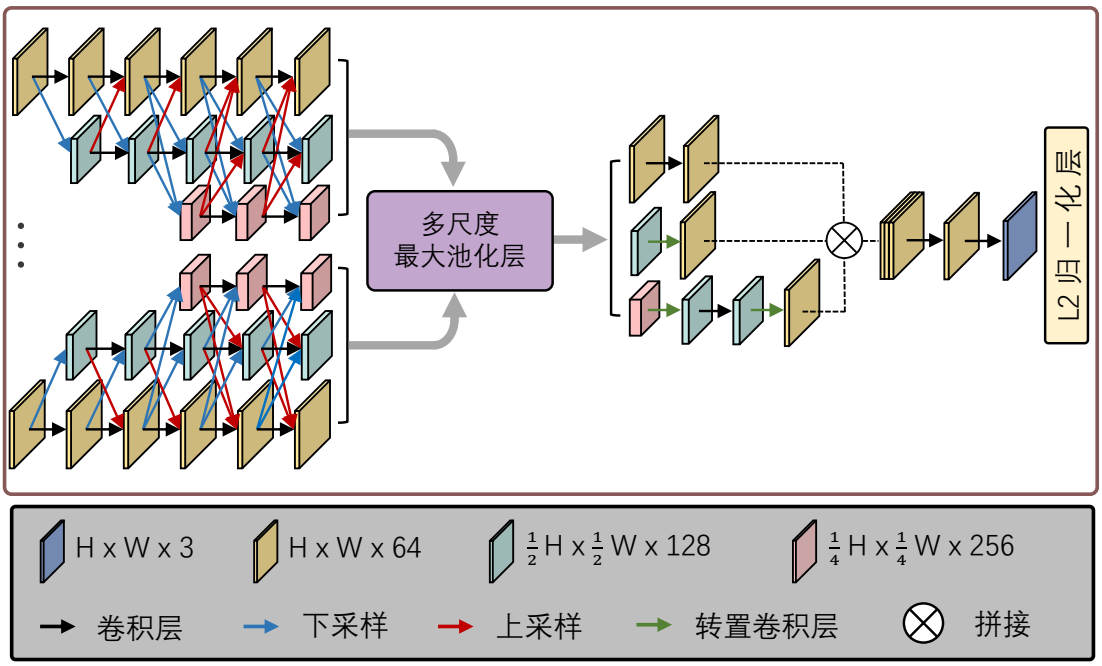


图 4-5 高分辨三维结构生成网络的高分辨特征提取器 f_{GE}、多尺度最大池化层特征聚合和特征回归器 f_{GR} 的详细结构

其中 θ_{GE} 表示高分辨特征提取器 f_{GE} 中的可学习参数，f_{GE} 同时提取三个不同尺度的特征，包括全分辨率特征 $\mathbf{\Psi}_i^{fr} \in \mathbb{R}^{64\times H\times W}$，二分之一分辨率特征 $\mathbf{\Psi}_i^{hr} \in \mathbb{R}^{128\times \frac{1}{2}H\times \frac{1}{2}W}$，四分之一分辨率特征 $\mathbf{\Psi}_i^{qr} \in \mathbb{R}^{256\times \frac{1}{4}H\times \frac{1}{4}W}$，即图 4-5的图例中黄、绿、红三种颜色的方块。为了实现三种不同尺度的特征提取，受到人体姿态估计任务[127] 中 HR-Net 启发，模型采用并行网络结构提取三个尺度的特征，而非从高分辨率层传递到低分辨率层，然后再增加分辨率的串联连接操作。并行的特征提取器可以保证高分辨率的特征始终不被降采样，因此可以同时保留表面法向的深层特征和高分辨率细节。实验（4.5.2）证明，提取高分辨率特征对于逐像素表面法向重建的准确性至关重要。

如图 4-5所示，下采样的操作通过卷积层执行，从全分辨率特征 $\mathbf{\Psi}_j^{fr}$ 至半二分之一辨率特征 $\mathbf{\Psi}_j^{hr}$ 和从二分之一分辨率特征 $\mathbf{\Psi}_j^{hr}$ 至四分之一分辨率特征 $\mathbf{\Psi}_j^{qr}$ 的降采样操作采用步长为 2 的卷积层，而从全分辨率特征 $\mathbf{\Psi}_j^{fr}$ 至四分之一分辨率特征 $\mathbf{\Psi}_j^{qr}$ 的降采样操作采用采用步长为 4 的卷积层。而上采样的操作通过双线性插值和 1×1 的卷积层来实现。其中双线性插值的作用是扩大分辨率，而随后的 1×1 的卷积层的作用则是调整到相应的特征通道数。相似地，存在两倍的上采样（$\mathbf{\Psi}_j^{qr}$ 到 $\mathbf{\Psi}_j^{hr}$、$\mathbf{\Psi}_j^{hr}$ 到 $\mathbf{\Psi}_j^{fr}$）和四倍的上采样（$\mathbf{\Psi}_j^{qr}$ 到 $\mathbf{\Psi}_j^{fr}$）两种操作。

此外，将高分辨率特征到低分辨率特征和从低分辨率特征到高分辨率特征的融合成相同分辨率的特征是通过捷径连接而不是拼接操作执行的。在特征提取器 f_{GE} 中，所有的不改变尺度的卷积层均采用步长为 1 的 3×3 卷积核，激活函数为 Leaky Relu。

为了使提出的高分辨三维结构生成网络能够处理任意数量的三种尺度的特征，本章提出了多尺度的最大池化层来聚合 n 个全分辨率、二分之一分辨率和四分之一分辨率特征，若以下标 i 表示像素位置上的索引，则多尺度的最大池化层可以表示为：

$$\boldsymbol{\Psi}_{\max}^{fr} = \bigcup_{i}^{H \times W} \max(\boldsymbol{\Psi}_{1,i}^{fr}, \boldsymbol{\Psi}_{2,i}^{fr}, \ldots, \boldsymbol{\Psi}_{n,i}^{fr}), \tag{4-5}$$

$$\boldsymbol{\Psi}_{\max}^{hr} = \bigcup_{i}^{\frac{1}{2}H \times \frac{1}{2}W} \max(\boldsymbol{\Psi}_{1,i}^{hr}, \boldsymbol{\Psi}_{2,i}^{hr}, \ldots, \boldsymbol{\Psi}_{n,i}^{hr}), \tag{4-6}$$

$$\boldsymbol{\Psi}_{\max}^{qr} = \bigcup_{i}^{\frac{1}{4}H \times \frac{1}{4}W} \max(\boldsymbol{\Psi}_{1,i}^{qr}, \boldsymbol{\Psi}_{2,i}^{qr}, \ldots, \boldsymbol{\Psi}_{n,i}^{qr}), \tag{4-7}$$

其中 $\boldsymbol{\Psi}_{\max}^{fr}$、$\boldsymbol{\Psi}_{\max}^{hr}$ 和 $\boldsymbol{\Psi}_{\max}^{qr}$ 为最大池化层后的三个不同尺度的聚合特征，包含从不同尺度中提取的最显着的特征信息。

随后，为了从三个不同尺度的聚合特征中重建物体的表面法向，本节设计了特征回归器 f_{GR}，即从 $\boldsymbol{\Psi}_{\max}^{fr}$、$\boldsymbol{\Psi}_{\max}^{hr}$ 和 $\boldsymbol{\Psi}_{\max}^{qr}$ 中生成预测的表面法向 $\tilde{\boldsymbol{N}}$，记作：

$$\tilde{\boldsymbol{N}} = f_{GR}(\boldsymbol{\Psi}_{\max}^{fr}, \boldsymbol{\Psi}_{\max}^{hr}, \boldsymbol{\Psi}_{\max}^{qr}; \theta_{GR}), \tag{4-8}$$

其中 θ_{GR} 表示特征回归器 f_{GR} 中可学习的参数。由于需要从三个不同尺度的特征中回归表面法向，因此 f_{GR} 的网络结构首先需要将三个尺度的特征通过转置卷积层处理至相同的分辨率。如图 4-5所示，f_{GR} 对全分辨率特征 $\boldsymbol{\Psi}_{\max}^{fr}$ 仅做一次步长为 1 的 3×3 的卷积，以进一步提取特征，对二分之一分辨率特征 $\boldsymbol{\Psi}_{\max}^{hr}$ 则进行一次转置卷积操作，以生成 $64 \times H \times W$ 大小的特征，而对四分之一分辨率特征 $\boldsymbol{\Psi}_{\max}^{qr}$ 进行两次转置卷积操作和一次步长为 1 的 3×3 的卷积操作，以更好从 $256 \times \frac{1}{4}H \times \frac{1}{4}W$ 生成 $64 \times H \times W$ 大小的特征。随后，f_{GR} 将生成的三个 $64 \times H \times W$ 特征沿着第一维拼接为 $192 \times H \times W$ 大小的特征，该特征包含三个尺度中最显著的特征。最后，以此特征经过两层卷积层和 L2 归一化层，得到重建的物体表面法向。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。

4.5 实验结果

本节对提出的归一化的高频区域增强光度立体模型进行了实验验证。首先，本节对提出的方法进行了消融实验与分析，包括超参数 λ 的选取、观察图像归一化操作有效性、特征聚合层的比较、高分辨率特征提取器的最优设置和模型训练时输入图像数量的影响。随后，在 DiLiGenT 数据集[31] 上，本节将提出的模型与先前的传统光度立体方法和基于深度学习的方法进行了对比。为了合理地评估预测的表面法向，本节采用的衡量指标包括 MAE、$err_{<10^\circ}$ 和 $err_{<30^\circ}$。

4.5.1 实验设置

本章提出的归一化的高频区域增强光度立体模型使用默认的 Adam 优化器进行优化（β_1 = 0.9 and β_2 = 0.999），初始的学习率则被设为 0.002 且每 5 个 epoch 减半。该模型具有 4.63M 的参数，使用 batchsize 大小为 32 的设置在单张的 RTX 3080Ti 训练了 30 个 epoch。训练的输入图像数量为 32。此外训练中输入光度立体图像的空间分辨率 $H \times W$ 被设置为 32 × 32。训练的数据集为采用 MERL BRDFs 数据集[116] 渲染的 Blobby 形状数据集[114] 和 Sculpture 形状数据集[115]，总计 84360 个用于训练的样本。

4.5.2 消融实验与分析

本节通过 DiLiGenT 数据集[31] 进行了定量的消融实验（对于消融研究中的所有实验，在 96 个光度立体图像输入的情况下进行了 3 次训练并取平均值作为汇报数据）。对于注意力权重损失消融实验及分析在 3.6.2中已经详细讨论，在此不做赘述。首先，本节对注意力权重损失中的保护性阈值 λ 的选取进行了实验，实验结果见图 4-6。

如图 4-6所示，可以看出超参数 λ 的适当选取对于提出模型性能至关重要。模型需要一个合适的 λ 的原因可以解释如下。梯度损失 $\mathcal{L}_{\text{Gradient}}$ 可以突出高频信息，在复杂区域提供更好的表面法线重建。但是，较大的权重 $\mathcal{L}_{\text{gradient}}$ 会冲淡对表面法向的误差的惩罚，因为梯度损失只提供了相邻像素之间的关系，而忽略了表面法向。在 3.6.2中的表 3-5也证明了仅使用梯度损失会导致光度立体网络的不收敛。

为了确定最优的 λ，本节使用从 1 到 10 的不同 λ 值对模型进行了实验测试，当 λ =6 时，三次训练的平均 MAE 为 7.036 度（越低越好），$err_{<10^\circ}$ 和 $err_{<30^\circ}$ 为

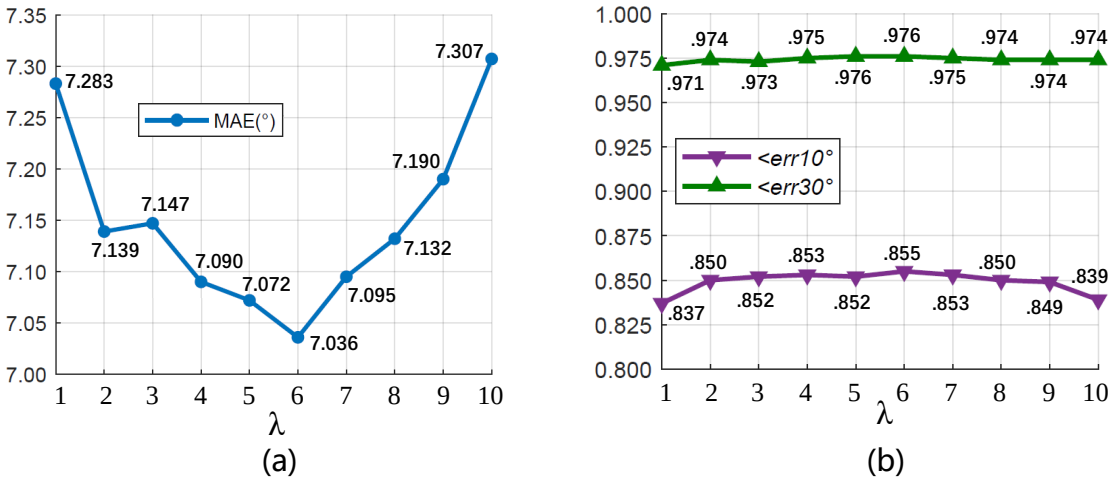


图 4-6 用不同的超参数 λ 值训练提出方法，(a) 在指标 MAE 上的结果，(b) 在指标 $err_{<10^\circ}$ 和 $err_{<30^\circ}$ 上的结果

0.855 和 0.976（越高越好）。请注意，在提出的归一化的高频区域增强光度立体模型中，即便超参数取任意的值（1 到 10），模型的性能也已经超过了之前基于深度学习的最佳方法 PS-FCN (Norm.)[69]。其 96 张输入图像下在 DiLiGenT 数据集[31] 上的 MAE 为 7.385 度。这也证明了本章模型的有效性。

表 4-1 模型在使用和不使用观察图像归一化的情况下的结果

操作	MAE ↓	$err_{<10^\circ}$ ↑	$err_{<30^\circ}$ ↑
使用观察图像归一化	**7.036**	**0.855**	**0.976**
不使用观察图像归一化	7.784	0.824	0.966

随后，本节在使用和不使用观察图像归一化的操作下评估了模型的性能，结果列在表 4-1 中。此外，图 4-7 显示了物体 Cat 的视觉结果，该物体的黄色框区域具有空间变化的表面材质。表 4-1和图 4-7的实验证明了观察图像归一化操作的有效性。通过使用观察图像归一化，生成的注意力图可以准确地反映具有真实高频结构的区域，而不受到空间变化的表面材质的激活。当然，提出模型的重建表面法向的精度也因此有着大幅度的提升。

其次，消融实验进一步比较了基于不同特征聚合方法下提出的归一化的高频区域增强光度立体模型的性能，结果在表 4-2中显示。表 4-2比较了多尺度前提下的最大池化层特征聚合、平均池化层特征聚合和最大池化层与平均池化层结合的特征聚合方式。为了保证除这三种特征聚合方式外其他网络结构保持完

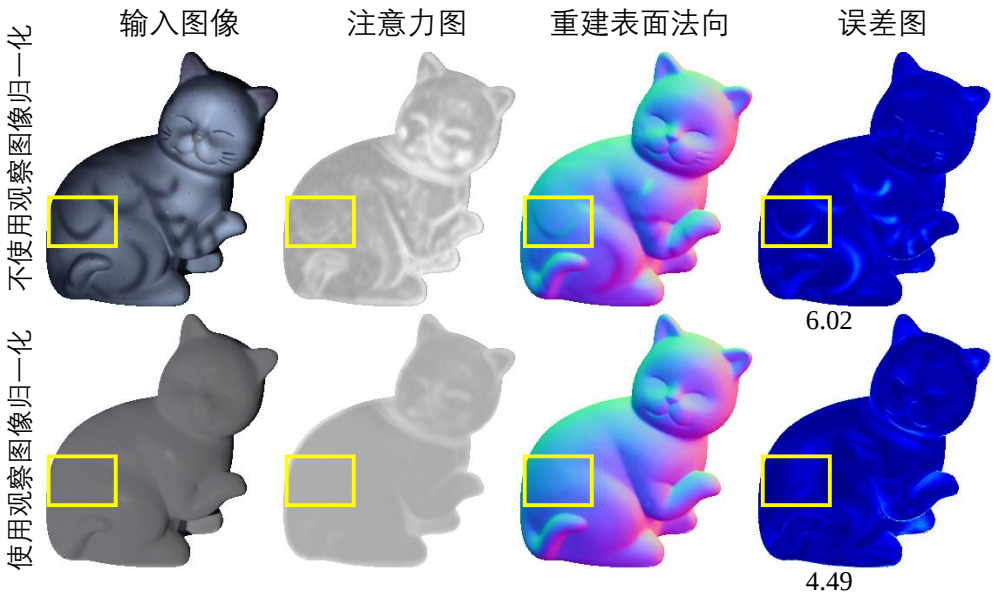


图 4-7 在物体 Cat 上使用和不使用观察图像归一化的可视化结果

全一致，实验在最大池化层与平均池化层结合的方式中，采取了先在第一维度上拼接，再使用 1×1 卷积层降维的方法，保证每一个尺度下的特征通道数与仅使用最大池化层特征聚合、平均池化层特征聚合的方式的通道数相同。

表 4-2 不同特征聚合方式的对比结果

特征聚合方法	MAE ↓	$err_{<10^\circ}$ ↑	$err_{<30^\circ}$ ↑
最大池化层	**7.036**	**0.855**	**0.976**
平均池化层	8.514	0.805	0.960
最大池化层 + 平均池化层	7.232	0.831	0.971

从表 4-2中，可以看到最大池化层的特征聚合在所有的三个指标上都达到了最佳性能。文献[77] 认为基于最大池化层和平均池化层的组合可以获得更好的性能，然而这与表 4-2中的实验结果相反。这可能是由于在本章模型中使用了观察图像归一化的操作。原因有二：第一，观察图像归一化的操作可以部分地视为平均池化层操作，因为每一幅归一化后的图像都包含来自所有原始图像的信息。因此，平均池化操作可能会导致特征冗余；第二，如图 4-4中所示，对于在某些光照方向下具有镜面高光的区域，在其他光照方向下并不具有镜面反射（即物体表面的高光随着光照方向的改变而改变）。归一化后的图像中，在其他光照方向下没有镜面反射的像素点的值会被抑制（分母中由于某个 m_j 存在镜面反射会变

大，因此归一化后像素值会变小），因此平均所有特征值的平均池化层的结果会受到影响，从而导致重建的表面法向出现偏差。

表 4-3进一步展示了高分辨特征提取器 f_{GE} 的消融实验。本章模型默认的结构编号标记为 ID (0)，其具有全分辨率特征 $\mathbf{\Psi}_j^{fr}$ + 二分之一分辨率特征 $\mathbf{\Psi}_j^{hr}$ + 四分之一分辨率特征 $\mathbf{\Psi}_j^{qr}$ 的三路并行高分辨率结构。对于编号 ID (1) ∼ ID (5)，实验调整了特征提取器 f_{GE} 的网络结构，以实现不同分辨率特征的不同组合。注意，ID (5) 中的 $\mathbf{\Psi}_j^{er} \in \mathbb{R}^{512\times\frac{1}{8}H\times\frac{1}{8}W}$ 表示八分之一分辨率的特征。ID (6) 表示提出的特征提取器 f_{GE} 使用四个残差块[123] 生成单一的四分之一分辨率特征 $\mathbf{\Psi}_j^{qr}$。在 ID (1) 和 (6) 中没有使用多尺度的最大池化层，因为在这两种消融方法中仅提取了单个分辨率的特征，因而特征回归器 f_{GR} 也做了相应的调整（仅采用全分辨率特征 $\mathbf{\Psi}_{\max}^{fr}$ 或四分之一分辨率特征 $\mathbf{\Psi}_{\max}^{qr}$ 的支路，且取消了不同尺度特征的拼接）对于 ID (0) ∼ ID (6) 的实验，特征提取器 f_{GE} 的结构也被相应的调整，以输出指定的特征组合方式。

表 4-3 不同特征聚合方式的对比结果

编号	网络结构	MAE ↓	$err_{<10^\circ}$ ↑	$err_{<30^\circ}$ ↑
ID (0)	默认结构（$\mathbf{\Psi}_j^{fr}$ + $\mathbf{\Psi}_j^{hr}$ + $\mathbf{\Psi}_j^{qr}$）	7.036	0.855	**0.976**
ID (1)	$\mathbf{\Psi}_j^{fr}$	7.157	0.852	0.974
ID (2)	$\mathbf{\Psi}_j^{fr}$+$\mathbf{\Psi}_j^{hr}$	7.071	0.853	**0.976**
ID (3)	$\mathbf{\Psi}_j^{fr}$+$\mathbf{\Psi}_j^{qr}$	7.083	0.852	0.975
ID (4)	$\mathbf{\Psi}_j^{hr}$+$\mathbf{\Psi}_j^{qr}$	7.106	0.852	0.975
ID (5)	$\mathbf{\Psi}_j^{fr}$ + $\mathbf{\Psi}_j^{hr}$ + $\mathbf{\Psi}_j^{qr}$ + $\mathbf{\Psi}_j^{er}$	**7.034**	**0.856**	**0.976**
ID (6)	残差模块[123]	7.101	0.854	0.975

如表 4-3所示，ID (2) ∼ ID (5) 的实验结果比较了使用不同特征分辨率组合的性能。当 ID (4) 中，高分辨特征提取器 f_{GE} 没有全分辨率特征 $\mathbf{\Psi}_j^{fr}$ 时性能明显更差，这说明了高分辨率的特征在逐像素的表面法向恢复任务中的性能具有至关重要的影响。注意 ID (1) 只有全分辨率特征，可以看成是一个没有上下采样的全卷积网络，难以取得优势。这是因为仅全尺寸特征的网络结构缺乏上下采样而导致感受野过小，难以利用图像上的邻域空间信息约束。从 ID (2)、ID (3) 和 ID (4) 的实验结果可以看出，结合具有更高分辨率的特征可以提供更好的性能。此外，实验也发现模型默认结构（$\mathbf{\Psi}_j^{fr}$ + $\mathbf{\Psi}_j^{hr}$ + $\mathbf{\Psi}_j^{qr}$）的性能不如具有额外的八分

之一分辨率特征 $\mathbf{\Psi}_j^{er}$ 的 ID (5)。但是在添加 $\mathbf{\Psi}_j^{er}$ 后，提升幅度很小，而 $\mathbf{\Psi}_j^{er}$ 的提取网络显著增加了模型参数和训练时间。这可能是因为这样的深度特征包含高级的语义信息较多，而这对于逐像素表面法向预测任务的用处有限。因此模型中高分辨特征提器 f_{GE} 采用 $\mathbf{\Psi}_j^{fr} + \mathbf{\Psi}_j^{hr} + \mathbf{\Psi}_j^{qr}$ 的组合方式。

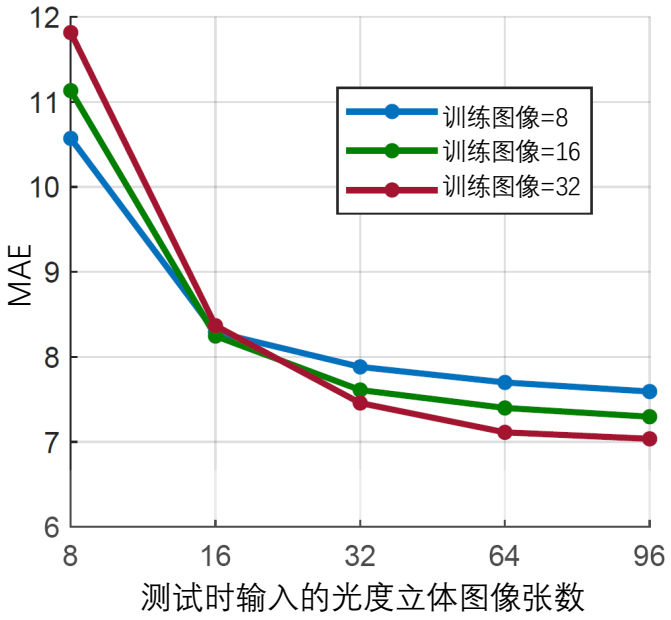


图 4-8 归一化的高频区域光度立体模型使用不同数量的输入光度立体图像训练和测试的结果

最后，图 4-8在 DiLiGenT 数据集[31] 展示了训练中使用不同输入数量的光度立体图像如何影响模型的性能。结果表明，当训练和测试的输入数量相同时，重建的表面法向精度最高。这表明，如果用于测试的输入图像数量已知且固定，则可以通过使用接近数量的输入图像进行训练和测试来进一步提高归一化的高频区域光度立体模型的性能。

4.5.3 真实拍摄数据集对比结果

首先，本节将第 3章中提出的自适应注意力光度立体模型与本章提出的归一化高频区域增强光度立体模型进行了比较，结果如图 4-9所示。图 4-9将提出的归一化高频区域增强光度立体模型、第 3章的自适应注意力光度立体模型、GPS-Net[77] 和 PS-FCN[25] 进行了比较。可以发现第 3章中提出的自适应注意力光度立体模型仅能处理复杂的表面结构，如褶皱边缘引起的高频表达区域，而归一化高频区域增强光度立体模型对复杂的表面结构和变化的表面材质区域都有高精度的重建效果，这正是观察图像归一化操作带来的改进。

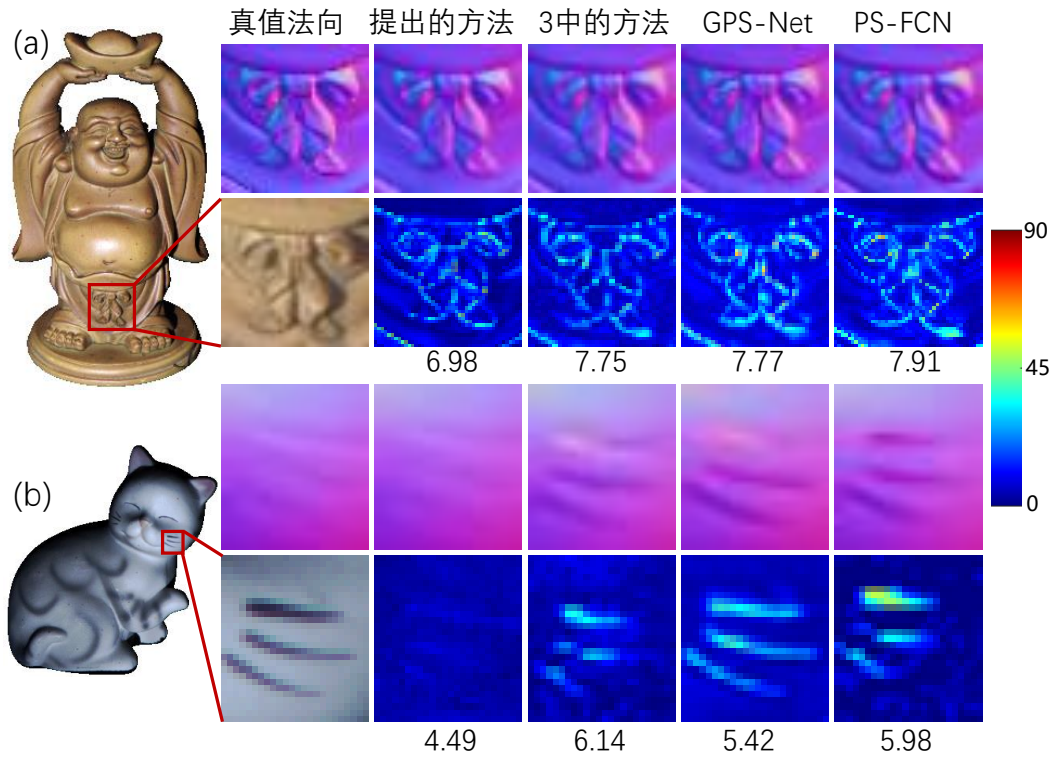


图 4-9 （a）复杂表面结构区域和（b）空间变化的表面材质区域的重建法向结果和误差图示例

其次，实验比较了提出的归一化的高频区域增强光度立体模型和其他最先进的方法在 DiLiGenT 数据集上进行了定量的比较。表 4-4展示了在以 96 张图像为输入下的结果。其中，传统方法以作者的姓氏的第一个字母 + 年份命名，LS 则代表最小二乘的基准方法[9]，基于深度学习的方法以网络简称命名，3的方法则表示自适应注意力光度立体模型。

表 4-4 将 DiLiGenT 数据集[31] 上所有以 96 张图像为输入的不同方法的 MAE 结果制成表格。粗体的值代表最佳性能，而下划线的值代表次佳性能。本章提出的归一化的高频区域增强光度立体模型在 10 个物体上实现了最小的平均表面法向角度误差，并在大多数物体上实现了最佳或次优的性能。图 4-10可视化了对比结果。其中第三列显示了使用[128] 方法利用重建的表面法向进行的三维重构和生成的注意力图。与 PS-FCN (Norm.)[69]、CNN-PS[75] 和 GPS-Net[77] 等方法相比，本章提出的方法在表面法向重建上实现了更好的性能。可以观察到，本章模型可以更准确地恢复那些具有投射阴影区域（白色框）的表面法向，例如物体 Harvest 的麻袋，以及那些具有高频皱纹的区域（橙色框），例如物体 Harvest 的汉字、物体 Pot1 的果实、物体 Buddha 的腰带。提出的归一化的高频区域增强光

图 4-10 DiLiGenT 数据集[31] 上物体 Harvest、Pot1 和 Buddha 的定量结果

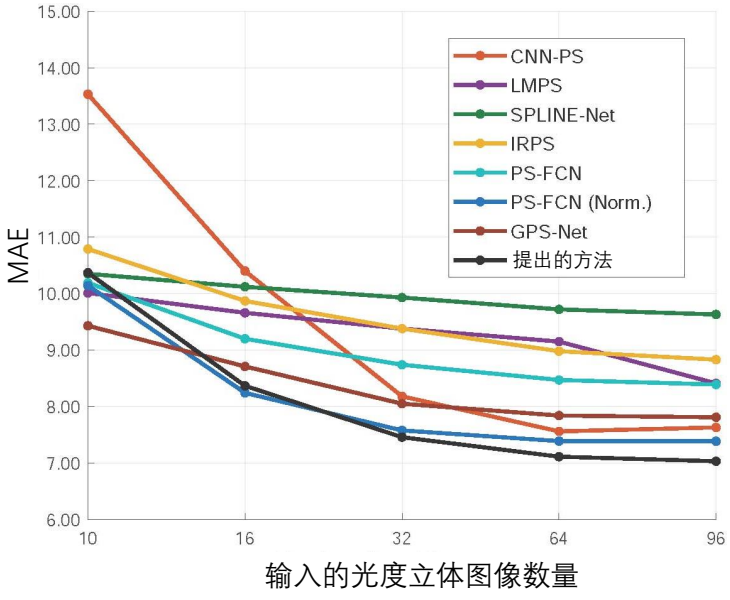


图 4-11 输入不同数量的图像下重建表面法向角度误差的比较结果

表 4-4 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均使用 96 幅图像进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
LS[9]	4.10	8.39	14.92	8.41	25.60	18.50	30.62	8.89	14.65	19.80	15.39
IW12[47]	2.54	7.32	11.11	7.21	25.70	16.25	29.26	7.74	14.09	16.17	13.74
WG10[46]	2.06	6.50	10.91	6.73	25.89	15.70	30.01	7.18	13.12	15.39	13.35
HM10[57]	3.55	11.48	13.05	8.40	14.95	14.89	21.79	10.85	16.37	16.82	13.22
GC10[52]	3.21	6.62	14.85	8.22	9.55	14.22	27.84	8.53	7.90	19.07	12.00
IA14[59]	3.34	7.11	10.47	6.74	13.05	9.71	25.95	6.64	8.77	14.19	10.60
ST14[58]	1.74	6.12	10.60	6.12	13.93	10.09	25.44	6.51	8.78	13.63	10.30
SPLINE-Net[27]	4.51	5.28	10.36	6.49	7.44	9.62	17.93	8.29	10.89	15.50	9.63
DPSN[24]	2.02	6.31	12.68	6.54	8.01	11.28	16.86	7.05	7.86	15.51	9.41
IRPS[26]	**1.47**	5.79	10.36	5.44	6.32	11.47	22.59	6.09	7.76	11.03	8.83
LMPS[76]	2.40	5.23	9.89	6.11	7.98	8.61	16.18	6.54	7.48	13.68	8.41
PS-FCN[25]	2.82	7.55	7.91	6.16	7.33	8.60	15.85	7.13	7.25	13.33	8.39
Manifold-PSN[73]	3.05	6.31	7.39	6.22	7.34	8.85	15.01	7.07	7.01	12.65	8.09
第 3章的方法	2.93	**4.86**	7.75	6.14	6.86	8.42	15.44	6.92	6.97	12.90	7.92
GPS-Net[77]	2.92	5.07	7.77	5.42	**6.14**	9.00	15.14	6.04	7.01	13.58	7.81
CNN-PS[75]	2.12	8.30	8.07	**4.38**	7.92	**7.42**	14.08	**5.37**	**6.38**	12.12	7.62
PS-FCN (Norm.)[69]	2.67	7.72	7.53	4.76	6.72	7.84	12.39	6.17	7.15	10.92	7.39
提出的方法	2.74	6.11	**6.98**	4.49	6.68	7.87	**12.38**	5.92	6.62	**10.51**	**7.03**

度立体模型可以在这些高频区域实现更准确的估计。

事实上，很多光度立体的实际应用中仅能输入比较稀疏的光度立体图像，很难有 96 张之多的图像输入。因此图 4-11实验在 DiLiGenT 数据集[31] 上展示了使用不同数量的输入图像评估本章提出的模型，并将其与最近最先进的基于深度学习的方法进行比较的结果，例如 PS-FCN (Norm.)[69]、GPS-Net[77]、CNN-PS[75]、PS-FCN[25]、IRPS[26]、LMPS[76] 和 SPLINE-Net[27]。

从图 4-11可以看出，当使用超过 32 张图像作为输入时，本章提出的归一化的高频区域增强光度立体模型优于所有其他方法。当使用 16 张输入图像时，提出的方法有第二好的性能，并且在仅使用 10 张输入图像时保持了有不错的表面法向重建精度。值得注意的是，有些方法只用 10 张输入图像训练，例如 LMPS[76] 和 SPLINE-Net[27]，而本章模型则是用 32 张输入图像训练的。当本章模型也用较少的输入图像（例如 8 张和 16 张）进行训练时，用 10 张图像输入的测试结果

比用 32 张输入图像训练的结果要好得多（如图 4-8所示）。

本节在更复杂的 Light Stage Data Gallery 数据集[119] 上进一步评估了提出的模型。图 4-12 显示了本章模型的定性结果。同样地，模型使用 32 张图像进行训练并使用从 253 张图像中随机选择的 100 张输入图像进行评估。对于物体 Helmet、Plant 和 Standing，输入图像被首先下采样到空间分辨率的一半，因为原始图像的分辨率太大而无法处理。

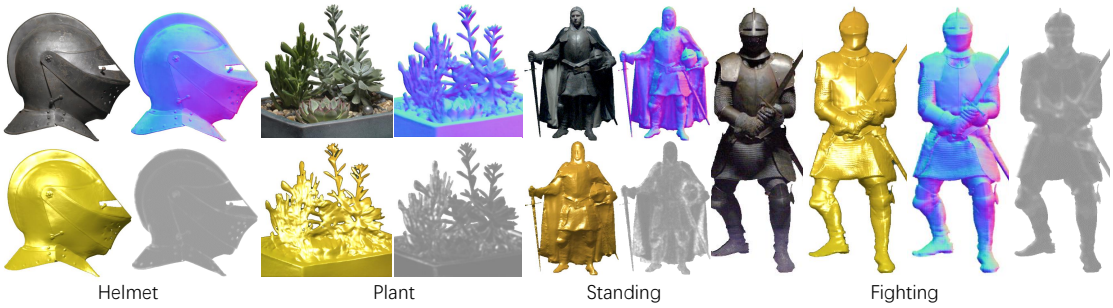


图 4-12 Light Stage Data Gallery 数据集[119] 中的四个物体 Helmet、Plant、Standing 和 Fighting 上的重建结果

如图 4-12所示，重建的表面法向保留了细节，避免了模糊，例如物体 Helmet 的螺丝，以及物体 Standing 和 Fighting 中的衣服。此外，可以看出，注意力图在高频区域被激活，例如边缘和皱纹。在训练期间物体 Plant 中植物的表面材质特性从未在 MERL BRDFs 材质数据集[116] 中出现，然而 Plant 的结果在视觉上相当准确，这表明本章提出模型的鲁棒性。

此外，本节还在 Gourd & Apple 数据集[55] 上定性的评估了模型的结果。该数据集中的物体表面既有褶皱的缝隙，又有变化的表面材质。图 4-13 显示了归一化的高频区域增强光度立体模型的结果，使用 32 张图像进行训练，并分别使用 16 张、48 张和 96 张输入图像进行表面法向重建。

可以看出，提出的归一化的高频区域增强光度立体模型具有处理任意数量输入图像的灵活性。在所有的输入图像数量下，在 Ground1 和 Ground2 上的重建的表面法向可以清楚地显示物体的褶皱。尽管在使用 16 张输入光度立体图像时可以在结果中发现一些噪声，但是基于不同数量的输入图像生成的注意力图可以显示复杂结构（皱纹）的正确表示。

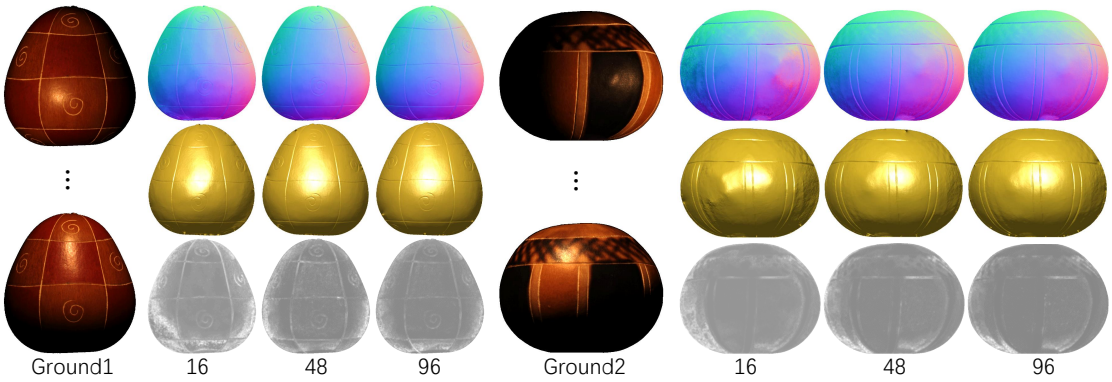


图 4-13 提出的模型对 Gourd & Apple 数据集[55] 中物体 Ground1 和 Ground2 的定性结果

4.5.4 合成数据集实验结果

由于真实拍摄的数据集中，表面材质的变化有限，为了进一步评估提出的归一化的高频区域增强光度立体模型在各种材质的物体表面的鲁棒性，本节在渲染的 Dragon 数据集上进行了实验。图 4-14展示了实验结果。Dragon 数据集中图片的分辨率大小为 384 × 384，并具有复杂的表面结构。Dragon 数据集的制作参照 2.6.1中训练集 Blobby[114] 和 Sculpture[115] 的渲染方法，使用基于物理的光线追踪器 Mitsuba[117]，利用 MERL BRDFs 数据集[116] 提供 100 种表面材质，并在物体上半球的空间内，随机选择 100 个光照方向。也就是说，Dragon 数据集中共有 100 种材质下的 Dragon 物体，每个材质的 Dragon 物体中，又有 100 张不同光照方向下的光度立体图像。

如图 4-14所示，使用 100 种不同的 MERL BRDF[116] 表面材质在物体 Dragon 上渲染，每种类型的材料在上半球用 100 个随机照明方向进行测试。重建的表面法向的性能在所有 100 种材料上实现了 4.65 度的平均 MAE，优于 PS-FCN(Norm.)[69] 的 4.89 度。

4.6 本章小结

本章提出了一种归一化的高频区域增强光度立体模型，显著改进了表面法向的重建结果，尤其是在那些高频区域，包括复杂的表面结构区域和剧烈变化的表面材质区域。注意力加权的法向重建损失为保留细节的梯度损失提供了一个自适应权重，以优化所提出的网络。观察图像归一化策略明确地消除了空间变化的表面材质的影响，并采用并行高分辨率结构来提取特征。消融实验证明

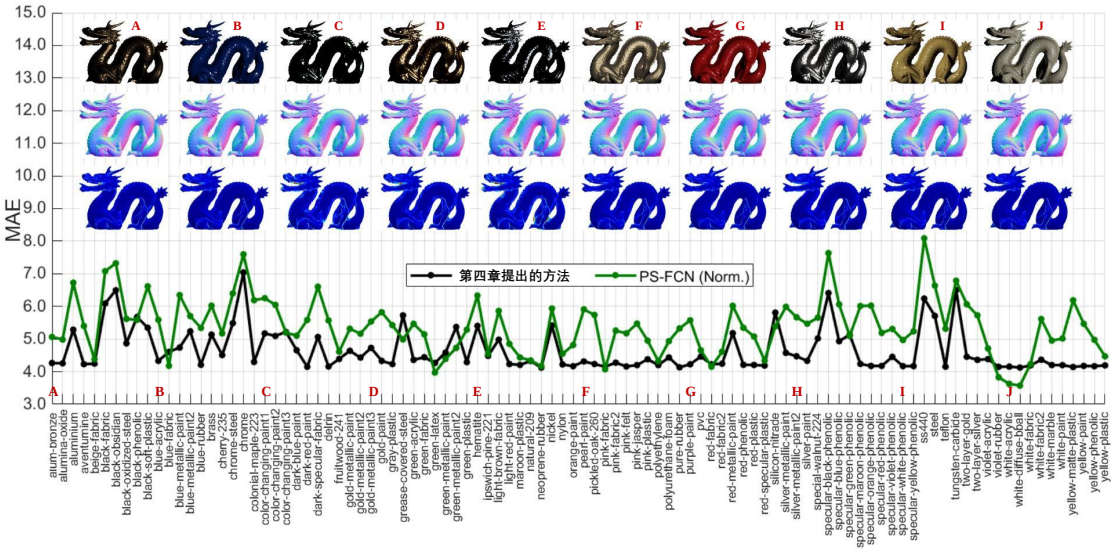


图 4-14 来自 MERL BRDFs 数据集[116] 的 100 种表面材质的 Dragon 物体上估计的重建表面法向结果

了提出模型的不同组成部分的有效性，这显著有利于提高表面法向的重建精度。对真实数据集（DiLiGenT[31]、Light Stage Data Gallery[119] 和 Gourd & Apple[55]）和合成数据集（Dragon）的广泛定量和定性实验表明，提出的归一化的高频区域增强光度立体模型优于其他最先进的方法。可视化的结果还表明，模型可以更好地重建高频区域的表面法向。此外，归一化的高频区域增强光度立体模型可以为其他低级和中级的视觉回归任务提供框架，例如深度估计和图像增强，注意力权重损失可以有利于结构细节的清晰恢复。

5 重光照-光度立体双重监督模型

5.1 研究背景

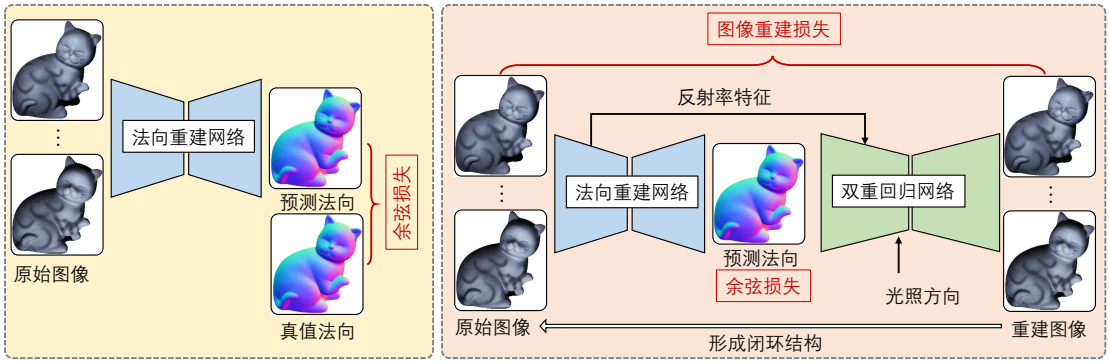
传统的光度立体方法面临具有一般反射率的非朗伯曲面的问题。利用深度神经网络，基于学习的光度立体算法[24,94,75,25] 能够改进非朗伯曲面下的表面法向的重建精度。但是以往的方法仅采用余弦损失这一单一约束。而继续增加模型的复杂度几乎无法提高估计表面法向重建精度，特别是在投射阴影、镜面反射和非凸结构等区域。究其原因，现有的基于深度学习的方法只关注物体表面法向，而忽略了图像重建也可以提供监督线索。因此本章利用常规的余弦损失和图像重建损失一起实现光度立体网络的优化。本章将重建物体的表面法向与重建光度立体图像两个任务相关联，以探索这种关联对表面法向重建的有益影响。事实上，在基于复杂反射模型的传统光度立体方法中，图像重建误差这一概念被精确地用作目标函数，以优化建模的表面反射特性。在某种程度上，本章的模型结合了学习和传统算法的目标，区别是提出的重光照-光度立体双重监督模型利用深度学习的方法逼近真实的渲染过程，而不是明确地建模图像成像公式。

为了在校准光度立体中实现重建表面法向和重建光度立体图像，本章提出了一种新颖的双重回归网络，并称之为重光照-光度立体双重监督模型。其使用双重回归网络从预测的表面法向中重建光度立体图像，使网络形成一个闭环，以减少表面法向的潜在学习空间并提供额外的监督。本章提出的模型在深度学习框架下统一了三维重建和渲染任务。实验表明提出的重光照-光度立体双重监督模型在校准光度立体任务中优于传统方法和基于学习的单一余弦损失的监督方法。其次，本章提出的重光照-光度立体双重监督模型可以生成任意光照下重建光度立体图像，不同照明方向下的准确重建图像直观地显示了表面的纹理和各向异性反射特性，这提供更直观的反射率感知，在许多视觉应用中有重要作用，例如虚拟现实和增强现实。

5.2 模型概述

本章提出了重光照-光度立体双重监督模型，利用深度学习探索如何利用额外的双重回归网络将重建的表面法向渲染回到光度立体图像，以进一步提高恢复表面法线的准确性。图 5-1展示了模型设计的思路，（a）图显示了常规的基于深度学习的光度立体框架，使用单一的余弦损失最小化重建的表面法向和真值

法向之间的误差，以优化网络。而（b）图则显示了本章提出的双重监督框架，在常规方法提供的余弦损失之外，通过设计的双重回归网络，可以将预测的法向再次重建回输入的光度立体图像，并在余弦损失之外额外提供原始输入图像和重建图像之间的图像重建损失。此外，依照成像模型式 (2-4)，若要从表面法向渲染一副图像，需要知道表面材质特性（反射率）以及照射的光照方向这些必要信息。



（a）常规模型的框架　　（b）提出的双重监督模型的框架

图 5-1 （a）先前的基于深度学习的光度立体模型的框架，（b）提出的重光照-光度立体双重监督模型的框架

依据图 5-1中框架，本章提出了重光照-光度立体双重监督模型，它结合了表面法向约束和重建图像约束。模型使用双重回归网络生成重光照光度立体图像，引入额外的约束以减少表面法向的潜在学习空间。如图 5-2所示，重光照-光度立体双重监督模型包含表面法向生成网络和双重回归网络，形成一个闭环，提供额外的监督。提出的模型在表面法向任务中应用了最大池化层特征聚合[25]，以确保任意数量的输入图像。表面法向生成网络重建物体的表面法向 $\tilde{\boldsymbol{N}}$，而双重回归网络重建光度立体图像 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n$ 。模型使用拼接的方式将来自表面法向生成网络的高维特征连接到双重回归网络中，以提供表面材质的信息。为了更好地回归指定光照方向下的图像模型还将光照方向与学习的特征相结合，模型还将编码的光照方向 $\boldsymbol{L}_j'$、$\boldsymbol{L}_j''$ 与回归器融合两次，以获得指定光照下的重光照光度立体图像。实验结果表明，本章模型在任意指定的光照方向下实现了准确的重光照光度立体图像重建，并且在校准光度立体中优于传统方法和基于学习的的单一余弦损失的监督方法。按照顺序，首先介绍表面法向生成网络的结构和双重回归网络的结构，再介绍双重监督的损失函数设置。

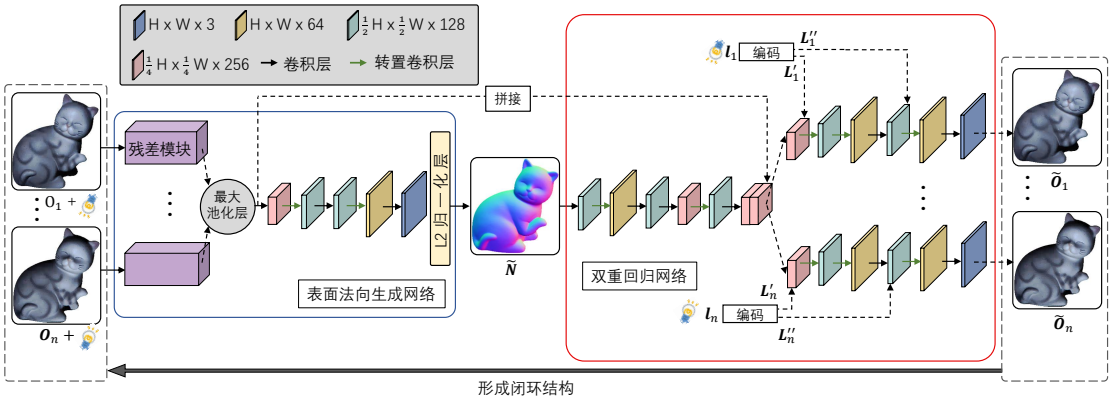


图 5-2 重光照-光度立体双重监督模型的详细结构

5.3 表面法向生成网络

表面法向生成网络旨在重建物体的表面法向，学习映射 $\tilde{\boldsymbol{N}} = f_{\text{Normal}}(\boldsymbol{\Phi}_j)$，使得估计重建的表面法向 $\tilde{\boldsymbol{N}}$ 逼近真值法向 $\boldsymbol{N}$。如图 5-2所示，表面法向生成网络包含三个部分，分别是特征提取器 f_{NE}、最大池化层特征聚合和特征回归器 f_{NR}。

对于在 n 个不同光照方向 $\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n \in \mathbb{R}^3$ 下拍摄的图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n \in \mathbb{R}^{3\times H\times W}$，与 3.3和 4.4的做法相同，首先将 $\boldsymbol{l}_j \in \mathbb{R}^3$ 沿 H 和 W 的方向复制，扩展至与图像具有相同分辨率大小的张量 $\boldsymbol{L}_j \in \mathbb{R}^{3\times H\times W}$，其中 $j \in \{1, 2, \cdots, n\}$。随后，将图像 $\boldsymbol{O}_j$ 与扩展得到的对应光照 $\boldsymbol{L}_j$ 利用拼接操作沿第一维度拼接起来，记为张量 $\boldsymbol{\Phi}_j \in \mathbb{R}^{6\times H\times W}$。

表面法向生成网络的特征提取器 f_{NE} 可以看作是 n 路分支的共享权重特征提取器，可以表示为：

$$\boldsymbol{\Psi}_j = f_{NE}(\boldsymbol{\Phi}_j; \theta_{NE}),\ j \in \{1, 2, \cdots, n\}, \tag{5-1}$$

其中 θ_{NE} 表示特征提取器 f_{NE} 中的可学习参数，$\boldsymbol{\Psi}_j \in \mathbb{R}^{256\times \frac{1}{4}H\times \frac{1}{4}W}$ 则是提取的深层特征。特征提取器 f_{NE} 的主体由六个残差模块组成[123]，激活函数设置为 Leaky ReLu。实际上本章也比较了使用 VGG[129] 的网络结构作为特征提取器，而残差模块[123] 的表现稍好一些，因此被选中。表 5-1展示了提出的特征提取器 f_{NE} 的具体结构。

随后，模型应用最大池化层特征聚合来处理任意数量的输入特征，以得到固定通道数的聚合特征。最大池化层特征聚合的方式可以从所有特征中提取最显著的信息，而忽略掉未激活的特征，能更好的处理物体表面投射阴影的区域。特

表 5-1 表面法向生成网络中特征提取器 f_{NE} 的网络结构。

输入	操作	卷积核大小	步长	输出
$3 \times H \times W$	卷积层 1	3×3	1	$64 \times H \times W$
$64 \times H \times W$	卷积层 2	3×3	1	$64 \times H \times W$
$64 \times H \times W$	卷积层 3	3×3	1	$64 \times H \times W$
$64 \times H \times W$	捷径连接 1（卷积层 2、卷积层 3）			$64 \times H \times W$
$64 \times H \times W$	卷积层 4	3×3	1	$64 \times H \times W$
$64 \times H \times W$	卷积层 5	3×3	1	$64 \times H \times W$
$64 \times H \times W$	捷径连接 2（卷积层 4、卷积层 5）			$64 \times H \times W$
$64 \times H \times W$	卷积层 6	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 7	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	捷径连接 3（卷积层 6、卷积层 7）			$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 8	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 9	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	捷径连接 4（卷积层 8、卷积层 9）			$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 10	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 11	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	捷径连接 5（卷积层 10、卷积层 11）			$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 12	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 13	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	捷径连接 6（卷积层 12、卷积层 13）			$256 \times \frac{1}{4}H \times \frac{1}{4}W$

征聚合的过程可以表示为：

$$\boldsymbol{\Psi}_{\max} = \bigcup_{i}^{\frac{1}{4}H \times \frac{1}{4}W} \max(\boldsymbol{\Psi}_{1,i}, \boldsymbol{\Psi}_{2,i}, \ldots, \boldsymbol{\Psi}_{n,i}), \tag{5-2}$$

其中 $\boldsymbol{\Psi}_{\max}$ 表示最大池化层聚合后的特征，下标 i 表示特征分辨率 $\frac{1}{4}H \times \frac{1}{4}W$ 中位置的索引。

之后，表面法向生成网络的特征回归器 f_{NR} 从 $\boldsymbol{\Psi}_{\max}$ 中学习表面法向 $\tilde{\boldsymbol{N}}$，可以表示为：

$$\tilde{\boldsymbol{N}} = f_{NR}(\boldsymbol{\Psi}_{\max}; \theta_{NR}), \tag{5-3}$$

其中 θ_{NR} 表示特征回归器 f_{NR} 中可学习的参数。表 5-2展示了特征回归器 f_{NR}

的具体网络结构。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。

表 5-2 表面法向生成网络中特征回归器 f_{NR} 的网络结构

输入	操作	卷积核大小	步长	输出
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 1	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 1	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 2	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 2	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 3	3×3	1	$3 \times H \times W$
$3 \times H \times W$	L2 归一化层			$3 \times H \times W$

5.4 双重回归网络

在表面法向生成网络之后，模型探索了用于学习重建图像的双重回归网络。双重回归网络旨在学习映射 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n = f_{\text{Dual}}(\tilde{\boldsymbol{N}})$，并使得重建的光度立体图像 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n$ 逼近原始的输入图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n$。如图 5-2所示, 提出的双重回归网络由一个特征提取器 f_{DE} 和一个特征回归器 f_{DR} 组成。

给定 5.3中重建的表面法线 $\tilde{\boldsymbol{N}}$ 和 5.3中聚合的特征 $\boldsymbol{\Psi}_{\max}$，双重回归网络的特征提取器 f_{DE} 学习到特征 $\boldsymbol{\Gamma} \in \mathbb{R}^{512 \times \frac{1}{4}H \times \frac{1}{4}W}$，记作：

$$\boldsymbol{\Gamma} = f_{DE}(\tilde{\boldsymbol{N}}, \boldsymbol{\Psi}_{\max}; \theta_{DE}), \tag{5-4}$$

其中 θ_{DE} 表示特征提取器 f_{DE} 中的可学习参数。特征提取器 f_{DE} 是由五个卷积层和两个转置卷积层组成的结构。表 5-3展示了双重回归网络的特征提取器 f_{DE} 的具体网络结构。

如表 5-3所示，模型在最后一层的卷积层后，拼接了来自表面法向生成网络的最大池化层聚合特征 $\boldsymbol{\Psi}_{\max}$。这是因为双重回归网络可以看作是一个从三维结构（表面法向）到二维图像（光度立体图像）的渲染过程。在光照成像模型中式(2-4)，想要渲染一幅图像必须需要知道其表面的材质信息（反射率 $\boldsymbol{\rho}$）和光照信息。而准确的表面法线 $\tilde{\boldsymbol{N}}$ 并不含有任何表面材质的信息，因此在双重回归网络的特征提取器 f_{DE} 中，模型将表面法向生成网络的聚合特征 $\boldsymbol{\Psi}'_{\max}$ 拼接至 f_{DE}，其提供了物体的表面材质信息。

表 5-3 双重回归网络中特征提取器 f_{DE} 的网络结构

输入	操作	卷积核大小	步长	输出
$3 \times H \times W$	卷积层 1	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 1	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 2	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 3	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 2	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 4	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 5	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	拼接（卷积层 5、$\boldsymbol{\Psi}_{\max}$）			$512 \times \frac{1}{4}H \times \frac{1}{4}W$

双重回归网络学习的映射 f_{Dual} 是从固定的表面法向 $\tilde{\boldsymbol{N}}$ 到任意张数 n 的重建图像 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n$。如上述讨论，模型还需要重建图像 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n$ 对应的光照方向信息才能使双重回归网络实现重建的过程。因此，在特征提取器 f_{DE} 之后，本章又设计了一个特征回归器 f_{DR}，以提供光照方向，并将提取到的深度特征 $\boldsymbol{\Gamma}$ 回归到重建的 n 张光度立体图像，记作：

$$\tilde{\boldsymbol{O}}_j = f_{DR}(\boldsymbol{\Gamma}, \boldsymbol{L}'_j, \boldsymbol{L}''_j; \theta_{DR}),\ j \in \{1, 2, \cdots, n\}, \tag{5-5}$$

其中 θ_{DR} 表示双重回归网络的特征回归器 f_{DR} 中的可学习参数，$\boldsymbol{L}'_j$ 和 $\boldsymbol{L}''_j$ 为编码的光照方向。其中，$\boldsymbol{L}'_j \in \mathbb{R}^{3 \times \frac{1}{4}H \times \frac{1}{4}W}$，$\boldsymbol{L}''_j \in \mathbb{R}^{3 \times \frac{1}{2}H \times \frac{1}{2}W}$。$\boldsymbol{L}'_j$ 和 $\boldsymbol{L}''_j$ 从光度立体图像 $\boldsymbol{O}_j$ 对应的光照方向 $\boldsymbol{l}_j$ 得到，具体方法是将 $\boldsymbol{l}_j \in \mathbb{R}^3$ 沿 H 和 W 的方向复制，分别扩展至 $\frac{1}{4}H \times \frac{1}{4}W$ 大小和 $\frac{1}{2}H \times \frac{1}{2}W$ 大小。为了更好地渲染到指定光照方向下的重建图像，模型将编码的光照方向 $\boldsymbol{L}'_j$、$\boldsymbol{L}''_j$ 与特征回归器融合两次。表 5-4展示了双重回归网络的特征回归器 f_{DR} 的具体网络结构。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。

在双重回归网络的训练中，为了形成重建光度立体图像 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n$ 与输入图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n$ 之间的图像重建损失，在特征回归器 f_{DR} 中输入的编码光照方向 $\boldsymbol{L}'_j$ 和 $\boldsymbol{L}''_j$ 应该与输入 $\boldsymbol{O}_j$ 的光照方向 l_j 完全对应。而在测试时，则可以输入任意的光照方向的 l 并编码至 $\boldsymbol{L}'$ 和 $\boldsymbol{L}''$，生成任意指定的光照方向下的重光照光度立体图像。这些重光照图像可以直观地显示表面的纹理和各向异性的反射特性，提供更直观的可视感知，并可以应用在虚拟现实，增强现实等任务。

表 5-4 双重回归网络中特征回归器 f_{DR} 的网络结构

输入	操作	卷积核大小	步长	输出
$512 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 1	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	拼接（卷积层 1、$\boldsymbol{L}_j'$）			$259 \times \frac{1}{4}H \times \frac{1}{4}W$
$259 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 1	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 2	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 2	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	拼接（卷积层 2、$\boldsymbol{L}_j''$）			$131 \times \frac{1}{2}H \times \frac{1}{2}W$
$131 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 3	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 3	3×3	1	$3 \times H \times W$

5.5 双重监督损失函数

为了实现对上述表面法向生成网络（5.3）和双重回归网络（5.4）的训练，学习到表面法向和光度立体图像的准确映射，本节提出了双重监督损失函数，以优化模型中的参数 θ_{NE}、θ_{NR}、θ_{DE} 和 θ_{DR}。双重监督损失函数 $\mathcal{L}_{\text{Dual}}$ 可以被表示为：

$$\mathcal{L}_{\text{Dual}} = \mathcal{L}_{\text{Cosine}}(\boldsymbol{N}, \tilde{\boldsymbol{N}}) + \lambda_t \mathcal{L}_{\text{Recons}}(\boldsymbol{O}_j, \tilde{\boldsymbol{O}}_j, \forall j), \tag{5-6}$$

其中第一项 $\mathcal{L}_{\text{Cosine}}(\boldsymbol{N}, \tilde{\boldsymbol{N}})$ 是广泛使用的余弦损失，第二项 $\mathcal{L}_{\text{Recons}}(\boldsymbol{O}_j, \tilde{\boldsymbol{O}}_j, \forall j)$ 则是重光照-光度立体双重监督模型通过双重回归网络额外形成的图像重建损失，λ_t 是图像重建损失的权重。

双重监督损失函数的第一部分 $\mathcal{L}_{\text{Cosine}}$，表示法向真值 $\boldsymbol{N}$ 和重建的表面法向 $\tilde{\boldsymbol{N}}$ 之间的余弦角度误差，由下式所示：

$$\mathcal{L}_{\text{Cosine}}(\boldsymbol{N}, \tilde{\boldsymbol{N}}) = \frac{1}{HW} \sum_{i}^{HW} \left(1 - \boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i\right), \tag{5-7}$$

其中 $\cdot$ 操作代表点乘。如果在像素位置 i 上重建的表面法向 $\tilde{\boldsymbol{N}}_i$ 与真值法向 $\boldsymbol{N}_i$ 越相似，则其点乘 $\boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i$ 将越接近 1，此时式 (5-7) 的值将越接近 0。

双重监督损失函数的第二部分 $\mathcal{L}_{\text{Recons}}(\boldsymbol{O}_j, \tilde{\boldsymbol{O}}_j, \forall j)$ 表示双重回归任务中重建的光度立体图像 $\tilde{\boldsymbol{O}}_1, \tilde{\boldsymbol{O}}_2, \cdots, \tilde{\boldsymbol{O}}_n$ 与输入图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n$ 之间的差异，可以被定义为：

$$\mathcal{L}_{\text{Recons}}(\boldsymbol{O}_j, \tilde{\boldsymbol{O}}_j, \forall j) = \frac{1}{j}\frac{1}{HW}\sum_{j}^{n}\sum_{i}^{HW}\|\boldsymbol{O}_{j,i} - \tilde{\boldsymbol{O}}_{j,i}\|_2^2, \tag{5-8}$$

其中 $\boldsymbol{O}_{j,i}$ 和 $\tilde{\boldsymbol{O}}_{j,i}$ 表示输入的第 j 张光度立体图像和重建光度立体图像中像素 i 位置上的值。

对于图像重建损失 $\mathcal{L}_{\text{Recons}}(\boldsymbol{O}_j, \tilde{\boldsymbol{O}}_j, \forall j)$，模型采用随着训练轮数（epoch）变化的 λ_t 来控制重建损失的权重。具体来说，λ_t 在训练时期会发生变化：设置 λ_t 在第一个训练 epoch 时值为 0，并在每个训练的 epoch 后增加 0.02（这里，用 Δ 表示增量，即 $\Delta = 0.02$）。设计这种线性变化的 λ_t 是基于这样一个事实：双重回归网络使用表面法向生成网络重建的表面法向、表面材质信息（来自表面法向生成网络聚合的特征 $\boldsymbol{\Psi}_{\max}$）和输入的照明方向来拟合成像模型 (式 2-4)。因此，双重回归网络的学习需要准确的表面法向，模型可以通过逐渐增加图像重建损失的权重来实现对重建表面法向的率先监督，以保持提出的端到端的重光照-光度立体双重监督模型的稳定训练。模型还将 λ_t 的最大设定值为 0.8。这可以防止图像重建损失的权重变得太大，而降低表面法向的重建精度。图 5-3展示了 λ_t 的可视化曲线。消融实验（5.6.2）表明，这种设计的 λ_t 能够为表面法向重建和光度立体图像重建提供最优的监督。

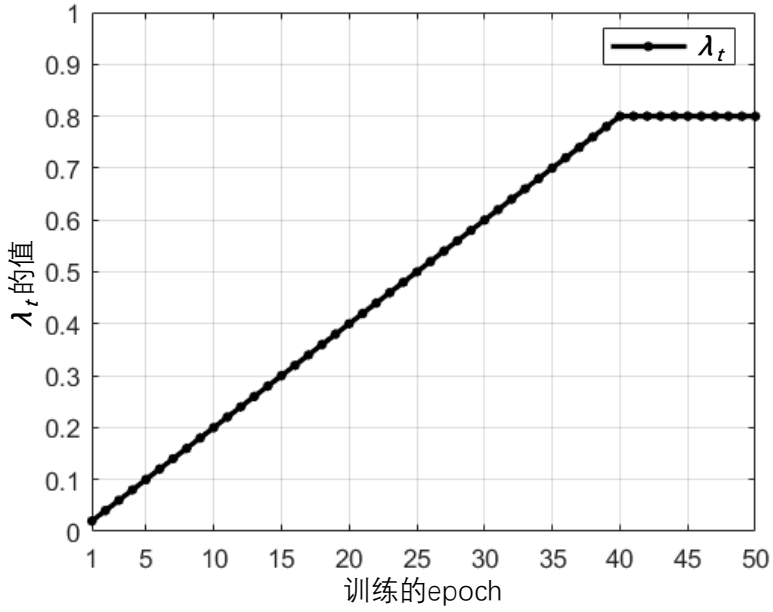


图 5-3 λ_t 值随着训练 epoch 变化的可视化曲线

5.6 实验结果

本节对提出的重光照-光度立体双重监督模型进行了实验验证。本节首先对提出的方法进行了消融实验与分析，包括双重回归网络对标定光度立体任务的作用和权重 λ_t 的选取。随后在 DiLiGenT 数据集[31] 上，将提出的模型与先前的传统光度立体方法和基于深度学习的方法进行了对比。为了合理地评估预测的表面法向，本节采用的衡量指标包括 MAE、$err_{<15^\circ}$。为了评估重建图像的准确性，本节采用了平均相对误差（REL），其定义为：

$$\text{REL} = \frac{1}{n}\frac{1}{T}\sum_{j}^{n}\sum_{i}^{T}\frac{\left|\boldsymbol{O}_{j,i} - \tilde{\boldsymbol{O}}_{j,i}\right|}{\boldsymbol{O}_{j,i}}, \tag{5-9}$$

其中 T 为图像掩模中物体表面所在位置的像素总数。此外，本节还使用了结构相似性指数（SSIM)[121] 来衡量重建图像与原始图像的相似度。SSIM 是一个 0 到 1 之间的数，越大表示重建图像和输入图像的差距越小，其可以被定义为：

$$\text{SSIM} = \frac{1}{n}\sum_{j}^{n}\frac{\left(2\mu_{x,j}\mu_{y,j} + c_1\right)\left(\sigma_{xy,j} + c_2\right)}{\left(\mu_{x,j}^2 + \mu_{y,j}^2 + c_1\right)\left(\sigma_{x,j}^2 + \sigma_{y,j}^2 + c_2\right)}, \tag{5-10}$$

其中 $mu_{x,j}$ 和 $\sigma_{x,j}$ 代表重建光度立体图像 $\tilde{\boldsymbol{O}}_j$ 的平均值和标准差，$\mu_{y,j}$、$\sigma_{y,j}$ 代表原始光度立体图像 $\boldsymbol{O}_j$ 的平均值和标准差，而 $\sigma_{xy,j}$ 则表示 $\tilde{\boldsymbol{O}}_j$ 和 $\boldsymbol{O}_j$ 的协方差，c_1 和 c_2 为默认设置常数，以防止式 (5-10) 中的分母为 0 带来的系统错误。

评估重建的表面法向和重建光度立体图像时，MAE、REL 和 SSIM 指标都仅衡量掩模内物体所占位置的像素，而不包括背景位置上的像素。此外在评价重建光度立体图像时，如式 (5-9) 和式 (5-10) 所示，REL 和 SSIM 衡量了所有重建图像 $\tilde{\boldsymbol{O}}_j$ $j \in \{1, 2, \cdots, n\}$ 的平均 REL 和 SSIM 值。

由于本章提出的重光照-光度立体双重监督模型可以在测试时生成任意光照方向下的重光照光度立体图像，为了更详细的评估提出的方法，进一步将重建的光度立体图像分为两类：属于原始输入光度立体图像的光照方向（BI）和不属于原始输入光度立体图像的光照方向（NBI）两种情况。

5.6.1 实验设置

本章提出的重光照-光度立体双重监督模型使用默认的 Adam 优化器进行优化（β_1 = 0.9 and β_2 = 0.999），初始的学习率最初设置为 0.001 且每过 5 个 epoch，学习率除以 2。模型使用的 batchsize 大小为 32 ，并在当个 RTX 3080Ti 上训练

了 50 个 epoch。训练的输入图像数量 n = 32。此外模型训练中将输入图像的分辨率 $H \times W$ 设置为 32 × 32。训练的数据集为采用 MERL BRDFs 数据集[116] 渲染的 Blobby 形状数据集[114] 和 Sculpture 形状数据集[115]，总计 84360 个用于训练的样本，即一个 epoch 含有 84360 个样本，每个样本输入 32 张不同光照方向下的光度立体图像。

5.6.2 消融实验与分析

重光照-光度立体双重监督模型的消融实验在 Blobby 数据集[114] 和 Sculpture 数据集[115] 的验证集中的全部 852 样本（每个样本采用 64 张不同光照的输入图像）上实施。如表 5-5所示，实验通过将完整的重光照-光度立体双重监督模型（ID (0)）与单一的表面法向生成网络（通过权重 λ_t 恒等于 0）（ID (1)）的重建结果进行了比较。之后，实验进一步探讨了图像重建损失 λ_t 对重建表面法向和重建光度立体图像精度的影响。实验通过将 λ_t 与不同固定的权重值（0.1、0.5 和 1）进行比较（ID (2) ∼ ID (4)）以验证了模型在训练中随 epoch 改变的 λ_t 策略的有效性。进一步地，实验评估了不同的增长速率（Δ）和不同的最大设定值对线性 λ_t 策略性能上的影响（ID (5) ∼ ID (8)）。最后，实验对 λ_t 随训练 epoch 线性增加和非线性增加（二次增长）做了相关的对比（ID (9) ∼ ID (10)）。在表 5-5中，所有重建的光度立体图像都属于原始输入光度立体图像的光照方向（BI），Δ 代表增加速率，PT 代表 λ_t 的最大设定值。对于 $< err_{15^\circ}$ 和 SSIM 指标来说，值越高越好。对于 MAE 和 REL 来说，值越低越好。

ID (0) 和 ID (1) 的实验表明，与单一的表面法向生成网络 ($\lambda = 0$) 相比，具有默认的线性增加的权重 λ_t 的重光照-光度立体双重监督模型在表面法向重建上取得了更好的性能。具体来说，本章提出的模型在验证集上取得了 11.47 度的 MAE，84.99% 的 $< err_{15^\circ}$，而对于单一余弦损失监督的网络来说，MAE 的值下降到 12.53 度，并且小于 15 度误差的比例（$< err_{15^\circ}$）仅为 81.55%。这证明了本章提出的模型增强了表面法向的学习能力。原因是额外的双重回归网络引入了额外的图像重建损失以提供额外的监督，有效地减少了表面法向学习的潜在空间。换句话说，它降低了表面法向的学习难度。

ID (2)、(3) 和 (4) 的实验展示了固定权重的 λ_t 对本章模型的影响。与固定权重值 λ_t 的双重监督模型相比，采用 ID (0) 提出的线性增长的 λ_t 策略可以取得最好的表面法向重建精度和重建图像的次优精度，这非常接近采用固定 λ_t = 1 时

表 5-5 提出的重光照-光度立体双重监督模型与单一表面生成网络的性能比较，以及不同 λ_t 加权策略对双重监督模型的影响

编号	λ_t 策略	重建的表面法向		重建的光度立体图像	
		MAE ↓	$< err_{15^\circ}$ ↑	SSIM ↑	REL ↓
ID (0)	线性 λ_t（Δ = 0.02, PT = 0.8）	**11.47**	84.99%	0.947	0.171
ID (1)	λ_t = 0	12.53	81.55%	-	-
ID (2)	λ_t = 0.1	11.64	84.61%	0.895	0.235
ID (3)	λ_t = 0.5	11.88	82.94%	0.939	0.182
ID (4)	λ_t = 1	12.50	81.79%	**0.963**	**0.166**
ID (5)	线性 λ_t（Δ = 0.02, PT=0.6）	11.57	**85.01%**	0.926	0.197
ID (6)	线性 λ_t（Δ = 0.02, PT = 1）	11.80	83.33%	0.951	0.169
ID (7)	线性 λ_t （Δ = 0.01, PT = 0.8*）	11.58	84.52%	0.914	0.209
ID (8)	线性 λ_t （Δ = 0.04, PT = 0.8）	11.55	84.39%	0.929	0.175
ID (9)	二次 λ_t （Δ = 0.001, PT= 0.8）	11.49	84.95%	0.916	0.197
ID (10)	二次 λ_t（Δ = 0.0005, PT = 0.8）	11.58	84.78%	0.934	0.188

的性能。虽然 λ_t = 1 实现了重建图像的最佳效果，但可以发现，与 ID (1) 单一的表面法向生成网络相比，其几乎没有提高重建表面法向的性能。这表明将双重回归网络的权重固定为 λ_t = 1 可能太大，无法对表面法向的监督产生足够的有益效果。注意，从实验结果中也可以看出，对于所有具有固定非零权重 λ_t，双重监督模型都能在表面法向重建上取得比单一表面法向生成网络更好的结果。如前所述，双重回归网络通过预测的表面法向、表面材质和光照方向来重建图像，其学习过程需要准确的表面法向，因此双重回归网络对表面法向重建提供了额外的约束。

ID (5) 和 (6) 对比了线性 λ_t 在相同的增加速率 Δ 下，不同最大设定值 PT 对重光照-光度立体双重监督模型的影响。可以发现在 ID (0) 中默认的最大设定值 PT = 0.8 时，表面法向重建和光度立体图像重建两个任务上都有着最佳的效果。虽然使用最大设定值 PT = 0.6 时，重建表面法向的性能几乎等于默认的 PT = 0.8（MAE 稍差，$< err_{15^\circ}$ 稍好），但是在这种情况下，光度立体图像重建的性能会有明显的下降。这表明太小的最大设定值可能无法为双重回归网络提供足够的监督。此外，使用最大设定值 PT = 1 时，虽然图像的重建结果稍好一些，但是重建表面法向的误差会大很多。这又表明最大设定值太大会导致很难对表面法

向重建任务提供有益的约束。根据实验结果，将 λ_t 的最大设定值 PT 设置为 0.8 时，既可以保护表面法向重建任务的准确性和优先级，也能对图像重建任务提供足够的约束，从而提供最佳的整体性能。

ID (7) 和 (8) 则对比了在相同最大设定值 PT = 0.8 的情况下，线性 λ_t 中不同增加速率 Δ 对性能的影响。注意 ID (7) 的实验中，由于 λ_t 的增加速率 Δ 过小，无法在训练结束前（50 个 epoch）达到最大设定值 PT = 0.8（最大权重仅能达到 0.5），因此在表 5-5中用 * 标记。可以看出 Δ = 0.02 比其他两个设置实现了更好的性能。ID (7) 中 λ_t 的性能较差是由于图像重建监督不足，而使用 Δ = 0.04 的增加速率可能太大（仅 20 个 epoch 后权重达到最大设定值 0.8）。这可能使得提出的模型学习不稳定，导致较大的表面法向和图像的重建误差。

最后，ID (9) 和 ID (10) 的实验将默认的线性增加权重 λ_t 与非线性（二次增加）的权重 λ_t 进行比较。可以看出，所提出的线性变化权重的策略明显优于二次变化权重的策略。这可能是因为二次增加的权重 λ_t 由于非线性权重在开始或结束时期的增长速度太慢或太快，导致训练不够稳定。因此，模型最终选择使用增加速率 Δ = 0.2，最大设定值 PT = 0.8 的线性权重 λ_t 作为默认设置，以获得最佳的表面法向重建和光度立体图像重建精度。

5.6.3 DiLiGenT 数据集对比结果

表 5-6列出了传统方法（以作者的姓氏的第一个字母 + 年份命名，LS 则代表最小二乘的基准方法[9]）和基于深度学习的方法（以网络简称命名）在 DiLiGenT 数据集[31] 上的 96 张输入光度立体图像的表面法向重建结果。图 5-4进一步展示了可视化的对比结果。粗体的值代表最佳性能，而下划线的值代表次佳性能。注意，为了公平比较，实验将 CNN-PS[75] 在相同的数据集（利用 MERL BRDFs 数据集[116] 渲染 Blobby 形状数据集[114] 和 Sculpture 形状数据集[115]）上重新训练，而非采用迪士尼 Disney BRDF 材质数据集[118] 进行渲染的 CyclesPS 数据集，并以 CNN-PS* 表示。因为除 CNN-PS 外，其他基于学习的方法都采用 MERL BRDFs 数据集[116] 提供材质信息。

表 5-6比较了我们提出的重光照-光度立体双重监督模型的表面法向重建结果。可以看出，我们的方法以 7.90 度的平均 MAE 排名第一（用 96 张图像进行测试）。在图 5-4中，我们在 96 张输入图像下展示了几种最先进的方法在物体 Buddha、Bear、Pot2 和 Harvest 上的可视化对比。重建误差图的亮度提升了 10 倍，

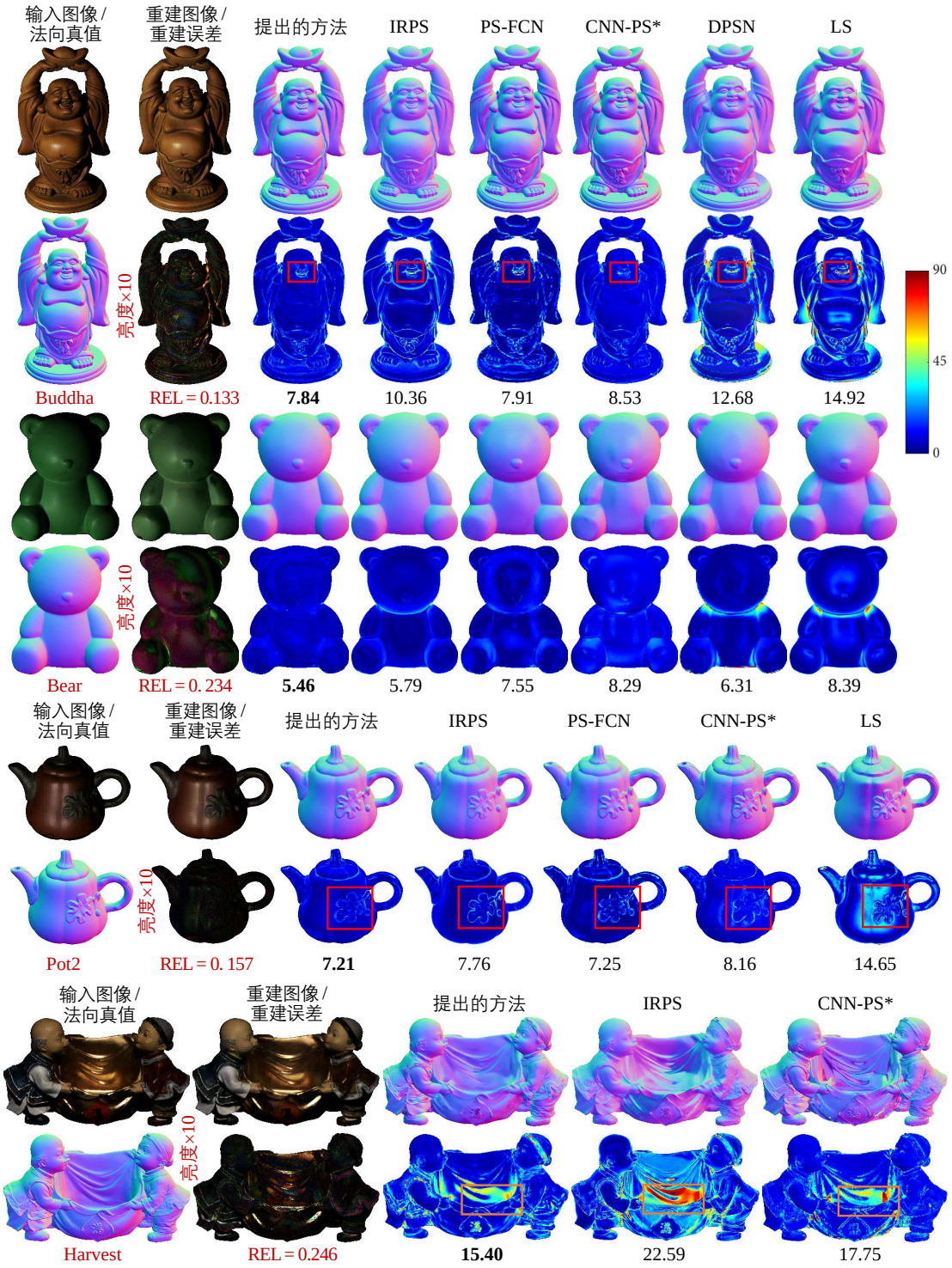


图 5-4 DiLiGenT 数据集[31] 中物体在 96 张输入图像下的可视化结果

表 5-6 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均以 96 幅图像为输入进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
LS[9]	4.10	8.39	14.92	8.41	25.60	18.50	30.62	8.89	14.65	19.80	15.39
IW12[47]	2.54	7.32	11.11	7.21	25.70	16.25	29.26	7.74	14.09	16.17	13.74
IA14[59]	3.34	7.11	10.47	6.74	13.05	9.71	25.95	6.64	8.77	14.19	10.60
ST14[58]	1.74	6.12	10.60	6.12	13.93	10.09	25.44	6.51	8.78	13.63	10.30
DPSN[24]	2.02	6.31	12.68	6.54	8.01	11.28	16.86	7.05	7.86	15.51	9.41
IRPS[26]	**1.47**	5.79	10.36	5.44	**6.32**	11.47	22.59	**6.09**	7.76	**11.03**	8.83
CNN-PS*[75]	2.23	8.29	8.53	5.75	9.74	8.66	17.75	5.91	8.16	11.61	8.66
LMPS[76]	2.40	**5.23**	9.89	6.11	7.98	8.61	16.18	6.54	7.48	13.68	8.41
PS-FCN[25]	2.82	7.55	7.91	6.16	7.33	8.60	15.85	7.13	7.25	13.33	8.39
提出的方法	2.27	5.46	**7.84**	**5.42**	7.01	**8.49**	**15.40**	7.08	**7.21**	12.74	**7.90**

以展示更好的细节。误差图中的红色框是结构复杂的区域，而误差图中的橙色框是具有强阴影和相互反射的区域。可以看出本章提出的模型的误差图在这些区域显示出较低的角度误差。与其他方法相比，本章提出的模型在结构复杂的区域也产生了更多的细节。这正归功于重光照-光度立体双重监督模型形成了一个闭环架构，以提供对表面法向重建的额外监督。虽然模型在平均 MAE 中获得了最好的性能，但它在一些物体上没有取得最佳的效果。我们推断，这是因模型采用的最大池化层来聚合特征[25]，而最大池化层会丢弃大量特征，只保留了最大响应值。因此，当输入图像增加时，模型的利用率会降低，这可能会在一定程度上影响其性能。此外，实验表明模型在物体 Ball 上并没有取得令人满意的结果。对于具有特别简单结构和几乎朗伯曲面的物体 Ball，模型并不优于传统的基于单一余弦损失监督的方法。因为在这种极端简单的情况下，余弦损失提供了很强的惩罚，而提出模型中的图像重建损失反而可能会削弱这种约束。

在实际应用中，很难获得同一视角下 96 张密集输入图像的光度立体图像，稀疏输入的光度立体输入则更常见。因此实验还将提出的重光照-光度立体双重监督模型与传统方法和基于深度学习的方法在只有 10 张图像输入的情况下进行比较，以测试提出的模型在较少输入图像下的鲁棒性。表 5-7展示了对比结果，其中粗体的值代表最佳性能。可以看出在稀疏的输入图像下，本章提出的模型依然能够取得最佳的表面法向重建精度。

表 5-7 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均以 10 幅图像为输入进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
IA14[59]	12.94	16.40	20.63	15.53	18.08	18.73	32.50	**6.28**	14.31	24.99	19.04
LS[9]	5.09	11.59	16.25	9.66	27.90	19.97	33.41	11.32	18.03	19.86	17.31
ST14[58]	5.24	9.39	15.79	9.34	26.08	19.71	30.85	9.76	15.57	20.08	16.18
IW12[47]	**3.33**	7.62	13.36	8.13	25.01	18.01	29.37	8.73	14.60	16.63	14.48
CNN-PS[75]	6.39	14.51	15.08	10.96	15.26	14.40	19.73	11.35	13.58	16.67	13.79
PS-FCN[25]	4.02	**7.18**	9.79	8.80	10.51	11.58	18.70	10.14	9.85	15.03	10.51
LMPS[76]	3.97	8.73	11.36	**6.69**	10.19	10.46	17.33	7.30	9.74	**14.37**	10.02
提出的方法	3.83	7.52	**9.55**	7.92	9.83	**10.38**	**17.12**	9.36	**9.16**	14.75	**9.94**

随后，表 5-8列出了在 DiLiGenT 数据集中 10 张和 96 张输入光度立体图像的情况下，本章提出的重光照-光度立体双重监督模型在法向重建和图像重建上的所有指标。对于 96 张输入图像的情况，因为所有的图像都输入了网络，所以所有重建的图像都属于 BI 类，而在 10 张输入图像的情况下，重建的图像中 BI 类有 10 张，NBI 类有 86 张。在 96 张图像输入时，模型的 SSIM 和 REL 分别为 0.948 和 0.167，而在测试 10 张图像输入时，模型的 SSIM 和 REL 分别为 0.944 和 0.166。可以看出，在输入光度立体图像较少的情况下，模型的光度立体图像重建精确度没有下降，具有鲁棒性。

表 5-8 DiLiGenT 数据集[31] 上测试 96 和 10 张输入光度立体图像的情况下重光照-光度立体双重监督模型的指标

输入图像数量	指标	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
96	MAE ↓	2.27	5.46	7.84	5.42	7.01	8.49	15.55	7.08	7.21	12.74	7.91
	$< err_{15^\circ}$ ↑	100%	98.26%	90.05%	97.11%	95.83%	88.50%	66.18%	92.79%	93.74%	83.02%	90.55%
	SSIM(BI) ↑	0.939	0.944	0.957	0.958	0.937	0.952	0.938	0.970	0.960	0.924	0.948
	SSIM(NBI) ↑	-	-	-	-	-	-	-	-	-	-	-
	REL(BI) ↓	0.070	0.244	0.133	0.067	0.255	0.154	0.256	0.141	0.167	0.185	0.167
	REL(NBI) ↓	-	-	-	-	-	-	-	-	-	-	-
10	MAE↓	3.83	7.52	9.55	7.92	9.83	10.38	17.12	9.36	9.16	14.75	9.94
	$< err_{15^\circ}$ ↑	100%	96.32%	85.16%	91.61%	83.72%	82.60%	62.65%	86.54%	88.87%	66.81%	88.43%
	SSIM(BI) ↑	0.943	0.939	0.951	0.974	0.945	0.948	0.930	0.962	0.931	0.917	0.944
	SSIM(NBI) ↑	0.960	0.933	0.964	0.966	0.940	0.955	0.931	0.969	0.936	0.914	0.947
	REL(BI) ↓	0.064	0.227	0.143	0.064	0.257	0.155	0.271	0.129	0.164	0.186	0.166
	REL(NBI) ↓	0.061	0.241	0.140	0.063	0.262	0.158	0.266	0.131	0.155	0.174	0.165

表 5-8中，BI 表示重建后的图像属于输入图像的光照方向，而 NBI 表示重建后的图像不属于输入图像的光照方向。对 10 张输入图像的实验表明，无论是否属于输入图像的照明方向，重光照图像重建的精度几乎相同。图 5-5 展示了模型在测试 96 张图像、48 张图像和 10 张图像的情况下对 DiLiGenT 数据集中物体 Cat、Reading 和 Cow 的表面法向重建结果和光度立体图像重建结果，重建误差图的亮度提升了 10 倍，以展示更好的细节。可以看出，在使用不同数量的输入图像进行测试时，重建的图像几乎没有显示出任何视觉差异。图 5-5中的黄色框表示具有变化的表面材质的区域。本章模型对这些区域进行了精确的重光照图像重建。

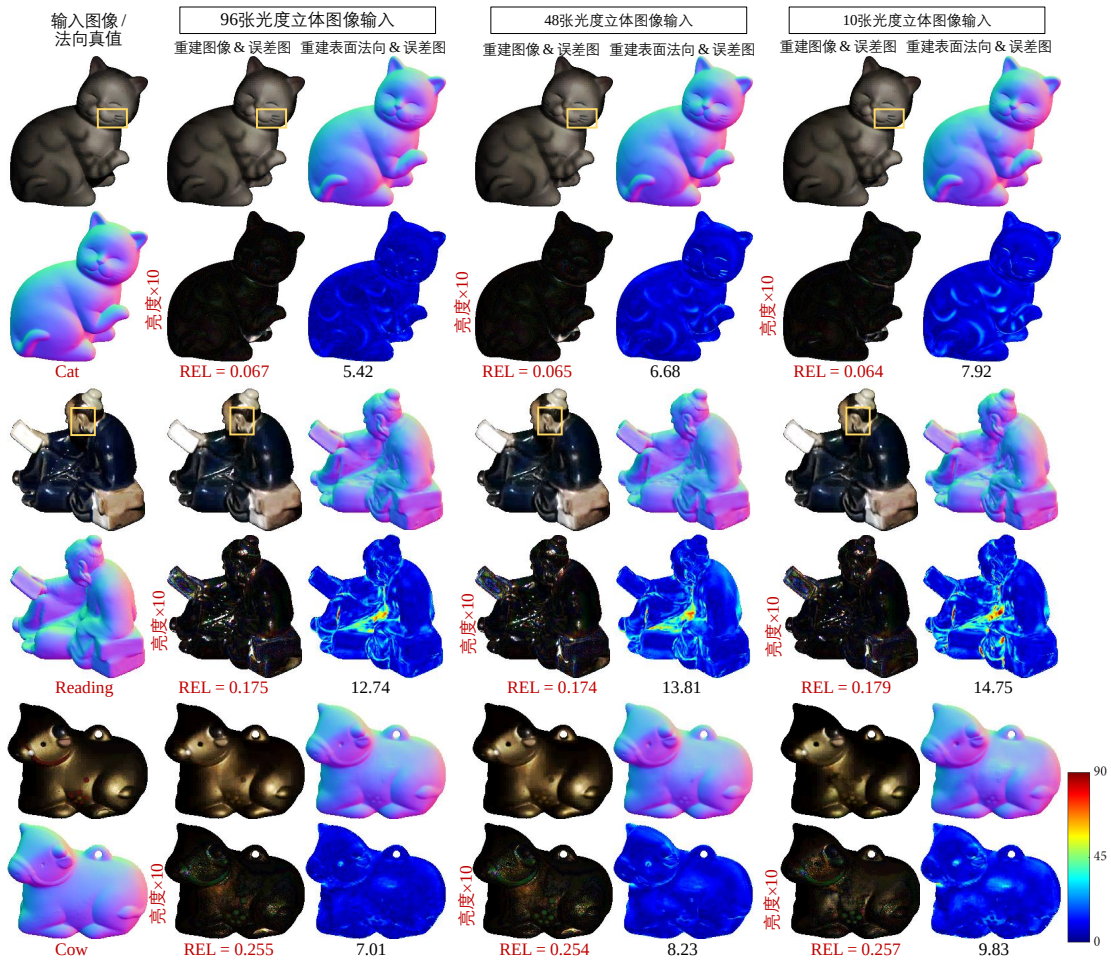


图 5-5 在 96、48 和 10 张输入图像情况下重光照-光度立体双重监督模型的可视化结果

此外图 5-6展示了 DiLiGenT 数据集中物体 Goblet 的示例，该示例使用 48 张输入图像进行了测试。具体来说，在测试中，实验选择了 96 张光度立体图像的

奇数序号（BI 组）作为输入。因此，偶数序号的图像（NBI 组）$2,4,\cdots,96$ 不在输入图像中。图 5-6分别展示了 BI 组和 NBI 组的可视化结果，这两个组具有相同数量的重建图像，重建的图像中 1、15、45、75 属于输入图像的光照方向（BI 组），而重建图像 30、60、90、96 不属于输入图像的光照方向（NBI 组）。如图 5-6所示，在 BI 组和 NBI 组中都可以准确估计镜面反射和阴影的位置。这说明编码的光照信息在双重回归网络中得到了很好的利用。提出的重光照-光度立体双重监督模型可以在任意光照方向下准确地生成指定的重光照光度立体重建图像。

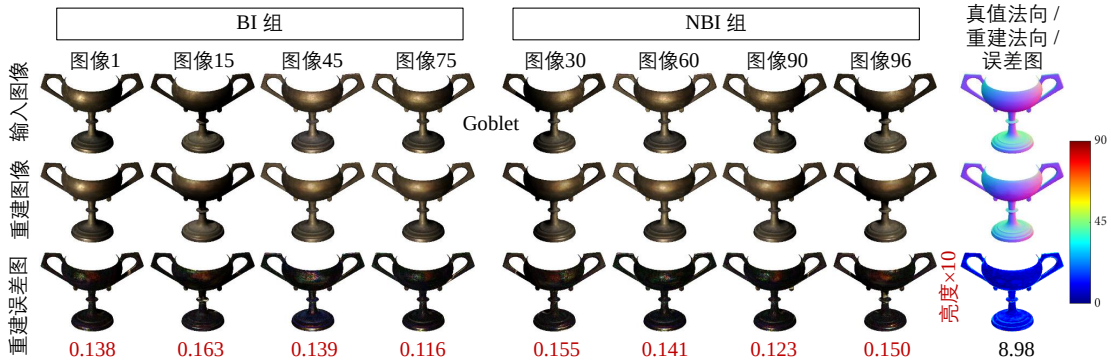


图 5-6 重光照-光度立体双重监督模型对物体 Goblet 的可视化结果

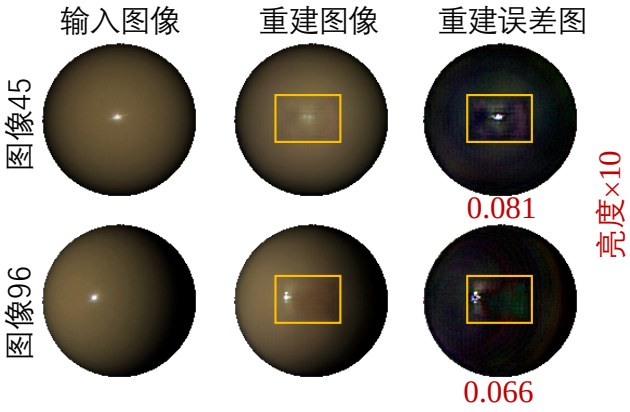


图 5-7 物体 Ball 的重光照光度立体图像重建结果

在上述实验中可以发现，重建图像的误差主要存在于镜面反射区域，其原因可能是在表面法向生成网络中使用了最大池化层特征聚合。最大池化层用于处理任意数量的输入并聚合来自多个输入的特征，从所有特征中提取最显著的信息[25]。然而最显著的信息总是包括镜面反射。在模型的双重回归任务中，最大池化层后的特征 $\boldsymbol{\Psi}_{\max}$ 提供了表面材质特征。不幸的是，它还带来了从所有输入中聚合的镜面反射信息，这可能会导致重建图像中的错误。图 5-7展示了一个明

显的示例物体 Ball，其在不同的输入图像下具有均匀分布的镜面反射。可以看出由于最大池化融合操作，几乎所有输入中存在的镜面反射都聚集在重建图像（黄色框）中，导致错误。

5.6.4 Light Stage Data Gallery 数据集实验结果

本节在更复杂的 Light Stage Data Gallery 数据集[119] 上进一步评估了本章提出的重光照-光度立体双重监督模型。Light Stage Data Gallery 中的图像分辨率远大于 DiLiGenT 数据集中的图像分辨率。由于 GPU 的内存限制，本节仅使用 72 张光度立体图像进行测试。图 5-8显示了实验的结果。由于该数据集中没有表面法向真值，实验仅能定性地评估本章模型。但是，重建图像仍然具有真值（即输入的光度立体图像），因此实验可以定量评估重建图像的精确度。

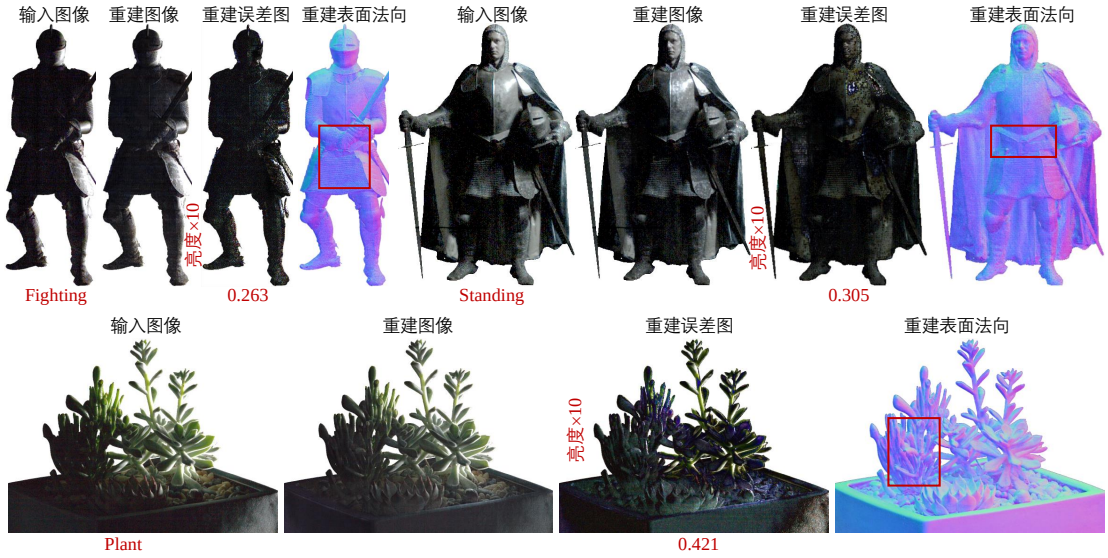


图 5-8 Light Stage Data Gallery 数据集[119] 中，采用 72 张光度立体图像输入模型以进行测试

如图 5-8所示，重建的表面法向可以准确地体现物体的形状。红色框代表具有复杂细节结构的区域，例如 Fighting 的纤维布裙和手，Standing 的腰带以及 Plant 的枝叶。可以看出，重建的表面法向在这些区域没有模糊。在物体 Fighting 中可以区分手指的形状，和衣服的粗糙纹理。这些例子说明了本章提出的重光照-光度立体双重监督模型的有效性。在图 5-8中还可以观察到，表面法向在某些应该光滑的地方存在一些噪声，例如物体 Standing 的盔甲，这可能由于输入的低质量图像中的噪声所引起。

图 5-8还展示了重建光度立体图像的结果。然而观察到重建图像的 REL 比为 DiLiGenT 数据集[31] 的值差。这可能是由于观察到的图像中的高频噪声影响了性能。显然噪声输入图像会影响重建图像的准确性。尽管如此，模型准确地生成了镜面反射和阴影的位置，例如在物体 Fighting 和 Standing 上表现的金属质感。

5.7 本章小结

本章提出了一种重光照-光度立体双重监督模型，用于同时重建表面法向和重光照光度立体图像。主要贡献是探索了重建图像的过程以进一步提高表面法向重建的准确性。这是通过对输入图像和重建图像引入额外的图像重建损失来实现的，从而形成一个闭环来提供额外的监督。此外，模型可以在任意光照方向下生成准确的重建图像（称之为重光照光度立体图像），以直观地显示表面的纹理信息和各向异性反射特性。对最广泛使用的 DiLiGenT 数据集的广泛定量比较表明，本章提出的模型优于传统和基于学习的校准光度立体方法。具体来说，实验结果表明模型重建的表面法向在 96 张和 10 张光度立体图像的输入下都能取得准确的结果，模型可以更好地处理复杂结构和强阴影区域。Light Stage Data Gallery 数据集上的额外定性实验进一步证实了提出的重光照-光度立体双重监督模型的鲁棒性。

尽管如此，本章模型也有一些缺陷。首先，如图 5-7所示，模型在密集的镜面反射区域会产生较大的重建图像误差。由于重建表面法向和重建图像是在提出的闭环的端到端模型中实现，重建图像存在的误差可能会进而影响到表面法向的重建，这导致提出的重光照-光度立体双重监督模型的性能弱于一些最新的方法，例如 PS-FCN (Norm.)[69] 和 GPS-Net[77]。其次由于重建图像需要预先知道表面法向、光照信息和表面材质，因此在双重回归网络中需要通过拼接操作输入表面法向生成网络的最大池化层聚合特征 $\mathbf{\Psi}_{\max}$，这就导致了表面材质不可改变。如图 5-6所示，尽管提出的模型可以生成物体 Goblet 在任意指定光照方向下的重光照光度立体图像，但是却不能改变 Goblet 的表面材质。因此，双重监督模型只能称之为重光照-光度立体，而非重渲染-光度立体。这严重限制了应用的场景，例如只有重渲染的光度立体图像才可以为光度立体的数据集扩充样本。在下一章中，本文进一步提出了重渲染-光度立体三重监督模型，以解决上述问题，并提高了表面法向重建的精度。

6 重渲染-光度立体三重监督模型

6.1 研究背景

上一章提出了一种重光照-光度立体双重监督模型，初步探索了重建图像这一双重回归过程以对输入图像和重建图像引入额外的图像重建损失，进一步提高表面法向重建的准确性。然而其在密集的镜面反射区域的重建精度并不是十分理想，导致表面法向的重建精度弱于一些最新的基于学习的光度立体方法。因此本章首先提出了局部-全局特征融合和深浅最大池化层聚合的网络结构。此外，由于反射率特征从表面法向生成网络固定获得的处理方式，导致重建的光度立体图像仅能改变光照方向而不能改变表面材质，因此只能称之为重光照的光度立体图像。然而，重光照的光度立体的应用场景有限。在光度立体等三维重建任务中，基于学习的方法面临的首要问题就是缺乏训练的数据集。这是因为三维重建任务中，我们很难获得准确的真实物体的三维真值。因此，如何在有限的光度立体数据集中进行数据的扩充，在有限的表面法向真值上渲染出尽量多的任意材质、任意光照的光度立体图像，就显得非常重要。

为此，本章提出了一种重渲染-光度立体三重监督模型，它从光度立体图像中学习表面法向，并在来自不同方向的光照和表面材质下重渲染光度图像。该模型由结构生成网络和重渲染网络的两个子网络组成，它们级联起来以端到端的方式执行表面法向重建和光度立体图像重渲染。重渲染网络为表面法向重建引入了额外的监督，与结构生成网络形成了一个闭环结构。模型还在重渲染网络中对光照方向和表面材质进行编码，以实现任意材质和光照方向的重渲染。在训练中我们建立了一个并行框架来同时学习一个物体的两种任意表面材质，提供额外的图像变换损失。因此本章提出的模型通过三种不同的损失函数，即余弦损失、图像重建损失和图像变换损失对网络进行优化训练。在训练中，模型将结构生成网络重建的表面法向和真值法向交替输入到重渲染网络中，以实现稳定的训练。实验表明，本章提出的模型可以准确地恢复具有任意输入数量图像的表面法向，并且可以使用任意表面材料重新渲染对象的图像。大量的实验结果表明，该模型优于基于单一表面恢复网络的方法，并在 100 种材料上显示出逼真的渲染结果。此外，通过对重渲染的光度立体图像再次输入模型，证明了提出的重渲染-光度立体三重监督模型作为光度立体数据集扩充方法的可行性。

6.2 模型概述

本章提出了一种重渲染-光度立体三重监督模型，同时重建物体表面法向和任意材质、任意光照方向下的重渲染光度立体图像。为了实现上述目标，本章提出了一个深度学习模型，如图 6-1所示，它包含两个关联的子网络，分别称为结构生成网络和重渲染网络。它们以级联的方式连接，通过最小化所示的三个损失函数（余弦损失、图像重建损失和图像变化损失）进行训练。图 6-1种，A 和 B 代表一个物体的两种任意材质。结构生成网络从校准的光度立体图像中重建物体的表面法向，而重渲染网络使用预测的表面法向再现物体在不同表面材质、不同光照条件下的光度立体图像。实际上，重渲染网络可以看作是表面法向重建的逆过程，为结构生成网络提供了额外的监督，以减少表面法向学习的潜在空间并形成闭环结构。

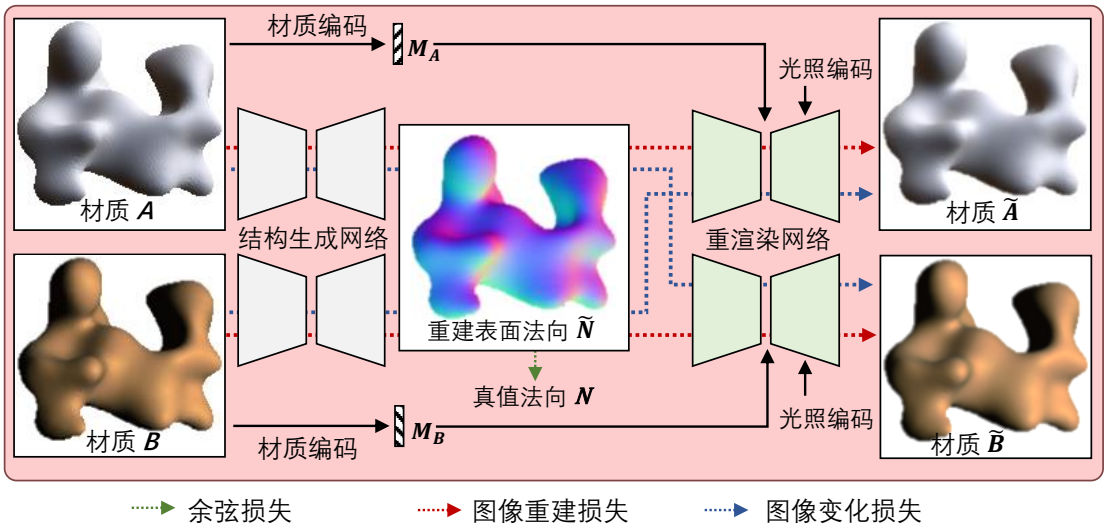


图 6-1 本章提出的重渲染-光度立体三重监督模型的整体结构

为了实现在任意的光照方向和表面材质下产生逼真的外观，模型以两种方式解决这个问题。如图 6-1所示，首先模型将编码的光照与高维特征相结合，在不同的光照方向下输出任意指定的渲染图像。其次模型按照条件 GAN[130] 中的编码方式，将表面材质信息视为一个条件，对 MERL BRDFs 数据集[116] 中的 100 种材料进一步编码形成 100 维 one-hot 特征，以使得重渲染网络可以渲染成任意表面材质下的光度立体图像。为了更好地学习表面材质属性的特征并提高数据集中训练数据的利用率，本章提出了一个并行框架来学习使用两种不同材料的对象的渲染。为了实现这一点，模型同时输入两组光度立体图像，并用交换的表

面材质编码特征渲染重建的物体。重渲染网络根据不同的材质编码对两种材质进行回归，形成图像重建损失和图像变化损失。按照顺序，首先介绍结构生成网络和重渲染网络，再介绍三重监督的损失函数和训练方法。

6.3 结构生成网络

结构生成网络可以看作是一个多输入单输出结构，如图 6-2所示，其由三个部分组成，分别是共享权重特征提取器 f_{GE} 和法向回归器 f_{GR}。通过 f_{GE} 中的深浅融合最大池化层特征聚合，结构生成网络可以使用任意输入数量的光度立体图像对其进行测试。

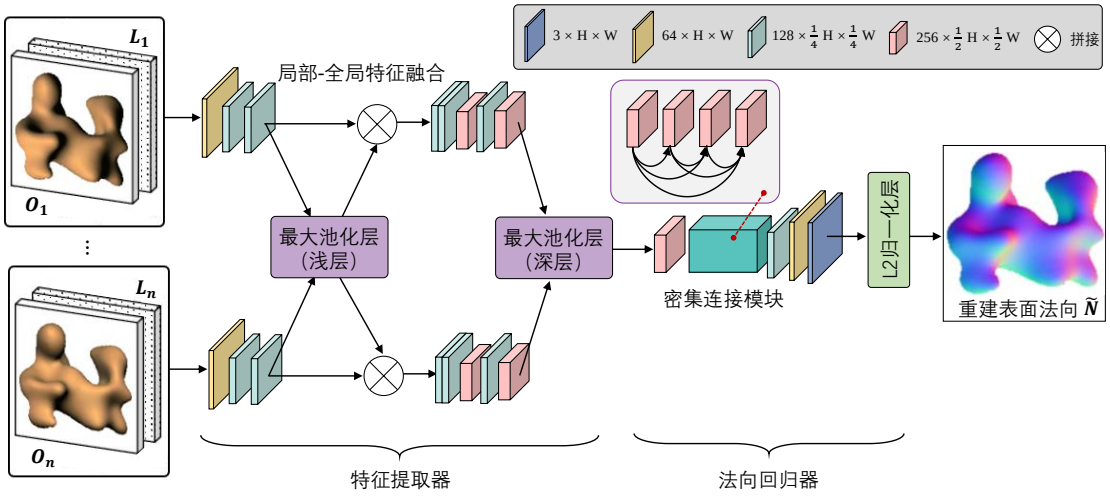


图 6-2 结构生成网络结构图，包括特征提取器 f_{GE} 和法向回归器 f_{GR}

作为标定的光度立体任务，首先将输入的光度立体图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n \in \mathbb{R}^{3\times H\times W}$ 与对应的光照方向 $\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n \in \mathbb{R}^3$ 进行融合。融合的方法是将 $\boldsymbol{l}_j \in \mathbb{R}^3$ 沿 H 和 W 的方向复制，扩展至与图像具有相同分辨率大小的张量 $\boldsymbol{L}_j \in \mathbb{R}^{3\times H\times W}$，其中 $j \in \{1, 2, \cdots, n\}$，并分别与图像 $\boldsymbol{O}_1, \boldsymbol{O}_2, \cdots, \boldsymbol{O}_n$ 在第一维度上拼接，形成张量 $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_n \in \mathbb{R}^{6\times H\times W}$。

在特征提取器 f_{GE} 中，本章提出了一种全局-局部特征融合和深浅最大池化层特征聚合的结构，以更好的提取重建表面法向所需要的特征。最大池化层特征聚合已被证明可以解决任意数量的输入特征提取问题[25]。因此模型的网络结构也采用了这种策略作为基础的模块。然而最大池化层特征聚合在同一位置上，从所有特征中仅选择最大响应值，这忽略了图像的一些非最大特征，而丢弃的局部特征（聚合前每个图像的特征）可能对表面法向的重建很重要。另一方面，以

前基于深度学习的光度立体模型未采用来自不同层的特征的聚合，即深层特征和浅层特征。深浅特征对模型的学习都有不可替代的影响，由于深浅特征具有不同的感受野，可能包含各自独特的信息，有利于重建物体的表面法向。因此特征提取器 f_{GE} 设计了全局-局部特征融合和深浅最大池化层特征聚合的结构，其可以被表示为：

$$\mathbf{\Psi} = f_{GE}(\mathbf{\Phi}_j; \theta_{GE}),\ j \in \{1, 2, \cdots, n\}, \tag{6-1}$$

其中 θ_{GE} 表示特征提取器 f_{GE} 中的可学习参数，$\mathbf{\Psi} \in \mathbb{R}^{256\times\frac{1}{4}H\times\frac{1}{4}W}$ 表示经过特征提取网络提取得到的深层特征。如图 6-2所示，模型使用两个最大池化层进行深浅多层特征聚合，第一个最大池化层用于从全局和局部提取特征（利用拼接操作合并最大池化层聚合特征和每个图像生成的特征）。这样，原始的局部特征信息与全局特征同时保留，有利于更充分的特征学习和训练，避免遗漏图像非最大特征的有用特征。本章认为在训练的早期，模型应该尽可能多地保留原始信息，而过早丢弃局部特征会丢失有用的信息，降低结果的准确性。第二个最大池化层则只聚合出全局特征。这是因为之前已经做了足够的特征学习，仅仅聚合全局特征就可以提高网络运行的效率。此外深浅最大池化层特征聚合可以融合不同感受野的特征，提高了结果的准确性。与原始框架，例如 PS-FCN[25] 相比，这些特征融合方法可以进一步提高表面法向的重建精度。表 6-1展示了本章提出的特征提取器 f_{GE} 的具体结构，其中卷积层的激活函数设置为均为 Leaky Relu。

表 6-1 结构生成网络中特征提取器 f_{GE} 的网络结构

输入	操作	卷积核大小	步长	输出
$3\times H\times W$	卷积层 1	3×3	1	$64\times H\times W$
$64\times H\times W$	卷积层 2	3×3	2	$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W3\times 3$	卷积层 3	3×3	1	$128\times\frac{1}{2}H\times\frac{1}{2}W$
n 个 $128\times\frac{1}{2}H\times\frac{1}{2}W$	最大池化层 1			$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W$	拼接（最大池化层 1、卷积层 3 ）			$256\times\frac{1}{2}H\times\frac{1}{2}W$
$256\times\frac{1}{2}H\times\frac{1}{2}W$	卷积层 4	3×3	2	$256\times\frac{1}{4}H\times\frac{1}{4}W$
$256\times\frac{1}{4}H\times\frac{1}{4}W$	转置卷积层 1	4×4	2	$128\times\frac{1}{2}H\times\frac{1}{2}W$
$128\times\frac{1}{2}H\times\frac{1}{2}W$	卷积层 5	3×3	2	$256\times\frac{1}{4}H\times\frac{1}{4}W$
n 个 $256\times\frac{1}{4}H\times\frac{1}{4}W$	最大池化层 2			$256\times\frac{1}{4}H\times\frac{1}{4}W$

随后是设计的法向回归器 f_{GR}，以从深层特征 $\boldsymbol{\Psi}$ 中回归物体的表面法向 $\tilde{\boldsymbol{N}}$，可以表示为：

$$\tilde{\boldsymbol{N}} = f_{GR}(\boldsymbol{\Psi}; \theta_{GR}), \tag{6-2}$$

其中 θ_{GR} 表示法向回归器 f_{GR} 中可学习的参数。为了实现更好的表面法向重建，法向回归器 f_{GR} 采用了基于密集连接（DenseNet）[131] 的模块。模型使用了三个密集连接模块，分别具有 2、4 和 3 层的卷积层。表 6-2展示了特征回归器 f_{GR} 的具体网络结构。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。

表 6-2 结构生成网络中法向回归器 f_{GR} 的网络结构

输入	操作	卷积核大小	步长	输出
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 1	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{2}H \times \frac{1}{2}W$	密集连接层 1	[1×1 卷积层] $\times 2$ [3×3 卷积层] $\times 2$ 1×1 过渡层		$256 \times \frac{1}{2}H \times \frac{1}{2}W$
$256 \times \frac{1}{2}H \times \frac{1}{2}W$	密集连接层 2	[1×1 卷积层] $\times 4$ [3×3 卷积层] $\times 4$ 1×1 过渡层		$256 \times \frac{1}{2}H \times \frac{1}{2}W$
$256 \times \frac{1}{2}H \times \frac{1}{2}W$	密集连接层 3	[1×1 卷积层] $\times 3$ [3×3 卷积层] $\times 3$ 1×1 过渡层		$256 \times \frac{1}{2}H \times \frac{1}{2}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 1	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 2	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 2	3×3	1	$3 \times H \times W$
$3 \times H \times W$	L2 归一化层			$3 \times H \times W$

6.4 重渲染网络

重渲染网络 f_{Render} 旨在从结构生成网络中重建的表面法向 $\tilde{\boldsymbol{N}}$ 渲染任意表面材质、任意光照方向的重渲染图像。图 6-3展示了重渲染网络 f_{Render} 的网络结构。

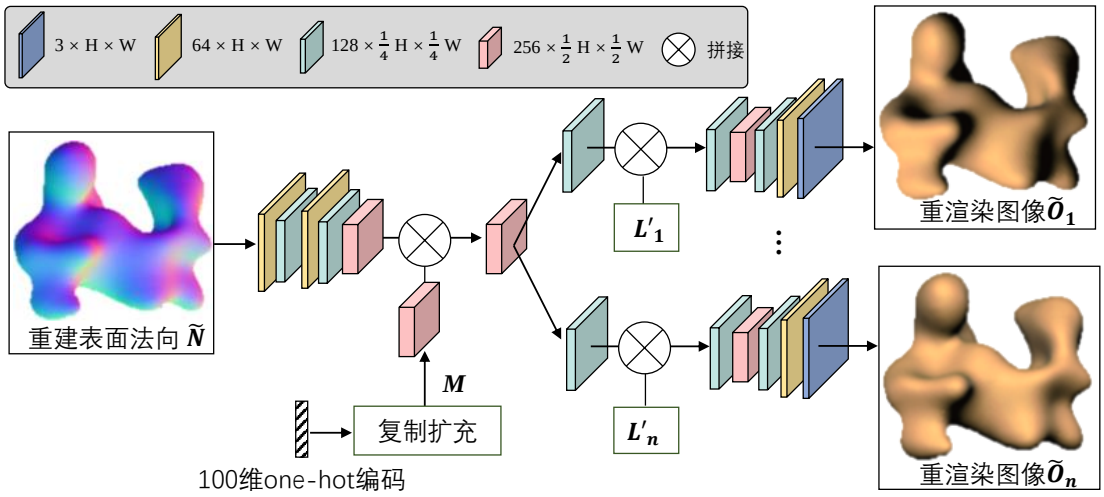


图 6-3 重渲染网络 f_{Render} 的结构图。

模型首先对 MERL BRDFs 数据集[116] 中的 100 种表面材质使用 one-hot 特征向量 $\boldsymbol{m} \in \mathbb{R}^{100}$ 进行编码。其中，每一种材质在 100 维的 one-hot 特征向量中都有唯一的表示方法。如图 6-3所示，先将 $\boldsymbol{m} \in \mathbb{R}^{100}$ 沿 H 和 W 的方向复制，扩展至 $\frac{1}{4}H \times \frac{1}{4}W$ 大小，记作 $\boldsymbol{M} \in \mathbb{R}^{100\times\frac{1}{4}H\times\frac{1}{4}W}$，以便与重渲染网络 f_{Render} 中的特征融合。相似地，也将输入的光照方向 $\boldsymbol{l} \in \mathbb{R}^3$ 沿 H 和 W 的方向复制，扩展至 $\frac{1}{2}H \times \frac{1}{2}W$ 大小，记作 $\boldsymbol{L'} \in \mathbb{R}^{3\times\frac{1}{2}H\times\frac{1}{2}W}$，以便与 f_{Render} 中的特征融合。表 6-3展示了重渲染网络 f_{Render} 的具体网络结构。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。

6.5 三重监督损失函数与训练方法

在上述 6.3结构生成网络和 6.4重渲染网络中，为了方便说明，这里仅拿一种材质的物体来阐述网络结构。实际上，如图 6-1所示，为了更好地学习表面材质的特征并提高数据集中训练数据的利用率，本章提出了一个并行框架来同时学习使用两种不同表面材质的对象的法向重建和图像重建过程。因此，本章模型总共有 4 条训练的路线，如果用 A、B 表示两种材质，那么有（1）$\boldsymbol{O}_1^A, \boldsymbol{O}_2^A, \cdots, \boldsymbol{O}_n^A \rightarrow \tilde{\boldsymbol{N}} \rightarrow \tilde{\boldsymbol{O}}_1^A, \tilde{\boldsymbol{O}}_2^A, \cdots, \tilde{\boldsymbol{O}}_n^A$；（2）$\boldsymbol{O}_1^A, \boldsymbol{O}_2^A, \cdots, \boldsymbol{O}_n^A \rightarrow \tilde{\boldsymbol{N}} \rightarrow \tilde{\boldsymbol{O}}_1^B, \tilde{\boldsymbol{O}}_2^B, \cdots, \tilde{\boldsymbol{O}}_n^B$；（3）$\boldsymbol{O}_1^B, \boldsymbol{O}_2^B, \cdots, \boldsymbol{O}_n^B \rightarrow \tilde{\boldsymbol{N}} \rightarrow \tilde{\boldsymbol{O}}_1^A, \tilde{\boldsymbol{O}}_2^A, \cdots, \tilde{\boldsymbol{O}}_n^A$；（4）$\boldsymbol{O}_1^B, \boldsymbol{O}_2^B, \cdots, \boldsymbol{O}_n^B \rightarrow \tilde{\boldsymbol{N}} \rightarrow \tilde{\boldsymbol{O}}_1^B, \tilde{\boldsymbol{O}}_2^B, \cdots, \tilde{\boldsymbol{O}}_n^B$。为了方便叙述，进一步将（1）、（2）、（3）和（4）分别简化为 $\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A$、$\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^B$、$\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^A$ 和 $\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^B$。可以看出，这些训练的路线可以分为两类，第一类是输入图像和重渲染的图像材质相同，包含 $\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A$ 和 $\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^B$（图

表 6-3 重渲染网络 f_{Render} 的网络结构

输入	操作	卷积核大小	步长	输出
$3 \times H \times W$	卷积层 1	3×3	1	$64 \times H \times W$
$64 \times H \times W$	卷积层 2	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 1	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 3	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 4	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	拼接（卷积层 4、材质编码 $\boldsymbol{M}$）			$356 \times \frac{1}{4}H \times \frac{1}{4}W$
$356 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 5	3×3	1	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	卷积层 6	3×3	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	拼接（卷积层 6、光照编码 $\boldsymbol{L'}_j$）			$131 \times \frac{1}{2}H \times \frac{1}{2}W$
$131 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 7	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 8	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 2	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 3	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 9	3×3	1	$3 \times H \times W$

6-1中红色箭头所示），第二类则是输入图像和重渲染的图像材质不同，包含 $\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^B$ 和 $\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^A$（图 6-1中蓝色箭头所示）。

对于上述两种分类的监督损失，本章称之为图像重建损失 $\mathcal{L}_{\text{Recons}}$ 和图像变化损失 $\mathcal{L}_{\text{Transf}}$。

图像重建损失 $\mathcal{L}_{\text{Recons}}$ 衡量相同材质的重渲染光度立体图像和输入光度立体图像之间的差异，可以被定义为：

$$\mathcal{L}_{\text{Recons}} = \mathcal{L}(\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A, \boldsymbol{O}^A) + \mathcal{L}(\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^B, \boldsymbol{O}^B), \tag{6-3}$$

其中 $\mathcal{L}(\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A, \boldsymbol{O}^A)$ 代表：

$$\mathcal{L}(\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A, \boldsymbol{O}^A) = \frac{1}{j}\frac{1}{HW}\sum_{j}^{n}\sum_{i}^{HW} \| \boldsymbol{O}_{j,i}^A - \tilde{\boldsymbol{O}}_{j,i}^{A \rightarrow A} \|_2^2, \tag{6-4}$$

其中 $\boldsymbol{O}_{j,i}^A$ 和 $\tilde{\boldsymbol{O}}_{j,i}^{A \rightarrow A}$ 表示第 j 张 A 材质下输入的光度立体图像和重建光度立体图像中像素 i 位置上的值。相似地，$\mathcal{L}(\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^B, \boldsymbol{O}^B)$ 代表：

$$\mathcal{L}(\boldsymbol{O}^B \to \tilde{\boldsymbol{O}}^B, \boldsymbol{O}^B) = \frac{1}{j}\frac{1}{HW}\sum_{j}^{n}\sum_{i}^{HW}\|\boldsymbol{O}_{j,i}^B - \tilde{\boldsymbol{O}}_{j,i}^{B\to B}\|_2^2, \tag{6-5}$$

其中 $\boldsymbol{O}_{j,i}^B$ 和 $\tilde{\boldsymbol{O}}_{j,i}^{B\to B}$ 表示第 j 张 B 材质下输入的光度立体图像和重建光度立体图像中像素 i 位置上的值。

图像变化损失 $\mathcal{L}_{\text{Transf}}$ 衡量不同材质的重渲染光度立体图像和输入光度立体图像之间的差异，可以被定义为：

$$\mathcal{L}_{\text{Transf}} = \mathcal{L}(\boldsymbol{O}^A \to \tilde{\boldsymbol{O}}^B, \boldsymbol{O}^B) + \mathcal{L}(\boldsymbol{O}^B \to \tilde{\boldsymbol{O}}^A, \boldsymbol{O}^A), \tag{6-6}$$

其中 $\mathcal{L}(\boldsymbol{O}^A \to \tilde{\boldsymbol{O}}^B, \boldsymbol{O}^B)$ 代表：

$$\mathcal{L}(\boldsymbol{O}^A \to \tilde{\boldsymbol{O}}^B, \boldsymbol{O}^B) = \frac{1}{j}\frac{1}{HW}\sum_{j}^{n}\sum_{i}^{HW}\|\boldsymbol{O}_{j,i}^B - \tilde{\boldsymbol{O}}_{j,i}^{A\to B}\|_2^2, \tag{6-7}$$

其中 $\boldsymbol{O}_{j,i}^B$ 和 $\tilde{\boldsymbol{O}}_{j,i}^{A\to B}$ 表示第 j 张 B 材质下输入的光度立体图像和由 A 材质的输入图像到 B 材质的重建光度立体图像中像素 i 位置上的值。相似地，$\mathcal{L}(\boldsymbol{O}^B \to \tilde{\boldsymbol{O}}^A, \boldsymbol{O}^A)$ 代表：

$$\mathcal{L}(\boldsymbol{O}^B \to \tilde{\boldsymbol{O}}^A, \boldsymbol{O}^A) = \frac{1}{j}\frac{1}{HW}\sum_{j}^{n}\sum_{i}^{HW}\|\boldsymbol{O}_{j,i}^A - \tilde{\boldsymbol{O}}_{j,i}^{B\to A}\|_2^2, \tag{6-8}$$

其中 $\boldsymbol{O}_{j,i}^A$ 和 $\tilde{\boldsymbol{O}}_{j,i}^{B\to A}$ 表示第 j 张 A 材质下输入的光度立体图像和由 B 材质的输入图像到 A 材质的重建光度立体图像中像素 i 位置上的值。

除了上述两个图像重建损失 $\mathcal{L}_{\text{Recons}}$ 和图像变化损失 $\mathcal{L}_{\text{Transf}}$，提出的模型还有传统的余弦损失 $\mathcal{L}_{\text{Cosine}}$ 提供监督，表示法向真值 $\boldsymbol{N}$ 和重建的表面法向 $\tilde{\boldsymbol{N}}$ 之间的余弦角度误差。由于模型采用并行训练的网络，实际上重建的表面法向有两个路线，即材质 A 下的输入光度立体图像和材质 B 下的输入光度立体图像，分别记作 $\tilde{\boldsymbol{N}}^A$ 和 $\tilde{\boldsymbol{N}}^B$，因此，重渲染-光度立体三重监督模型的余弦损失 $\mathcal{L}_{\text{Cosine}}$ 的定义由下式所示：

$$\mathcal{L}_{\text{Cosine}} = \frac{1}{HW}\sum_{i}^{HW}\left(2 - \boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i^A - \boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i^B\right), \tag{6-9}$$

其中 $\cdot$ 操作代表点乘。如果在像素位置 i 上重建的表面法向 $\tilde{\boldsymbol{N}}_i^A$ 和 $\tilde{\boldsymbol{N}}_i^B$ 与真值法向 $\boldsymbol{N}_i$ 越相似，则其点乘 $\boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i^A$ 和 $\boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i^B$ 都将越接近 1，此时式 (6-9) 的值将越接近 0。

在训练中，模型通过最小化以下联合损失函数 $\mathcal{L}$ 来优化提出的结构生成网络和重渲染网络，联合损失函数 $\mathcal{L}$ 包含上述讨论的三个损失函数，即图像重建损失 $\mathcal{L}_{\text{Recons}}$、图像变化损失 $\mathcal{L}_{\text{Transf}}$ 和余弦损失 $\mathcal{L}_{\text{Cosine}}$，可以被表示为：

$$\mathcal{L} = \mathcal{L}_{\text{Cosine}} + \lambda(\mathcal{L}_{\text{Recons}} + \mathcal{L}_{\text{Transf}}), \tag{6-10}$$

其中 λ 为图像重建损失和图像变化损失的权重，设置为 0.1。在模型中，重渲染网络的学习需要表面法向作为输入。然而，结构生成网络重建的表面法向在训练开始时是不准确的。使用不准确的输入会使重渲染网络收敛到不正确的局部最小值。不同于 5.5中采用随训练 epoch 增长而变化的权重，本章提出了一种交替训练策略来训练重渲染网络，它交替使用重建的表面法向和法向真值作为重渲染网络的输入。具体来说，对于每个样本，在训练完一次上述描述的（1）、(2)、(3)和（4）路线后，又使用真值法向 $\boldsymbol{N}$ 额外训练重渲染网络两次：一次用于生成材质为 A 的重渲染光度立体图像，另一次用于生成材质为 B 的重渲染光度立体图像。因此实际上，式 (6-10) 中的联合损失函数 $\mathcal{L}$ 应该写成如下形式：

$$\mathcal{L} = \mathcal{L}_{\text{Cosine}} + \lambda(\mathcal{L}_{\text{Recons}} + \mathcal{L}_{\text{Transf}} + \mathcal{L}_{\text{Assist}}), \tag{6-11}$$

其中 $\mathcal{L}_{\text{Assist}}$ 为辅助重建损失，其衡量采用真值法向 $\boldsymbol{N}$ 作为输入的重渲染网络渲染的 A、B 材质下重建光度立体图像 $\tilde{\boldsymbol{O}}^{N\to A}$、$\tilde{\boldsymbol{O}}^{N\to B}$ 与输入光度立体图像 $\boldsymbol{O}^A$、$\boldsymbol{O}^B$ 之间的误差，可以表示为：

$$\mathcal{L}_{\text{Assist}} = \frac{1}{j}\frac{1}{HW}\sum_{j}^{n}\sum_{i}^{HW}(\|\boldsymbol{O}^A_{j,i} - \tilde{\boldsymbol{O}}^{N\to A}_{j,i}\|_2^2 + \|\boldsymbol{O}^B_{j,i} - \tilde{\boldsymbol{O}}^{N\to B}_{j,i}\|_2^2), \tag{6-12}$$

其中 $\boldsymbol{O}^B_{j,i}$、$\boldsymbol{O}^A_{j,i}$ 和 $\tilde{\boldsymbol{O}}^{N\to A}_{j,i}$、$\tilde{\boldsymbol{O}}^{N\to B}_{j,i}$ 分别表示第 j 张 A、B 材质下输入的光度立体图像和由法向真值渲染得到得 A、B 材质的重建光度立体图像中像素 i 位置上的值。消融实验（6.6.2）表明，该策略有利于结构生成网络和重渲染网络的收敛，以获得更精确的重建结果。

6.6 实验结果

本节对提出的重渲染-光度立体三重监督模型在多个数据集上与最先进的基于深度学习的方法以及传统方法进行了比较。首先，本节对模型进行了消融实验

和分析，包括结构生成网络的结构设计、重渲染网络对表面法向重建的影响、辅助重建损失 $\mathcal{L}_{\text{Assist}}$（即交替训练策略）的作用、权重 λ 的选取等。本文采用 MAE 指标衡量重建表面法向的精度，采用 REL 指标衡量重渲染光度立体图像的精度。注意，对于重渲染为相同材质的图像（$\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A$、$\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^B$），其指标用 REL (O) 表示，对于重渲染为不同材质的图像（$\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^B$、$\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^A$），其指标用 REL (C) 表示。

6.6.1 实验设置

使用默认的 Adam 优化器进行优化（$\beta_1 = 0.9$ and $\beta_2 = 0.999$），初始的学习率最初设置为 0.001 且每过 5 个 epoch 学习率除以 2，来训练提出的模型。使用大小为 128 的 batchsize，epoch 为 20 的设置在 8 张 RTX 3090 上进行训练。此外，在训练中将输入图像的分辨率 $H \times W$ 设置为 32×32。用于训练的合成数据集与 PS-FCN[25] 使用的相同，包括两个形状数据集，分别是 Blobby 数据集[114] 和 SculptureBlobby 数据集[115]，由 MERL BRDFs 数据集[116] 渲染得到。但是，使用数据集的方式不同。在实验中，我们使用两种随机选择的材料（来自 MERL BREDFs 数据集）渲染对象的每个样本，在图 6-1中表示为材料 A 和 B。总计 84360 个（42180× 2）用于训练的样本，即一个 epoch 含有 84360 个样本，每个样本输入 32 张不同光照方向下的光度立体图像。

6.6.2 消融实验与分析

为了定量评估本章提出的重渲染-光度立体三重监督模型的网络设计和训练策略的有效性。在基于合成测试集[69] 的物体 Armadillo 和 Dragon 上（如图 6-4所示），使用来自 MERL BRDFs 数据集[116] 的所有 100 种表面材质和上半球空间内的 100 个随机分布的光照下渲染。我们通过计算两个物体在平均 100 种材质下所有重建表面法向的 MAE 和 REL 来衡量本章提出的模型的表面重建性能，并用 REL (O) 衡量当前材质下的重渲染图像的效果，用 REL (C) 衡量渲染为其他 99 种重渲染图像的效果。

表 6-4展示了消融实验的结果。实验首先对结构生成网络中提出的全局-局部特征融合和深浅最大池化层特征聚合的结构进行消融实验。编号 ID (0) 表示本章提出的模型。ID (G1) 表示无全局-局部特征融合，即在第一次最大池化层特征聚合后不再拼接局部特征（此时表 6-1中应取消拼接层，并且卷积层 4 的输入应为 $128 \times \frac{1}{2}H \times \frac{1}{2}W$）。ID (G2) 表示仅采用最后一次的深层最大池化层特征聚合

图 6-4 消融实验中使用的测试物体 Armadillo 和 Dragon

（由于取消第一次最大池化层，同时也无全局-局部特征融合）。ID (G3) 则表示在法向回归器中取消三层密集连接模块[131]。其次，实验在 ID (R4) 中对比了重渲染网络的作用，即单独使用提出的结构生成网络是否可以重建出误差更小的物体表面法向。最后，实验在 ID (T5) 中讨论了使用重建法向和真值法向交替训练策略的效果（当联合损失 $\mathcal{L}$ 中不含有辅助重建损失 $\mathcal{L}_{\text{Assist}}$ 时，即等价于模型不使用交替训练策略）。

表 6-4 消融实验的结果。

编号	消融方法	MAE ↓	REL(O) ↓	REL(C) ↓
ID (0)	提出的方法	**5.96**	**0.073**	**0.072**
ID (G1)	无全局-局部特征融合	6.07	0.077	0.077
ID (G2)	无浅层最大池化层	6.20	0.083	0.081
ID (G3)	无密集连接模块[131]	6.12	0.078	0.077
ID (R4)	无重渲染网络	6.71	-	-
ID (T5)	无辅助重建损失 $\mathcal{L}_{\text{Assist}}$	6.49	0.084	0.085

如表 6-4所示，ID (0) 与 ID (G1) 的实验表明全局-局部特征融合的有效性，但保留了两个最大池化层的特征聚合。仅使用最大值池化进行全局特征聚合，可能导致缺少一些重要的非最大响应值特征。实验表明，使用全局和局部特征融合的网络可以全面提升在表面法向重建和重渲染图像重建上的精度。ID (G2) 表示在结构生成网络的特征提取器 f_{GE} 中仅在最后使用一个最大池化层来聚合特征。可以发现，仅使用一个最大池化层聚合特征也会使重建表面法向和重建图像的性能下降。这是因为，经过不同卷积层处理的特征含有不同的信息，在深层的最

大池化层中被丢弃的信息可能在浅层中有着最大的激活值，而本章提出的深浅最大池化层特征聚合的结构可以保留这些先前被丢弃的信息，有利于重建物体的表面法向。ID (G3) 则验证了在结构生成网络的法向回归器 f_{GR} 中添加密集连接模块[131] 的作用，在添加 3 层的密集连接模块后，本章提出的模型的表面重建误差减小了 0.18 度，在重渲染光度立体图像上的误差也有所减小。这是因为密集连接模块减轻了梯度消失的现象，并通过加强不同层特征的传递拼接更有效地利用了提取的特征。

此外，如表 6-4所示，为了验证提出的重渲染-光度立体三重监督模型中，图像重建为表面法向重建任务带来的提升，本节进行了 ID (R4) 的实验。在 ID (R4) 中，仅保留相同的结构生成网络，而取消了重渲染网络。因此 ID (R4) 的网络仅采用单一的余弦损失进行优化。可以发现其表面法向重建的误差比提出的多监督模型大很多。这说明重渲染网络可以看作是表面法向重建的逆过程，为结构生成网络提供了额外的监督并使模型形成闭环结构，减少了表面法向学习的潜在空间，提升了重建的精度。

在表 6-4中，实验最后评估了提出的交替训练策略，即以 $\mathcal{L}_{\text{Cosine}}+\lambda(\mathcal{L}_{\text{Recons}}+\mathcal{L}_{\text{Transf}}+\mathcal{L}_{\text{Assist}})$ 替代 $\mathcal{L}_{\text{Cosine}}+\lambda(\mathcal{L}_{\text{Recons}}+\mathcal{L}_{\text{Transf}})$。由于用两个级联网络学习整个过程是非常困难的（$\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^A$ 、$\boldsymbol{O}^A \rightarrow \tilde{\boldsymbol{O}}^B$、$\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^A$ 和 $\boldsymbol{O}^B \rightarrow \tilde{\boldsymbol{O}}^B$），而重渲染网络的辅助重建损失 $\mathcal{L}_{\text{Assist}}$ 使用法向真值作为输入，可以看作是一项简单的监督任务，指导重渲染网络快速达到最优性能。实验结果证明了提出的交替训练策略和辅助重建损失 $\mathcal{L}_{\text{Assist}}$ 的有效性。

此外为了确定式 (6-11) 中联合损失函数 $\mathcal{L}$ 中的最佳权重 λ，图 6-5评估了从 0 到 1 的不同 λ 值对重渲染-光度立体三重监督模型的影响，其左侧 Y 轴代表重建表面法向的 MAE，而右侧 Y 轴代重表渲染图像的 REL。如图 6-5所示，当 λ = 0.1 时性能最好，重渲染-光度立体三重监督模型达到了最小的 MAE、REL(O) 和 REL(C)。值得注意的是，当权重 λ = 0 时，模型仅使用结构生成网络进行训练，而没有图像重建产生的额外监督（等价于表 6-4中的 ID (R4)）。在此时，没有重建的重渲染光度立体图像，因此缺少 λ = 0 时的 REL(O) 和 REL(C) 指标。此结果进一步反映了提出重渲染网络的有效性，它可以为表面法向重建提供额外的监督。有趣的是，当 λ 大于 0.2 后，重渲染的图像误差也在逐渐增大，而并非随着图像重建损失和图像变化损失增强而变得更准确。这可能是因为结构生成网

络和重渲染网络，作为一个级联在一起的整体进行端到端的训练，过高的权重 λ 会首先导致表面法向重建误差的增大，进而影响重渲染网络的性能。

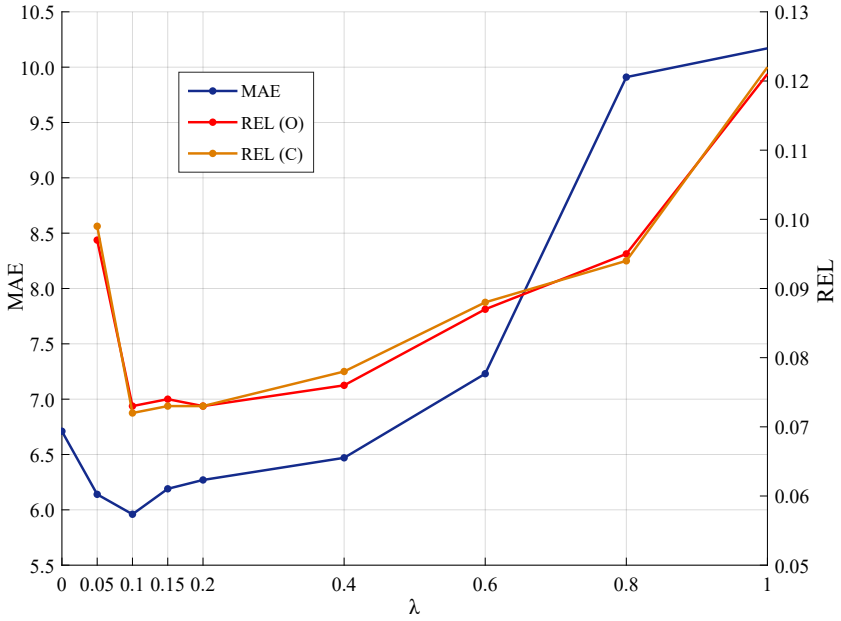


图 6-5 不同权重 λ 训练重渲染-光度立体三重监督模型的结果

6.6.3 DiLiGenT 数据集对比结果

本节在 DiLiGenT 数据集[31] 上将重渲染-光度立体三重监督模型与其他最先进的方法进行了对比。表 6-5 和 6-6分别在以 96 张和 10 张光度立体图像为输入的情况下将本章模型与标定光度立体的传统方法（以作者的姓氏的第一个字母 + 年份命名，LS 则代表最小二乘的基准方法[9]）和深度学习方法进行了比较（以网络简称命名）。粗体的值代表最佳性能，而下划线的值代表次佳性能。

对于基于深度学习的方法，如表 6-5所示，在 96 张密集的输入图像下，本章提出的模型取得了最佳的表面法向重建结果。然而稀疏输入图像在实际应用中更为常见。因此如表 6-6所示，本节在 10 张稀疏的输入图像下，测试了不同的方法，其中 SPLINE-Net[27] 和 LMPS[76] 是专门为稀疏输入图像设计的。粗体的值代表最佳性能，而下划线的值代表次佳性能。此外在 DiLiGenT 数据集[31] 中，本节测试了不同数量的输入图像下的性能，以验证提出的重渲染-光度立体三重监督模型的鲁棒性，如图 6-6 所示。我们还分别在图 6-7和图 6-8中展示了重建表面法向和重渲染光度立体图像的可视化结果。

如表 6-5、表 6-6 和图 6-6所示，本章提出的重渲染-光度立体三重监督模型在不同输入数量的光度立体图像下都实现了平均最小的表面法向重建误差。对

表 6-5 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均以 96 幅图像为输入进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
LS[9]	4.10	8.39	14.92	8.41	25.60	18.50	30.62	8.89	14.65	19.80	15.39
IW12[47]	2.54	7.32	11.11	7.21	25.70	16.25	29.26	7.74	14.09	16.17	13.74
ST14[58]	1.74	6.12	10.60	6.12	13.93	10.09	25.44	6.51	8.78	13.63	10.30
SPLINE-Net[27]	4.51	5.28	10.36	6.49	7.44	9.62	17.93	8.29	10.89	15.50	9.63
DPSN[24]	2.02	6.31	12.68	6.54	8.01	11.28	16.86	7.05	7.86	15.51	9.41
IRPS[26]	**1.47**	5.79	10.36	5.44	6.32	11.47	22.59	6.09	7.76	11.03	8.83
CNN-PS*[75]	2.23	8.29	8.53	5.75	9.74	8.66	17.75	5.91	8.16	11.61	8.66
LMPS[76]	2.40	5.23	9.89	6.11	7.98	8.61	16.18	6.54	7.48	13.68	8.41
PS-FCN[25]	2.82	7.55	7.91	6.16	7.33	8.60	15.85	7.13	7.25	13.33	8.39
第 5章的方法	2.27	5.46	7.84	5.42	7.01	8.49	15.40	7.08	7.21	12.74	7.90
GPS-Net[77]	2.92	**5.07**	7.77	5.42	**6.14**	9.00	15.14	6.04	7.01	13.58	7.81
CNN-PS[75]	2.12	8.30	8.07	**4.38**	7.92	7.42	14.08	**5.37**	**6.38**	12.12	7.62
PS-FCN (Norm.)[69]	2.67	7.72	7.53	4.76	6.72	7.84	**12.39**	6.17	7.15	**10.92**	7.39
提出的方法	2.23	5.29	**7.03**	5.56	6.68	**7.04**	14.04	6.51	6.72	12.25	**7.34**

CNN-PS* 代表 5.6.3节中所述采用相同数据集训练的 CNN-PS 的结果。

表 6-6 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均以 10 幅图像为输入进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
IA14[59]	12.94	16.40	20.63	15.53	18.08	18.73	32.50	**6.28**	14.31	24.99	19.04
LS[9]	5.09	11.59	16.25	9.66	27.90	19.97	33.41	11.32	18.03	19.86	17.31
ST14[58]	5.24	9.39	15.79	9.34	26.08	19.71	30.85	9.76	15.57	20.08	16.18
IW12[47]	**3.33**	7.62	13.36	8.13	25.01	18.01	29.37	8.73	14.60	16.63	14.48
CNN-PS[75]	9.11	14.08	14.58	11.71	14.04	15.48	19.56	13.23	14.65	16.99	14.34
PS-FCN[25]	4.02	7.18	9.79	8.80	10.51	11.58	18.70	10.14	9.85	15.03	10.51
SPLINE-Net[27]	4.96	5.99	10.07	7.52	**8.80**	10.43	19.05	8.77	11.79	16.13	10.35
PS-FCN(Norm.)[69]	4.38	**5.92**	8.98	**6.30**	14.66	10.96	18.04	7.05	11.91	**13.23**	10.14
LMPS[76]	3.97	8.73	11.36	6.69	10.19	10.46	17.33	7.30	9.74	14.37	10.02
5中的方法	3.83	7.52	9.55	7.92	9.83	10.38	17.12	9.36	9.16	14.75	9.94
GPS-Net[77]	4.33	6.34	8.87	6.81	9.34	10.79	**16.92**	7.50	**8.38**	15.00	9.43
提出的方法	4.33	6.05	**8.12**	6.87	9.38	**10.33**	17.51	7.71	9.52	13.43	**9.32**

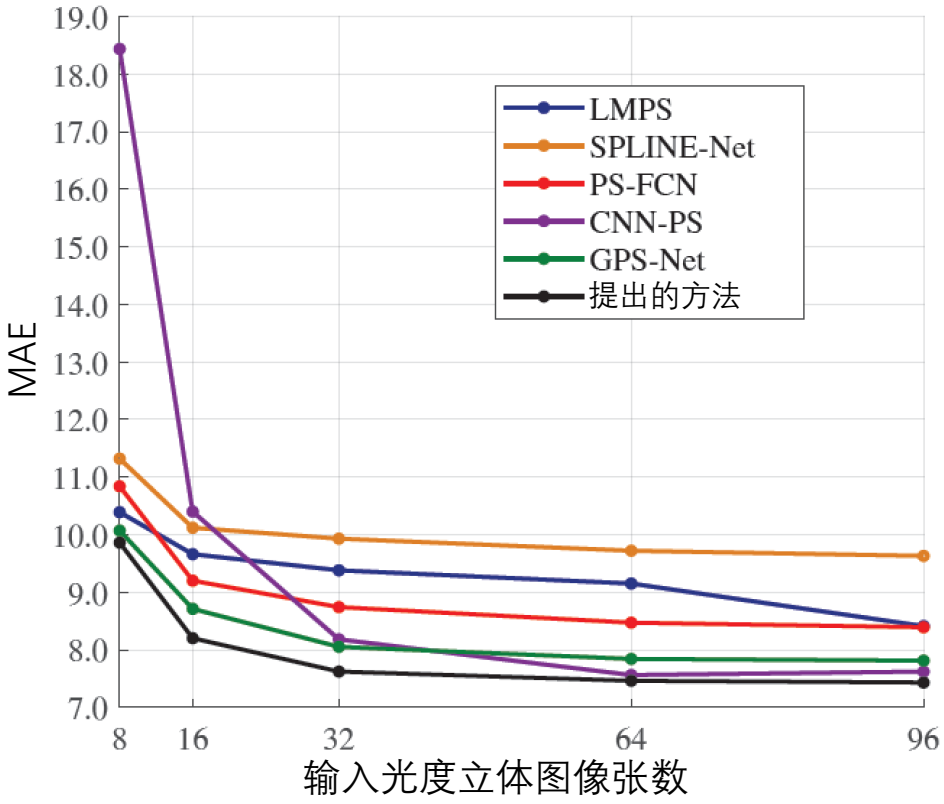


图 6-6 不同数量的输入光度立体图像的比较结果

于复杂的物体，例如 Buddha、Harvest 和 Pot2，以及那些包含阴影和相互反射的强非朗伯物体，例如 Goblet，如图 6-7所示，所提出的模型实现了最佳或次优的性能。可以看出，本章模型在那些有投射阴影的区域（红色框）取得了更好的效果，例如对象 Harvest 的麻袋，以及有褶皱的区域（黄色框），例如对象 Pot2 的花朵浮雕。这些结果说明了本章提出的模型的有效性。

此外图 6-8展示了 DiLiGenT 数据集[31] 中的重渲染光度立体图像结果，其前两行显示了具有不同光照方向下的重渲染光度立体图像。后三行则展示了不同表面材质渲染的光度立体图像。重渲染图像下方的英文代表 MERL BRDFs 数据集[116] 中表面材质的名称。图 6-8首先展示物体 Buddha 和 Cat 在不同光照方向下的重渲染图像，但保持相同的材质。图 6-8展示了用不同的表面材质在物体 Harvest 和 Reading 上测试。可以看出，在结构复杂的物体上，用金属材质渲染时，高光和阴影可以很明显的展现，并且渲染的图像不受原始光度立体图像表面上空间变化的表面材质的影响。尽管在 DiLiGent 数据集[31] 中没有使用 MERL BRDF[116] 渲染的真值图像，但提出模型的重渲染图像可以显示出物体的细节。这证明了重渲染网络实现的优异性能。

图 6-7 DiLiGenT 数据集[31] 上 96 张输入图像下的定量结果

图 6-8 DiLiGenT 数据集中物体的重渲染结果

6.6.4 Light Stage Data Gallery 数据集实验结果

本节在更复杂的 Light Stage Data Gallery 数据集[119] 上使用一般非朗伯表面材质进一步评估了本章提出的模型。由于缺乏表面法向真值，图 6-9 中展示了提出模型的定性结果，包括重建的表面法向和重渲染图像。由于缺乏真值法向，这里使用[128] 进一步展示了重建的表面法线的三维重构结果，以清楚地显示物体的重建细节。重渲染-光度立体三重监督模型使用 32 张光度立体图像进行训练，并使用从 Light Stage Data Gallery 数据集中所有 253 个图像中随机选择的 64 张输入图像进行测试。请注意，物体 Helmet、Plant 和 Fighting 的光度立体图像首先被下采样到空间分辨率的一半，因为原始分辨率 (1024 × 1024) 太大。

如图 6-9所示，重建的表面法向及其三维重建可以准确地显示物体的形状，例如物体 Helmet 的螺丝，物体 Standing 的裙子。此外可以看出，不同表面材料的重渲染图像显示出合理的高光和阴影。重渲染的图像上的细节仍然清晰，例如物体 Plant 的叶子，以及物体 Helmet 的铆钉。在图 6-9种可以发现，使用一些类似金属材料的渲染图像非常暗（另可见图 6-8 中物体 Buddha 的重渲染图像）。这

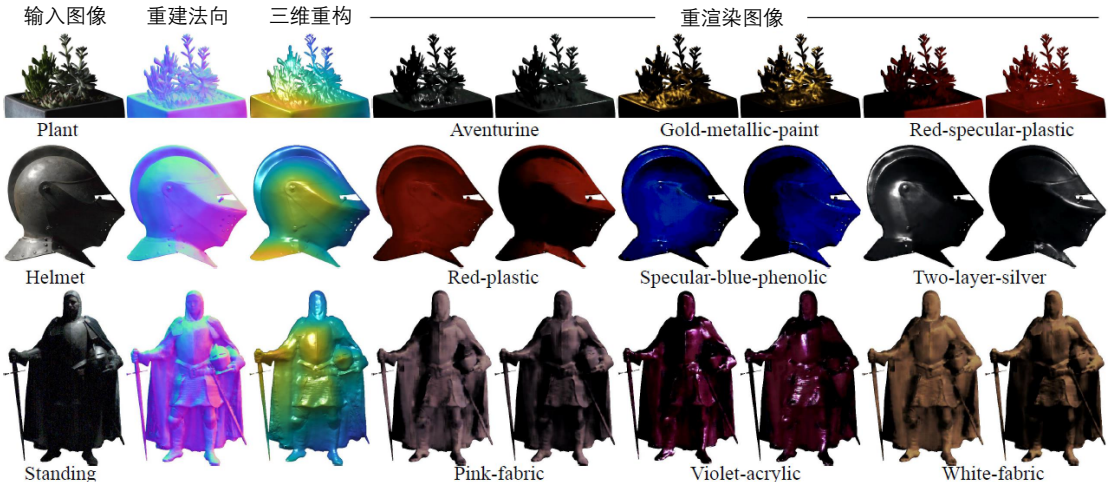


图 6-9 采用 64 张输入图像对 Light Stage Data Gallery 数据集的评估

是因为模型的重渲染网络从表面法向、编码表面材质和光照方向直接生成重渲染图像，而不受全局噪声的影响。相比之下，真实照片不能完全避免全局噪声，例如自然光照和来自其他对象（例如背景）的相互反射，这可能会影响表面法向的计算。

6.6.5 合成数据集实验结果

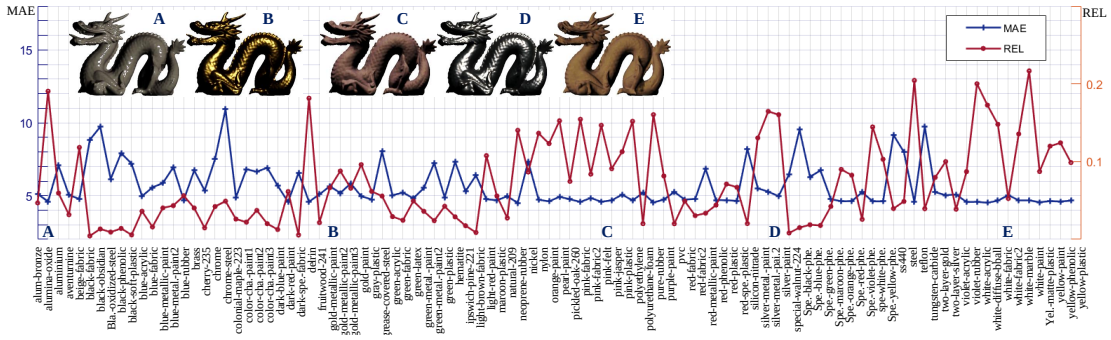


图 6-10 MERL BRDFs 数据集[116] 中 100 种表面材质的物体 Dragon 上重建的表面法向和重渲染图像的 MAE 和 REL

由于 6.6.3和 6.6.4节中的测试数据集都为真实拍摄数据集，缺乏各种材质下的光度立体图像。本节进一步在实验中使用合成测试数据[69] 的物体 Dragon，其使用 MERL BRDFs 数据集[116] 渲染了 100 种表面材质，每种材质在上半球都有 100 个随机的光照方向下的图像。图 6-10 显示了来自 MERL BRDFs 数据集[116] 的 100 种表面材质下重建的表面法向和 Dragon 的渲染图像的性能，其左侧 Y 坐标表示重建表面法向的 MAE，而右侧 Y 坐标表示重渲染图像的 REL，一些渲染

示例也显示在左上角。在图 6-11 和图 6-12中，分别可视化了重建的表面法向和重渲染光度立体图像，输入图像上方的英文代表表面材质的名称。

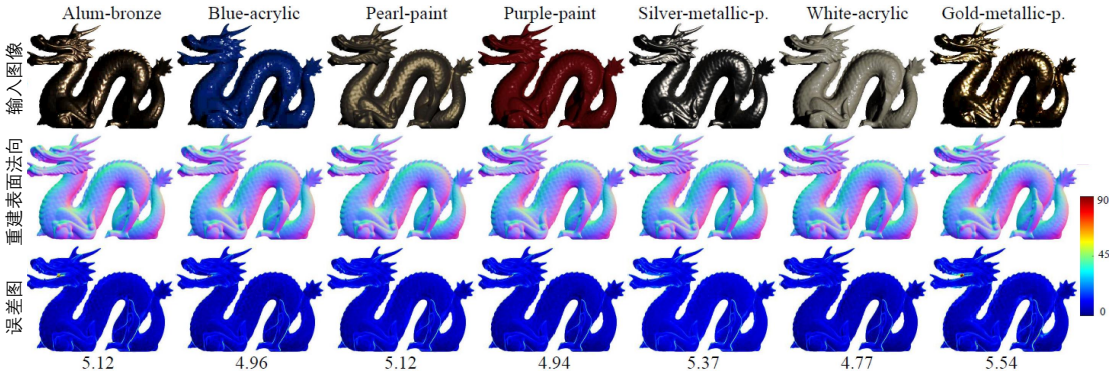


图 6-11 物体 Dragon 的重建表面法向结果

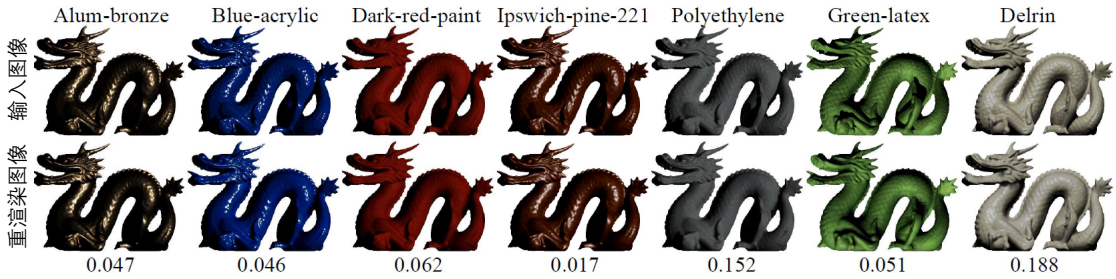


图 6-12 物体 Dragon 的重渲染结果

图 6-10 表明模型在包括金属材料和漫反射材料在内的各种表面材质上始终取得了可靠的性能。这是因为重渲染网络可以进一步为表面法向重建提供额外的监督。当遇到非朗伯曲面时，这种有效性将变得更加明显。有趣的是，重渲染图像的 REL，如图 6-12所示，表明具有强非朗伯材质的表面，例如 Alum-bronze 和 Blue-acrylic，可能比那些漫反射的表面材质，例如 Polyethylene and Delrin 有更好的重建精度。

6.6.6 作为数据扩充方法的验证实验

在光度立体等三维重建任务中，基于学习的方法往往由于缺少训练的数据而难以获得令人满意的性能。因此本章提出的重渲染-光度立体三重监督模型的另一重要作用就是作为光度立体数据集的扩充方法。因为提出的重渲染网络可以在的表面法向上渲染出尽量多的任意材质、任意光照下的光度立体图像。

为了验证重渲染的光度立体图像的精确度是否能够作为光度立体数据集的扩充，本节将重渲染的光度立体图像作为我们模型的输入，再次对其进行二次表

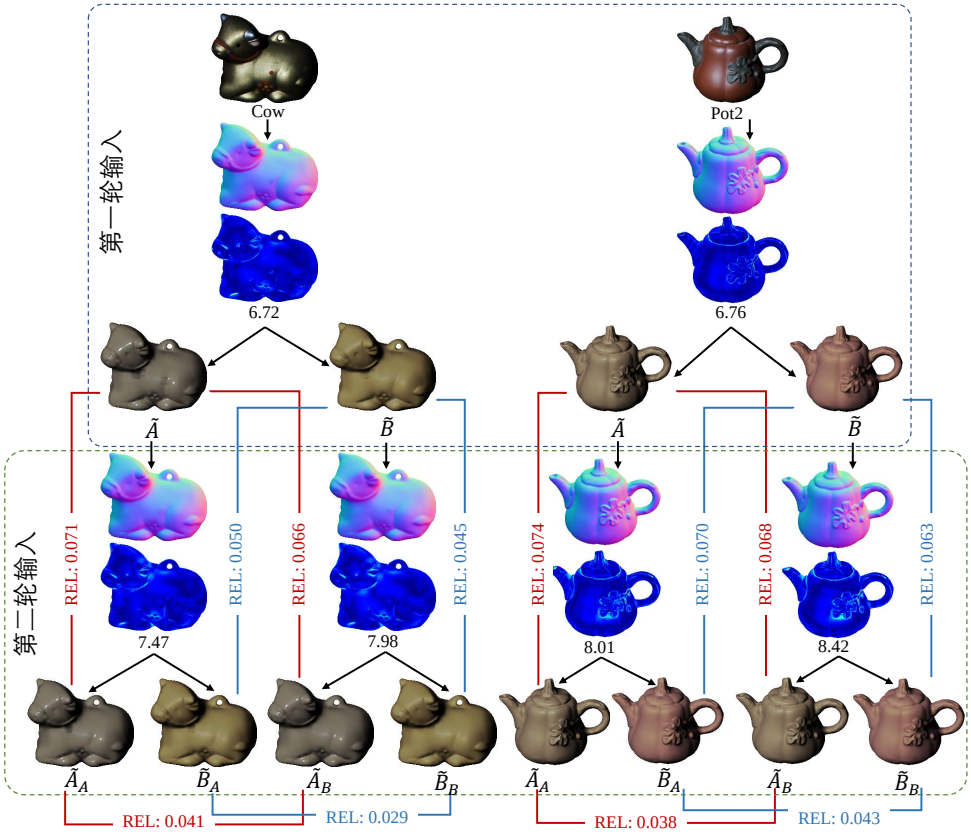


图 6-13 多次重渲染的光度立体图像和表面法向重建的结果

面法向的重建和二次重渲染。结果如图 6-13所示，在 DiLiGenT 数据集中的物体 Cow 和 Pot2 上，将第一次重渲染-光度立体三重监督模型重渲染的材质 A、B 下的光度立体图像 $\tilde{A}$ 和 $\tilde{B}$（各 96 张），再次使用提出的模型获得重建的表面法向和重渲染图像 $\tilde{A}_A$、$\tilde{B}_A$、$\tilde{A}_B$ 和 $\tilde{B}_B$。也就是说，共有以下生成的路径：原始图像 $to\ \tilde{A} \rightarrow \tilde{A}_A$、原始图像 $\rightarrow \tilde{A} \rightarrow \tilde{B}_A$、原始图像 $\rightarrow \tilde{B} \rightarrow \tilde{A}_B$、原始图像 $\rightarrow \tilde{B} \rightarrow \tilde{B}_B$。在图 6-13中，对于物体 Cow，材质 A 设置为 Alumina-oxide，材质 B 设置为 White-diffuse-bball，而对于物体 Pot2，材质 A 设置为 Neoprene-rubber，材质 B 设置为 Pink-felt，在第二轮输入中，下标代表使用的第一轮重渲染图像的材质。

图 6-13的实验结果表明，与使用原始输入图像重建的表面法向相比，通过输入第二轮渲染图像 $\tilde{A}$、$\tilde{B}$ 重建的表面法向仅显示出细微的差异。此外，具有相同材质的渲染图像之间的 REL（来自不同的路径，即 $\tilde{A}$ 和 $\tilde{A}_A$、$\tilde{A}_B$，$\tilde{B}$ 和 $\tilde{B}_A$、$\tilde{B}_B$）也表明不同路径的重渲染图像间的差异微小。这些实验证明了本章提出的重渲染-光度立体三重监督模型的有效性和鲁棒性。因此，模型也可以用作光度立体的数据增强方法，以解决真实拍摄物体的表面法线真值很难获得的问题。

6.7 本章小结

本章提出了重渲染-光度立体三重监督模型，它是一个级联的框架，分别使用结构生成网络和重渲染网络来学习物体的表面法向和重渲染的光度立体图像。重渲染网络为表面法向重建任务提供了额外的监督，形成了一个闭环结构，因而可以提高表面法向重建任务的性能。此外模型可以渲染具有不同表面材质的光度立体图像，可用于光度立体数据增强。消融研究表明了附加的重渲染网络以及提出的网络架构的有效性。在最广泛使用的 DiLiGenT 数据集上进行的大量实验表明，本章提出的模型优于其他的校准光度立体方法。合成测试数据和真实拍摄数据集的实验也表明，本章提出的模型可以生成逼真的重渲染光度立体图像。

7 融合物理先验的光度立体模型

7.1 研究背景

在先前的章节中，本文分别介绍了高频增强的光度立体模型和多重监督的光度立体模型，这些模型依靠注意力权重损失和多重损失来进提升表面法向的重建精度。然而对于基于深度学习的光度立体任务而言，现有的方法都遵循着学习从图像域到法向域的映射这一框架，即仅使用光度立体图像和对应的光照方向作为模型的输入，将重建的表面法向作为模型的输出。由于非朗伯表面镜面反射和投射阴影区域产生的非线性反射特性，先前这种从图像域到法向域的跨域映射学习很难取得十分准确的结果。首先，这些区域在训练集中只占很少的比例，因此其模式很难被充分的学习。其次，镜面反射区域和投射阴影等区域会存在过曝和过暗的像素观察值，这导致网络很难从单张输入的光度立体图像中学习到有用的特征。因此，如何能利用光度立体图像自身的物理先验信息来辅助跨域的映射，以提升重建效果，是本章亟需解决的一个问题。

为此，本章提出了一种融合物理先验的光度立体深度模型，利用物理模型下最小二乘法得到的初始法向（即朗伯假设下的光度立体方法[9]）作为先验信息重新参数化网络权重，利用深度神经网络强大的拟合能力来纠正这些由一般反射特性引起的误差。在融合物理先验法向的框架之上，本章又提出了局部亲和力特征模块，以更好的重建高精度的表面法向。局部亲和力特征模块通过显式揭示相邻特征的关系来学习丰富的结构表示，提升表面法向重建精度。大量实验验证了提出的融合物理先验的光度立体模型在具有挑战性的数据集上有着准确的表面法向重建结果。

7.2 模型概述

本章提出了一种融合物理先验的光度立体模型，以更好地重建非朗伯材质的表面法向，如图 7-1所示。模型利用输入的光度立体图像修正朗伯假设下最小二乘法得到的初始法向[9] 以获得准确的表面法向，也就是说，本章提出的模型在相同的表面法线空间 $\{\mathbb{Y} \rightarrow \tilde{\mathbb{Y}}\}$ 中学习映射，而之前的深度学习的方法则学习从 RGB 的光度立体图像空间到表面法线空间 $\{\mathbb{X} \rightarrow \tilde{\mathbb{Y}}\}$ 的映射。利用物理模型下的最小二乘法得到的初始法向[9] 和法向真值，在漫反射区域理论上是相同的，因此本章提出的模型相当于扩大了非朗伯材质的误差在总误差中的比例，模型

会更倾向于优化这部分非朗伯材质导致的初始法向的错误。融合物理先验的光度立体模型学习了差分的特征，放大了非朗伯材质导致的初始表面法向的误差，减少了可学习的参数空间并改善了表面法向的重建结果。

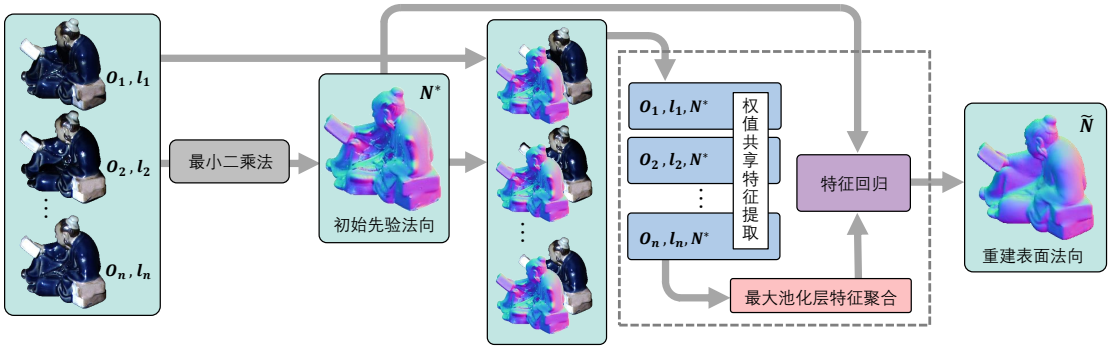


图 7-1 本章提出的融合物理先验的光度立体模型的整体结构

为了进一步提升表面法向重建的精确度，在融合物理先验法向的框架之上，本章提出了局部亲和力特征模块。局部亲和力模块弥补了传统卷积层的特征倾向于捕获外观信息，且不能明确地揭示相邻像素的特征关系的缺陷。这丰富了提取特征的结构信息。

在本章中，首先介绍提出的局部亲和力特征模块，随后介绍提出的融合物理先验法向的光度立体模型的网络结构。

7.3 局部亲和力特征模块

本章提出了一个局部亲和力特征模块（LAF），它通过计算邻域像素的特征的弦相似度以衡量结构关系。通过这种方式，该模块可以通过揭示相邻像素的亲和力来学习丰富的结构特征。卷积层提取的特征倾向于捕获外观信息，但不能明确地揭示相邻变化关系。然而由于光照方向的变化，输入的光度立体图像的颜色外观通常是不同的。为了弥补结构信息的不足，本节采用局部亲和特征模块来丰富原始卷积特征，其具体的操作是计算每一个点与周边邻域点的相似性关系，这类方法在很多的先前工作中有所探讨[132,133,134]。在这里，我们采用与文献[134]相同的衡量方法，即采用余弦相似度计算邻域关系。余弦相似度可以灵活度量邻域的结构关系，因为它可以计算任意数量的特征通道，并为每个相邻点输出一个固定维数的标量。这种特性对卷积神经网络的结构十分有益，因为其反向传播的优化过程需要固定的特征通道数。如图 7-2所示，红色圆圈代表图像中一个像素点，而两点之间的连线代表其余弦相似度，其值被填入中心点位置。局部亲和力

特征模块计算每个点与其 8 个邻居（3 × 3 窗口）之间的余弦相似度，这扩展了额外的 8 个特征通道。这样，每个附加通道都代表了中心点与其中一个相邻点之间的关系。局部亲和力特征模块可以表述为：

$$f_{LAF}^{p} = sim(f(x,y), f(x+p_x, y+p_y)), \tag{7-1}$$

其中 f_{LAF}^{p} 表示局部亲和力特征模块在 (x,y) 位置上附加的 p 个通道，$f(x,y)$ 是在 (x,y) 位置上的输入特征值，(p_x,p_y) 表示第 p 个相邻像素点到位置 (x,y) 的偏移量，因此 $f(x+p_x,y+p_y)$ 即是第 p 个相邻相邻像素上的输入特征值。$sim(\boldsymbol{a},\boldsymbol{b})$ 为余弦相似度函数，计算公式为：

$$sim(\boldsymbol{a},\boldsymbol{b}) = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|\cdot\|\boldsymbol{b}\|}. \tag{7-2}$$

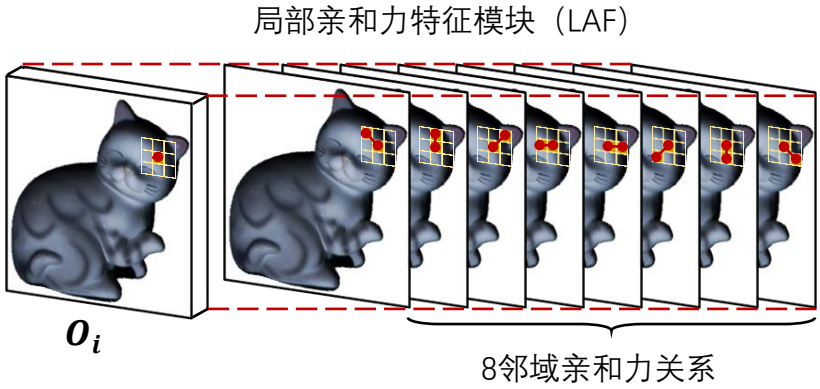


图 7-2 局部亲和力特征模块的操作示意图

在本章提出的融合物理先验的光度立体模型中，模型将输入的初始表面法向和光度立体图像分别进行局部亲和力特征模提取操作，并进一步将初始法向的八通道亲和力特征和输入图像的八通道亲和力关系特征做点乘操作，以更好地学习输入的光度立体图像中存在变化的表面材质的特性。如图 7-3所示，在（a）、（b）中，亲和力特征有不同的响应模式。当拍摄的物体存在空间变化的表面材质时，该区域的输入图像中必然会提取高响应的局部亲和力特征。而此时物理先验的初始表面法向是平坦的（表面法向不受表面材质变化的影响），所以其提取的局部亲和力特征的值是低响应的。而在存在复杂结构的表面，如褶皱和边缘，初始表面法向和输入图像提取的局部亲和力特征都会高响应。因此，点乘后的特征可以提供明显的判别信息，提高融合物理先验的光度立体模型的性能。

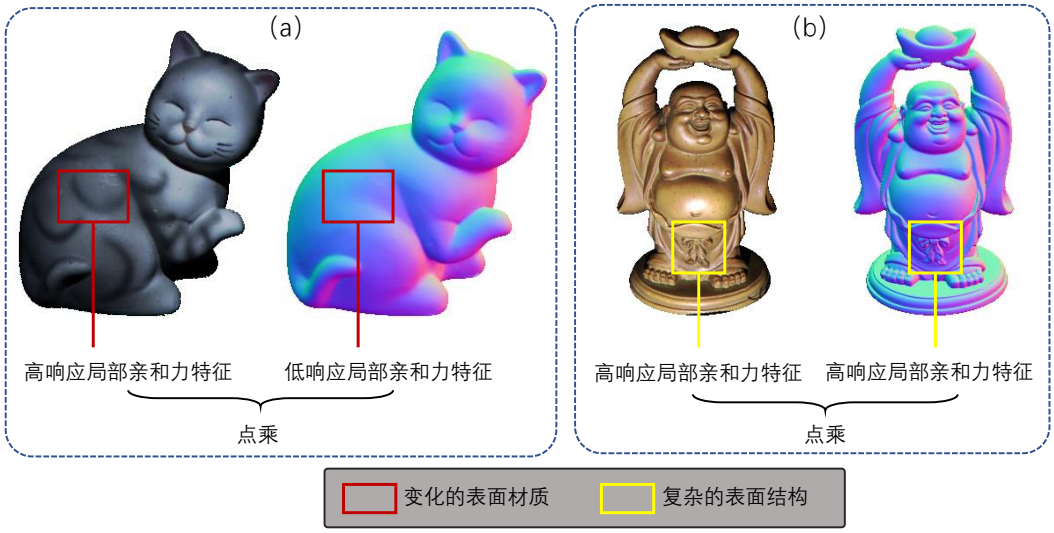


图 7-3 （a）变化的表面材质的例子,，（b）复杂的表面结构的例子

7.4 融合物理先验的网络结构

若物体上一点有真值法向 $\boldsymbol{n} \in \mathbb{R}^3$，当其被 j 个（$j \in \{1,2,\cdots,n\}$）光照方向 $\boldsymbol{l_j} \in \mathbb{R}^3$ 照射时，该点的像素强度为 $\boldsymbol{o}_j \in \mathbb{R}^3$。假设拍摄的像素强度 $\boldsymbol{o}_j$ 可以被划分为两个部分[49]：漫反射强度 $\boldsymbol{o}_j^d$ 和其他强度 $\boldsymbol{o}_j^o$，即 $\boldsymbol{o}_j = \boldsymbol{o}_j^d + \boldsymbol{o}_j^o$。其中漫反射强度 $\boldsymbol{o}_j^d$ 代表物体表面材质中理想的朗伯分项反射的强度，其他强度 $\boldsymbol{o}_j^o$ 则代表一些非朗伯分项反射的强度，例如镜面反射、投射阴影以及物体间的相互反射等噪声。

假如可以准确的在非朗伯材质的物体上，从观测的像素强度 $\boldsymbol{o}_j$ 中提取漫反射分项 $\boldsymbol{o}_j^d$，那么根据理想的朗伯假设光度立体方法[9]，可以通过最小二乘求解，表面法向 $\boldsymbol{n}$，如下式：

$$\boldsymbol{n} = \frac{\boldsymbol{L}^{-1}\boldsymbol{O}^d}{|\boldsymbol{L}^{-1}\boldsymbol{O}^d|}, \tag{7-3}$$

其中 $\boldsymbol{O}^d = [\boldsymbol{o}_1^d, \boldsymbol{o}_2^d, \ldots, \boldsymbol{o}_n^d]'$，$\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_j]'$。

然而无论是异常值剔除的策略(2.4.1)还是建模复杂反射模型的策略(2.4.2)，都没有算法能十分准确地从像素总强度中提取漫反射分项 $\boldsymbol{O}^d$。因此我们不再使用复杂的模型或深度学习来进一步减少拟合误差，而是重新考虑非朗伯条件下的朗伯假设，即利用物理模型下的最小二乘法[9]先求解一个初始表面法向 $\boldsymbol{n}^*$：

$$\boldsymbol{n}^* = \frac{\boldsymbol{L}^{-1}\boldsymbol{O}}{|\boldsymbol{L}^{-1}\boldsymbol{O}|}, \tag{7-4}$$

其中 $\boldsymbol{O} = [\boldsymbol{o}_1^d + \boldsymbol{o}_1^o, \boldsymbol{o}_2^d + \boldsymbol{o}_2^o, \ldots, \boldsymbol{o}_n^d + \boldsymbol{o}_n^o]'$。可以看出，初始表面法向 $\boldsymbol{n}^*$ 的误差是

由于由非朗伯和噪声的分项 $\boldsymbol{O}^o = [\boldsymbol{o}_1^o, \boldsymbol{o}_2^o, \dots, \boldsymbol{o}_n^o]'$ 造成的。因此可以通过初始表面法向 $\boldsymbol{n}^*$ 在像素总强度 $\boldsymbol{O} = \boldsymbol{O}^d + \boldsymbol{O}^o$ 的条件下学习域内映射：

$$\frac{\boldsymbol{L}^{-1}\boldsymbol{O}^d}{|\boldsymbol{L}^{-1}\boldsymbol{O}^d|} = Mapping\,(\frac{\boldsymbol{L}^{-1}\boldsymbol{O}}{|\boldsymbol{L}^{-1}\boldsymbol{O}|}, \boldsymbol{O}^d + \boldsymbol{O}^o). \tag{7-5}$$

不同于先前单独从输入图像中映射表面法向的深度学习方法，本章提出的模型使用输入图像 $\boldsymbol{O}^d + \boldsymbol{O}^o$ 作为指导初始表面法向 $\boldsymbol{n}^*$ 校正的条件。对于基于深度学习的光度立体方法，本章认为上述映射是重建表面法向的更好选择。首先，相比从图像域到法向域的跨域映射，法向域的域内映射可以减少求解可能的学习函数的空间。其次，基于最大池化层聚合的深度学习方法本质上是单独提取每一张输入的光度立体图像，因此割裂了不同光照下图像间的关系。而单一光度立体图像中，出现的镜面反射、投射阴影等过曝和过暗像素导致其很难被单独提取出有用的特征，因此通过物理先验得到的初始法向 $\boldsymbol{n}^*$ 可以在一定程度上预先提供图像间的关系从而辅助特征提取的过程。最后，初始法向 $\boldsymbol{n}^*$ 在朗伯材质区域求解的部分是正确的，其由于非朗伯表面材质和噪声引起的误差可以用 $\frac{\boldsymbol{L}^{-1}\boldsymbol{O}^d}{|\boldsymbol{L}^{-1}\boldsymbol{O}^d|}$ - $\frac{\boldsymbol{L}^{-1}\boldsymbol{O}}{|\boldsymbol{L}^{-1}\boldsymbol{O}|}$ 来表示。因此，本章模型相当于放大了非朗伯误差在总误差中的比例，并学习了差分特征，更关注学习表面法向的误差而不是重建整个表面法向，从而实现的高精度的表面法向重建。

融合物理先验的网络结构可以分为三个部分，包括特征回归阶段 f_{Ext}、最大池化层特征聚合阶段和特征回归阶段 f_{Reg}。

如图 7-4所示，特征提取阶段 f_{Ext} 可以看作是 n 路分支的共享权重特征提取网络，可以表示为：

$$\boldsymbol{\Psi}_j = f_{Ext}(\boldsymbol{O}_j, \boldsymbol{l}_j, \boldsymbol{N}^*; \theta_{Ext}),\ j \in \{1, 2, \cdots, n\}, \tag{7-6}$$

其中 θ_{Ext} 表示特征提取阶段 f_{Ext} 中的可学习参数，$\boldsymbol{\Psi}_j \in \mathbb{R}^{256\times\frac{1}{4}H\times\frac{1}{4}W}$。

如 7-4所示（操作下方的红色数字代表特征或者图像的通道数），首先对输入的光度立体图像 $\boldsymbol{O}_j$ 和物理先验的初始表面法向 $\boldsymbol{N}^*$（整幅分辨率为 $H \times W$ 图像的初始表面法向）进行局部亲和力特征提取，并将 $\boldsymbol{O}_j$ 和 $\boldsymbol{N}^*$ 中提取的局部亲和力特征分别记作 $\boldsymbol{\Omega}_j$ 和 $\boldsymbol{\Gamma} \in \mathbb{R}^{8\times H\times W}$。模型首先计算了 $\boldsymbol{\Omega}_j$ 和 $\boldsymbol{\Gamma}$ 点乘结果，记为 $\boldsymbol{\Upsilon}_j$ 并融合进后续的网络结构中，这为评判物体时复杂表面结构还是变化的表面材质提供了信息（见 7.3中的讨论）。

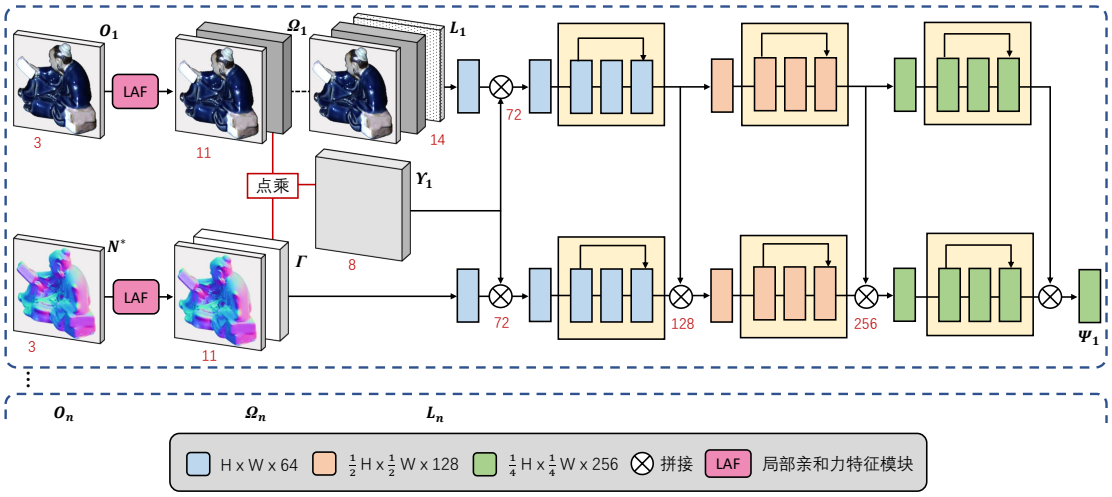


图 7-4 特征提取阶段的具体网络结构

对于标定的光度立体任务，参照传统的操作，将光照方向 $\boldsymbol{l}_j$ 沿 H 和 W 的方向复制，扩展至与图像具有相同分辨率大小的张量 $\boldsymbol{L}_j \in \mathbb{R}^{3\times H\times W}$，以更好的与图像融合。注意这里融合后的输入图像特征的第一维度通道数为 14，其中前 3 维为输入光度立体图像 $\boldsymbol{O}_j$ 的 RGB 通道，4 至 11 维为局部亲和力特征 $\boldsymbol{\Omega}_j$ 的 8 邻域余弦相似性特征通道，12 至 14 维为对应光照方向的 xyz 分量。对于融合后的初始法向特征，其前 3 维为初始表面法向 $\boldsymbol{N}^*$ 的 xyz 方向，4 至 11 维为局部亲和力特征 $\boldsymbol{\Gamma}$ 的 8 邻域余弦相似性特征。

随后模型将融合后的光度立体图像特征和初始表面法向特征分别输入两个网络中，如图 7-4所示，每一个网络的主干结构都包括三个残差模块[123]，不同的是，在每一次的残差模块后，模型将来自光度立体图像的特征拼接到初始法向提取的特征中，以实现特征融合，并最终得到提取的融合特征 $\boldsymbol{\Psi}_j \in \mathbb{R}^{256\times\frac{1}{4}H\times\frac{1}{4}W}$。在网络中，所有的卷积层都采用 3×3 大小的卷积核，激活函数选取 Leaky ReLu，在第二和第三个残差模块前的降采样时其步长为 2，其余都为 1。

为了更好地处理任意数量下的光度立体图像和初始表面法向中提取的任意数量特征，模型应用最大池化层特征聚合和平均池化层特征聚合拼接融合的方式来处理任意数量的输入特征，以得到固定通道数的聚合特征：

$$\boldsymbol{\Psi}_{\max} = \bigcup_{i}^{\frac{1}{4}H\times\frac{1}{4}W} \max(\boldsymbol{\Psi}_{1,i}, \boldsymbol{\Psi}_{2,i}, \ldots, \boldsymbol{\Psi}_{n,i}), \tag{7-7}$$

$$\boldsymbol{\Psi}_{\text{avg}} = \bigcup_{i}^{\frac{1}{4}H\times\frac{1}{4}W} \text{avg}(\boldsymbol{\Psi}_{1,i}, \boldsymbol{\Psi}_{2,i}, \ldots, \boldsymbol{\Psi}_{n,i}), \tag{7-8}$$

其中 $\boldsymbol{\Psi}_{\text{max}}$ 和 $\boldsymbol{\Psi}_{\text{avg}}$ 分别表示最大池化层聚合后的特征和平均池化层聚合后的特征，下标 i 表示特征分辨率 $\frac{1}{4}H \times \frac{1}{4}W$ 中位置的索引。尽管单独使用平均池化层特征聚合，会导致未激活的特征和显著特征被平滑[25]，但是，本章模型的特征提取阶段不但从光度立体图像 $\boldsymbol{O}_j$ 中提取特征，还从朗伯先验的初始法向 $\boldsymbol{N}^*$ 中提取特征。因此，提取的特征 $\boldsymbol{\Psi}_j$ 不仅仅包含不同光照方向下的图像分解特征[69]，还包含全局的初始法向特征，需要采用平均池化层聚合以保留全局的信息。消融实验（7.5.2）证明，采用最大池化层 + 平均池化层聚合的特征将会提升表面法向重建的性能。

在得到聚合的特征 $\boldsymbol{\Psi}_{\text{max}}$ 之后，融合物理先验的光度立体模型利用特征回归阶段 f_{Reg} 重建物体的表面法向，如下式所示：

$$\tilde{\boldsymbol{N}} = f_{Reg}(\boldsymbol{\Psi}_{\text{fused}}, \boldsymbol{N}^*; \theta_{Reg}), \tag{7-9}$$

其中 θ_{Reg} 表示特征回归阶段 f_{Reg} 中的可学习参数，$\boldsymbol{\Psi}_{\text{fused}} \in \mathbb{R}^{512\times\frac{1}{4}H\times\frac{1}{4}W}$ 为聚合的特征 $\boldsymbol{\Psi}_{\text{max}}$ 和 $\boldsymbol{\Psi}_{\text{avg}}$ 在第一维度上拼接后的融合特征。为了实现更好的重建效果，模型在特征回归阶段中继续融合了初始的表面法向 $\boldsymbol{N}^*$。特征回归阶段 f_{Reg} 的具体网络结构如表 7-1所示。模型将物理先验的初始法向 $\boldsymbol{N}^* \in \mathbb{R}^{3\times H\times W}$ 做下采样处理至 $\boldsymbol{N}^*_h \in \mathbb{R}^{3\times\frac{1}{2}H\times\frac{1}{2}W}$，并两次拼接至特征回归网络。此外在 f_{Reg} 中，特征的分辨率被上采样 3 次，下采样 1 次，这种设计可以扩大感受野并使 $\boldsymbol{N}^*_h$ 初始法向更好地与特征融合。除最后一层的卷积层激活函数设置为 Tanh 外，其余层的激活函数均为 Leaky ReLu。

本章提出的融合物理先验的光度立体模型中可学习的参数 θ_{Ext}、θ_{Reg} 的优化过程由重建的表面法向和法向真值的角度误差的最小化来实现，即采用余弦损失函数 $\mathcal{L}_{\text{Cosine}}$：

$$\mathcal{L}_{\text{Cosine}} = \frac{1}{HW}\sum_{i}^{HW}\left(1 - \boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i\right), \tag{7-10}$$

其中 $\cdot$ 操作代表点乘。当像素位置 i 上重建的表面法向 $\tilde{\boldsymbol{N}}_i$ 与真值法向 $\boldsymbol{N}_i$ 越相似，其点乘 $\boldsymbol{N}_i \cdot \tilde{\boldsymbol{N}}_i$ 将越接近 1，此时式 (7-10) 的值将越接近 0。

表 7-1 特征回归阶段 f_{Reg} 的网络结构

输入	操作	卷积核大小	步长	输出
$512 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 1	4×4	2	$256 \times \frac{1}{2}H \times \frac{1}{2}W$
$256 \times \frac{1}{2}H \times \frac{1}{2}W$	拼接（转置卷积层 1、降采样初始法向 $\boldsymbol{N}_h^*$ ）			$259 \times \frac{1}{2}H \times \frac{1}{2}W$
$259 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 1	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 2	3×3	2	$256 \times \frac{1}{4}H \times \frac{1}{4}W$
$256 \times \frac{1}{4}H \times \frac{1}{4}W$	转置卷积层 2	4×4	2	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	拼接（转置卷积层 2、降采样初始法向 $\boldsymbol{N}_h^*$ ）			$131 \times \frac{1}{2}H \times \frac{1}{2}W$
$131 \times \frac{1}{2}H \times \frac{1}{2}W$	卷积层 3	3×3	1	$128 \times \frac{1}{2}H \times \frac{1}{2}W$
$128 \times \frac{1}{2}H \times \frac{1}{2}W$	转置卷积层 3	4×4	2	$64 \times H \times W$
$64 \times H \times W$	卷积层 4	3×3	1	$3 \times H \times W$
$3 \times H \times W$	L2 归一化层			$3 \times H \times W$

7.5 实验结果

为了验证本章提出的融合物理先验的光度立体模型的定量性能，我们采用了平均角度误差（MAE）来评估表面法向重建的精度，还采用 $< err_{15^\circ}$ 和 $< err_{30^\circ}$ 来测量法向角度误差小于 15° 和 30° 的像素所占物体表面总像素的比例。本节首先对提出的模型进行了详细的消融实验和分析，随后在 DiLiGenT 数据集[31] 上评价了提出的融合物理先验的光度立体模型的性能。

7.5.1 实验设置

本章提出的融合物理先验的光度立体模型是使用 PyTorch 实现的，并且利用 Adam 优化器的默认设置（β_1= 0.9 和 β_2=0.999）。初始学习率设置为 0.002，并且每过 5 个 epoch 学习率除以 2。在 batchsize = 24 的情况下，模型在单张 RTX 3080Ti 上训练了 40 个 epoch 以达到最佳效果。用于训练的合成数据集与 PS-FCN[25] 使用的相同，其包括两个形状数据集，分别是 Blobby 数据集[114] 和 SculptureBlobby 数据集[115]。在训练时，先采用最小二乘法[9] 在每一个样本中利用随机的 32 张光度立体图像计算物理先验下的初始表面法向 $\boldsymbol{N}^*$，并记录选取的 32 张图像和 $\boldsymbol{N}^*$ 一起输入提出的模型。训练中将输入图像的分辨率 $H \times W$ 设置为 32×32。

7.5.2 消融实验与分析

融合物理先验的光度立体模型的消融实验在 Blobby 数据集[114] 和 Sculpture 数据集[115] 的验证集中的全部 852 样本（每个样本采用 32 张不同光照的输入图像）上实施。如表 7-2所示，数字代表验证集所有样本的平均值，对于 MAE，其值越低越好。对于 $< err_{15^\circ}$ 和 $< err_{30^\circ}$，其值越高越好。ID (0) 代表提出的模型的默认结构；ID (1) 表示取消模型中的局部亲和力特征模块，此时网络的特征回归阶段 f_{Ext} 将直接对输入图像 $\boldsymbol{O}_j$ 和对应光照方向 $\boldsymbol{L}_j$、初始法向 $\boldsymbol{N}^*$ 两部分提取特征；ID (2) 表示取消对局部亲和力特征模块提取的 $\boldsymbol{\Omega}_j$ 和 $\boldsymbol{\Gamma}$ 进行点乘获得特征 $\boldsymbol{\Upsilon}_j$，此时拼接 $\boldsymbol{\Upsilon}_j$ 的操作也被取消；ID (3) 表示采用平坦的网络结构代替 f_{Ext} 中三个残差模块[123]；ID (4) 代表在特征回归阶段 f_{Reg} 中取消两次融合下采样的初始法向 $\boldsymbol{N}_h^*$；ID (5) 代表物理先验采用异常值剔除法中的 WG10[46] 方法而不是最小二乘法[9] 处理输入的图像，以获得初始法向 $\boldsymbol{N}^*$；ID (6) 则代表不采用初始法向 $\boldsymbol{N}^*$ 的网络，此时特征提取阶段仅对输入图像 $\boldsymbol{O}_j$ 和对应光照方向 $\boldsymbol{L}_j$ 提取特征（含局部亲和力特征模块），特征回归阶段也取消初始法向 $\boldsymbol{N}_h^*$ 的拼接融合；ID (7) 和 ID (8) 则分别表示在特征提取阶段后对 n 个特征仅提取最大池化层特征聚合 $\boldsymbol{\Psi}_{\text{max}}$ 和平均池化层特征聚合 $\boldsymbol{\Psi}_{\text{avg}}$。为了进一步评估融合物理先验的光度立体模型的性能和泛化性，图 7-6还展示了用 MERL BRDFs 数据集[116] 渲染了 100 种表面材质的合成物体 Dragon 的实验结果。在图 7-6中，比较了提出的方法的性能（黑线，对应表 7-2中 ID (0)），不融和初始法向（红线，对应 ID (6)）、不使用残差模块（紫线，对应 ID (3)）和最小二乘法[9] 求解的初始法向（蓝线）。

如表 7-2所示，ID (0)、ID (1) 和 ID (2) 的实验表明提出的局部亲和力特征模块以及特征 $\boldsymbol{\Upsilon}_j$ 融合的有效性。这是因为输入的光度立体图像的颜色外观经常随光照方向变化而变化，因此通过邻域余弦相似度可以丰富结构特征的表达。而 ID (2) 进一步分析了融合特征 $\boldsymbol{\Upsilon}_j$ 的作用。$\boldsymbol{\Upsilon}_j$ 由光度立体图像的局部亲和力特征和初始法向的局部亲和力特征点乘得来，以判断光度立体图像和初始法向的局部结构是否相似，这可以为变化的表面材质情况提供判别。因此融合 $\boldsymbol{\Upsilon}_j$ 特征进一步提升了模型的性能。

ID (0) 和 ID (3) 的实验显示了残差模块[123] 和普通的无残差模块网络在特征提取阶段的性能。参考表 7-2，实验表明应用残差模块具有较低的法向角度误差。另外，如图 7-5 所示，可以看出本章模型（黑线）的误差在大多数表面材质上都

表 7-2 对 32 张输入光度立体图像下的验证集进行消融分析的结果

编号	消融方法	MAE ↓	$< err_{15^\circ}$ ↑	$< err_{30^\circ}$ ↑
ID (0)	提出的方法	11.51	**84.68%**	94.89%
ID (1)	无局部亲和力特征模块	11.87	83.50%	94.84
ID (2)	无 $\boldsymbol{\Upsilon}_j$	11.82	84.02%	94.86%
ID (3)	无残差模块[123]	11.79	84.16%	94.86%
ID (4)	f_{Reg} 无 $\boldsymbol{N}_h^*$ 融合	11.92	83.95%	94.85%
ID (5)	初始法向采用 WG10[46]	**11.46**	84.66%	**94.90%**
ID (6)	无初始法向 $\boldsymbol{N}^*$	12.28	82.31%	94.79%
ID (7)	无平均池化层特征聚合特征 $\boldsymbol{\Psi}_{\text{avg}}$	11.60	84.52%	94.87%
ID (8)	无最大池化层特征聚合特征 $\boldsymbol{\Psi}_{\text{max}}$	12.08	83.16%	94.85%

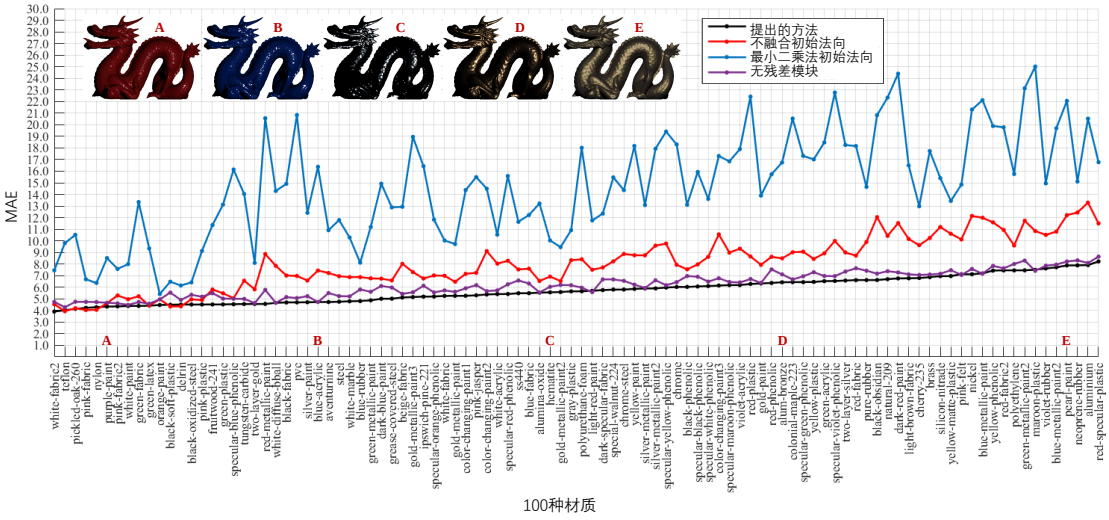


图 7-5 使用来自 MERL BRDFs 数据集[116] 的 100 种表面材质渲染的物体 Dragon 的表面法向重建的结果

比平坦网络（紫线）的结构小。结果表明，残差模块提高重建表面法向的准确度。原因是因为残差块可以有效地避免深度网络中的梯度消失。ID (4) 的实验则检测了在特征回归阶段多次融合初始法向信息的作用。这表明在深层特征层中融合初始法向将丰富细节并提高准确性。

ID (0) 和 ID (5) 的实验显示了融合不同物理先验下的初始法向对模型的影响。在 ID (5) 中，模型使用秩最小化方法（WG10）[46] 的结果作为先验。如表 7-2所示，我们发现秩最小化的初始法向在 MAE 和 $< err_{30^\circ}$ 两个指标上略好于朗伯假设下最小二乘法[9] 求解的初始法向。这是因为秩最小化方法[46] 比最小二乘

法[9] 具有更好的性能。但是我们仍然选择使用最小二乘法求解的初始法向，原因有以下两点：首先，秩最小化方法需要大量时间来检测和去除异常值，使用秩最小化的融合物理先验光度立体模型需要 65 个小时才能完成 40 个 epoch 的训练，而使用最小二乘的融合物理先验光度立体模型仅需 19 个小时即可完成训练；其次，诸如秩最小化之类的异常值提出法通常仅对有限类别的表面材质有效，且需要较多的光度立体图像才能实现较高的重建精度。因此，在重建的表面法向精度差别很小的情况下，本章模型依然使用最小二乘法求解的初始法向进行融合。

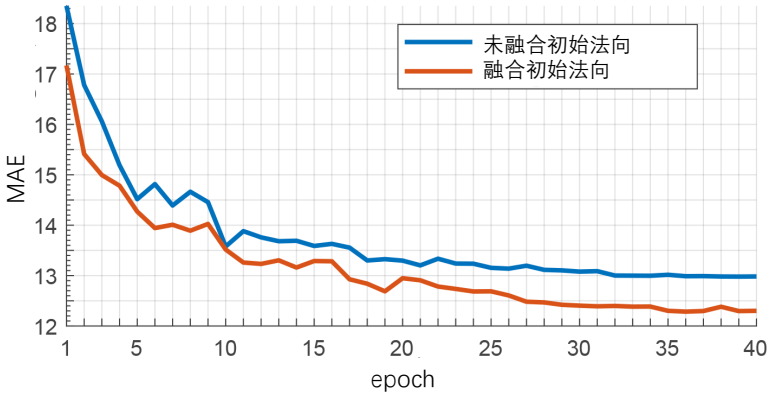


图 7-6 验证集上的收敛情况的比较

ID (0) 和 ID (6) 的实验的显示了融合初始法向对模型性能的提升。可以看出，融合初始法向的方法在所有指标上始终比传统的跨域映射策略表现更好。这是因为提出的模型在相同的法向空间中学习映射。因此，模型可以更快地收敛并实现准确的估计。另外，如图 7-5 所示，可以看出，在大多数材料上，提出的模型（黑线）明显优于没有融合初始法向（红线）和融合的最小二乘法[9] 求解的初始法向。请注意，融合物理先验的光度立体模型在最小二乘法的初始法向中有较大误差的材质上有鲁棒的表现，而未融合初始法向的方法却随着表面材质的变化有较大变化。此外，在图 7-6中我们展示了融合与不融合初始法向的模型训练时在验证集上的收敛情况，其蓝色曲线表示未融合初始法向的网络，而橙色曲线表示融合物理先验初始法向的网络。这两个网络都使用相同的参数和 32 张图像作为输入进行训练。如图 7-6所示，融合初始法向的模型在测试集中实现了更低的收敛误差并且有更快的收敛速度，这表明本章模型比以前的仅从输入图像学习表面法向的模型更有效。

ID (0)、ID (7) 和 ID (8) 的实验则比较了特征聚合方式对表面法向重建的精

度影响，可以看出添加平均池化层特征聚合可以提升重建的性能，任意单独使用一种特征聚合的效果都不如融合两种方法的性能，特别是在 ID (8) 仅使用平均池化层聚合特征时，重建精度有较大的下降。

7.5.3 DiLiGenT 数据集对比实验结果

我们在 DiLiGenT 数据集[31] 上报告了结果，表 7-3展示了 96 张输入图像下的结果，表 7-4展示了 10 张输入图像下的结果。本节将提出的模型与传统方法（以作者的姓氏的第一个字母 + 年份命名，LS 则代表最小二乘的基准方法[9]）和深度学习方法（以网络简称命名）进行了比较。粗体的值代表最佳性能，而下划线的值代表次佳性能。此外，表 7-3和表 7-4还展示了提出的融合物理先验的光度立体模型分别使用 10、32 和 64 张输入光度立体图像下进行训练的结果。

表 7-3 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均以 96 幅图像为输入进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
LS[9]	4.10	8.39	14.92	8.41	25.60	18.50	30.62	8.89	14.65	19.80	15.39
IW12[47]	2.54	7.32	11.11	7.21	25.70	16.25	29.26	7.74	14.09	16.17	13.74
IA14[59]	3.34	7.11	10.47	6.74	13.05	9.71	25.95	6.64	8.77	14.19	10.60
ST14[58]	1.74	6.12	10.60	6.12	13.93	10.09	25.44	6.51	8.78	13.63	10.30
SPLINE-Net[27]	4.51	5.28	10.36	6.49	7.44	9.62	17.93	8.29	10.89	15.50	9.63
DPSN[24]	2.02	6.31	12.68	6.54	8.01	11.28	16.86	7.05	7.86	15.51	9.41
IRPS[26]	**1.47**	5.79	10.36	5.44	6.32	11.47	22.59	6.09	7.76	11.03	8.83
LMPS[76]	2.40	5.23	9.89	6.11	7.98	8.61	16.18	6.54	7.48	13.68	8.41
PS-FCN[25]	2.82	7.55	7.91	6.16	7.33	8.60	15.85	7.13	7.25	13.33	8.39
GPS-Net[77]	2.92	**5.07**	7.77	5.42	**6.14**	9.00	15.14	6.04	7.01	13.58	7.81
CNN-PS[75]	2.12	8.30	8.07	**4.38**	7.92	**7.42**	14.08	**5.37**	**6.38**	12.12	7.62
PS-FCN (Norm.)[69]	2.67	7.72	**7.53**	4.76	6.72	7.84	**12.39**	6.17	7.15	**10.92**	7.39
提出的方法（32 张训练）	2.10	5.19	7.54	5.93	6.60	7.75	13.16	6.41	6.93	11.02	**7.26**
提出的方法（10 张训练）	2.44	5.71	8.04	6.79	7.03	8.45	15.41	7.24	7.12	11.56	7.98
提出的方法（64 张训练）	2.09	5.04	7.39	5.74	6.55	7.72	13.06	6.09	6.82	10.85	7.14

如表 7-3 和表 7-4所示，本章提出的融合物理先验的光度立体模型在 96 和 10 张输入光度立体图像的情况下（默认设置下使用 32 张图像训练）在 MAE 指标上都优于其他最先进的方法。图 7-7进一步可视化了重建结果。可以看出模型缩小了具有强非朗伯和非凸表面的误差。红色框是具有镜面反射的区域，白色

表 7-4 DiLiGenT 数据集[31] 上不同方法的比较，所有方法均以 10 幅图像为输入进行评估

方法	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	平均值
IA14[59]	12.94	16.40	20.63	15.53	18.08	18.73	32.50	**6.28**	14.31	24.99	19.04
LS[9]	5.09	11.59	16.25	9.66	27.90	19.97	33.41	11.32	18.03	19.86	17.31
ST14[58]	5.24	9.39	15.79	9.34	26.08	19.71	30.85	9.76	15.57	20.08	16.18
IW12[47]	**3.33**	7.62	13.36	8.13	25.01	18.01	29.37	8.73	14.60	16.63	14.48
CNN-PS[75]	9.11	14.08	14.58	11.71	14.04	15.48	19.56	13.23	14.65	16.99	14.34
PS-FCN[25]	4.02	7.18	9.79	8.80	10.51	11.58	18.70	10.14	9.85	15.03	10.51
SPLINE-Net[27]	4.96	5.99	10.07	7.52	**8.80**	**10.43**	19.05	8.77	11.79	16.13	10.35
PS-FCN(Norm.)[69]	4.38	**5.92**	8.98	**6.30**	14.66	10.96	18.04	7.05	11.91	**13.23**	10.14
LMPS[76]	3.97	8.73	11.36	6.69	10.19	10.46	17.33	7.30	9.74	14.37	10.02
GPS-Net[77]	4.33	6.34	8.87	6.81	9.34	10.79	**16.92**	7.50	**8.38**	15.00	9.43
提出的方法（32 张训练）	3.98	6.86	**8.33**	6.32	8.90	10.45	17.05	7.44	8.58	13.94	**9.19**
提出的方法（10 张训练）	3.95	6.72	8.01	6.24	8.63	9.91	16.80	7.39	8.54	13.41	8.99
提出的方法（64 张训练）	4.16	7.71	9.54	7.55	9.90	11.07	18.05	8.38	9.09	14.80	10.03

框是具有投射阴影的区域，黄色框代表具有非凸面（皱纹）的区域。与其他方法相比，我们的方法在这些区域产生了更准确的估计。可以观察到，模型在具有高光、阴影和褶皱的区域重建了更准确的表面法向，例如 Buddha 的衣领、Reading 的褶皱衣服和 Harvest 的口袋。

请注意，CNN-PS[75] 在物体 Bear 上的结果是使用所有 96 个输入图像计算的，其 MAE 为 8.30，而在 CNN-PS 原文中，作者丢弃了 Bear 的 96 张光度立体图像中的前 20 张，仅用后 76 张进行表面法向的重建，其所述丢弃物体 Bear 的前 20 张图像的原因是该物体胃的区域光度值是错误的。在仅使用后 76 张光度立体图像进行表面法向的重建时，CNN-PS 可以达到 4.20 度，然而其他所有方法报告的结果都采用了 Bear 的全部 96 张图像。为了公平比较，实验展示了基于相同数量的测试图像（所有 96 张图像）的结果。如图 7-8所示，当融合物理先验的光度立体模型在使用全部图像进行训练时精度仅有微小的下降，而 CNN-PS 的精度却有非常显著的下降，这表明本章模型对噪声非常的鲁棒。

此外，本节还探讨了不同数量的输入光度立体图像对训练的影响。如表 7-3和表 7-4中展示了使用 10 和 64 张图像训练的融合物理先验的光度立体模型的结果（64 表示训练数据集中图像的最大数量）。可以看出，使用 64 张图像训练

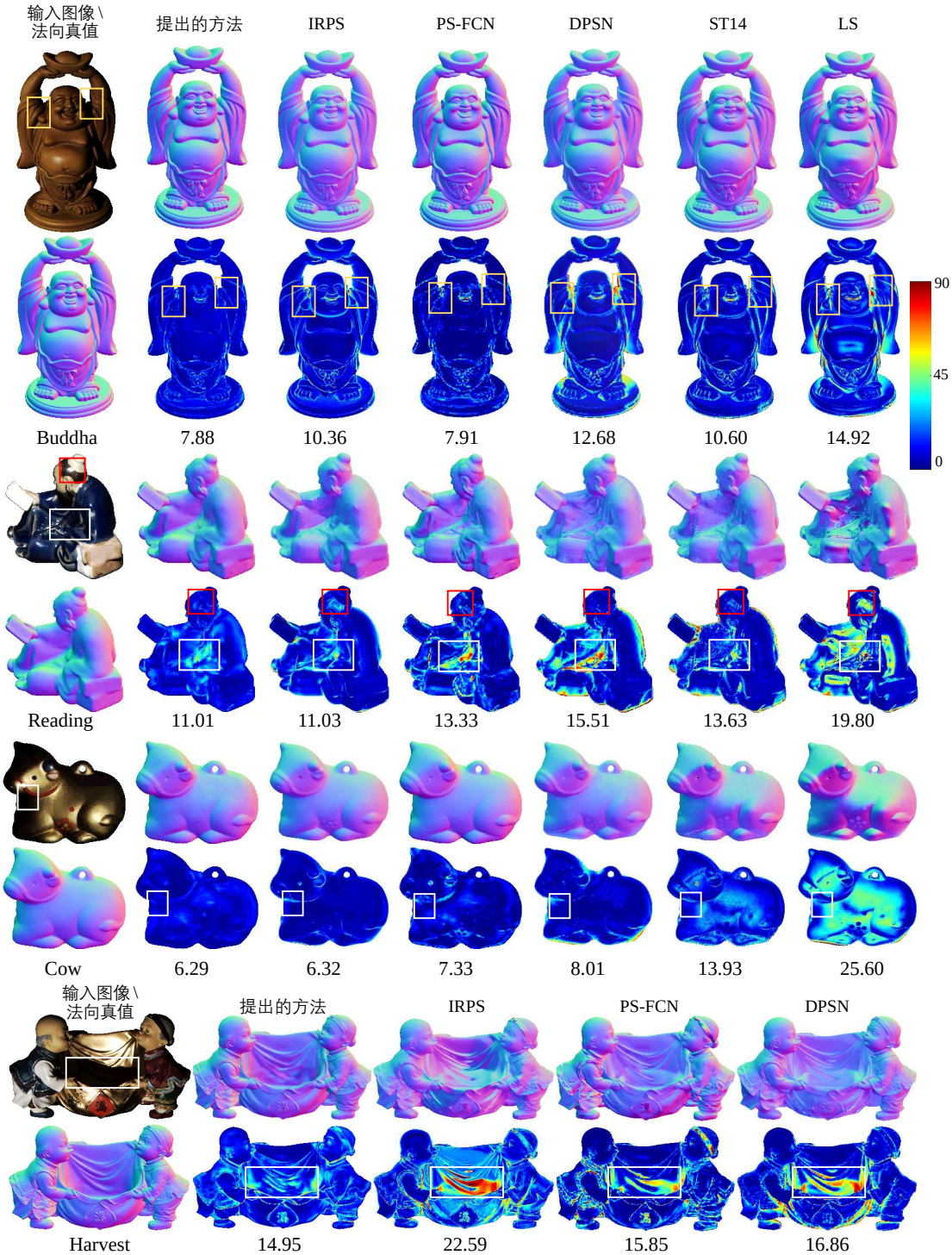


图 7-7 来自 DiLiGenT 数据集[31] 中的强非朗伯物体的定量结果

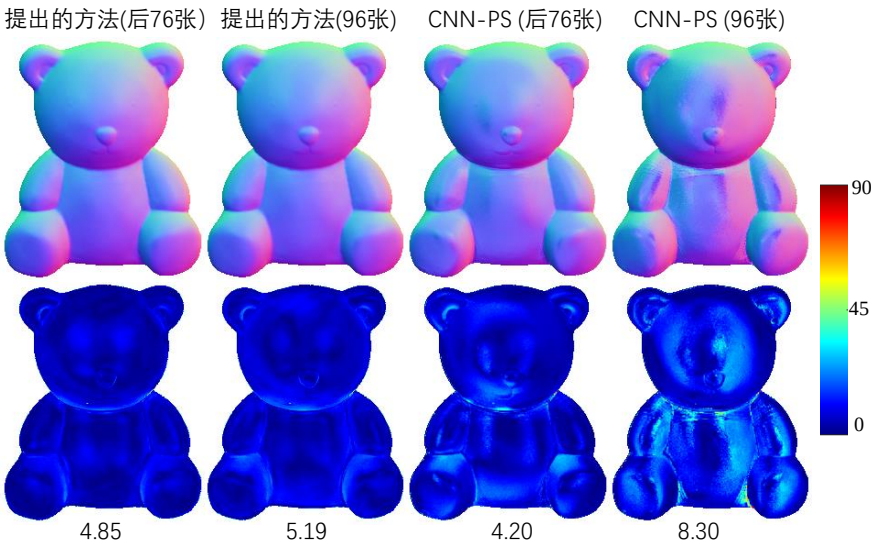


图 7-8 提出的融合朗伯先验的光度立体模型和 CNN-PS[75] 使用全部的 96 张和后 76 张光度立体图像时在物体 Bear 上的结果

的方法在使用 96 张图像下测试时比使用 32 张图像训练的方法取得了更好的性能。相比之下，使用 10 张图像训练的方法在使用 10 张图像下测试时也优于默认的 32 张图像训练的模型。换句话说，训练和测试之间相似的输入图像数量将有利于表面法向的重建，原因是物理模型下的初始法向也与输入图像得数量有关，输入图像数量不同导致的初始法向差异会在一定程度上影响模型的学习。因此为了获得最佳性能，本章建议在训练和测试期间使用相似数量的输入图像。尽管如此，本章模型的默认设置（用 32 张图像训练）在 96 张和 10 张光度立体图像的测量中仍取得了最先进的性能。

7.6 本章小结

本章提出了一种融合物理先验的光度立体模型，该模型从物理模型中最小二乘法求解的初始法向中重建精确的表面法向。与之前从 RGB 图像空间求解法向空间的深度学习方法相比，本章提出的模型在相同的法向空间中进行映射，并且更加关注初始法向中的错误，即学习放大的差分特征。消融实验表明提出的模型有更准确的重建精确度。此外，模型的收敛速度比仅使光度立体图像学习表面法向的传统深度学习方法更快。对真实数据集（DiLiGenT）和合成数据集（Dragon）的广泛定量比较表明，本章提出的模型优于最先进的方法。可视化结果也表明融合物理先验的光度立体模型可以更好地重建强非朗伯表面材质的表面法向。

此外，本章提出的融合物理先验的光度立体模型还可以支持更广泛的应用。例如，本章模型可扩展用于非理想光照环境的光度立体（照明不是平行光或具有额外自然光光照等）。在这些任务中，初始法向可以在理想光照假设下被计算，然后通过模型进行细化和矫正，从而重建出准确的物体表面法向。

8 总结与展望

8.1 工作总结

本文围绕非朗伯光度立体的深度学习模型对 1.2节中提出的四个科学问题展开研究。本文从物体表面高频区域的增强、增加额外约束监督和融合先验信息的角度，提出了一系列有效的基于深度学习的非朗伯光度立体模型，并为光度立体数据集的样本扩充提出了方法。本文的研究工作可以概括为以下几个方面：

（1）在第 3章中，首先针对现有方法在物体表面复杂结构处存在误差大、不清晰的问题，提出一种自适应注意力光度立体模型，利用注意力加权的法向重建损失，为复杂结构区域施加高权重的细节保护损失，避免了此前方法采用的单一欧氏距离损失带来的模糊问题。实验结果表明提出的模型具有良好的性能，在复杂区域能重建更清晰的表面法向。

（2）在第 4章中，针对第 3章中提出的模型难以在变化的表面材质区域中有效重建的问题，提出了归一化的高频区域增强光度立体模型，利用光度立体图像的归一化操作，解决了物体表面复杂结构区域和物体表面材质变化区域的重建误差，实现准确的表面法向重建。所提出的方法在公开数据集上取得了最佳的重建精度。

（3）在第 5章中，针对现有方法采用单一的余弦损失优化模型的问题，提出了重光照-光度立体模型，利用双重回归网络将重建的表面法向渲染回光度立体图像，使模型形成闭环，提供额外图像重建损失的监督。实验结果表明提出的模型具有较好的表面法向重建精度，在镜面反射、投射阴影等区域有着更准确的结果，并且提出的模型可以生成任意指定的重光照光度立体图像。

（4）针对现有的光度立体数据集样本不足的问题，且第 5章提出的模型难以生成任意材质的重建图像，在第 6章中提出了重渲染-光度立体三重监督模型，利用编码的材质信息，通过重渲染网络生成任意表面材质、光照下的重渲染光度立体图像，并为模型提供了余弦损失、图像重建损失和图像变化损失三种监督。实验结果表明提出的模型可以同时重建高精度的物体表面法向和重渲染光度立体图像，实现对少量样本的光度立体数据集的扩充。

（5）在第 7章中，针对现有方法仅从光度立体图像中跨域学习物体表面法向这一问题，提出了融合物理先验的光度立体模型，将物理模型下的初始法向与光

度立体图像融合，以学习差分特征，使其转变为学习法向空间的域内映射任务。此外，在该框架之上，本文提出了局部亲和力特征模块。实验结果表明提出的模型实现了高精度的表面法向重建，更好地处理具有强非朗伯表面材质的物体重建。

8.2 未来展望

本文针对非朗伯光度立体的深度学习模型进行了多方面的研究，取得了一些成果。但是，在研究过程中，也发现了一些需要进一步提高和深入的问题，在这里举例：

（1）在提出的自适应光度立体模型和归一化的高频区域增强光度立体模型中，当拍摄的物体具有特别简单的表面结构时，注意力生成网络可能会将唯一的镜面反射区域误判为高频的信息，导致这些区域余弦约束不足，造成一定误差。解决镜面反射误激活注意力图的问题可能需要在所有输入的光度立体图像中判断高频的信息是否发生空间的改变，因为物体表面复杂的结构和变化的表面材质激活的注意力权重几乎不随不同图像光照方向改变而改变，而镜面反射则随着光照方向的变化而变化。因此未来的研究工作将探索如何减少镜面反射对注意力图的影响，以进一步提升重建的精度。

（2）在提出的重渲染-光度立体三重监督模型中，采用了 100 维的 one-hot 编码处理表面材质数据集的 100 种材质信息。然而真实世界的物体表面材质远不止 100 种，因此需要更丰富，更真实地材质数据集对模型进行训练。与之衍生的另一个问题既是，如何更全面的编码更多的表面材质的特征信息。在提出的模型中，重渲染的光度立体提图像会在投射阴影和镜面反射区域存在稍大的误差，这正是因为目前 one-hot 的特征编码形式难以涵盖丰富的物体表面材质特性。因此未来的研究工作将探索寻找更合理的方式将表面材质信息融合重渲染网络，以生成更准确、更多材质种类的重渲染光度立体图像。

参考文献

[1] Forsyth D, Ponce J. Computer vision: A modern approach.[M]. Prentice hall, 2011.

[2] 龙霄潇, 程新景, 朱昊, 等. 三维视觉前沿进展[J]. 中国图象图形学报, 2021, 26(6):1389-1428.

[3] Wu S, Rupprecht C, Vedaldi A. Unsupervised learning of probably symmetric deformable 3d objects from images in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020: 1-10.

[4] Agarwal S, Furukawa Y, Snavely N, et al. Building rome in a day[J]. Communications of the ACM, 2011, 54(10):105-112.

[5] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(8):1362-1376.

[6] Yao P, Zhang H, Xue Y, et al. As-global-as-possible stereo matching with adaptive smoothness prior[J]. IET Image Processing, 2019, 13(1):98-107.

[7] 高天寒, 杨子艺. 图像序列的增量式运动结构恢复[J]. 中国图象图形学报, 2019, 24(11): 1952-1961.

[8] Horn B K. Shape from shading; a method for obtaining the shape of a smooth opaque object from one view[D]. Massachusetts Institute of Technology, 1970.

[9] Woodham R J. Photometric method for determining surface orientation from multiple images [J]. Optical Engineering, 1980, 19(1):139-144.

[10] Kontsevich L, Petrov A, Vergelskaya I. Reconstruction of shape from shading in color images [J]. Journal of the Optical Society of America A, 1994, 11(3):1047-1052.

[11] 邓学良, 何扬波, 周建丰. 基于光度立体的三维重建方法综述[J]. 现代计算机, 2021, 27 (23):133-143.

[12] Chen L, Zheng Y, Shi B, et al. A microfacet-based reflectance model for photometric stereo with highly specular surfaces[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 3162-3170.

[13] Shi B, Tan P, Matsushita Y, et al. Elevation angle from reflectance monotonicity: Photometric stereo for general isotropic reflectances[C]//Proceedings of the European Conference on Computer Vision. 2012: 455-468.

[14] Chandraker M, Bai J, Ramamoorthi R. On differential photometric reconstruction for unknown, isotropic brdfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(12):2941-2955.

[15] DePiero F W, Trivedi M M. 3-d computer vision using structured light: Design, calibration,

and implementation issues[M]//Advances in Computers: volume 43. 1996: 243-278.

[16] Salvi J, Pages J, Batlle J. Pattern codification strategies in structured light systems[J]. Pattern Recognition, 2004, 37(4):827-849.

[17] Zhou Z, Wu Z, Tan P. Multi-view photometric stereo with spatially varying isotropic materials [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1482-1489.

[18] Fan H, Qi L, Ju Y, et al. Refractive laser triangulation and photometric stereo in underwater environment[J]. Optical Engineering, 2017, 56(11):113101.

[19] 简振雄, 王晰, 任杰骥, 等. 基于近场光度立体视觉的金属表面纹理重构[J]. 光学学报, 2021, 41(11):109-119.

[20] Wu B, Li Y, Liu W C, et al. Centimeter-resolution topographic modeling and fine-scale analysis of craters and rocks at the chang'e-4 landing site[J]. Earth and Planetary Science Letters, 2021, 553:116666.

[21] 韩海燕, 张静. 基于虚拟现实的三维动态场景重建[J]. 现代电子技术, 2018, 41(2):170-173.

[22] Shen L, Tan P. Photometric stereo and weather estimation using internet images[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2009: 1850-1857.

[23] Sattler M, Sarlette R, Klein R. Efficient and realistic visualization of cloth[C]//Rendering Techniques. 2003: 167-178.

[24] Santo H, Samejima M, Sugano Y, et al. Deep photometric stereo network[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017: 501-509.

[25] Chen G, Han K, Wong K Y K. Ps-fcn: A flexible learning framework for photometric stereo [C]//Proceedings of the European Conference on Computer Vision. 2018: 3-18.

[26] Taniai T, Maehara T. Neural inverse rendering for general reflectance photometric stereo[C]// Proceedings of the International Conference on Machine Learning. 2018: 4857-4866.

[27] Zheng Q, Jia Y, Shi B, et al. Spline-net: Sparse photometric stereo through lighting interpolation and normal estimation networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 8549-8558.

[28] Chen G, Waechter M, Shi B, et al. What is learned in deep uncalibrated photometric stereo? [C]//Proceedings of the European Conference on Computer Vision. 2020: 745-762.

[29] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.

[30] Gondal M W, Schölkopf B, Hirsch M. The unreasonable effectiveness of texture transfer

for single image super-resolution[C]//Proceedings of the European Conference on Computer Vision. 2018: 80-97.

[31] Shi B, Mo Z, Wu Z, et al. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo.[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(2):271-284.

[32] Nicodemus F E. Directional reflectance and emissivity of an opaque surface[J]. Applied optics, 1965, 4(7):767-775.

[33] 章毓晋. 图像工程[M]. 清华大学出版社有限公司, 2005.

[34] Chen G. Single view analysis of non-lambertian objects based on deep learning[D]. The University of Hong Kong, 2020.

[35] Nayar S K, Ikeuchi K, Kanade T. Shape from interreflections[J]. International Journal of Computer Vision, 1991, 6(3):173-195.

[36] Ikeuchi K, Horn B K. An application of the photometric stereo method[C]//Proceedings of the International Joint Conference on Artificial Intelligence-Volume 1. 1979: 413-415.

[37] Smith J, Lin T L, Ranson K, et al. The lambertian assumption and landsat data[J]. Photogrammetric Engineering and Remote Sensing, 1980, 46(9):1183-1189.

[38] Ackermann J, Goesele M, et al. A survey of photometric stereo techniques[J]. Foundations and Trends® in Computer Graphics and Vision, 2015, 9(3-4):149-254.

[39] Herbort S, Wöhler C. An introduction to image-based 3d surface reconstruction and a survey of photometric stereo methods[J]. 3D Research, 2011, 2(3):4.

[40] Zheng Q, Shi B, Pan G. Summary study of data-driven photometric stereo methods[J]. Virtual Reality & Intelligent Hardware, 2020, 2(3):213-221.

[41] Solomon F, Ikeuchi K. Extracting the shape and roughness of specular lobe objects using four light photometric stereo[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(4):449-454.

[42] Chandraker M, Agarwal S, Kriegman D. Shadowcuts: Photometric stereo with shadows[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2007: 1-8.

[43] Wu T P, Tang K L, Tang C K, et al. Dense photometric stereo: A markov random field approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(11): 1830-1846.

[44] Mukaigawa Y, Ishii Y, Shakunaga T. Analysis of photometric factors based on photometric linearization.[J]. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 2007, 24(10):3326-3334.

[45] Verbiest F, Van Gool L. Photometric stereo with coherent outlier handling and confidence

estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8.

[46] Wu L, Ganesh A, Shi B, et al. Robust photometric stereo via low-rank matrix completion and recovery[C]//Proceedings of the Asian Conference on Computer Vision. 2010: 703-717.

[47] Ikehata S, Wipf D, Matsushita Y, et al. Robust photometric stereo using sparse regression[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012: 318-325.

[48] Georghiades A S. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo[C]//Proceedings of the IEEE International Conference on Computer Vision: volume 3. 2003: 816.

[49] Tozza S, Mecca R, Duocastella M, et al. Direct differential photometric stereo shape recovery of diffuse and specular surfaces[J]. Journal of Mathematical Imaging and Vision, 2016, 56 (1):57-76.

[50] Yeung S K, Wu T P, Tang C K, et al. Normal estimation of a transparent object using a video [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(4):890-897.

[51] Chung H S, Jia J. Efficient photometric stereo on glossy surfaces with wide specular lobes [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8.

[52] Goldman D B, Curless B, Hertzmann A, et al. Shape and spatially-varying brdfs from photometric stereo[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32 (6):1060-1071.

[53] Ackermann J, Langguth F, Fuhrmann S, et al. Photometric stereo for outdoor webcams[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012: 262-269.

[54] Hui Z, Sankaranarayanan A C. A dictionary-based approach for estimating shape and spatially-varying reflectance[C]//Proceedings of the IEEE International Conference on Computational Photography. 2015: 1-9.

[55] Alldrin N, Zickler T, Kriegman D. Photometric stereo with non-parametric and spatially-varying reflectance[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8.

[56] Li S, Shi B. Photometric stereo for general isotropic reflectances by spherical linear interpolation[J]. Optical Engineering, 2015, 54(8):083104.

[57] Higo T, Matsushita Y, Ikeuchi K. Consensus photometric stereo[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010: 1157-1164.

[58] Shi B, Tan P, Matsushita Y, et al. Bi-polynomial modeling of low-frequency reflectances[J].

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(6):1078-1091.

[59] Ikehata S, Aizawa K. Photometric stereo using constrained bivariate regression for general isotropic surfaces[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2179-2186.

[60] Iwahori Y, Woodham R J, Tanaka H, et al. Neural network to reconstruct specular surface shape from its three shading images[C]//Proceedings of the International Conference on Neural Networks: volume 2. 1993: 1181-1184.

[61] Cheng W C. Neural-network-based photometric stereo for 3d surface reconstruction[C]// Proceedings of the IEEE International Joint Conference on Neural Network. 2006: 404-410.

[62] Elizondo D, Zhou S M, Chrysostomou C. Surface reconstruction techniques using neural networks to recover noisy 3d scenes[C]//Proceedings of the International Conference on Artificial Neural Networks. 2008: 857-866.

[63] Hertzmann A, Seitz S M. Example-based photometric stereo: Shape reconstruction with general, varying brdfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8):1254-1264.

[64] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.

[65] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[66] Kuznietsov Y, Stuckler J, Leibe B. Semi-supervised deep learning for monocular depth map prediction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6647-6655.

[67] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2):295-307.

[68] Santo H, Samejima M, Sugano Y, et al. Deep photometric stereo networks for determining surface normal and reflectances[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1):114-128.

[69] Chen G, Han K, Shi B, et al. Deep photometric stereo for non-lambertian surfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1):129-142.

[70] Wang X, Jian Z, Ren M. Non-lambertian photometric stereo network based on inverse reflectance model with collocated light[J]. IEEE Transactions on Image Processing, 2020, 29: 6032-6042.

[71] Ju Y, Peng Y, Jian M, et al. Learning conditional photometric stereo with high-resolution features[J]. Computational Visual Media, 2022, 8(1):105-118.

[72] Liu Y, Ju Y, Jian M, et al. A deep-shallow and global–local multi-feature fusion network for photometric stereo[J]. Image and Vision Computing, 2021:104368.

[73] Ju Y, Jian M, Dong J, et al. Learning photometric stereo via manifold-based mapping[C]// Proceedings of the IEEE International Conference on Visual Communications and Image Processing. 2020: 411-414.

[74] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. science, 2000, 290(5500):2319-2323.

[75] Ikehata S. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces[C]// Proceedings of the European Conference on Computer Vision. 2018: 3-18.

[76] Li J, Robles-Kelly A, You S, et al. Learning to minify photometric stereo[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 7568-7576.

[77] Yao Z, Li K, Fu Y, et al. Gps-net: Graph-based photometric stereo network[C]//Proceedings of the Advances in Neural Information Processing Systems. 2020: 33.

[78] Logothetis F, Budvytis I, Mecca R, et al. Px-net: Simple and efficient pixel-wise training of photometric stereo networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2021: 12757-12766.

[79] Weber M, Cipolla R. A practical method for estimation of point light-sources.[C]// Proceedings of the British Machine Vision Conference: volume 2001. 2001: 471-480.

[80] Takai T, Maki A, Niinuma K, et al. Difference sphere: an approach to near light source estimation[J]. Computer Vision and Image Understanding, 2009, 113(9):966-978.

[81] Aoto T, Taketomi T, Sato T, et al. Position estimation of near point light sources using a clear hollow sphere[C]//Proceedings of the International Conference on Pattern Recognition. IEEE, 2012: 3721-3724.

[82] Hayakawa H. Photometric stereo under a light source with arbitrary motion[J]. Journal of the Optical Society of America A, 1994, 11(11):3079-3089.

[83] Del Bue A, Xavier J, Agapito L, et al. Bilinear factorization via augmented lagrange multipliers[C]//Proceedings of the European Conference on Computer Vision. 2010: 283-296.

[84] Miyazaki D, Ikeuchi K. Photometric stereo under unknown light sources using robust svd with missing data[C]//Proceedings of the IEEE International Conference on Image Processing. 2010: 4057-4060.

[85] Belhumeur P N, Kriegman D J, Yuille A L. The bas-relief ambiguity[J]. International journal of computer vision, 1999, 35(1):33-44.

[86] Chandraker M K, Kahl F, Kriegman D J. Reflections on the generalized bas-relief ambiguity[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: volume 1. 2005: 788-795.

[87] Drbohlav O, Šára R. Specularities reduce ambiguity of uncalibrated photometric stereo[C]// Proceedings of the European Conference on Computer Vision. 2002: 46-60.

[88] Shi B, Matsushita Y, Wei Y, et al. Self-calibrating photometric stereo[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010: 1118-1125.

[89] Tan P, Mallick S P, Quan L, et al. Isotropy, reciprocity and the generalized bas-relief ambiguity [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2007: 1-8.

[90] Papadhimitri T, Favaro P. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima[J]. International journal of computer vision, 2014, 107(2):139-154.

[91] Sato I, Okabe T, Yu Q, et al. Shape reconstruction based on similarity in radiance changes under varying illumination[C]//Proceedings of the IEEE International Conference on Computer Vision. 2007: 1-8.

[92] Okabe T, Sato I, Sato Y. Attached shadow coding: Estimating surface normals from shadows under unknown reflectance and lighting conditions[C]//Proceedings of the IEEE International Conference on Computer Vision. 2009: 1693-1700.

[93] Lu F, Matsushita Y, Sato I, et al. Uncalibrated photometric stereo for unknown isotropic reflectances[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1490-1497.

[94] Chen G, Han K, Shi B, et al. Self-calibrating deep photometric stereo networks[C]// Proceedings of the Conference on Computer Vision and Pattern Recognition. 2019: 8739-8747.

[95] 谢利民, 黄心汉, 宋展. 基于不同光照条件的三维重建算法[J]. 华中科技大学学报 (自然科学版), 2013, 41(S1):403-406.

[96] Clark J J. Active photometric stereo.[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: volume 92. 1992: 29-34.

[97] Iwahori Y, Sugie H, Ishii N. Reconstructing shape from shading images under point light source illumination[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: volume 1. 1990: 83-87.

[98] Ahmad J, Sun J, Smith L, et al. An improved photometric stereo through distance estimation and light vector optimization from diffused maxima region[J]. Pattern Recognition Letters, 2014, 50:15-22.

[99] Bony A, Bringier B, Khoudeir M. Tridimensional reconstruction by photometric stereo with near spot light sources[C]//Proceedings of the European Signal Processing Conference. 2013: 1-5.

[100] Nie Y, Song Z. A novel photometric stereo method with nonisotropic point light sources[C]// Proceedings of the International Conference on Pattern Recognition. 2016: 1737-1742.

[101] Mecca R, Wetzler A, Bruckstein A M, et al. Near field photometric stereo with point light sources[J]. SIAM Journal on Imaging Sciences, 2014, 7(4):2732-2770.

[102] Mecca R, Quéau Y. Unifying diffuse and specular reflections for the photometric stereo problem[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. 2016: 1-9.

[103] Blinn J F. Models of light reflection for computer synthesized pictures[J]. ACM Transactions on Computer Graphics, 1977, 11(2):192-198.

[104] Santo H, Waechter M, Matsushita Y. Deep near-light photometric stereo for spatially varying reflectances[C]//Proceedings of the European Conference on Computer Vision. 2020: 137-152.

[105] Logothetis F, Budvytis I, Mecca R, et al. A cnn based approach for the near-field photometric stereo problem[C]//Proceedings of the British Machine Vision Conference. 2020.

[106] Kim H, Wilburn B, Ben-Ezra M. Photometric stereo for dynamic surface orientations[C]// Proceedings of the European Conference on Computer Vision. 2010: 59-72.

[107] Fyffe G, Yu X, Debevec P. Single-shot photometric stereo by spectral multiplexing[C]// Proceedings of the IEEE International Conference on Computational Photography. 2011: 1-6.

[108] Smith M L, Smith L N. Dynamic photometric stereo—a new technique for moving surface analysis[J]. Image and Vision Computing, 2005, 23(9):841-852.

[109] Hernández C, Vogiatzis G, Brostow G J, et al. Non-rigid photometric stereo with colored lights[C]//Proceedings of the IEEE International Conference on Computer Vision. 2007: 1-8.

[110] Anderson R, Stenger B, Cipolla R. Augmenting depth camera output using photometric stereo.[C]//Proceedings of the Conference on Machine Vision Applications. 2011: 369-372.

[111] Lu L, Qi L, Luo Y, et al. Three-dimensional reconstruction from single image base on combination of cnn and multi-spectral photometric stereo[J]. Sensors, 2018, 18(3):764.

[112] Ju Y, Qi L, He J, et al. Mps-net: Learning to recover surface normal for multispectral photometric stereo[J]. Neurocomputing, 2020, 375:62-70.

[113] Ju Y, Dong X, Wang Y, et al. A dual-cue network for multispectral photometric stereo[J]. Pattern Recognition, 2020, 100:107162.

[114] Johnson M K, Adelson E H. Shape estimation in natural illumination[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011: 2553-2560.

[115] Wiles O, Zisserman A. Silnet: Single-and multi-view reconstruction by learning from sil-

houettes[C]//Proceedings of the British Machine Vision Conference. 2017: 99.1-99.13.

[116] Matusik W, Pfister H, Brand M, et al. A data-driven reflectance model[J]. ACM Transactions on Graphics, 2003, 22(3):759-769.

[117] Jakob W. Mitsuba renderer[Z]. 2010.

[118] McAuley S, Hill S, Hoffman N, et al. Practical physically-based shading in film and game production[M]//ACM SIGGRAPH 2012 Courses. 2012: 1-7.

[119] Einarsson P, Chabert C F, Jones A, et al. Relighting human locomotion with flowed reflectance fields[C]//Proceedings of the Eurographics Conference on Rendering Techniques. 2006: 183-194.

[120] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436-444.

[121] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4):600-612.

[122] Blau Y, Michaeli T. The perception-distortion tradeoff[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6228-6237.

[123] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[124] Hartmann W, Galliani S, Havlena M, et al. Learned multi-patch similarity[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1586-1594.

[125] Choy C B, Xu D, Gwak J, et al. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction[C]//Proceedings of the European Conference on Computer Vision. 2016: 628-644.

[126] Ummenhofer B, Zhou H, Uhrig J, et al. Demon: Depth and motion network for learning monocular stereo[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5038-5047.

[127] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5693-5703.

[128] Simchony T, Chellappa R, Shao M. Direct analytical methods for solving poisson equations in computer vision problems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(5):435-446.

[129] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the International Conference on Learning Representations. 2015: 1-14.

[130] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[131] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.

[132] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7):971-987.

[133] Gan Y, Xu X, Sun W, et al. Monocular depth estimation with affinity, vertical pooling, and label enhancement[C]//Proceedings of the European Conference on Computer Vision. 2018: 224-239.

[134] Liu B, Yu H, Long Y. Local similarity pattern and cost self-reassembling for deep stereo matching networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022.

致　谢

时光飞逝，转眼间，我在海大六年的硕博连读生活即将画上句号。

在无数个失眠的夜里，我曾一遍一遍想过，致谢这部分到底该怎么写，该用如何绚丽的辞藻开头，又以如何押韵的排比结束。然而真到了这一天，我却发现华丽的语言，却无法形容我内心的感受。可以说除了感慨时间的快速流逝，便是对老师、亲人、朋友的真诚感谢。

首先，我要向我的导师、中国海洋大学计算机科学与技术学院院长董军宇教授表示衷心的感谢和敬意。硕博连读六年期间，董老师在生活及学习方面等给予了我很大的帮助。从刚入学开始，董老师就一直努力培养我、鼓励我，让我明确了博士期间的学习方向和应有的科研态度；给我提供机会去提升、展示自己，六年如一日。在我面对困难时，董老师总能及时的开导我，帮助我。董老师严谨治学的态度、待人接物的理念、对前沿问题敏锐的洞察力，都令我受益终身。另外，我还要特别感谢实验室的亓琳老师、高峰老师、张述老师、范浩老师、王改革老师、董兴辉老师、于彦伟老师，感谢各位老师对我的悉心教导，让我明白如何一步步展开科研工作。是他们的支持和帮助伴我度过了难忘的博士生涯。

我还要特别感谢香港理工大学 Kennth Kin-Man Lam 教授，山东财经大学蹇木伟教授，北京大学施柏鑫教授、彭宇新教授，英国莱斯特大学 Huiyu Zhou 教授，英国南安普顿大学 Sheng Chen 教授对我学术和生活上的指导和帮助，他们高深的学术造诣和治学态度令我受益匪浅，引领我前进。

此外，我还要感谢香港理工大学韩钰先生、肖均先生、王聪先生，天津商业大学姚鹏博士，清华大学深圳国际研究生院陈扬先生，英国南安普顿大学冯晓萌女士，澳大利亚悉尼科技大学王英宇先生，实验室饶源、胡业琦、郭少翔等硕博同学，以及所有关心、帮助我的师生和朋友们，感谢他们在学术上的探讨和生活上对我的支持，使我在人生的道路上走出更加坚定、精彩的自己。

最后，我要格外感谢我的父母和家人，感谢他们默默的支持和无私的关爱，对我 28 年来的养育和培育。我取得的所有成绩都是属于他们的。

亦感谢将论文看到最后的您，您的认可或批评都将成为我今后成长的无限动力。

个人简历、在学期间发表的学术论文与研究成果

个人简历

1994 年 7 月 2 日出生于山东省青岛市。

2012 年 9 月考入四川大学机械学院工业设计专业，2016 年 6 月本科毕业并获得工学学士学位。

2016 年 9 月考入中国海洋大学信息科学与工程学部计算机应用技术专业攻读硕士学位。2018 年硕博连读，转入中国海洋大学信息科学与工程学部计算机应用技术专业攻读博士学位至今。

2021 年 1 月至 2021 年 7 月在香港理工大学工程学院电子及资讯工程系进行博士研究生联合培养。

发表的学术论文

[1] **Yakun Ju**, Junyu Dong, Sheng Chen. Recovering surface normal and arbitrary images: A dual regression network for photometric stereo[J]. *IEEE Transactions on Image Processing, 2021, 30: 3676-3690.* **（CCF-A 期刊，SCI 中科院分区一区，影响因子 10.856，第一作者）**

[2] **Yakun Ju**, Kin-Man Lam, Yang Chen, *et al.* Pay attention to devils: A photometric stereo network for better details[C]. *Proceedings of the International Conference on International Joint Conferences on Artificial Intelligence (IJCAI 2020), 694-700.* **（CCF-A 会议，第一作者）**

[3] **Yakun Ju**, Xinghui Dong, Yingyu Wang, *et al.* A Dual-cue Network for Multispectral Photometric Stereo [J]. *Pattern Recognition, 2020, 100: 107162.* **（CCF-B 期刊, SCI 中科院分区一区，影响因子 7.740，第一作者）**

[4] **Yakun Ju**, Lin Qi, Jichao He, *et al.* MPS-Net: Learning to Recover Surface Normal for Multispectral Photometric stereo [J]. *Neurocomputing, 2020, 375: 62-70.* **（CCF-C 期刊, SCI 中科院分区二区，影响因子 5.719，第一作者）**

[5] **Yakun Ju**, Muwei Jian, Shaoxiang Guo, *et al.* Incorporating Lambertian Priors into Surface Normals Measurement [J]. *IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.* **（SCI 中科院分区二区，影响因子 4.016，第一作者）**

[6] **Yakun Ju**, Yuxin Peng, Muwei Jian, *et al.* Learning Conditional Photometric Stereo with High-resolution Features [J]. *Computational Visual Media, 2022, 8(1): 105-118.* **（SCI 中科院分区二区，影响因子 3.365，中国科技期刊卓越行动计划期刊，第一作者）**

[7] **Yakun Ju**, Muwei Jian, Junyu Dong, *et al.* Learning Photometric Stereo via Manifold-based Mapping [C]. *Proceedings of the IEEE International Conference on Visual Communications and Image Processing (IEEE VCIP 2020), 2020: 411-414.* **（EI 会议，第一作者）**

[8] **举雅琨**, 蹇木伟, 饶源, 等. MASR-PSN：低分光度立体图像的高分法向重建深度学习模型 [J]. 中国图象图形学报, 已录用. **（CCF 计算领域高质量科技期刊分级目录-T2，第一作者）**

[9] Yanru Liu, **Yakun Ju**, Muwei Jian, *et al.* A Deep-shallow and Global-local Multi-feature Fusion Network for Photometric Stereo [J]. *Image and Vision Computing, 2022, 118: 104368.* **（CCF-C 期刊, SCI 中科院分区三区，影响因子 2.818，通讯作者）**

[10] Yingyu Wang, **Yakun Ju**, Muwei Jian, *et al.* Self-supervised Depth Completion with Attention-based Loss [C]. *Proceedings of the International Workshop on Advanced Imaging Technology (IWAIT 2020), 11515.* **（EI 会议，第二作者）**

[11] Shaoxiang Guo, Eric Rigall, **Yakun Ju**, *et al.* 3D Hand Pose Estimation from Monocular RGB with Feature Interaction Module [J]. *IEEE Transactions on Circuits and Systems for Video Technology, 2022, Early Access.* **（CCF-B 期刊，SCI 中科院分区一区，影响因子 4.685，第三作者）**

[12] Yuan Rao, Jian Yang, **Yakun Ju**, *et al.* Learning General Feature Descriptor for Visual Measurement with Hierarchical View Consistency [J]. *IEEE Transactions on Instrumentation and Measurement, 2022, Early Access.* **（SCI 中科院分区二区，影响因子 4.016，第三作者）**

授权的发明专利

[1] **举雅琨**，董军宇，亓琳，卢亮。一种基于深度学习的单帧图像三维重建装置及方法，专利号：201711302400，授权日期：2021 年 2 月。**第一发明人**

[2] **举雅琨**，董军宇，高峰。基于深度学习的高频区域增强的光度立体三维重建

方法，专利号：202111524515，授权日期：2022 年 3 月。**第一发明人**

参与的主要课题

[1] 国家重大科研仪器研制项目“水下光学高分辨率三维成像仪研发”。（项目编号：41927805）

[2] 国家国际科技合作项目“水下高精度三维实时检测分析系统合作研发”。（项目编号：2014DFA10410）

攻读博士学位期间主要获奖情况

[1] 2020 年研究生国家奖学金

[2] 2022 山东省优秀毕业生

[3] 2020 年山东省研究生优秀成果奖

[4] 2021 年浪潮奖学金

[5] 2018 年歌尔声学奖学金

[6] 中国海洋大学 2020-2021 年优秀研究生

[7] 中国海洋大学 2019-2020 年优秀研究生

[8] 中国海洋大学 2018-2019 年优秀研究生