

深度学习技术在光度立体三维重建中的应用

王国琿*, 卢彦汀

西安工业大学光电工程学院, 陕西 西安 710021

摘要 光度立体三维(3D)重建是机器视觉和光度学领域中研究的热点问题,由于设备简单、成本低廉、分辨率高,得到广泛的应用。近年来,伴随着人工智能与深度学习技术的蓬勃发展,光度立体技术的发展进入一个崭新的时代。对深度学习技术在光度立体3D重建中的研究进展进行综述。首先,介绍光度学3D重建的研究背景和基本原理;其次,对光度立体3D重建方法的类型进行概述;接着,简要介绍常用的合成与实际拍摄数据集;然后,详细阐述深度学习技术在光度立体3D重建中的应用,它将基于物理模型的光度立体技术变为一种“数据驱动”下的技术,从而实现较高的预测精度;最后,进行分析与总结,并指出深度学习技术在光度立体领域所面临的挑战以及未来的研究趋势。

关键词 光度立体; 三维重建; 光度学; 人工智能; 深度学习; 数据驱动

中图分类号 O432

文献标志码 A

DOI: 10.3788/LOP230431

Application of Deep Learning Technology to Photometric Stereo Three-dimensional Reconstruction

Wang Guohui*, Lu Yanting

School of Optoelectronic Engineering, Xi'an Technological University, Xi'an 710021, Shaanxi, China

Abstract Photometric stereo three-dimensional (3D) reconstruction is a hot topic in the fields of machine vision and photometry. This method is widely used because of its simple equipment, low cost, and high resolution. With the rapid advancement of artificial intelligence and deep learning technology in recent years, the development of photometric stereo technology has entered a new era. This paper reviews the progress in the application of depth learning technology to photometric stereo 3D reconstruction. First, the research background and the basic principles of photometric 3D reconstruction are introduced. Next, various types of photometric stereo 3D reconstruction methods are summarized. Then, the commonly used synthetic and real-photoed datasets are briefly introduced. Further, a detailed description of the applications of depth learning technology in photometric stereo 3D reconstruction is provided, wherein the physical model-based photometric stereo technology is transformed into a “data-driven” technology to achieve high prediction accuracy. Finally, the paper analyzes and summarizes the challenges and opportunities for future research in the application of deep learning technology to photometric stereo reconstruction.

Key words photometric stereo; three-dimensional reconstruction; photometry; artificial intelligence; deep learning; data-driven

1 引言

关于物体的三维(3D)形貌重建是人类视觉的主要功能之一,也是机器视觉与计算机视觉领域中一个十分重要的研究方向,从明暗恢复形状(SFS)和光度立体方法是光度学3D重建物体表面形貌的两种关键方式,其目标是利用单个或多个光源照射下获得的单

幅或多幅图像的灰度变化(明暗)来重建物体表面各点的法向量,进而由法向量通过积分获得物体表面的3D形貌^[1-2]。

SFS方法最早是由美国麻省理工学院(MIT)的学者Horn^[3-4]对月球表面进行3D形貌重建时提出的。他认为SFS方法可以看作是成像过程的逆过程,图像的灰度信息与成像时物体表面3D形貌造成的反射光强

收稿日期: 2023-01-01; 修回日期: 2023-02-16; 录用日期: 2023-02-22; 网络首发日期: 2023-03-02

基金项目: 陕西省教育厅科学研究计划项目资助(22JY025)、陕西省自然科学基金基础研究计划资助项目(2022JM-318)

通信作者: *booler@126.com

弱密切相关,因而可由物体表面亮度的明暗变化来获得其高度信息。Horn用图像辐照度方程解释了物体表面3D形貌与图像的灰度信息之间的关系,SFS方法实质上就是建立和求解图像辐照度方程。然而图像辐照度方程是关于两个未知量的非线性偏微分方程(PDE),显然是病态的。如果不引入附加约束条件,仅由方程本身是无法正确求解的。研究者们为了解决SFS问题的病态性,相继引入一些约束条件,并提出许多算法。依据文献[5-6]的建议,这些算法大体上分为4类:线性化反射图方法、局部分析方法、最小化能量方法以及PDE方法。

这些方法中最值得一提的是Woodham^[7]在研究SFS方法的基础上提出的光度立体方法,这种方法实质上是通过增加光照约束条件来解决SFS方法的病态性。他利用不同方向的 $m(m \geq 3)$ 个光源依次照射同一被测物体,获得 m 幅图像,进而联立 m 幅图像的辐照度方程求解物体表面的法向量。由于在成像过程中摄像机和被测物体的位置均保持不变,Woodham的这一做法避免了3D重建过程中多幅图像对应点的匹配问题。

光度学3D重建作为一种典型的被动非接触式测量手段,兼顾非接触式测量和被动式方法的优点,它能够快速、有效地提取被测物体表面的几何形状信息,其设备简单、成本低廉、分辨率高、适用性强,具有非常广阔的发展前景。其中,光度立体方法的优势在于多幅图像中对应点是完全配准的,避免了多视角图像或图像序列中对应点的匹配问题,同时可以实现对无纹理表面的3D重建,而这是基于纹理特征匹配的立体视觉方法所无法达到的。并且该方法能对漫反射或非漫反射物体表面进行3D重建,细节重建能力突出,可获得稠密点云结构和显著的高频信息特征,因而已经成为机器视觉和光度学领域中研究的热点问题,并广泛应用于工农业产品检测^[8-10]、医学图像分析与重建^[11-13]、人脸与指纹等生物特征识别^[14-15]、文化遗产保护^[16-17]以及月面3D形貌重建^[18-19]等领域。

2016年3月,谷歌旗下DeepMind公司所开发的围棋软件AlphaGo战胜了韩国围棋九段棋手李世石^[20]。自此,以深度学习为代表的人工智能(AI)技术开始全面进入大众的视野,并且业界普遍认为,它很可能带来下一次科技革命,在未来深刻地改变人们的生产生活。当前,人工智能已经在机器视觉、图像工程、自然语言处理(NLP)等多个领域的技术上取得了全面的突破^[21-27]。近年来,伴随着人工智能技术的蓬勃发展,深度学习技术开始在光学成像、计算成像、光学测量等场合逐渐渗透,并展现出惊人的潜力^[28-34]。对于光度立体3D重建方法而言,深度学习技术已经开始成功应用于光照方向估计、反射率重建、法向量重建等方面。这些应用表明了人工智能技术的辅助下,光度立体技术的发展进入了一个崭新的时代,其重建精度、适用性

等方面取得了巨大的突破。

本文首先介绍光度学3D重建的两种方式——SFS和光度立体方法的基本原理;其次,介绍光度立体3D重建方法的类型以及常用的合成与实际拍摄数据集;接着,阐述深度学习技术在光度立体3D重建中的应用,它将基于物理模型的光度立体技术变为一种“数据驱动”下的技术;最后,指出光度立体3D重建面临的挑战以及未来的研究趋势。

2 光度学3D重建理论

2.1 SFS 3D重建基本原理

建立如图1所示的SFS成像模型^[35],设摄像机成像平面为 xy 平面,镜头的光轴与 z 轴重合, θ_i 为物体表面3D形貌 $z = z(x, y)$ 的法向量 $\mathbf{n}(x, y)$ 与光照方向向量 \mathbf{L} 之间的夹角, θ_r 为 $\mathbf{n}(x, y)$ 与摄像机方向向量 \mathbf{V} 之间的夹角。

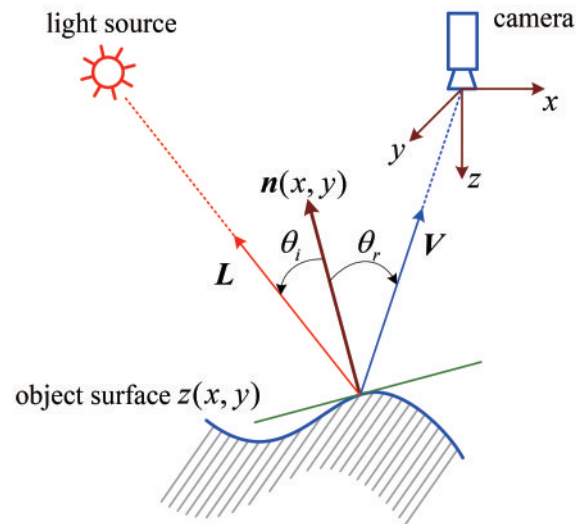


图1 SFS成像模型^[35]

Fig. 1 The imaging model of SFS^[35]

SFS和光度立体3D重建问题中往往假定光源 \mathbf{L} 为点光源且位于无穷远处,因而其光照方向向量可记为 $\mathbf{L} = (p_s, q_s, -1)^T$,摄像机的方向向量 \mathbf{V} ,物体表面法向量 $\mathbf{n}: \mathbf{R}^2 \rightarrow \mathbf{R}^3, (x, y) \mapsto [p(x, y), q(x, y), -1]^T$,其中

$$\begin{cases} p(x, y) = \frac{\partial z(x, y)}{\partial x} \\ q(x, y) = \frac{\partial z(x, y)}{\partial y} \end{cases} \quad (1)$$

在摄像机遵循正交投影条件下(关于透视投影条件下图像辐照度方程的推导可参见文献[1, 2, 36]),SFS 3D重建问题实质上就是求解图像辐照度方程^[3-4, 35]:

$$I(x, y) = R[\mathbf{n}(x, y)] = R[p(x, y), q(x, y)], \quad (2)$$

式中: $I(x, y)$ 为摄像机观测到的图像灰度值; $R(p, q)$ 是由反射模型确定的反射图。

通常, SFS 和光度立体 3D 重建问题中假定物体表面反射特性遵循 Lambert 反射模型。此时, 根据光度学理论, 反射图 $R(p, q)^{[3-4,7]}$ 为

$$R(p, q) = \rho(x, y) \cos \theta_i, \quad (3)$$

式中: $\rho(x, y)$ 为 Lambert 物体表面 $z = z(x, y)$ 处的反射率,

$$\cos \theta_i = \frac{\mathbf{n} \cdot \mathbf{L}}{\|\mathbf{n}\| \|\mathbf{L}\|} = \frac{p_s p + q_s q + 1}{\sqrt{p_s^2 + q_s^2 + 1} \sqrt{p^2 + q^2 + 1}}. \quad (4)$$

将式(3)和(4)代入式(2), 整理可得到 SFS 图像辐照度方程:

$$I(x, y) = \rho(x, y) \frac{p_s p(x, y) + q_s q(x, y) + 1}{\sqrt{p_s^2 + q_s^2 + 1} \sqrt{p^2(x, y) + q^2(x, y) + 1}}. \quad (5)$$

为了求解图像辐照度方程(5), SFS 中一般设 $\rho(x, y) \equiv \rho_0$, 并令 $E(x, y) = I(x, y)/\rho_0$, 此时有

$$E(x, y) = \frac{p_s p(x, y) + q_s q(x, y) + 1}{\sqrt{p_s^2 + q_s^2 + 1} \sqrt{p^2(x, y) + q^2(x, y) + 1}}. \quad (6)$$

显然, SFS 图像辐照度方程(6)是关于两个未知量 $p(x, y), q(x, y)$ 的非线性 PDE, 如果不引入附加约束条件, 仅由方程本身是无法正确求解的。研究者们为了解决辐照度方程(6)的病态性, 相继引入一些约束条件, 并提出许多算法, 如线性化反射图方法、局部分析方法、最小化能量方法以及 PDE 方法, 具体可参见文献[5-6]。

2.2 光度立体 3D 重建基本原理

对于图像辐照度方程(5)来说, 其存在 3 个未知量 $\rho(x, y), p(x, y), q(x, y)$, 一种最为直接的办法就是通过增加光照约束条件去除病态性。换句话说, 为了求解这 3 个未知量, 需至少构造 3 个不相关的图像辐照度方程, 这也正是 Woodham 的研究思路^[7]。他提出了如图 2 所示的光度立体方法, 利用不同方向的 $m(m \geq 3)$ 个光源 L_1, L_2, \dots, L_m 依次照射同一被测物体, 获得 m 幅图像。联立 $I_1(x, y), I_2(x, y), \dots, I_m(x, y)$ 可得:

$$\begin{cases} I_1(x, y) = \rho(x, y) \frac{p_{s1} p(x, y) + q_{s1} q(x, y) + 1}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1} \sqrt{p^2(x, y) + q^2(x, y) + 1}} \\ I_2(x, y) = \rho(x, y) \frac{p_{s2} p(x, y) + q_{s2} q(x, y) + 1}{\sqrt{p_{s2}^2 + q_{s2}^2 + 1} \sqrt{p^2(x, y) + q^2(x, y) + 1}} \\ \vdots \\ I_m(x, y) = \rho(x, y) \frac{p_{sm} p(x, y) + q_{sm} q(x, y) + 1}{\sqrt{p_{sm}^2 + q_{sm}^2 + 1} \sqrt{p^2(x, y) + q^2(x, y) + 1}} \end{cases}. \quad (7)$$

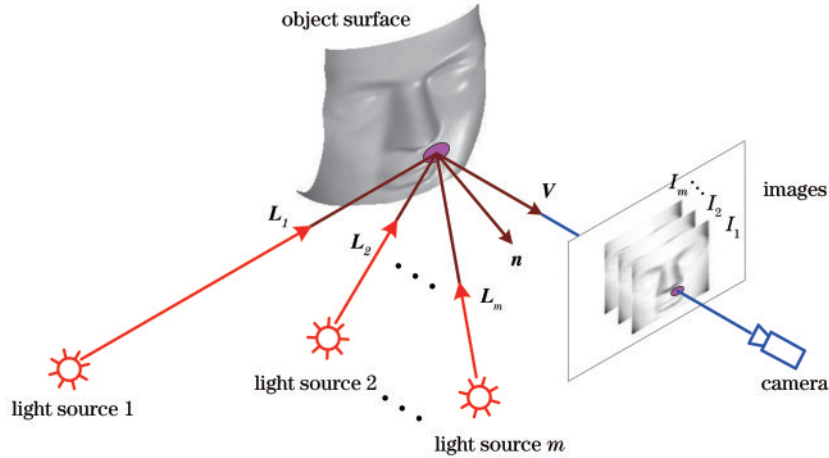


图 2 光度立体成像模型

Fig. 2 The imaging model of photometric stereo

为了方便计算, 记光照单位方向向量 $S_i = L_i / \|L_i\|$, 物体表面单位法向量 $N(x, y) = \mathbf{n}(x, y) / \|\mathbf{n}(x, y)\|$, $\mathbf{g}(x, y) = \rho(x, y) \mathbf{N}(x, y)$, 整理式(7)可得光度立体图像辐照度方程组:

$$\begin{cases} I_1(x, y) = \mathbf{g}(x, y) \cdot \mathbf{S}_1 \\ I_2(x, y) = \mathbf{g}(x, y) \cdot \mathbf{S}_2 \\ \vdots \\ I_m(x, y) = \mathbf{g}(x, y) \cdot \mathbf{S}_m \end{cases} \triangleq \mathbf{i}(x, y) = \mathbf{S} \mathbf{g}(x, y), \quad (8)$$

式中: $\mathbf{i}(x, y) = [I_1(x, y), I_2(x, y), \dots, I_m(x, y)]^T$; $\mathbf{S} = [\mathbf{S}_1^T, \mathbf{S}_2^T, \dots, \mathbf{S}_m^T]^T$ 。

显然, $\mathbf{g}(x, y)$ 可以通过求解线性系统(8)得到。典型地, 当 $m \geq 4$ 时, 可以求解得到线性系统(8)的最小二乘解(9), 进而可重建得到 $\rho(x, y)$ 、 $\mathbf{N}(x, y)$ 。

$$\mathbf{g}(x, y) = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{i}(x, y), \quad (9)$$

$$\rho(x, y) = \|\mathbf{g}(x, y)\|, \quad (10)$$

$$\mathbf{N}(x, y) = \frac{\mathbf{g}(x, y)}{\rho(x, y)} = \frac{\mathbf{g}(x, y)}{\|\mathbf{g}(x, y)\|}. \quad (11)$$

进一步, 如果需要重建物体表面 3D 形貌 $z(x, y)$, 可由法向量 $\mathbf{N}(x, y)$ 通过积分获得, 具体积分方法可

参见文献[37-42]。其中, Horn-Brooks 方法^[37]和 Frankot-Chellappa 方法^[38]是最早提出的、且具代表性的积分方法。

图 3 给出了 $m=5$ 时模拟摄像机观测到的图像和光度立体 3D 重建的结果^[43]。图 3(a)~(e) 为利用不同方向的 5 个光源 L_1, L_2, \dots, L_5 依次照射同一球体, 通过如图 2 所示的成像模型模拟生成的 5 幅灰度图像(具体可参见第 3.1 节); 图 3(f) 为光度立体方法式(10)重建的反射率 $\rho(x, y)$; 图 3(g) 为光度立体方法式(9)重建的 $\mathbf{g}(x, y)$; 图 3(h) 为光度立体方法式(11)重建的球体表面单位法向量 $\mathbf{N}(x, y)$; 图 3(i) 为通过积分方法重建的球体表面 3D 形貌 $z(x, y)$ 。

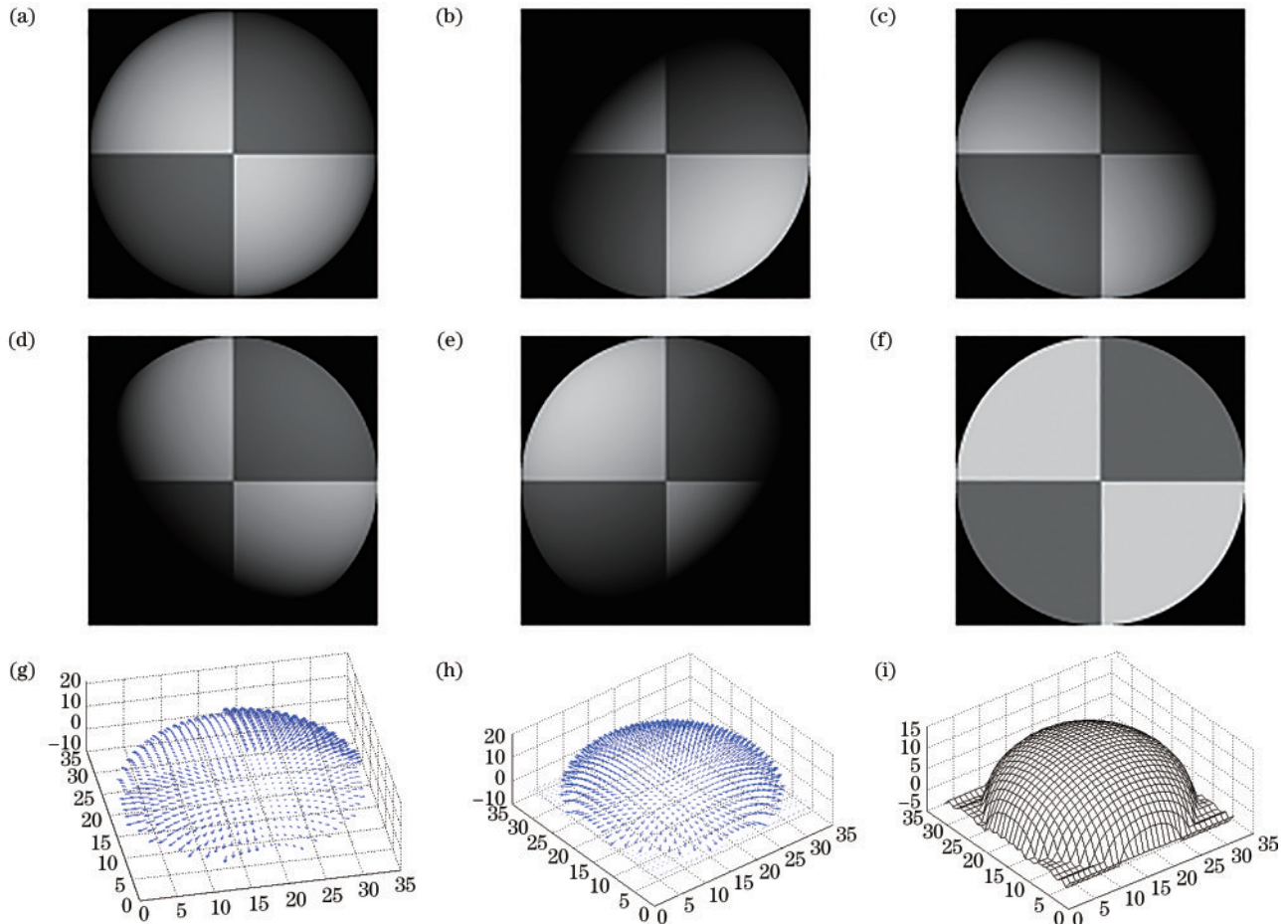


图 3 光度立体 3D 重建示意图^[43]。(a)~(e) 球体合成图像 $I_1(x, y) \sim I_5(x, y)$; (f) 重建的反射率 $\rho(x, y)$; (g) 重建的向量 $\mathbf{g}(x, y)$; (h) 重建的法向量 $\mathbf{N}(x, y)$; (i) 重建的球体形貌 $z(x, y)$

Fig. 3 The illustration of photometric stereo 3D reconstruction^[43]. (a)~(e) The synthetic images $I_1(x, y) \sim I_5(x, y)$ of a sphere; (f) the reconstructed albedo $\rho(x, y)$; (g) the reconstructed vector field $\mathbf{g}(x, y)$; (h) the reconstructed normal field $\mathbf{N}(x, y)$; (i) the reconstructed shape $z(x, y)$ of the sphere

2.3 光度立体 3D 重建方法类型

在 Woodham 研究成果的基础上, 涌现了众多的光度立体方法。根据文献[44-45]的分类原则——按照光源几何(方向或位置)与强度信息是否已知, 光度立体 3D 重建方法大致可分为 3 类: 未标定的光度立体(几何与强度信息均未知)、半标定的光度立体(几何信

息已知、强度信息未知)和标定的光度立体(几何与强度信息均已知)。

其中, 未标定光度立体方法中具有里程碑意义的是 Hayakawa^[46]提出的基于奇异值分解的方法, 然而该方法得到的法向量并不唯一, 存在着 3×3 个参数的线性歧义。Belhumeur 等^[47]对 Hayakawa 提出的方法进

行了深入的研究,证明了借助于表面可积性约束,线性歧义可以降为仅含 3 个参数的 generalized bas-relief (GBR) 歧义。为了进一步解决 GBR 歧义问题, Papadimitri 等^[48]利用局部漫反射最大值(此时物体局部表面的法向量与光照的方向向量一致)得到了 GBR 参数空间中的闭合解,该方法简单,鲁棒性和一致性较好。Sengupta 等^[49]联合秩约束和可积性约束,并利用 alternating direction method of multipliers (ADMM) 算法^[50]解决了较少(4~6 个)光源下 GBR 歧义和数据丢失问题。

针对半标定光度立体方法,Cho 等^[44,51]在最小二乘框架下利用交替最小化方法解决了非均匀光源强度以及摄像机处于自动曝光状态下的光度立体问题,然而该方法采用的是非鲁棒的最小二乘框架,故缺乏鲁棒性。Quéau 等^[52]在变分框架下利用鲁棒的交替加权最小二乘法解决了不精确光照方向信息下的半标定光度立体问题。

一直以来,众多学者关注标定光度立体方法的研究。标定光度立体涉及光源几何信息和强度信息的标定,通常利用一些外形、空间位置和反射率已知的几何物体辅助计算光源的方向/位置和强度。早期的标定光度立体大多假定点光源位于无穷远处或者为方向光源(平行光出射),同时假定光源的强度均匀一致。Zhang 等^[53]利用一个已知尺寸的 Lambert 球,结合关键点(球体表面法向量与光照方向向量垂直处)信息,可以获得光源的方向和强度,然而实际中纯 Lambert 反射球体和尺寸已知的条件限制了该方法的应用。为了解除上述条件的限制,Zhou 等^[54]利用已标定的双摄像机和一个尺寸与位置均未知的标定球获得光源的参数,标定球为含有 Lambert 反射和镜面反射的表面,其中,镜面反射用于光源方向的计算,Lambert 反射用于光源强度的计算。近些年来,为了去除点光源位于无穷远处的假设,标定光度立体大都采用近点光源(如发光面较小的 LED 光源),并开始考虑光源强度的角度衰减。针对近点光源位置的标定研究,Powell 等^[55]利用 3 个相对位置已知的反射球来实现光源标定,其中,表面具有漫反射特性的标定球用来确定摄像机的成像参数,通过另外两个反射球所形成的镜面反射点基于三角几何关系来确定光源的位置。Hara 等^[56]对物体表面的漫反射和镜面反射特性分别利用 Lambert 模型和 Torrance-Sparrow 模型建模,利用最小二乘拟合和非线性优化算法得到点光源的位置和物体表面的反射参数。由于采用镜面反射的标定物不能获得光源的光轴方向,Park 等^[57]采用 Lambert 平面标定板,结合光源的空间光强分布关于光轴方向径向对称的假设,实现了光源位置、光轴方向以及光强分布情况的标定。Visentini-Scarzanella 等^[58]去除了文献[57]方法中复杂的对称轴提取步骤,仅利用 Lambert 平面标定板的图像灰度最大值推导出了光源位置、光轴方向以及光强

分布情况的闭合解。Quéau 等^[59]借鉴文献[60]的想法,提出两步标定法:1)利用镜面反射球标定光源位置;2)利用 Lambert 平面标定板标定光源的光轴方向以及最大光强,并通过最小化重投影误差对上述参数进行优化,取得了较好的效果。笔者亦提出了一种基于立体视觉的 LED 光源位置参数标定方法,可获得较高的光源位置标定精度^[61]。

如前所述,传统的光度立体方法通常假定物体表面反射特性遵循 Lambert 反射模型。近年来,采用非 Lambert 模型来描述物体表面的反射特性也引起了研究者的兴趣。Shi 等^[62-63]全面分析比较了 8 种非 Lambert 光度立体方法,包括 2 种基于异常值丢弃的方法、1 种基于 Ward 反射模型的方法、5 种基于各向同性双向反射分布函数(BRDF)的方法,并对各种方法的优缺点进行了总结说明。Mecca 等^[64-65]提出了一种用于描述非均质材料的反射特性的单瓣 BRDF,大大减少了光度立体图像辐射度方程组的复杂性,可转化为拟线性 PDEs,并通过变分法求解。

至此,笔者介绍了 SFS 和光度立体 3D 重建技术的基本原理及方法类型。这些基本原理构成了光度学 3D 重建技术的物理模型,传统的光度立体 3D 重建技术大都是基于物理模型的技术。当前,通过深度学习技术,借助于光度立体图像数据集,光度立体 3D 重建技术可变为一种“数据驱动”下的技术,并且在光照方向估计、BRDF 重建、法向量重建、适用性等方面展现出了超越传统基于物理模型技术的能力。

3 光度立体 3D 重建中常用数据集

数据是深度学习技术的输入,基于深度学习的光度立体 3D 重建训练需要物体在不同光照 L_1, L_2, \dots, L_m 下的图像 I_1, I_2, \dots, I_m 和对应的表面法向量 N 的真值(ground truth)。下面将介绍基于深度学习的光度立体 3D 重建中常用的公开数据集。

3.1 合成图像数据集

合成的光度立体图像数据集实质上是利用计算机图形学知识对 3D 数据模型进行渲染(rendering)^[66]而来的。因而,需要生成/采集 3D 数据模型(法向量),同时需要指定摄像机投影方式、物体材质(反射特性)和光源类型(方向向量)。

对于简单的物体,如球体、花瓶等,可以直接采用物体 3D 形貌的解析式 $z = z(x, y)$ 生成 3D 数据模型^[5-6],进而可通过求导方式获得物体表面法向量 N ,然后根据指定的摄像机投影方式(正交投影、透视投影等)、物体表面反射率 ρ 和光照方向向量 L ,利用光度立体成像模型式(7)生成合成图像 I 。

对于复杂的物体,可以利用计算机辅助设计软件,或使用 3D 扫描仪配合逆向工程方法构建 3D 数据模型,结合特定的物体表面材质和光照方向,通过

渲染器(如 Cycles^[67]、Mitsuba^[68]等)生成合成图像 I 。例如,文献[69-80]使用 Blobby^[81]、Sculpture^[82]、Stanford 3D Scanning^[83]以及 Web 3D(互联网上下载)^[76]等形貌数据集作为 3D 数据模型,利用 MERL BRDF 数据集(100 种 BRDFs)^[84]或 Disney BRDF 数据集(11 个参数控制)^[85]提供的不同物体表面材质的

BRDFs,使用 DiLiGenT 数据集(96 种光照方向)^[62-63]提供的或半球空间内随机/均匀选择的光照方向,通过基于物理的光线追踪将 3D 数据模型、BRDF、光照方向渲染成深度学习训练/测试所需的光度立体图像。表 1 给出了常用的合成光度立体图像数据集的具体情况。

表 1 常用的合成光度立体图像数据集的具体情况

Table 1 Details of the commonly used synthetic photometric stereo datasets

Used datasets	3D shape (number)	BRDF (number)	Light (number)	Number of images
Reference [69-70]	Blobby ^[81] (8)	MERL ^[84] (100)	DiLiGenT ^[62-63] (96)	76800, 28800
Reference [72-74, 71, 77-78]	Blobby ^[81] (8)+Sculpture ^[82] (8)	MERL ^[84] (2 of 100)	Random (64)	1658880+3794688
Reference [75]	Blobby ^[81] (9)	MERL ^[84] (8 of 100)	Random (144)	10368
Reference [79-80]	CyclesPS ^[79] (15)	Disney ^[85] (~15000) ^[86]	Uniform (1000/1280)	53400

其中,文献[72]合成的 Blobby、Sculpture 光度立体数据集以及文献[79]合成的 CyclesPS 光度立体图像数据集是目前最为常用的基于深度学习的光度立体训练数据集,图 4 展示了 3 个数据集的部分数据。其中,第 1、2 行为不同表面材质和不同光照方向下的合成图像,为了方便显示,对第 1、2 行图像的亮度进行了

适当调整,第 3 行为对应的表面法向量。此外,对于原始的 Blobby 形貌数据集^[81],虽有 10 种不同形貌,但是均较为简单且充分光滑。而对于 Sculpture 形貌数据集^[82]、CyclesPS 数据集^[79],3D 数据模型相对较为复杂,能提供充分的形貌细节信息。

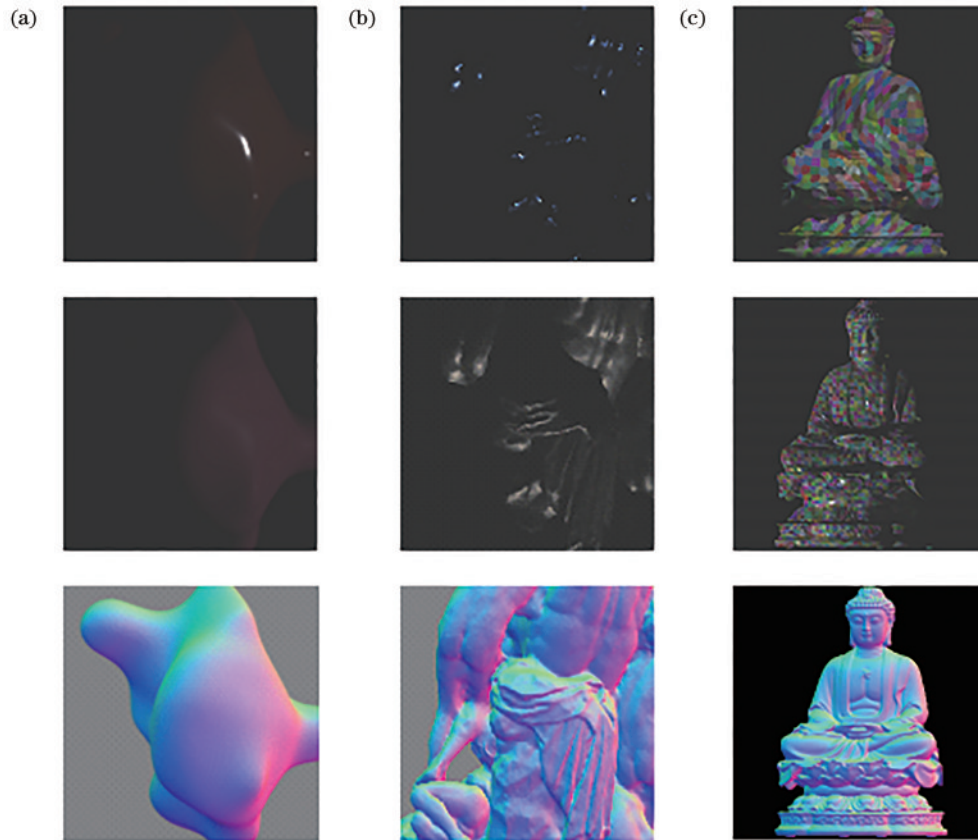


图 4 常用的合成光度立体图像数据集示意图。(a) Blobby 光度立体数据集^[72]; (b) Sculpture 光度立体数据集^[72]; (c) CyclesPS 光度立体数据集^[79]

Fig. 4 The illustration of the commonly used synthetic photometric stereo datasets. (a) Blobby photometric stereo dataset^[72]; (b) Sculpture photometric stereo dataset^[72]; (c) CyclesPS photometric stereo dataset^[79]

3.2 实际拍摄图像数据集

DiLiGenT 数据集^[62-63]是目前最为常用的光度立体 3D 重建标准测试数据集,图 5 给出了 DiLiGenT 数

据集中的部分数据。该数据集包含 10 个物体,依次命名为 ball、bear、buddha、cat、cow、goblet、harvest、pot1、pot2 和 reading。这些物体均具有较强的非 Lambert 表

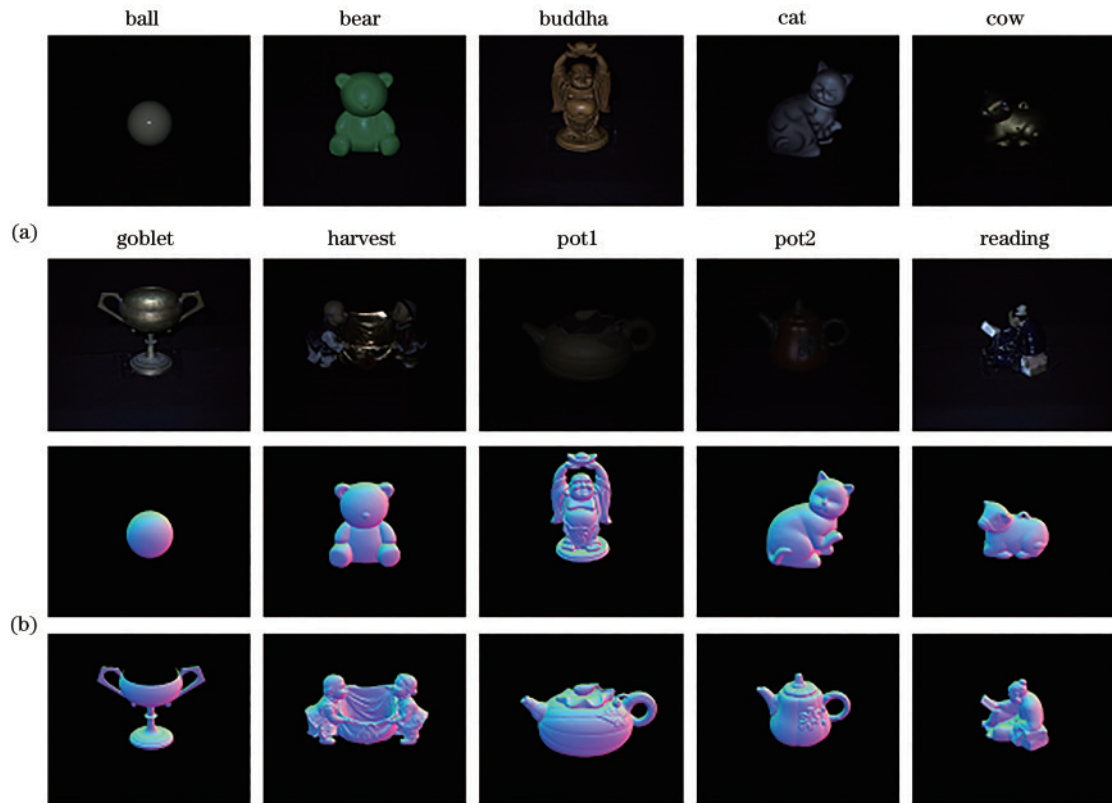


图 5 DiLiGenT 数据集示意图^[62-63]。(a)某一光照方向下的光度立体图像;(b)物体表面法向量
Fig. 5 Schematic diagrams of DiLiGenT dataset^[62-63]. (a) Photometric stereo images under a certain light direction;
(b) surface normal of the objects

面反射特性和较复杂的结构信息。对于每一个物体,使用 96 种不同方向的光依次照射,因而可拍摄得到 96 幅光度立体图像,每幅图像的分辨率为 612×512 。图 5(a)为某一光照方向下实际拍摄的 10 个物体的光度立体图像。此外,作者使用高精度的 3D 扫描仪获取每一个物体的 3D 网格,并计算得到表面法向量真值,如图 5(b)所示。

除了 DiLiGenT 数据集以外,常用的光度立体 3D 重建测试数据集还有 Gourd & Apple 数据集^[87]、Light Stage Data Gallery 数据集^[88]和 Harvard 数据集^[89]。其中:Gourd & Apple 数据集^[87]包含 3 个物体,依次命名为 Gourd1、Gourd2 和 Apple,分别使用 102、98 和 112 个光照方向获取光度立体图像;Light Stage Data Gallery

数据集^[88]由 6 个物体组成,使用高达 253 个光照方向获取光度立体图像;Harvard 数据集^[89]由 7 个物体组成,使用 20 个光照方向获取光度立体图像。近年来,伴随着人工智能与深度学习技术在光度立体 3D 重建中的蓬勃发展,还出现了 DiLiGenT-MV 数据集^[90]、Kaya 数据集^[45]、LUCES 数据集^[91-92]以及 DiLiGenT10²数据集^[93]。其中:为了去除传统数据集通常只有一个视角的局限,DiLiGenT-MV 数据集^[90]在 DiLiGenT 数据集^[62-63]的基础上扩展到 20 个视角;LUCES 数据集^[91-92]提供 14 个物体在 52 个近场 LED 光源照射下的光度立体图像。此外,DiLiGenT10²数据集^[93]给出了 10 个物体在 10 种不同材质下的光度立体图像。表 2 总结了常用的实际拍摄光度立体图像数据集的具体情况。

表 2 常用的实际拍摄光度立体图像数据集的具体情况

Table 2 Details of the commonly used real-photoed photometric stereo datasets

Datasets	Number of 3D shapes	Number of views	Number of lights	Number of images
DiLiGenT ^[62-63]	10	1	96	960
Gourd & Apple ^[87]	3	1	102, 98, 112	312
Light Stage Data Gallery ^[88]	6	1/2	253	2277
Harvard ^[89]	7	1	20	140
DiLiGenT-MV ^[90]	5	20	96	9600
Kaya ^[45]	3	1	260	780
LUCES ^[91-92]	14	1	52	728
DiLiGenT10 ² ^[93]	10 ($\times 10$)	1	100	10000

4 基于深度学习的光度立体 3D 重建

4.1 深度学习技术简介

深度学习(DL)技术是一种基于人工神经网络(ANN)的机器学习技术。2006年,Hinton等^[94-95]提出深度置信网络(DBN),利用预训练(pretraining)和微调(fine-tuning)方法初步缓解神经网络中存在的局部极值和梯度消失问题。之后,深度学习技术引发学术界关注,并在往后的十多年中取得了突破性的进展。

随着深度学习技术的不断发展,各种更加强大的深度学习模型不断涌现。目前,被广泛应用的深度学习模型主要有DBN^[94-95]、循环神经网络(RNN)^[96]和卷积神经网络(CNN)^[97]等。其中,CNN由于其强大的图像特征提取能力和分类预测能力,被广泛应用于机器视觉与计算机视觉领域。1998年,LeCun等^[97]首次提出应用于手写字识别的CNN架构,该架构被称之为LeNet-5,具体包括卷积层、下采样层(又称之为池化层)和全连接层。其中,卷积层用于图像的特征提取,下采样层用于已提取特征的二次提取,全连接层用于图像的分类预测。2012年,Krizhevsky等^[98]首次将CNN模型(该模型也被称之为AlexNet)用于ILSVRC-2012竞赛中,该模型在图像分类领域的错误率仅为15.3%,相较于之前最佳的方法,错误率降低了十个百分点,引起了巨大的轰动,也掀起了国内外学者对CNN模型的研究热潮。此后,出现了越来越多的模型,如VGG^[99]、ResNet^[100]、DenseNet^[101]等,CNN模型的精度也得到进一步提高。

4.2 深度学习在光度立体 3D 重建中的应用

1993年,Iwahori等^[102]采用单隐藏层神经网络,在反射率参数已知和未知两种情况下,成功使用3幅光度立体图像完成物体表面法向量的重建。这是神经网络与光度立体 3D 重建的首次结合。之后,Rajaram等^[103]、Cheng^[104]和 Elizondo等^[105]等采用多隐藏层神经网络和光度立体的方法进行法向量的重建。然而,上

述基于神经网络的光度立体 3D 重建方法大都只适用于 Lambert 模型^[102,104-105],在实际应用中有一定的局限性。随着深度学习技术的蓬勃发展以及CNN在图像识别和 3D 重建^[106-108]领域取得的巨大成功,大量的深度学习模型被应用于光度立体的 3D 重建问题。深度学习技术的应用开辟了解决非 Lambert 光度立体问题的新方向,也使得光度立体 3D 重建技术更加适用于实际拍摄场景。

4.2.1 DPSN

2017年,Santo等^[69]第一次将深度学习技术应用于光度立体 3D 重建,提出一种有监督的深度光度立体网络(DPSN)。该网络在复杂的反射率观测值与物体表面法向量之间建立灵活的映射关系,在处理非 Lambert 表面法向量的预测问题上取得了优于传统方法的结果,首次应用深度学习技术便获得了巨大的成功。

DPSN 由光度立体图像输入、阴影层(shadow layer)、全连接层(fully connected layers)和法向量图输出等 4 部分组成,如图 6 所示。该网络中,光度立体图像的光照方向是预先设定的,图像中的测量向量(measurement vectors)以逐像素(per-pixel)的形式输入 DPSN 中用于训练。在光度立体成像时,由于全局光照效应导致的投射阴影(cast shadow)问题是不可避免的,且用于训练的基于局部光照效应的 BRDF 模型无法对该效果进行建模。因此,为了在训练阶段模拟投射阴影效果,DPSN 引入基于 Dropout 技术的阴影层,随机丢弃一部分输入测量向量,即将它们设置为 0。通过阴影层来训练网络,DPSN 可以有效地学习从观测值到表面法向量的映射关系,并综合考虑了各种 BRDFs 和投射阴影。在预测阶段,给定与训练阶段相同的光照方向,DPSN 能以逐像素的方式估计物体表面的法向量,将估计出的 3 个通道的表面法向量分别归一化,通过取均值方式进行合并,接着进一步归一化,最终获得表面法向量的估计值,具体过程可参见文

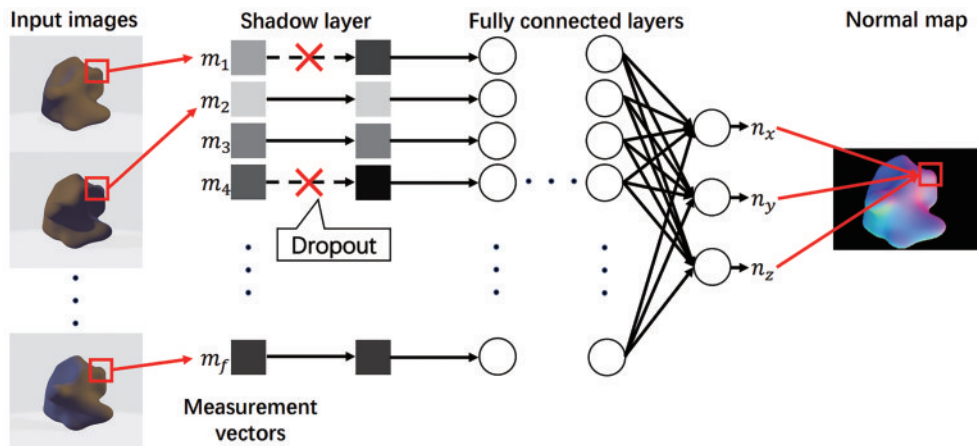


图 6 DPSN 架构^[69]

Fig. 6 The architecture of DPSN^[69]

献[69]。

图 7 给出了 DPSN 方法和传统的 L1 方法^[109]、L2 方法(即 BASELINE^[7])对 DiLiGenT 数据集^[62-63]中的 6 个较为复杂的物体表面法向量重建的结果和相应的误差图^[69]。其中:第 1 行为真值;第 2、3、4 行上方为不同方法重建的表面法向量,中间为相应的误差图,下方数字代表平均角度误差(MAE)。从图中可以看到,对于 buddha 和 reading 这类由陶瓷材料制成的物体,传统的基于 Lambert 模型的 L1、L2 方法可以获得较好的重建结果。对于 cow 和 goblet 这类由金属材料制成的非 Lambert 物体,DPSN 方法预测的结果明显优于传统方法,对于具有强镜面反射的非 Lambert 表面法向量重

建有着较好的效果。同时,阴影层中 Dropout 策略的运用和对投射阴影的有效模拟也显著提高了该方法的精度,其使用的深度神经网络(DNN)更是将光度立体 3D 重建技术真正变为一种“数据驱动”下的技术。随后,Santo 等^[70]对该工作进行了进一步的扩展,结合数据驱动的 BRDF^[110-111],使得 DPSN 可以对物体表面材质的 BRDF 进行预测。但是,该方法依然存在一些明显的问题,由于在网络训练过程中对光照方向信息的利用不充分^[86],DPSN 需要假定训练数据和预测数据的光照方向相同,不能够处理未知光照方向的光度立体 3D 重建问题,这限制了该方法在实际拍摄场景中的应用。

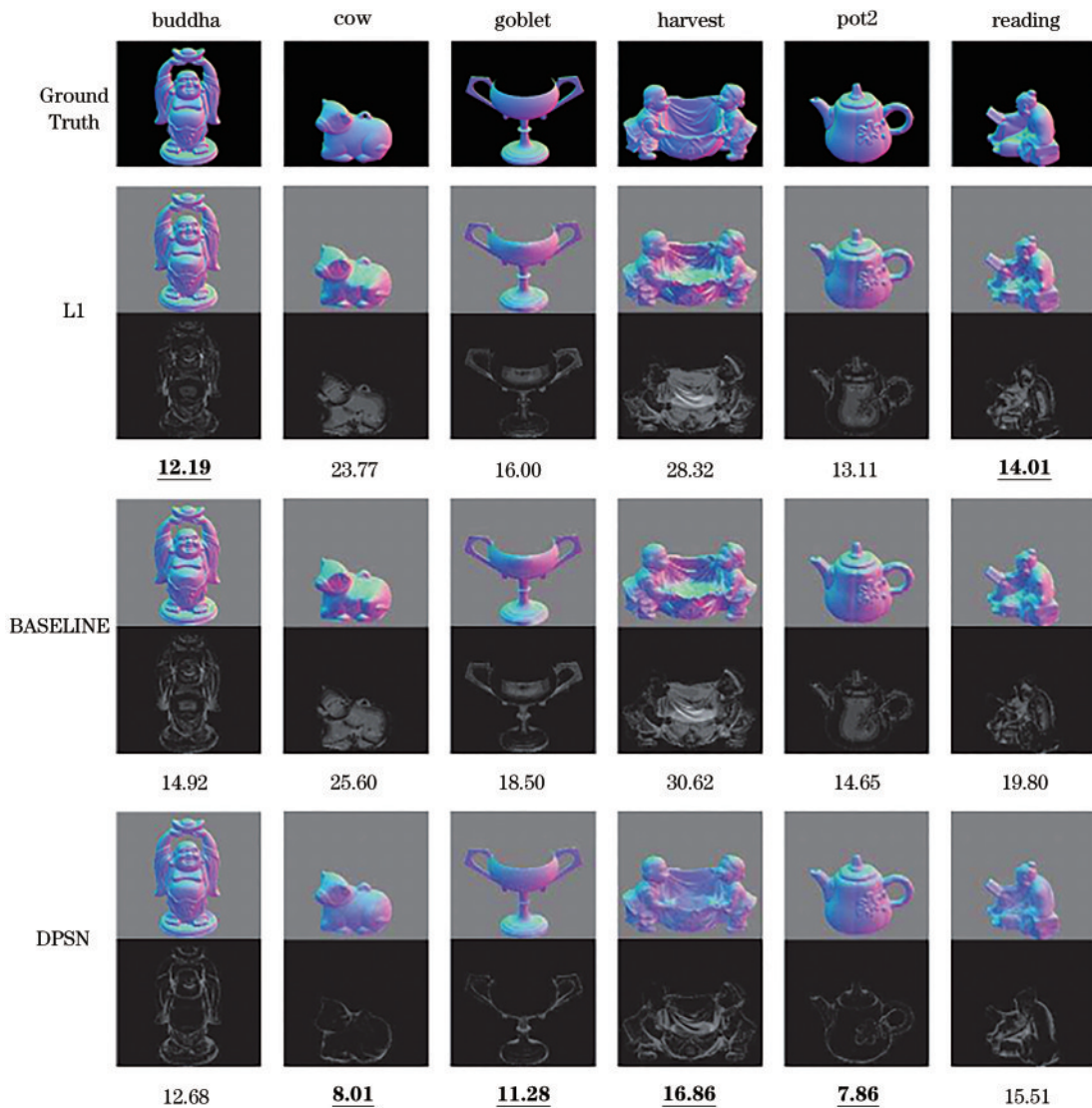


图 7 对 DiLiGenT 数据集中部分物体重建结果^[69]

Fig. 7 Reconstruction results for several objects from DiLiGenT dataset^[69]

4.2.2 PS-FCN、UPS-FCN 和 SDPS-Net

DPSN 的提出掀起了深度学习技术在光度立体 3D 重建领域的研究热潮。针对 DPSN 中光照方向需要预先定义等问题,各国学者们展开了广泛的研究。2018 年,Chen 等^[72]提出一种基于全卷积网络

(FCN)^[112]的适用于光度立体的深度神经网络,称之为 PS-FCN。该网络不仅解决了 DPSN 中光照方向需要预先定义的问题,使深度学习在光度立体 3D 重建领域的泛化能力大大提升,而且在非 Lambert 表面的处理上得到了更好的结果。

PS-FCN 是一个多分支连体网络,由共享权重特征提取器(shared-weight feature extractor)、特征聚合层(fusion layer)和法向量回归网络(normal regression network)等 3 部分组成,如图 8 所示。其中:共享权重特征提取器用于提取输入光度立体图像的特征;特征聚合层主要用于融合提取的特征;法向量回归网络用于重建表面法向量。相较于 DPSN,该网络架构有着显著的优势:首先,由于共享权重特征提取器和聚合层中最大池化操作(max-pooling)的应用,PS-FCN 的训

练过程和测试过程中均不依赖预先定义好的光照方向,训练阶段和预测阶段所使用的光照方向也可以不同;再者,输入端可以输入任意数量的光度立体图像用于特征的提取,并且输入顺序也可以是任意的;最后,最大池化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(non-activated)的特征,使得该方法在非 Lambert 表面,尤其是投射阴影场景下具有很强的鲁棒性。

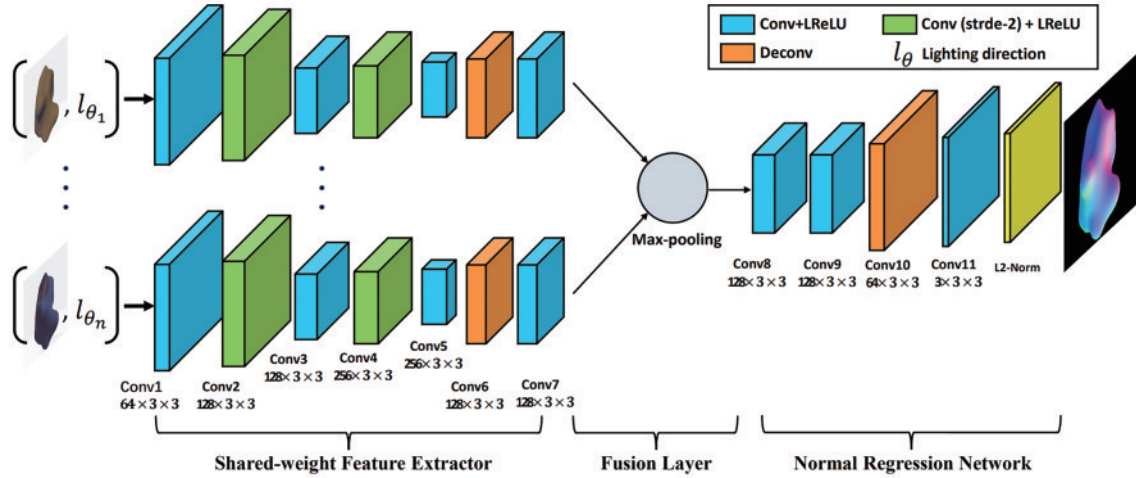


图 8 PS-FCN 架构^[72]

Fig. 8 The architecture of PS-FCN^[72]

图 9 展示了 PS-FCN 中融合后特征图的可视化效果^[72]。从图中可以看出,具有相似法向量的不同区域在不同通道中被激活。每个通道都可以被解释为法向量属于某个特定方向的概率(或者也可以解释为物体的明暗是在某个特定光照方向下渲染的)。因此,可以从这些概率分布中精确地推断出物体表面的法向量。

表 3 给出了 PS-FCN 重建 DiLiGenT 数据集^[62-63]的平均角度误差,并与 DPSN 进行了比较。尽管 DPSN

使用的 Dropout 技术已经良好地解决了投射阴影的问题,但 PS-FCN 所使用的最大池化操作对阴影问题的处理效果更加优异,对实际拍摄的非 Lambert 物体表面法向量的重建效果更好。此外,PS-FCN 用于训练的数据集是由 Blobby 形貌数据集^[81]、Sculpture 形貌数据集^[82]和 MERL BRDF 数据集^[84]合并创建而成的新数据集(具体情况可参见表 1),这也使得 PS-FCN 方法更适用于复杂形貌信息的物体。

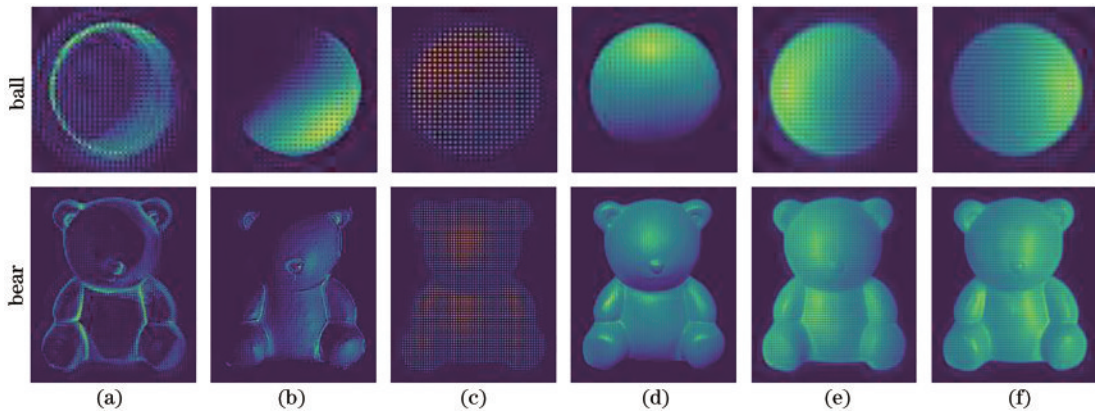


图 9 融合后特征图的可视化^[72]。(a)~(f) 128 个融合后特征图通道中的某 6 个通道

Fig. 9 Visualization of the feature map after fusion^[72]. (a)-(f) 6 of 128 channels of the fused feature map

尽管 PS-FCN 只需要在训练过程中去除光照方向即可用于解决未标定光度立体问题(即 Uncalibrated PS-FCN 版本,UPS-FCN^[72]),但在实际预测过程中性

能表现并不是特别优异。2019 年,Chen 等^[73]在 PS-FCN 的基础上加以拓展,提出了一个名为 self-calibrating deep photometric stereo networks (SDPS-

Net)的两阶段模型,致力于解决未标定的非 Lambert 光度立体问题。SDPS-Net 使用两阶段深度学习架构,有效利用中继监督(intermediate supervision)较好地解决了上述问题,具体重建结果如表 3 所示。

SDPS-Net 由光源标定网络(LCNet)和法向量预测网络(NENet)构成,如图 10 所示。LCNet 可以将任意数量的光度立体图像作为输入,将每一幅图像分别输入一个共享权重的特征提取器进行特征提取,提取相应的局部特征(local feature)。随后,通过最大池化

操作将提取到的所有局部特征融合为全局特征(global feature),全局特征包含物体的形状以及反射特性等信息,可以帮助解决光照的歧义性问题。之后,综合考虑局部特征和全局特征,估计出相应的光照方向和光强。NENet 使用 LCNet 预测的光照方向和光强对输入图像进行表面法向量的预测。因此,从功能上看,NENet 与 PS-FCN 很类似,区别在于,PS-FCN 在已知光照方向情况下进行训练,NENet 则使用 LCNet 预测的光照方向,在处理未标定光度立体问题上表现得更好。

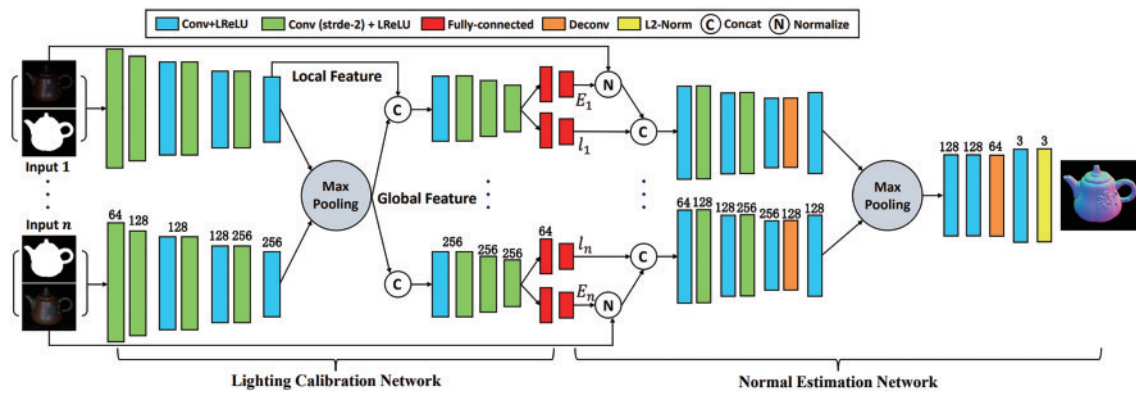


图 10 SDPS-Net 架构^[73]

Fig. 10 The architecture of SDPS-Net^[73]

4.2.3 CNN-PS

由于 DPSN 需要假定训练数据和预测数据的光照方向相同,因此该方法不适用于任意数量光度立体图像输入的情况。针对这一问题,2018 年, Ikehata^[79] 提出另一种基于 CNN 的名为 CNN-based photometric stereo (CNN-PS) 的方法,并首次引入观察图(observation map)的概念,它是在光照方向已知的情况下,由观测值进行逐像素半球面投影(pixelwise hemi-spherical projection)所生成的^[79],用于克服固定数量光度立体图像输入的问题。图 11 中,由于光照方向可以是整个单位半球的向量,因而光照方向 L_j ($j = 1, 2, \dots, m$) 与 x - y 坐标系[由于 x - y 平面垂直于摄

像机方向向量 V ,故 $V = (0, 0, 1)^T$ 之间存在映射关系,此时可以计算得到观察图(观察图示意可参见图 12 第 2 列),具体计算过程可参见文献[79]。

考虑光滑凸表面(如球体)上某一点的观察图,随着光照方向偏离真实表面法向越远,观察图的值越小且逐步减小为零。此外,观察图在局部范围内集中出现高强度值,表明光滑表面上存在窄镜面反射区域。反过来,若是考虑粗糙非凸表面(如纸碗)上某一点的观察图,与光滑凸表面会存在明显差异,观察图中没有局部范围内集中出现高强度值,可以推断该表面可能是粗糙的,存在投射阴影和互反射情况。因此,通过这种方式,观察图可以合理地编码表面上某点的几何形状、材质和光照属性。

CNN-PS 以观察图作为输入特征输入网络进行学习,因此可接受非结构化的光度立体图像输入(即输入图像的数量和顺序均可改变),并适用于非 Lambert 反射与全局光照效应(投射阴影、互反射)等结合在一起的实际拍摄场景。

但是,一般情况下得到的观察图通常是稀疏的。已有研究表明,丢失的数据通常被认为是 CNN 输入的问题,并且经常被插值处理^[113]。然而, Ikehata 发现平滑插值处理会降低性能,因为观察图通常是不平滑的,并且零值具有重要意义(如阴影)。为此, CNN-PS 使用各向同性约束下旋转伪不变性(rotation pseudo-invariance)的处理方式解决了该问题。给定光度立体图像和光照,生成观察图,通过围绕摄像机的方向向量

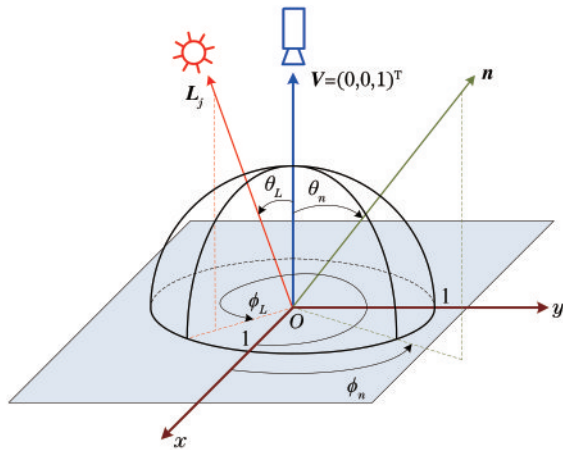


图 11 观察图生成时的坐标系

Fig. 11 The coordinate system for observation map

V 旋转光照方向和表面法向量 K 个特定角度, 便得到 K 张观察图, 实现训练集扩充以便让网络学习旋转伪不变性。此外, 观察图的尺寸应谨慎选择, 随着尺寸的增加, 观察图将变得稀疏。Ikehata 凭经验发现, 当图像数量少于 1000 时, 观察图尺寸取 32 时性能最佳。

由于 CNN-PS 输入数据的尺寸相对较小, 因此

CNN 体系结构的选择并不是 CNN-PS 中至关重要的问题。Ikehata 使用的网络结构是 DenseNet^[101] 网络的一种变体 (具体结构亦可参见图 13 中的 normal estimation network 部分)。给定观察图, 经过训练后的网络可以对表面法向量进行精确预测, 如图 12 所示。

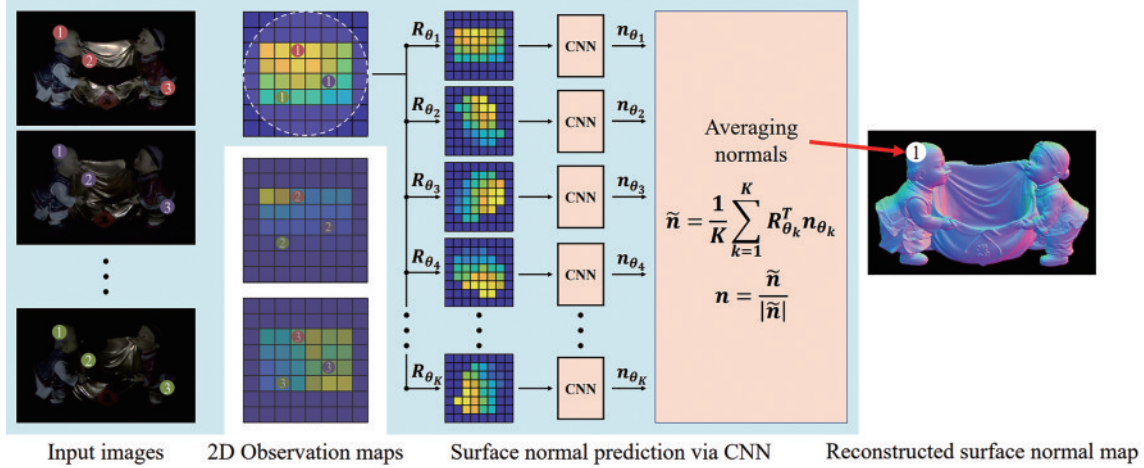


图 12 CNN-PS 预测模块示意图^[79]

Fig. 12 The framework of the prediction module of CNN-PS^[79]

实际中, 由于不可能训练出理想的旋转伪不变网络, 因此不同旋转程度的观察图预测的表面法向量并不相同。为了进一步强化旋转伪不变性, Ikehata 通过图 12 所示的特定角度 $\theta \in \theta_1, \theta_2, \dots, \theta_k$ 旋转光照方向来扩充输入数据, 对每个像素进行法向量预测, 然后将不同角度下的预测结果取平均, 得到最终的预测结果。实验结果表明, CNN-PS 对表面法向量的重建效果非

常好 (具体重建结果可参见表 3), 特别适用于非凸和非 Lambert 表面。此外, Ikehata 还创建了一个名为 CyclesPS 的合成光度立体图像数据集 (具体情况可参见表 1), 该数据集中的 3D 数据模型几何结构复杂, 包含了投射阴影和互反射在内的各种复杂场景, 为后续相关研究提供了良好的训练数据。

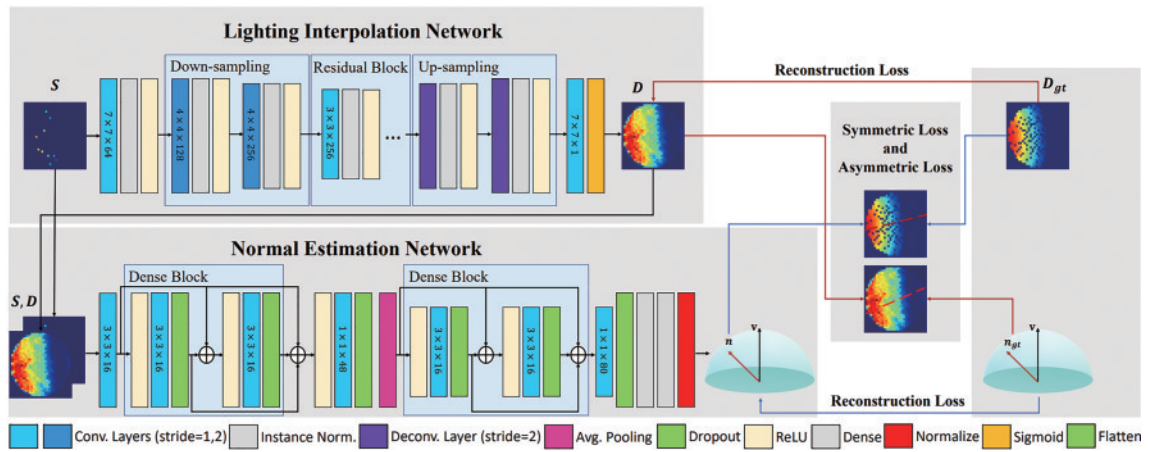


图 13 SPLINE-Net 架构^[80]

Fig. 13 The architecture of SPLINE-Net^[80]

4.2.4 SPLINE-Net 和 LMPS

通过对 CNN-PS 的研究可知, 一般情况下得到的观察图是稀疏的, 若将稀疏的观察图直接输入网络中会导致表面法向量预测精度的显著下降。因此, 需要采取一定的措施对稀疏的观察图进行处理, 而 CNN-

PS 使用各向同性约束下的旋转伪不变性对训练集进行扩充。为了解决稀疏光度立体 (sparse photometric stereo) 问题, 在观察图思想的基础上, Zheng 等^[80] 和 Li 等^[75] 分别提出了 SPLINE-Net 方法和 LMPS 方法, 利用插值回归和连接表的方式将观察图的方法应用到稀

疏输入的情况,使其在较少的光度立体图像输入下依然能够取得不错的重建精度^[66](具体重建结果可参见表 3)。

但实际上,这两种方法对该问题的解决思路是不同的,如图 13 所示:SPLINE-Net 首先建立了光照插值网络(LI-Net),将稀疏的观察图视为损坏的绘画处理,通过图像修复(inpainting)的方法生成密集的观察图;然后,将稀疏的观察图和由 LI-Net 生成的密集观察图同时送入法向量预测网络(NE-Net)进行表面法向量的预测。需要注意的是,NE-Net 与 CNN-PS 的网络架构保持一致,因为该网络能够有效地通过观察图进行表面法向量的预测。同时,为了进一步减小密集观察图的光照插值的求解空间,以促进对表面法向量的精确估计,SPLINE-Net 提出了对称和非对称损失函数(symmetric and asymmetric loss functions)。对于各向同性 BRDF 而言:在没有全局光照效应的情况下,观察图会呈现出对称的样式,此时可引入对称损失函数;当存在全局光照效应(投射阴影、互反射)的情况时,观察图样式的对称性会被破坏,此时可引入非对称损失函数^[86]。此外,LI-Net 和 NE-Net 的损失函数均由重建损失函数、对称损失函数和非对称损失函数构成。

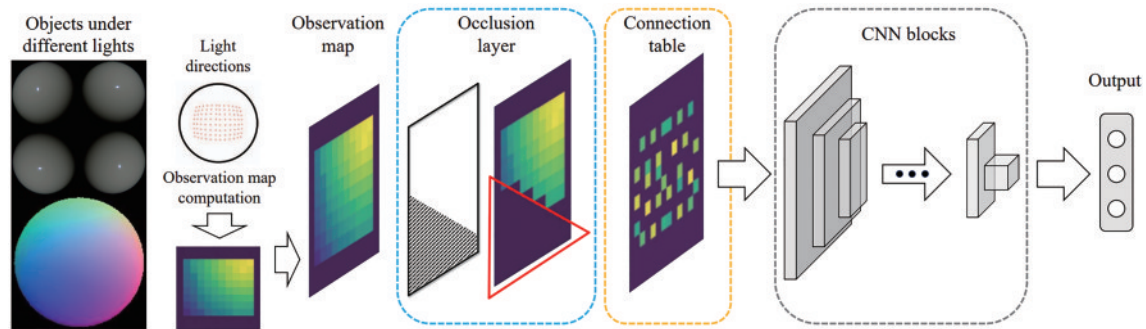


图 14 LMPS 架构^[75]

Fig. 14 The architecture of LMPS^[75]

4.2.5 DR-PSN

前面着重介绍了几种深度学习技术在光度立体中的应用,绝大多数都止步于表面法向量的重建,没有将预测的表面法向量与图像重建结合起来提供进一步的监督线索。2021 年, Ju 等^[77]将重建物体表面法向量与重建光度立体图像两个任务相关联,提出一种应用于光度立体问题的双重回归网络(DR-PSN)。该网络不仅能够实现表面法向量的预测,还能够从预测的表面法向量中实现光度立体图像的重建,并结合表面法向量约束和重建图像约束减小表面法向量的潜在学习空间^[66],使得表面法向量的重建精度得到提高。

DR-PSN 由法向量回归网络(normal regression task)和双重回归网络(dual regression task)两部分构成,如图 15 所示。法向量回归网络用于重建物体表面的法向量,由特征提取器、最大池化层和回归器构成,这与 PS-FCN 很相似。由于特征聚合部分使用的

LMPS 基于连接表(connection table)的思想,通过合适的损失函数和排序选择过程,训练好的连接表可以选择与表面法向量预测过程最相关的光照方向,在该光照方向下采用逐像素的方式对表面法向量进行预测。需要注意的是,与 CNN-PS 中观察图的生成采用笛卡尔坐标方式不同,LMPS 中观察图的生成采用球面坐标方式,并且观察图的尺寸变为 14。为了解决投射阴影的问题,LMPS 中还采用了一种新的处理方式,即遮挡层(occlusion layer)。与 DPSN^[69]中的阴影层将一些输入数据随机设置为零的思路不同, Li 等^[75]注意到投射阴影的产生通常与光照方向有关,而不是随机产生的。具体做法是,遮挡层随机选择观察图的两侧,并在每一侧随机选择一个点。随后,连接这两个点的直线将观察图分成两个区域,将这两个区域中较小的区域设置为零,以此来模拟投射阴影的情况。使用上述遮挡层训练网络可以有效地学习投射阴影的样式并能更加鲁棒地预测表面法向量。将经过遮挡层“过滤”处理后的观察图输入连接表,连接表匹配与表面法向量预测过程最相关的光照方向,并通过逐元素乘积得到稀疏观察图,作为 CNN(采用 DenseNet^[101]的一种变体)的输入对表面法向量进行预测,具体过程如图 14 所示。

也是最大池化技术,因此回归网络的输入端可以是任意数量的光度立体图像,在处理物体表面的投射阴影区域时也拥有更好的效果。双重回归网络由一个特征提取器和一个特征回归器构成,用于重建光度立体图像。可以注意到,在双重回归网络中,特征提取器最后一层卷积层后拼接了来自法向量回归网络中的最大池化层的聚合特征。这是因为双重回归网络可以看作是一个从 3D 数据模型(法向量)到 2D 图像(光度立体图像)的渲染过程。如前所述,想要渲染一幅图像必须知道物体表面材质(BRDF)和光照方向。因此,在双重回归网络特征提取器中,将法向量回归网络的聚合特征拼接过来为后续光度立体图像的重建提供了物体表面的材质信息^[66]。重建后的光度立体图像可以使整个网络形成闭环(closed-loop),为表面法向量的预测提供额外的监督,进一步提高表面法向量的预测精度。

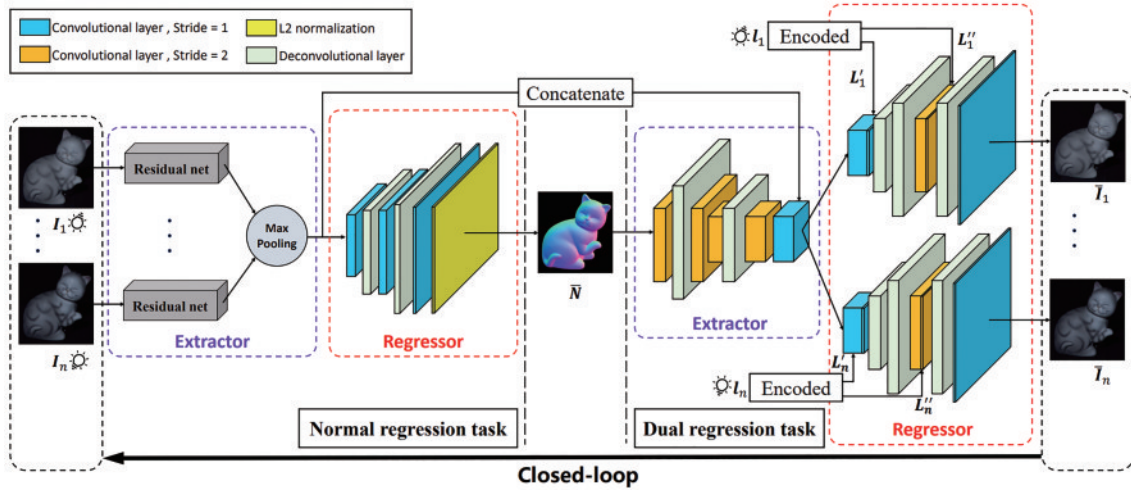


图 15 DR-PSN 架构^[77]

Fig. 15 The architecture of DR-PSN^[77]

DR-PSN 是目前唯一一个采用双重回归网络的深度学习光度立体方法,创新性地运用深度学习技术完成光度立体图像的重建,并且进一步提高了表面法向量预测的精确性,也给后续对提高法向量预测精度的研究提供了除增加网络结构复杂性之外的另一条道路,即引入额外的辅助监督。经过大量的实验验证,DR-PSN 对表面法向量的重建效果优于传统的和几乎所有基于深度学习的光度立体方法(具体重建结果可参见表 3)。而且,DR-PSN 在解决稀疏光度立体、具有复杂结构和投射阴影情况、任意光照方向下生成指定的重建图像等问题上都有很好的表现。但是,这种全新的双重回归网络也存在一些不足,如在密集的镜面反射区域会产生较大的重建图像误差、无法用于未标定光度立体。

4.2.6 IRPS

近些年来,几乎所有应用于光度立体 3D 重建领域的深度学习技术都采用监督学习(supervised learning)

的方式,该方式需要使用大量的真实可靠数据进行训练。然而,实际拍摄场景中大规模准确测量物体的表面法向量非常困难,因为需要高精度的 3D 几何形状来可靠地计算物体表面梯度,所以基于监督学习的方法都采用合成光度立体图像数据集训练(具体情况可参见表 1)。为了解决这一问题,Taniai 等^[114]提出一种名为 IRPS 的适于通用 BRDF 光度立体的 CNN,该网络基于无监督学习(unsupervised learning)方式,给定观测到的光度立体图像和相应的光照方向,IRPS 会根据图像逆向渲染(inverse render)物体表面法向量和表面材质的 BRDFs,将其输入图像辐照度方程(或渲染方程)中重建观测到的图像。合成图像用于定义重建损失(reconstruction loss),以此实现无监督学习。

IRPS 由光度立体网络(PSNet)和图像重建网络(IRNet)构成,如图 16 所示。将观测图像分别输入 PSNet 和 IRNet,PSNet 从观测图像中重建物体表面的法向量,IRNet 预测出与投射阴影和 BRDFs 相关的

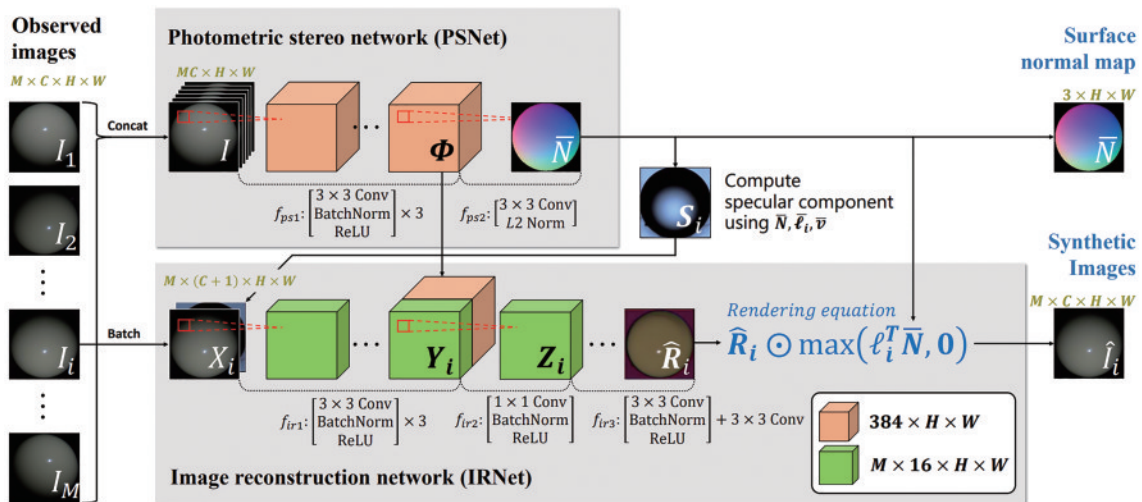


图 16 IRPS 架构^[114]

Fig. 16 The architecture of IRPS^[114]

\hat{R}_i , 并同预测的表面法向量一起输送到渲染方程 (rendering equation), 最终输出合成图像。通过最小化观测到的图像和输出的合成图像之间的重建损失函数来优化网络权重, 从而实现不需要法向量真值的无监督学习。此外, 学习过程是在测试阶段直接在测试场景上进行的, 无需进行任何预训练。

IRPS 相较于其他基于监督学习的网络有着不错的表现。然而, 其缺点也较为明显, 如对于复杂场景 (如 harvest) 的表面法向量重建效果不好 (具体重建结果可参见表 3), 主要原因是此场景存在强烈的投射阴影以及互反射。除此之外, 无监督的学习方式尽管无需真值数据, 但代价是更加庞大和复杂的计算, 较基于监督学习的方式运行速度也更缓慢。由于实际拍摄场景中很难大规模获取真值数据, 因此, 基于无监督学习的光度立体方法仍然会是未来研究的一大主流方向。2021 年, Kaya 等^[45]对 IRPS 进行了扩展, 使其可以应用于未标定的光度立体问题。但是, IRPS 中存在的问题依然无法得到有效解决。目前, 针对基于无监督学习的光度立体 3D 重建问题的研究仍然相对较少, 相信其在未来会有更大的发展潜力。

5 结论与展望

伴随着人工智能与深度学习技术的蓬勃发展以

及光度立体数据集的不断完善, 光度立体技术的发展进入了一个崭新的时代。由于使用强大的深度神经网络, 结合大规模的训练数据, 同时利用 Dropout、最大池化、观察图等技术手段, 与传统的方法相比, 基于深度学习的光度立体方法能够处理复杂 BRDF、存在投射阴影以及互反射等实际拍摄的场景, 因而适用性更强。

表 3 为不同基于深度学习的光度立体方法对 DiLiGenT 标准数据集^[62-63]重建结果的比较, 所有结果以平均角度误差作为评价标准, 关于传统方法对 DiLiGenT 标准数据集重建结果的比较可参见文献 [115]。可以看出与传统的方法相比, 基于深度学习的光度立体方法重建精度更高。进一步可以发现, 在具有较复杂 3D 几何形状的对象 (如 buddha、harvest、reading) 和较强烈的非 Lambert 反射 (如 cow、goblet、harvest、reading) 时, 基于深度学习的光度立体方法比传统的方法通常具有更好的性能, 充分体现了深度学习技术强大的泛化能力。然而, 另一方面也可以看出, 大多数基于深度学习的光度立体方法在 3D 几何形状简单、接近 Lambert 反射的物体上 (如 ball) 性能表现一般, 可能是庞大的神经网络结构以及拥有复杂 BRDF 的训练数据集导致的。

表 3 不同方法对 DiLiGenT 标准数据集^[62-63]重建结果的比较

Table 3 Comparison of reconstruction results for different methods on the DiLiGenT benchmark dataset^[62-63]

Method	ball	bear	buddha	cat	cow	goblet	harvest	pot1	pot2	reading	average
BASELINE ^[7]	4.10	8.39	14.92	8.41	25.60	18.50	30.62	8.89	14.65	19.80	15.39
DPSN ^[69]	2.02	6.31	12.68	6.54	8.01	11.28	16.86	7.05	7.86	15.51	9.41
PS-FCN ^[72]	2.82	7.55	7.91	6.16	7.33	8.60	15.85	7.13	7.25	13.33	8.39
UPS-FCN ^[72]	6.62	11.23	15.87	14.68	11.91	20.72	27.79	13.98	14.19	23.26	16.02
SDPS-Net ^[73]	2.77	6.89	8.97	8.06	8.48	11.91	17.43	8.14	7.50	14.90	9.51
CNN-PS ^[79]	2.2	4.1	7.9	4.6	8.0	7.3	14.0	5.4	6.0	12.6	7.2
SPLINE-Net ^{[80]*}	4.96	5.99	10.07	7.52	8.80	10.43	19.05	8.77	11.79	16.13	10.35
LMPS ^{[75]*}	3.97	8.73	11.36	6.69	10.19	10.46	17.33	7.30	9.74	14.37	10.02
DR-PSN ^[77]	2.27	5.46	7.84	5.42	7.01	8.49	15.40	7.08	7.21	12.74	7.90
IRPS ^[114]	1.47	5.79	10.36	5.44	6.32	11.47	22.59	6.09	7.76	11.03	8.83

Note: *there are 10 light directions for SPLINE-Net^[80] and LMPS^[75] methods and 96 light directions for other methods.

综上所述, 基于深度学习的光度立体方法相较于传统的方法拥有众多的优势, 并且在这一领域逐渐取得了支配地位, 但是也存在如下问题:

1) 目前, 绝大多数应用于光度立体领域的深度神经网络只能用于处理标定的光度立体问题, 无法处理未标定的光度立体问题或对未标定情况下的重建效果不好。在实际应用中, 想要完全获取光源几何 (方向或位置) 与强度信息有时是很困难的, 因此, 针对能更好处理未标定光度立体问题的深度学习模型的研究是很有必要的。

2) 基于监督学习的深度学习技术在光度立体领域

已经得到了长足的发展, 但是需要大量物体表面法向量的真值作为训练数据。基于无监督学习的方式可以克服监督学习存在的这一问题, 但需要更加庞大的计算量且会导致运行速度变慢。如何优化这两种不同的方式以及能否将二者进一步结合也是未来主要的研究方向。

3) 对于基于深度学习的光度立体 3D 重建技术来说, 训练和测试数据集是非常重要的。随着该技术的不断发展, 目前现有的数据集已逐渐无法满足该技术的需要。训练样本不足会在一定程度上降低深度学习模型的准确性; 较少的测试场景 (如 DiLiGenT

数据集^[62-63]仅有 10 个物体)也不能完全准确地评估表面法向量重建的效果。对数据集的扩充同样是一个热门的研究方向。

4)在处理非 Lambert 光度立体问题上,深度学习技术的表现明显优于传统方法,并且已经实现了巨大的突破,如阴影层、最大池化、观察图和遮挡层都能很好地处理一般的投射阴影问题。但在更加复杂的场景(如 DiLiGenT 数据集^[62-63]中的 harvest 和 reading),表面法向量的重建效果并不是很好,适合复杂场景以及强烈非 Lambert 反射问题的新方法是深度学习在光度立体领域的发展趋势。

作者综述了近年来深度学习技术在光度立体 3D 重建中的应用情况。首先介绍了光度学 3D 重建的研究背景和基本原理,其次概述了光度立体 3D 重建方法的类型,接着介绍了光度立体中常用的合成与实际拍摄数据集,然后着重阐述了深度学习技术在光度立体 3D 重建中的应用,最后进行了分析与总结,指出了深度学习技术在光度立体领域面临的挑战以及未来的研究趋势。总体而言,深度学习为解决光度立体 3D 重建问题提供了新的技术手段,并取得了显著效果,但其仍然存在不少问题,未来基于深度学习的光度立体 3D 重建仍然是一个值得重点研究的热门方向。

参 考 文 献

- [1] 王国琿, 张璇. 透视投影下混合表面 3D 重建的快速 SFS 算法[J]. 光学学报, 2021, 41(12): 1215003.
Wang G H, Zhang X. Fast shape-from-shading algorithm for 3D reconstruction of hybrid surfaces under perspective projection[J]. Acta Optica Sinica, 2021, 41(12): 1215003.
- [2] 王国琿, 宋玉贵. 一种非朗伯表面 SFS 的快速粘性解算法[J]. 仪器仪表学报, 2015, 36(7): 1577-1583.
Wang G H, Song Y G. A fast viscosity solution algorithm of shape from shading for non-Lambertian surfaces[J]. Chinese Journal of Scientific Instrument, 2015, 36(7): 1577-1583.
- [3] Horn B K P. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view[D]. Cambridge: Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1970.
- [4] Horn B K P. Robot vision[M]. Wang L, Jiang X L, Transl. Beijing: China Youth Publishing Group, 2014.
- [5] Zhang R, Tsai P S, Cryer J E, et al. Shape from shading: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690-706.
- [6] Durou J D, Falcone M, Sagona M. Numerical methods for shape-from-shading: a new survey with benchmarks [J]. Computer Vision and Image Understanding, 2008, 109(1): 22-43.
- [7] Woodham R J. Photometric method for determining surface orientation from multiple images[J]. Optical Engineering, 1980, 19(1): 139-144.
- [8] Cao Y L, Ding B J, Chen J X, et al. Photometric-stereo-based defect detection system for metal parts[J]. Sensors, 2022, 22(21): 8374.
- [9] Liu H Y, Yan Y H, Song K C, et al. Efficient optical measurement of welding studs with normal maps and convolutional neural network[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 5000614.
- [10] Bernotas G, Scorza L C T, Hansen M F, et al. A photometric stereo-based 3D imaging system using computer vision and deep learning for tracking plant growth[J]. GigaScience, 2019, 8(5): giz056.
- [11] Hao Y, Li J, Meng F, et al. Photometric stereo-based depth map reconstruction for monocular capsule endoscopy[J]. Sensors, 2020, 20(18): 5403.
- [12] Kim J, Al Faruque H, Kim S, et al. Multimodal endoscopic system based on multispectral and photometric stereo imaging and analysis[J]. Biomedical Optics Express, 2019, 10(5): 2289-2302.
- [13] Sun J A. Evaluation on an optical scanning device for skin profile measurement[J]. Journal of Cosmetic Science, 2018, 69(1): 9-17.
- [14] Chen Z, Ji Y, Zhou M Y, et al. 3D face reconstruction using color photometric stereo with uncalibrated near point lights[C]//2020 IEEE International Conference on Computational Photography (ICCP), April 24-26, 2020, St. Louis, MO, USA. New York: IEEE Press, 2020.
- [15] Cheng K H M, Kumar A. Contactless biometric identification using 3D finger knuckle patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 1868-1883.
- [16] Yeh C K, Matsuda N, Huang X, et al. A streamlined photometric stereo framework for cultural heritage[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9913: 738-752.
- [17] Jian M W, Dong J Y, Gong M G, et al. Learning the traditional art of Chinese calligraphy via three-dimensional reconstruction and assessment[J]. IEEE Transactions on Multimedia, 2020, 22(4): 970-979.
- [18] Liu W C, Wu B, Wöhler C. Effects of illumination differences on photometric stereo shape-and-albedo-from-shading for precision lunar surface reconstruction[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 136: 58-72.
- [19] Peng M, Di K C, Wang Y X, et al. A photogrammetric-photometric stereo method for high-resolution lunar topographic mapping using Yutu-2 rover images[J]. Remote Sensing, 2021, 13(15): 2975.
- [20] Borowiec S. AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol[EB/OL]. [2022-10-09]. <https://www.ft.com/content/f6b90460-aaa5-11e5-9fca-fb0f946df1f0>.
- [21] 冯世杰, 左超, 尹维, 等. 深度学习技术在条纹投影三维成像中的应用[J]. 红外与激光工程, 2020, 49(3): 0303018.
Feng S J, Zuo C, Yin W, et al. Application of deep learning technology to fringe projection 3D imaging[J]. Infrared and Laser Engineering, 2020, 49(3): 0303018.

- [22] Wang G H, Zheng H, Zhang X. A robust checkerboard corner detection method for camera calibration based on improved YOLOX[J]. *Frontiers in Physics*, 2022, 9: 819019.
- [23] Laga H, Jospin L V, Boussaid F, et al. A survey on deep learning techniques for stereo-based depth estimation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(4): 1738-1764.
- [24] Zheng H, Wang G H, Li X C. Identifying strawberry appearance quality by vision transformers and support vector machine[J]. *Journal of Food Process Engineering*, 2022, 45(10): 14132.
- [25] Zheng H, Wang G H, Li X C. YOLOX-Dense-CT: a detection algorithm for cherry tomatoes based on YOLOX and DenseNet[J]. *Journal of Food Measurement and Characterization*, 2022, 16(6): 4788-4799.
- [26] Houssein E H, Mohamed R E, Ali A A. Machine learning techniques for biomedical natural language processing: a comprehensive review[J]. *IEEE Access*, 2021, 9: 140628-140653.
- [27] Otter D W, Medina J R, Kalita J K. A survey of the usages of deep learning for natural language processing [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(2): 604-624.
- [28] Li R J, Zeng T, Peng H C, et al. Deep learning segmentation of optical microscopy images improves 3-D neuron reconstruction[J]. *IEEE Transactions on Medical Imaging*, 2017, 36(7): 1533-1541.
- [29] 孙正, 王树雁. 深度学习在血管内光学相干层析成像中的应用现状[J]. *激光与光电子学进展*, 2022, 59(22): 2200002.
Sun Z, Wang S Y. Application of deep learning in intravascular optical coherence tomography[J]. *Laser & Optoelectronics Progress*, 2022, 59(22): 2200002.
- [30] 王飞, 王昊, 卞耀明, 等. 深度学习在计算成像中的应用[J]. *光学学报*, 2020, 40(1): 0111002.
Wang F, Wang H, Bian Y M, et al. Applications of deep learning in computational imaging[J]. *Acta Optica Sinica*, 2020, 40(1): 0111002.
- [31] 左超, 冯世杰, 张翔宇, 等. 深度学习下的计算成像: 现状、挑战与未来[J]. *光学学报*, 2020, 40(1): 0111003.
Zuo C, Feng S J, Zhang X Y, et al. Deep learning based computational imaging: status, challenges, and future[J]. *Acta Optica Sinica*, 2020, 40(1): 0111003.
- [32] Wang W H, Zhao X, Jiang Z X, et al. Deep learning-based scattering removal of light field imaging[J]. *Chinese Optics Letters*, 2022, 20(4): 041101.
- [33] Nguyen H, Novak E, Wang Z Y. Accurate 3D reconstruction via fringe-to-phase network[J]. *Measurement*, 2022, 190: 110663.
- [34] Boukhtache S, Abdelouahab K, Berry F, et al. When deep learning meets digital image correlation[J]. *Optics and Lasers in Engineering*, 2021, 136: 106308.
- [35] 王国琿, 韩九强, 张新曼, 等. 一种新的基于从明暗恢复形状的月球表面三维形状恢复算法[J]. *宇航学报*, 2009, 30(6): 2265-2269.
Wang G H, Han J Q, Zhang X M, et al. A new three-dimensional reconstruction algorithm of the lunar surface based on shape from shading method[J]. *Journal of Astronautics*, 2009, 30(6): 2265-2269.
- [36] 王国琿, 韩九强, 贾洪海, 等. 一种从明暗恢复形状的快速黏性解算法[J]. *西安交通大学学报*, 2009, 43(6): 43-47.
Wang G H, Han J Q, Jia H H, et al. A fast viscosity solution algorithm for shape from shading[J]. *Journal of Xi'an Jiaotong University*, 2009, 43(6): 43-47.
- [37] Horn B K P, Brooks M J. The variational approach to shape from shading[J]. *Computer Vision, Graphics, and Image Processing*, 1986, 33(2): 174-208.
- [38] Frankot R T, Chellappa R. A method for enforcing integrability in shape from shading algorithms[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1988, 10(4): 439-451.
- [39] Agrawal A, Raskar R, Chellappa R. What is the range of surface reconstructions from a gradient field? [M]//Leonardis A, Bischof H, Pinz A. *Computer vision-ECCV 2006. Lecture notes in computer science*. Berlin: Springer, 2006, 3951: 578-591.
- [40] Harker M, O'Leary P. Regularized reconstruction of a surface from its measured gradient field[J]. *Journal of Mathematical Imaging and Vision*, 2015, 51(1): 46-70.
- [41] Quéau Y, Durou J D, Aujol J F. Variational methods for normal integration[J]. *Journal of Mathematical Imaging and Vision*, 2018, 60(4): 609-632.
- [42] Quéau Y, Durou J D, Aujol J F. Normal integration: a survey[J]. *Journal of Mathematical Imaging and Vision*, 2018, 60(4): 576-593.
- [43] Forsyth D, Ponce J. *Computer vision: a modern approach* [M]. Upper Saddle River: Prentice Hall, 2003.
- [44] Cho D, Matsushita Y, Tai Y W, et al. Semi-calibrated photometric stereo[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(1): 232-245.
- [45] Kaya B, Kumar S, Oliveira C, et al. Uncalibrated neural inverse rendering for photometric stereo of general surfaces[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 3803-3813.
- [46] 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.
- [47] Belhumeur P N, Kriegman D J, Yuille A L. The bas-relief ambiguity[J]. *International Journal of Computer Vision*, 1999, 35(1): 33-44.
- [48] Papadimitri 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.
- [49] Sengupta S, Zhou H, Forkel W, et al. Solving uncalibrated photometric stereo using fewer images by jointly optimizing low-rank matrix completion and integrability[J]. *Journal of Mathematical Imaging and Vision*, 2018, 60(4): 563-575.
- [50] Boyd S, Parikh N, Chu E, et al. *Distributed*

- optimization and statistical learning via the alternating direction method of multipliers[J]. *Foundations and Trends in Machine Learning*, 2011, 3(1): 1-122.
- [51] Cho D, Matsushita Y, Tai Y W, et al. Photometric stereo under non-uniform light intensities and exposures [M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9906: 170-186.
- [52] Quéau Y, Wu T, Lauze F, et al. A non-convex variational approach to photometric stereo under inaccurate lighting[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 350-359.
- [53] Zhang Y F, Yang Y H. Illuminant direction determination for multiple light sources[C]//Proceedings IEEE Conference on Computer Vision and Pattern Recognition, June 15, 2000, Hilton Head Island, SC. New York: IEEE Press, 2002: 269-276.
- [54] Zhou W, Kambhampettu C. Estimation of illuminant direction and intensity of multiple light sources[M]//Heyden A, Sparr G, Nielsen M, et al. *Computer vision-ECCV 2002. Lecture notes in computer science*. Berlin: Springer, 2002, 2353: 206-220.
- [55] Powell M W, Sarkar S, Goldgof D. Calibration of light sources[C]//Proceedings IEEE Conference on Computer Vision and Pattern Recognition, June 15-15, 2000, Hilton Head, SC, USA. New York: IEEE Press, 2002: 263-269.
- [56] Hara K, Nishino K, Ikeuchi K. Light source position and reflectance estimation from a single view without the distant illumination assumption[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(4): 493-505.
- [57] Park J, Sinha S N, Matsushita Y, et al. Calibrating a non-isotropic near point light source using a plane[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 2267-2274.
- [58] Visentini-Scarzanella M, Kawasaki H. Simultaneous camera, light position and radiant intensity distribution calibration[C]//Image and Video Technology: 7th Pacific-Rim Symposium, November 25-27, 2015, Auckland, New Zealand, New York: ACM Press, 2015: 557-571.
- [59] Quéau Y, Durix B, Wu T, et al. LED-based photometric stereo: modeling, calibration and numerical solution[J]. *Journal of Mathematical Imaging and Vision*, 2018, 60(3): 313-340.
- [60] Xie L M, Song Z, Jiao G H, et al. A practical means for calibrating an LED-based photometric stereo system[J]. *Optics and Lasers in Engineering*, 2015, 64: 42-50.
- [61] 刘嘉缘, 王国琿. 基于立体视觉的光度立体 LED 光源位置参数标定方法[J]. *激光与光电子学进展*, 2022, 59(16): 1615001.
- Liu J Y, Wang G H. Stereo vision-based calibration approach for position parameters of LED in photometric stereo[J]. *Laser & Optoelectronics Progress*, 2022, 59(16): 1615001.
- [62] Shi B X, Wu Z, Mo Z P, et al. A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3707-3716.
- [63] Shi B X, Mo Z P, 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.
- [64] Mecca R, Quéau Y. Unifying diffuse and specular reflections for the photometric stereo problem[C]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV), March 7-10, 2016, Lake Placid, NY, USA. New York: IEEE Press, 2016.
- [65] Mecca R, Quéau Y, Logothetis F, et al. A single-lobe photometric stereo approach for heterogeneous material [J]. *SIAM Journal on Imaging Sciences*, 2016, 9(4): 1858-1888.
- [66] 举雅琨. 非朗伯光度立体的深度学习模型[D]. 青岛: 中国海洋大学, 2022.
- Ju Y K. Depth learning model of non-lambert photometric stereo[D]. Qingdao: Ocean University of China, 2022.
- [67] Cycles open source production rendering[EB/OL]. [2022-12-11]. <https://www.cycles-renderer.org/>.
- [68] Mitsuba 3 physically based renderer[EB/OL]. [2022-12-11]. <https://www.mitsuba-renderer.org/>.
- [69] Santo H, Samejima M, Sugano Y, et al. Deep photometric stereo network[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 501-509.
- [70] 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*, 2022, 44(1): 114-128.
- [71] Santo H, Waechter M, Matsushita Y. Deep near-light photometric stereo for spatially varying reflectances[M]//Vedaldi A, Bischof H, Brox T, et al. *Computer vision-ECCV 2020. Lecture notes in computer science*. Cham: Springer, 2020, 12353: 137-152.
- [72] Chen G Y, Han K, Wong K Y K. PS-FCN: a flexible learning framework for photometric stereo[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11213: 3-19.
- [73] Chen G Y, Han K, Shi B X, et al. Self-calibrating deep photometric stereo networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 8731-8739.
- [74] Chen G Y, Han K, Shi B X, et al. Deep photometric stereo for non-Lambertian surfaces[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 129-142.
- [75] Li J X, Robles-Kelly A, You S D, et al. Learning to

- minify photometric stereo[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 7560-7568.
- [76] Ju Y K, Qi L, He J C, et al. MPS-Net: learning to recover surface normal for multispectral photometric stereo[J]. *Neurocomputing*, 2020, 375: 62-70.
- [77] Ju Y K, Dong J Y, Chen S. Recovering surface normal and arbitrary images: a dual regression network for photometric stereo[J]. *IEEE Transactions on Image Processing*, 2021, 30: 3676-3690.
- [78] Ju Y K, Shi B X, Jian M W, et al. NormAttention-PSN: a high-frequency region enhanced photometric stereo network with normalized attention[J]. *International Journal of Computer Vision*, 2022, 130(12): 3014-3034.
- [79] Ikehata S. CNN-PS: CNN-based photometric stereo for general non-convex surfaces[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018*. Lecture notes in computer science. Cham: Springer, 2018, 11219: 3-19.
- [80] Zheng Q, Jia Y M, Shi B X, et al. SPLINE-net: sparse photometric stereo through lighting interpolation and normal estimation networks[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8548-8557.
- [81] Johnson M K, Adelson E H. Shape estimation in natural illumination[C]//CVPR, June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 2553-2560.
- [82] Wiles O, Zisserman A. SilNet: single- and multi-view reconstruction by learning from silhouettes[C]//Proceedings of the British Machine Vision Conference 2017, September 4-7, 2017, London, UK. London: British Machine Vision Association, 2017: 1-13.
- [83] Curless B, Levoy M. A volumetric method for building complex models from range images[C]//Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, August 4-9, 1996, New Orleans, LA, USA. New York: ACM Press, 1996: 303-312.
- [84] Matusik W, Pfister H, Brand M, et al. A data-driven reflectance model[J]. *ACM Transactions on Graphics*, 2003, 22(3): 759-769.
- [85] Brent B. Physically-based shading at Disney[EB/OL]. [2022-10-02]. <https://disneyanimation.com/publications/physically-based-shading-at-disney/>
- [86] Zheng Q, Shi B X, Pan G. Summary study of data-driven photometric stereo methods[J]. *Virtual Reality & Intelligent Hardware*, 2020, 2(3): 213-221.
- [87] Alldrin N, Zickler T, Kriegman D. Photometric stereo with non-parametric and spatially-varying reflectance[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008.
- [88] Einarsson P, Chabert C F, Jones A, et al. Relighting human locomotion with flowed reflectance fields[C]//ACM SIGGRAPH 2006 Sketches on-SIGGRAPH '06, July 30-August 3, 2006, Boston, Massachusetts. New York: ACM Press, 2006: 183-194.
- [89] Xiong Y, Chakrabarti A, Basri R, et al. From shading to local shape[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(1): 67-79.
- [90] Li M, Zhou Z L, Wu Z, et al. Multi-view photometric stereo: a robust solution and benchmark dataset for spatially varying isotropic materials[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4159-4173.
- [91] Mecca R, Logothetis F, Budvytis I. et al. LUCES: a dataset for near-field point light source photometric stereo [EB/OL]. (2021-04-27)[2022-10-08]. <https://arxiv.org/abs/2104.13135>.
- [92] Logothetis F, Mecca R, Budvytis I, et al. A CNN based approach for the point-light photometric stereo problem [J]. *International Journal of Computer Vision*, 2023, 131(1): 101-120.
- [93] Ren J J, Wang F S, Zhang J H, et al. DiLiGenT10²: a photometric stereo benchmark dataset with controlled shape and material variation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 12571-12580.
- [94] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [95] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.
- [96] Williams R J, Zipser D. A learning algorithm for continually running fully recurrent neural networks[J]. *Neural Computation*, 1989, 1(2): 270-280.
- [97] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [98] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), December 3-6, 2012, Lake Tahoe, NV, USA. Canada: NIPS, 2012: 1097-1105.
- [99] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2022-10-09]. <https://arxiv.org/abs/1409.1556>.
- [100] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [101] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [102] Iwahori Y, Woodham R J, Tanaka H, et al. Neural network to reconstruct specular surface shape from its three shading images[C]//Proceedings of 1993 International

- Conference on Neural Networks (IJCNN-93-Nagoya, Japan), October 25-29, 1993, Nagoya, Japan. New York: IEEE Press, 2002: 1181-1184.
- [103] Rajaram K V, Parthasarathy G, Faruqi M A. A neural network approach to photometric stereo inversion of real-world reflectance maps for extracting 3-D shapes of objects[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1995, 25(9): 1289-1300.
- [104] Cheng W C. Neural-network-based photometric stereo for 3D surface reconstruction[C]//The 2006 IEEE International Joint Conference on Neural Network Proceedings, July 16-21, 2006, Vancouver, BC, Canada. New York: IEEE Press, 2006: 404-410.
- [105] Elizondo D, Zhou S M, Chrysostomou C. Surface reconstruction techniques using neural networks to recover noisy 3D scenes[M]//Kůrková V, Neruda R, Koutník J. *Artificial neural networks-ICANN 2008*. Lecture notes in computer science. Berlin: Springer, 2008, 5163: 857-866.
- [106] Wang X L, Fouhey D F, Gupta A. Designing deep networks for surface normal estimation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 539-547.
- [107] Choy C B, Xu D F, Gwak J, et al. 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016*. Lecture notes in computer science. Cham: Springer, 2016, 9912: 628-644.
- [108] Wang N Y, Zhang Y D, Li Z W, et al. Pixel2Mesh: generating 3D mesh models from single RGB images [M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018*. Lecture notes in computer science. Cham: Springer, 2018, 11215: 55-71.
- [109] Ikehata S, Wipf D, Matsushita Y, et al. Robust photometric stereo using sparse regression[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 318-325.
- [110] Nielsen J B, Jensen H W, Ramamoorthi R. On optimal, minimal BRDF sampling for reflectance acquisition[J]. *ACM Transactions on Graphics*, 2015, 34(6): 1-11.
- [111] Xu Z X, Nielsen J B, Yu J Y, et al. Minimal BRDF sampling for two-shot near-field reflectance acquisition[J]. *ACM Transactions on Graphics*, 2016, 35(6): 1-12.
- [112] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 24, 2016, New York: IEEE Press, 2016: 640-651.
- [113] Kim K, Gu J W, Tyree S, et al. A lightweight approach for on-the-fly reflectance estimation[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 20-28.
- [114] Tani ai T, Maehara T. Neural inverse rendering for general reflectance photometric stereo[EB/OL]. (2018-02-28)[2022-12-08]. <https://arxiv.org/abs/1802.10328>.
- [115] Summary of benchmarking results[EB/OL]. [2022-12-11]. <https://sites.google.com/site/photometricstereodata/single/summary-of-benchmarking-results>.