

文章编号 1004-924X(2025)15-2424-17

高频信息物体多层多元特征权重自适应 融合三维重建网络

王 标, 李 影, 融百川, 刘 璟, 张 进, 王永红^{*}
(合肥工业大学 仪器科学与光电工程学院, 安徽 合肥 230000)

摘要:为解决光度立体深度学习在实现物体表面法向量重建过程中存在的表面纹理的高频信息丢失而导致重建精度偏差大的问题,提出了一种多层多元特征权重自适应融合三维重建网络(MMF-Net)。网络主要将PS-FCN作为基准模型,引入对称式编码-解码结构,增强网络的学习和特征表达能力,提升了不同层级之间特征整合能力;设计独立层间权重自适应调节的多元卷积层,通过增加额外的可训练权重,兼顾具有形状信息与纹理信息,并且能够更好地捕捉到表面纹理的细节变化信息,使得该网络在高频信息较密集场景下的表现更加稳定和准确;辅助增加跳跃连接结构,通过中间层特征跨层连接至后续层级,保留物体高频信息且强化低频信息,以实现物体高低频特征信息的融合性应用。利用DiLiGenT基准数据集进行了相关测试,实验结果表明,MMF-Net能够实现平均MAE达到 6.94° ,对比PS-FCN(Norm)的 7.39° 提升了6%,在其中两幅含有高频信息物体的平均重建误差为 11.03° ,对比先前方法FUPS-Net的 12.52° 提升了12%。MMF-Net实现了光度立体物体表面低频信息和高频信息的有效性获取,为以物体表面法向量为基础的三维高精度重建提供参考。

关键词:深度学习;光度立体视觉;多元卷积;特征融合;自适应权重

中图分类号:TP394.1;TP183 **文献标识码:**A

doi:10.37188/OPE.20253315.2424

CSTR:32169.14.OPE.20253315.2424

Multi-layer multi-feature adaptive weight fusion network for 3D reconstruction of objects with high-frequency information

WANG Biao, LI Ying, RONG Baichuan, LIU Jing, ZHANG Jin, WANG Yonghong^{*}

(School of Instrument Science and Opto-electronics Engineering, Hefei University of
Technology, Hefei 230000, China)

^{*} Corresponding author, E-mail: yhwang@hfut.edu.cn

Abstract: To mitigate the loss of high-frequency surface texture information and the resulting reduction in reconstruction accuracy in deep learning-based photometric stereo, a multi-layer multi-element feature weight adaptive fusion 3D reconstruction network (MMF-Net) is proposed. The network architecture builds on PS-FCN and incorporates a symmetric encoder-decoder to enhance feature learning, representation capacity, and multi-level feature integration. A novel multi-element convolution layer with independent inter-layer adaptive weights is introduced; by incorporating additional trainable weights, both shape

收稿日期:2025-05-28;修订日期:2025-06-25.

基金项目:国家自然科学基金资助项目(No. 52375536);长三角科技创新共同体联合攻关资助项目(No. 2022CS-JGG1300);合肥市关键技术研发揭榜挂帅资助项目(No. 2023SGJ020)

and texture cues are jointly leveraged to better capture fine surface texture variations, thereby improving stability and accuracy in scenes containing dense high-frequency information. An auxiliary skip-connection mechanism is also employed to propagate intermediate-layer features to later stages, preserving high-frequency details while reinforcing low-frequency structure, and enabling effective fusion of multi-band (high- and low-frequency) surface information. The method was evaluated on the DiLiGenT benchmark. MMF-Net attains an average mean angular error (MAE) of 6.94° , representing a 6% improvement over PS-FCN (Norm) at 7.39° . For objects exhibiting pronounced high-frequency surface detail, the average reconstruction error is 11.03° , a 12% improvement relative to FUPS-Net at 12.52° . The results demonstrate that MMF-Net effectively captures both low- and high-frequency surface information in photometric stereo, offering a viable approach for high-precision 3D reconstruction from surface normals.

Key words: deep learning; photometric stereo vision; multivariate convolution; feature fusion; adaptive weighting

1 引 言

物体表面法向量的准确获取在计算机视觉、三维重建、机器人视觉以及虚拟现实等领域具有极其重要的应用价值。随着科技的不断进步,对于复杂物体表面细节精确感知的需求日益增长,尤其是在处理高频信息表面和多材质物体时,传统的法向量预测方法面临着诸多挑战。当固定相机在不同光线方向下拍摄静态物体的多幅图像时,可利用光度立体技术估算物体的表面法线。自光度立体理论诞生以来,因其低成本且能快速有效提取被测物体表面几何形状信息等优势,备受中外学者广泛关注。

从明暗恢复形状(Shape Recovery From Shading, SFS)的方法由 Horn 在 1970 年提出^[1-2],引入图像辐照度方程,结合物体表面反射系数,在单张图像中代入各点光亮度数值,通过表面可微分性、曲率约束和光滑度约束等条件,求解各点深度信息。然而,图像辐照度方程本质上是一个含两个未知量的非线性偏微分方程,仅靠方程无法得出正确解。因此,为消除病态、简化问题,SFS 方法做出两个假设:光源为无限远点光源;物体反射模型为朗伯体表面反射模型。但最终结果仍不理想。

针对传统 SFS 方法的不适定问题,Woodham^[3]等通过引入多光源约束条件建立了光度立体数学模型。该方法通过线性代数方法联立求解,当物体表面在 $m \geq 3$ 个独立方向光源下获得观测图像集时,构建的辐照度方程组可解析出表面梯度场的唯一解。

基于 Woodham 的研究成果,研究人员提出了许多光度立体方法,按光源方向和强度信息是否已知大致分为三类^[4-5]。未标定(光源方向、强度均未知)、半标定(光源几何信息已知、强度未知)和标定(光源方向、强度均已知)光度立体方法。对于未标定光度立体技术,Hayakawa^[6]等提出基于奇异值分解的方法,但存在法向量线性歧义。Papadimitriou 等^[7]利用局部漫反射最大值求得参数空间闭合解,兼具简单性与稳定性。Sengupta 等^[8]结合秩约束与可积性约束,通过 AD-MM 算法^[9]解决 4-6 光源下的参数歧义和数据丢失问题。对于半标定光度立体,Cho 等^[4,10]在最小二乘框架下用交替最小化处理非均匀光源及自动曝光场景。Quéau 等^[11]通过鲁棒交替加权最小二乘法,在变分框架中解决光照方向不精确问题。Zhang 等^[12]利用已知尺寸 Lambert 球与关键点信息标定光源方向和强度。Powell 等^[13]通过 3 个相对位置已知的反射球实现标定。Hara 等^[14]结合 Lambert 与 Torrance-Sparrow 模型拟合光源位置与反射参数。Quéau 等^[15-16]提出两步标定法提升效果。上述传统光度立体方法通常假定被测物体为 Lambertian 体,但现实世界中很多物体并遵循 Lambertian 体反射规律,故研究人员提出大致四类方法对非 lambertian 体光度立体进行处理,即基于异常值拒绝的方法,基于复杂反射率模型的方法,基于样本的方法和基于学习的方法。基于异常值拒绝通过识别排除光度图像异常值处理光照不均,如基于秩最小化^[17]、期望值最大化^[18]、稀疏贝叶斯回归^[19]等方法,但需大量

输入图像,仅能粗略估计表面形状。基于复杂反射率模型考虑复杂光反射特性提升估计精度,采用 Blinn-Phong 模型, Cook-Torrance 模型^[20]等分析模型,可模拟高光、阴影等,但对光照和场景条件要求高,适用材料有限。基于范例的方法依赖反射度测量、BRDF 数据等样本数据,如 Hertzmann 和 Seitz^[21]等利用参考球体将问题转化为点匹配,能准确模拟真实材质,但受样本库限制,泛化性不足。近年来,在人工智能持续创新的驱动下,以深度学习为代表的前沿技术正深度渗透至光学传感领域,尤其在光场成像、计算式光学重建及高精度光学检测等领域实现了技术革新。目前在光度立体视觉和深度学习的融合领域,前人已经开展了大量的研究工作,将传统的光度视觉方法与深度学习技术相结合,以克服传统方法的局限性,并实现更精确和稳定的三维重建和深度估计。

在光度立体三维重建的深度学习方法探索中,Santo 团队^[22]于 2017 年取得突破性进展,设计出首个监督式深度网络 DPSN。该方法结合数据驱动的 BRDF^[23-24]可对物体表面材质的 BRDF 进行预测。但由于网络训练中对光照方向信息利用不充分^[25],DPSN 在重建过程中其算法设计基于固定光照方向(96 个固定方向)的先验假设,从而对非预设光照条件下的三维重建精度无法保证,一定程度上限制了实际场景应用。为突破这一限制,Chen 团队^[26]于 2018 年开发了基于全卷积架构(FCN^[27])的 PS-FCN 模型,通过特征通道融合技术将彩色图像信息(RGB)与空间光照向量(XYZ)联合编码为 6 维输入特征。该设计有效突破了传统方法对预设光源方向的依赖性,在提升未知光照场景适应性的同时,显著增强了复杂反射物体(如非 Lambertian 体)表面几何重建的稳定性。但 PS-FCN 大量使用 3×3 大小的卷积核提取特征,可能过度平滑图像特征,从而导致部分物体预测时丢失细节信息。同年,Ikehata 提出基于 CNN 的 CNN-PS^[28]方法,首次引入“观察图”概念,在光照方向抑制条件下,通过观测值逐像素半球面投影,克服了固定数量光度立体图像输入的限制。

Wang 等为了提升光度立体网络在具有微小凹痕或突起的反射表面上的预测精度,提出了 SDPS-M2CAN 网络模型^[29],通过采用多尺度特

征提取模块、自适应归一化策略等方法,网络模型能够更加准确地恢复出表面的微小结构。Fang 等则设计了一种基于自注意力和多特征融合的法线恢复网络^[30]用于帮助网络理解图像长距离像素之间的依赖关系。Yao 等提出了一种适用于真实照明场景的基于非结构化神经网络的深度光度立体模型^[31],该方法基于函数逼近理论实现,针对光度立体视觉在少量光照情形下的数据稀疏问题提出了一种解决方案。Liu 等提出一种新的多特征融合光度立体网络,即 MF-PSN^[32],专注于局部-全局和深浅层特征融合,有效地利用了每个输入图像的信息且有效提高了模型的准确性和鲁棒性。Ju 等提出了一种针对未知光方向下非 lambertian 体的单级深度未校准光度立体网络(UPS),即 FUPS-Net^[33],克服了两级 UPS 方法的局限性,提供了更好的训练稳定性、简洁的端到端结构,并避免了在脱节网络中累积的错误。

本文在研究前人工作的基础上,提出了一种用于含有高频信息物体三维重建的多层多元特征权重自适应融合网络(Multi-layer Multi-element feature weight adaptive Fusion 3D reconstruction network, MMF-Net)。Zhang 等^[34]种基于 OKG 算法的二级特征匹配方法 SURF-OKG,将图片比例空间划分为多个网格进行特征点自适应匹配处理,提高了三维重建精度。Deng 等^[35]提出了一种相位测量的不连续位置自适应分割方法,可用于高精度的三维重建。为提升网络架构对于含有高频信息输入的提取和重建能力,本文提出了一种多层多元卷积模块用来提取输入特征,并通过人为额外设置可训练的自适应权重调整每个卷积层中每个像素对于整体图像的贡献,以实现权衡。其次,还提出一种对称式编码-解码结构以提高网络的学习和表征能力,此外,通过辅助增加跳跃连接结构以注重保留高频信息和强化低频信息,为网络在处理细节时提供了更加全面和有效的能力。网络使用来自 blobby shape^[28]数据集和 sculpture shape^[29]数据集创建的合成数据集来训练网络。实验表明,MMF-Net 相较于单一尺度的卷积核结构,具有更好的准确性和稳定性,在实现含有高频信息物体表面法向量的重建方面表现出色。并且在 DiLiGenT^[30-31]基准数据集上进行消融实验,验证了方法的有效

性。此外,为了更加有力地证明 MMF-Net 的泛化性,本文在 DiLiGenT-MV^[36] 公开数据集以及自建的光度立体数据集上进行与基准模型 PS-FCN 的对比实验,证明了 MMF-Net 在面对不同情况具有优秀的稳定性和泛化性。

2 多层多元特征权重自适应融合网络设计

2.1 网络结构

PS-FCN 证明了最大池化操作可以解决顺序不可知的问题,一般来说,最大池化操作可以从所有特征中提取最显著的信息,它融合了对任意数量的输入以随机顺序提取特征,自动保留最显著的特征。因此,与 PS-FCN 类似,本文所提出的网络也采用了这种策略。本文将空间坐标系下光照方向 $L_i \in \mathbb{R}^3$ 扩展至与所拍摄图像具有相同分辨率大小的张量 $L_i \in \mathbb{R}^{H \times W \times 3}$ (H, W 为图像分辨率),随后将归一化后的图像与对应的 L_i 拼接为张量 $\varnothing_i \in \mathbb{R}^{H \times W \times 6}$,其中第三维度前三项为图像 RGB 对应值,后三项为光照的 x, y, z 方向的分量。本文将提出的网络分为三个部分,即多层多元特征提取结合层,最大池化聚合层和对称式解码重建层。

2.1.1 多元特征提取结合层

给定 n 个输入张量 $\varnothing_1, \varnothing_2, \dots, \varnothing_m$, PS-FCN 从每个输入张量中提取特征,逐层进入深度卷积,然而,就本文观点认为,使用大量 3×3 尺寸的卷积核来进行特征提取,会导致特征信息的过度平滑化,难以有效捕获输入图像中各种特征及其细节,最终造成该网络在所含处理细节丰富的高频信息时表现不佳。为了提升模型对输入图像所含高频信息的提取和重建能力,就不同大小的卷积核尺寸对模型提取重建能力的影响进行了研究,研究发现 1×1 尺寸的卷积核提升了网络的细节感知能力,但会导致较多的信息丢失, 3×3 尺寸卷积核鼓励网络利用图像各个邻域特征,覆盖更广泛的图像区域,使网络能够捕获更多的全局上下文信息,但在细节感知能力上则有所欠缺。为了充分发挥两种尺寸卷积核的优势并实现互补,本研究提出了一种独立层间权重自适应调节的多层多元卷积模块。模块结构如图 1 所示。

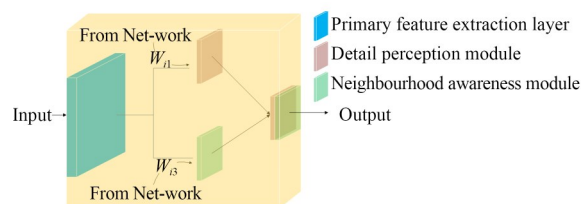


图 1 多层多元卷积模块结构

Fig. 1 Multilayer multivariate convolutional module structure

多元特征提取结合层由数个多层多元卷积模块堆叠而成,每个模块包含两个关键层。第一层为初级特征提取层,而第二层则为多元特征结合层。在初级特征提取层中,输入张量经过初级特征提取,生成初级特征。这些初级特征被传递至多元特征结合层,该层通过使用不同尺寸的卷积核对初级特征进行特征元提取,并将这些特征进行结合。即:

$$Out = Conv_{1 \times 1} + Conv_{3 \times 3}. \quad (1)$$

值得强调的是,实验结果表明,尽管两种尺度的卷积均呈现出有益的信息,但它们的简单叠加并未达到更为优越的效果。故认为特征之间存在一定的相关性,而这种相关性并非简单叠加,且其难以通过人为手段准确衡量。为了克服这一问题,引入自适应权重机制并对权重进行严格约束。每个多元特征结合层均包含两个独立的权重参数,这些参数参与网络的训练,并由网络自主学习得出。

$$Out = \left(\frac{W_{1 \times 1}}{W_{1 \times 1} + W_{3 \times 3}} \right) Conv_{1 \times 1} + \left(\frac{W_{3 \times 3}}{W_{1 \times 1} + W_{3 \times 3}} \right) Conv_{3 \times 3}, \quad (2)$$

式中 $Conv_{1 \times 1}$ 和 $Conv_{3 \times 3}$ 的系数即为权重,初始值均为 1,取值为 $[0, 1]$ 。如此,引入自适应权重机制的设计允许模型在每一层自动学习各个特征的权重比例。通过动态调整这些权重,模型能够更灵活地适应数据的复杂性,避免对某些特定特征的过度依赖,进而实现对高频信息更为有效的融合。这种机制的作用在于模型能够根据输入数据的不同特性,智能地分配注意力,使得网络更有针对性地利用 1×1 尺寸卷积核提取细节感知能力,同时通过 3×3 尺寸卷积核广泛捕获全局上下文信息。通过这种方式,自适应权重机制在融合高频信息的同时,有效防止了模型对某一

特定类型信息的过度专注,从而提升了模型的泛化能力。这样的设计策略不仅使得网络更具适应性,还加强了它在处理各类细节丰富表面上的性能,使它在复杂场景中更为强大和稳定。

同时,对于全卷积神经网络来说,通过增加网络层数,可以提高特征图的抽象程度,但却可能因此失去一些重要的低级细节。一般的,全局的深层特征图通常包含更高级的语义信息,而局部的浅层特征图则保留更多细节信息。由于信息的逐层传递,在深度增加的网络中可能会导致一些细节信息的丢失,尤其是对于那些本来就较为稀疏的低频信息而言。此外,对于所含细节丰富的高频信息来说,这种信息丢失也可能导致重建精度下降等问题,为了尽可能对这些高频细节信息进行保留,同时强化模型对低频信息的重建能力,引入了跳跃连接结构来对模型在特征提取阶段进行辅助。跳跃连接结构允许模型直接连接融合局部低级特征和全局高级特征,从而缓解信息传递的瓶颈问题。鉴于深层特征与浅层特征的通道数不匹配的情况,采用额外的特征提取操作,具体为采用 256 个 3×3 卷积核,每个核覆盖输入的 64 个通道并生成一个新通道,最终输出 256 通道,确保通道数量与深层特征匹配,并增强了局部特征的代表能力。这一策略有助于平衡不同深度的信息,并使网络能够更有效地传播特征信息,同时提高了模型的高低频信息重建性能。这种设计不仅考虑了特征的抽象程度,而且注重了保留高频信息和强化低频信息,为网络在处理细节时提供了更加全面和有效的能力。

设有 n 个输入张量,每个张量通过多层多元卷积模块进行处理,在每一模块中,使用两种尺寸的卷积核来提取特征。设第 i 层的 3×3 卷积

核的权重为 $W_{3 \times 3}^i$, 1×1 卷积核的权重为 $W_{1 \times 1}^i$, 第一层两种卷积核的权重均为 1, 初始值输入张量为 $\varnothing$, 则第 i 个模块的输出可以表示为:

$$Feature^i = \alpha^i \cdot Conv(\varnothing^i, W_{3 \times 3}^i) + \beta^i \cdot Conv(\varnothing^i, W_{1 \times 1}^i), \quad (3)$$

其中: α^i 和 β^i 是第 i 个模块的偏置量, i 初始值为 1, 最大为 4, $Conv$ 表示卷积操作。在多层多元卷积层中,每一层的输出作为下一层的输入,从初始状态开始一共进行 4 次 1×1 尺寸和 3×3 尺寸的卷积核自适应权重调整,同时图片通道数由 32 经 3 次翻倍递增至 256。因此整个模块的输出可以表示为:

$$Out = f\left(\sum_{i=1}^4 (\alpha^i \cdot Conv(\varnothing^i, W_{3 \times 3}^i) + \beta^i \cdot Conv(\varnothing^i, W_{1 \times 1}^i))\right), \quad (4)$$

其中两个压缩(下采样)阶段的目的是提高网络感受野,降低运算和内存需求,此外还包含一个上采样层帮助恢复细节信息。相较于 PS-FCN, MMF-Net 的多层多元特征提取融合层允许网络同时学习多个不同复杂度的特征并且将多个特征图组合在一起以创建更丰富的表达,使模型能够更好地理解复杂的输入数据。由于采用了辅助跳跃连接结构,网络拥有深浅层特征融合的能力,从而拥有更佳的网络性能。

本文所提出的多层多元特征提取融合层由 4 个多层多元卷积模块级联,如图 2 所示。其中,两个压缩阶段的目的是提高网络感受野,降低运算和内存需求,此外还包含一个上采样层帮助恢复细节信息。相较于 PS-FCN, 多元特征提取结合层允许网络同时学习多个不同复杂度的特征,并

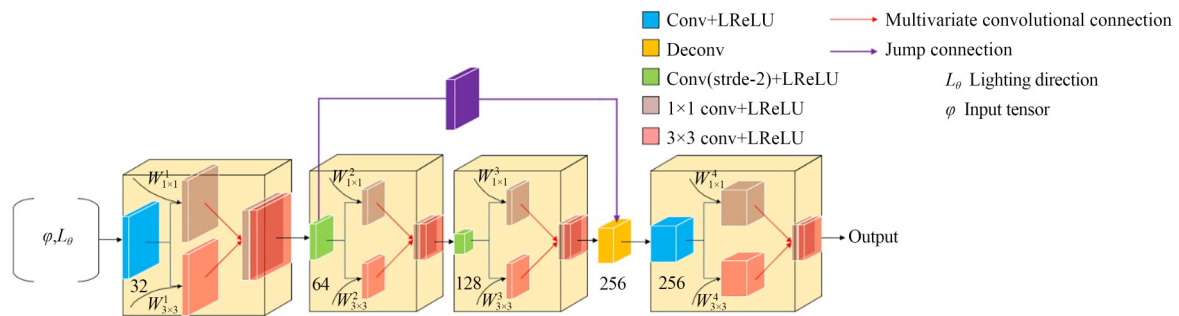


图 2 多层多元特征提取融合层结构

Fig. 2 Multi-layer multivariate feature extraction fusion layer structure

且将多个特征图组合在一起,以创建更丰富的表达,使网络能够更好地理解复杂的输入数据,同时,由于采用了辅助跳跃连接结构,使得网络拥有深浅层特征融合的能力,从而拥有更佳的网络性能。

2.1.2 最大池化聚合层

最大池化聚合层的主要作用是将来不同输入图像的特征图进行特征融合。在光度立体视觉中,某些光照方向下的高亮度区域往往包含丰富的几何信息,这些信息能够为表面法线的推断提供关键线索。最大池化可以自然地聚合这些强特征,此外,最大池化在训练过程中可以忽略非激活的特征,这使得模型对阴影区域更加稳定。

2.1.3 对称式解码重建层

解码阶段在整个模型中扮演着重要的角色,其功能在于解读编码器参数,这个任务需要解码器能够从编码器提供的潜在表示中恢复出物体表面的法向量信息。因此,为了提升整体模型性能,对解码器结构的设计至关重要。在所提出的网络中,设计了对称式编码-解码结构,

这种结构与编码器有类似但相反的层次和模式,减少了模型的复杂性,使其可以更好地捕捉编码阶段提取的特征,并在解码阶段进行逆向操作,避免丢失关键细节,使模型的整体行为更加可控和一致。

2.2 损失函数

MMF-Net通过预测法线图和真实法线图之间的误差来监督学习。本文使用余弦相似度作为网络的损失函数,如下:

$$S_{\text{los}} = \frac{1}{HW} \sum_{i,j} (1 - N_i \cdot \tilde{N}_i), \quad (5)$$

其中: H, W 为分辨率, i, j 为图像像素索引, N_i 为真实法线, $\tilde{N}_i$ 为预测的表面法线。预测的法线越接近真实法线,则 $N_i \cdot \tilde{N}_i$ 越接近1,损失函数就越小。

2.3 网络整体结构

整体网络结构如图3,网络使用NVIDIA GeForce RTX 3090进行了Adam的优化训练,学习率衰减设置为0.5,初始学习率为0.001,正则化参数beta1为0.9,beta2为0.999,使用32的批次大小训练了60轮,默认输入图像数量为32张。

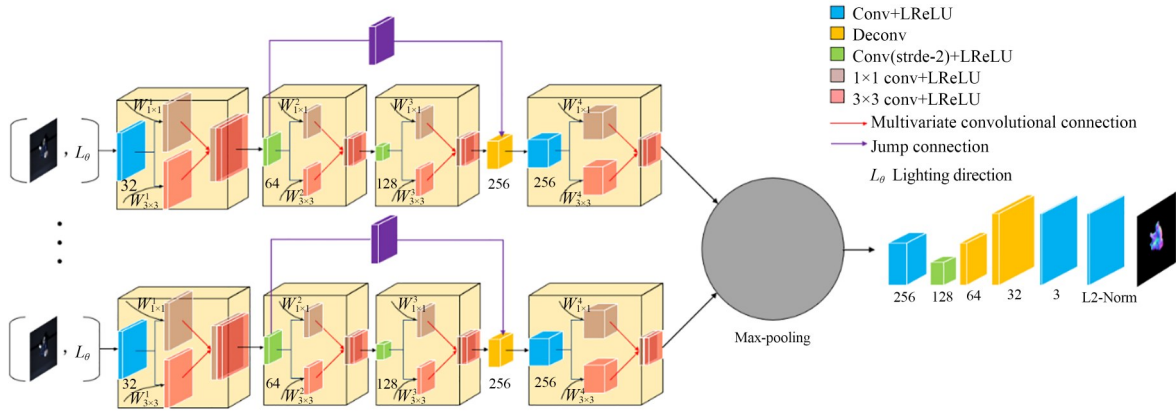


图3 MMF-Net模型的完整网络结构

Fig. 3 Complete network structure of MMF-Net model

3 实验与评估结果

3.1 数据集及评价指标

在实验中,使用MERL BRDF数据集渲染的Bobby数据集^[38]和Sculpture数据集^[40]用于训练。基于DiliGenT数据集^[41-42]对训练好的模型进行评估。该数据集包含10种具有96个不同光照方向的物体。采用平均角度误差(Mean Angle Er-

ror, MAE)作为模型的评价指标,在三维重建模型评价中,MAE能直接量化方向向量的角度偏差,准确反映方向一致性,且对尺度波动和异常值更稳健,符合三维重建对方向正确性的核心需求。而其他误差评定方法,如最小二乘法,无法有效区分方向与数值误差,对尺度敏感,易因非方向问题或异常值导致评价失真,故不适合此类任务。具体计算公式如下:

$$E_{\text{MA}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (1 - N_{ij} \cdot \hat{N}_{ij}), \quad (6)$$

其中: H, W 分别为输入图像的高和宽, 即图像分辨率大小, N_{ij} 是输入图像中第 i 行第 j 列像素包含的真实法向量, $\hat{N}_{ij}$ 是模型预测的图像第 i 行第 j 列像素的表面法向量。MAE 越小, 代表模型预测的法向量越接近真实值, 模型精度越高。

3.2 消融实验

为了证明提出方法的有效性, 本文首先对所

提出的网络中各个模块进行消融实验, 并使用 DiLiGenT benchmark 数据集^[40-41]中全部 10 个物体的平均 MAE 作为测试结果(96 幅输入图像), 表 1 展示了各个模块的消融实验结果, 其中序号 0 为本文提出的完整的网络, 序号 1 为取消多元特征提取结合层和对称式解码重建层的对称结构的消融实验, 序号 2 为取消自适应权重策略的消融实验, 序号 3 为取消多元卷积层的消融实验, 序号 4 为取消跳跃连接结构的消融实验。

表 1 消融实验

Tab. 1 Ablation experiments

Serial number	Ablation method	MAE/(°)										
		Ball	Cat	Pot1	Bear	Pot2	Buddha	Goblet	Reading	Cow	Hawvest	Avt
0	MMF-Net	2.22	4.55	6.19	7.27	6.15	7.35	7.56	10.04	6.09	12.02	6.95
1	Elimination of symmetrical structures	2.40	4.67	6.00	7.89	6.51	7.19	7.73	11.28	6.96	12.33	7.30
2	Elimination of adaptive weights	2.27	4.32	6.04	6.94	6.79	7.24	7.80	11.11	6.47	12.11	7.11
3	Elimination of multiscale convolution	2.59	4.78	6.31	6.27	6.40	7.26	7.81	10.58	7.09	12.22	7.13
4	Elimination of the jump structure	2.43	4.19	6.05	7.56	6.58	7.21	7.81	10.90	6.43	12.41	7.16

实验首先探讨了多元特征提取结合层和对称式解码重建层的对称结构对网络性能带来的影响(序号 1)。序号 1 取消了这种对称结构, 实验发现网络对于数据集中几乎所有物体的表面法线预测角度误差都有明显的增大, 印证了先前的想法。

如图 4 所示, 以 DiLiGenT benchmark 中的物体 reading 为例, 观察到(红框为高频区域, 黄框为低频区域)相较于采用对称式结构(彩图见期刊电子版), 序号 1 在高频区域和低频区域性能表现均有所下降, 原因是在模型的解码节点, 使用对称结构提高了整合来自不同层级之间特征信息

的能力, 从而能够有效提升模型的整体性能。

在实验 2 中, 实验探讨了自适应权重策略对网络性能的影响, 本次实验取消了权重策略, 仅采用 1×1 卷积与 3×3 卷积结果的简单叠加。实验结果表明, 虽然表面较为平滑的物体的法线重建误差对比序号 0 有所缓和, 但对含有高频信息的物体表面, 法线重建误差均有所上升, 平均 MAE 上升了 0.17° 。这种现象的解释在于平滑的表面通常不包含过多复杂的纹理和细节, 其法线变化相对较为均匀, 自适应权重机制的引入增加了模型的复杂性, 在模型内部引入了不必要的变化, 从而使模型估计的法线变化不稳定, 增加了误差。而对富含高频信息的物体来说, 它们往往具有复杂的纹理和细节, 自适应权重策略可以更好地捕捉这些纹理信息。此外, 高频信息涉及局部细节的变化, 自适应权重策略可以有选择性地强调或抑制这些局部特征, 使预测精度上升。在实际应用中, 需要重建的物体往往带有多种程度的高频信息, 引入自适应权重策略加强了模型在高频信息重建方面的性能, 提升了模型整体性能和实用性。

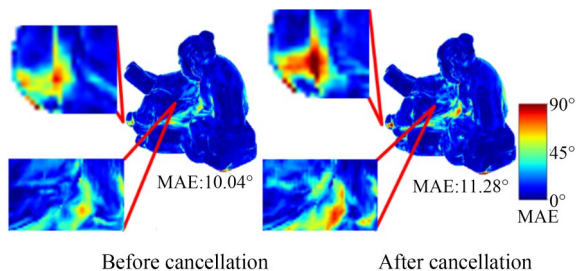


图 4 对称式编码-解码重构层对模型性能的影响

Fig. 4 Impact of symmetric encoding-decoding reconstruction layer on model performance

在对序号3进行分析前,需要说明的是,由于自适应权重策略是应用在 1×1 卷积和 3×3 卷积结合的结构上,所以取消 1×1 卷积结构后,自适应权重策略也随之消失。对于序号3,实验仅保留了 3×3 尺寸的卷积核,大量的 3×3 卷积层导致了图像的过度平滑,从而导致细节丢失,造成表面法线重建的模糊。与序号2相比,实验误差进一步增大,这表明在取消 1×1 卷积结构的情况下,过度依赖 3×3 卷积核导致了性能下降。这进一步验证了之前观察到的自适应权重策略与多尺度卷积结合策略相互促进的结果。同时也证明了自适应权重策略和多元卷积的协同作用对于提升表面法线重建精度的重要性。二者的配合使用不仅在保留细节的同时降低了过度平滑的风险,而且在模型整体性能上取得了显著的提升。

最后,在序号4中,实验删除了特征提取阶段的跳跃连接结构。结果显示,舍弃跳跃连接结构后,平均MAE增加了 0.22° 。这表明跳跃连接结构对于特征提取至关重要。在神经网络中,不同深度的层专注于捕获不同尺度的特征,而跳跃连接的引入使得这些层能够协同工作。其价值在于允许中间层的特征传递到后续层级,从而有助于保留和传递有用的信息,这对于处理图像的纹理和其他高低频特征至关重要。跳跃连接的存在使得网络能够更好地理解整体图像结构,并在不同层次上整合信息,有利于更全面地把握图像的细节与整体特征。更值得关注的是,跳跃连接结构的使用在图像处理中对于细微的纹理和低频特征的保存起到重要作用。这种结构在中间层之间建立了通道,促使信息在各层之间自由流动,最终对模型的性能提升具有显著影响。以DiLiGenT benchmark中的物体pot1为例,图5中

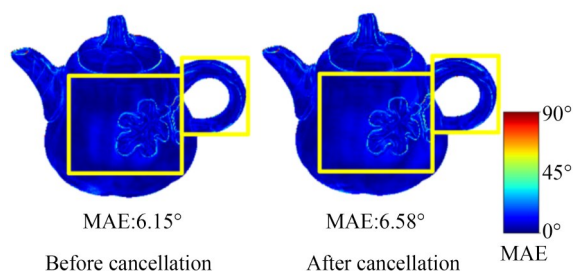


图5 跳跃连接结构对模型性能的影响

Fig. 5 Effect of jump connection structure on model performance

黄色框中的误差可视化呈现了跳跃连接结构对网络性能的影响,清晰地展示了其在信息传递和特征整合方面的关键作用。

3.3 对比实验及误差可视化对比

本文在DiLiGenT benchmark数据集上进行了对比实验,实验将MMF-Net与传统光度立体方法和基于深度学习方法的性能进行了广泛的对比。表2展示了在96幅密集输入图像下的重建误差。如表2所示,MMF-Net在DiLiGenT benchmark数据集的结果对比中取得了最佳平均MAE,对比2022年的MF-PSN领先 0.32° ,对比2023年的MASK-FCN领先 0.36° ,对比2024年的SDPS-MIX和FS24分别领先 2.85° 和 1.55° ,对比2025年的FUPS-Net领先 0.1° ,在与近三年的方法对比中取得较优结果。此外,在广泛对比下可以看到,MMF-Net在大多数图像下的性能都取得了较为优秀的MAE结果(红色字体为最优,蓝色字体为次最优)。图6为用于对比分析的折线图,其数据来源于表2。分别描绘了各对比组的重建MAE随物体变化的折线,通过不同的线条样式和颜色区分各组数据,能够方便地识别出不同网络间表面法向量重建性能差异。

图7~图8示了本文方法与其他方法的误差可视化对比结果。从图中可以看出,本文提出的网络在投射阴影处(黑色框),平坦表面处(黄色框),复杂结构处(红色框)都拥有更好的性能。尤其是在表面较为复杂,褶皱区域较多的物体上,本文网络表现较为突出,例如pot1, pot2中的花雕处,reading中人物的衣袍褶皱处,harvest中的满字等,本文所提出的对称式结构和自适应权重策略,对于复杂结构的表面来说,通过训练,自适应权重策略有选择性地强调或抑制这些局部特征,使得网络可以更好地捕捉这些纹理信息,从而在复杂结构的表面获得更好的重建效果。网络在投射阴影处也比其他方法拥有更好的重建结果,例如harvest右边小孩衣服与布袋的接缝处和reading中人物的袖口垂下处等,这可能是由于多元卷积特征结合策略起到了效果。多元卷积充分融合了 1×1 卷积和 3×3 卷积各自的优点,使网络既能关注到每像素细节信息,又可以利用图像的上下文信息,保留了这些阴影特征信息。此外,网络在平坦表面处也有着优秀的表

表 2 各光度立体方法的 MAE 对比实验

Tab. 2 Comparison experiments of MAE of photometric stereo methods (°)

Method	Ball	Cat	Pot1	Bear	Pot2	Buddha	Goblet	Reading	Cow	Hervest	AVG
Baseline ^[3] (1992)	4.10	8.41	8.89	8.39	14.65	14.92	18.50	19.80	25.60	30.62	15.39
WG10 ^[17] (2010)	2.06	6.73	7.18	6.50	13.12	10.91	15.70	15.39	25.89	30.01	13.35
AZ08 ^[41] (2008)	2.71	6.53	7.23	5.96	11.03	12.54	13.93	14.17	21.48	30.50	12.61
IA14 ^[42] (2014)	3.34	6.74	6.64	7.11	8.77	10.47	9.71	14.19	13.05	25.95	10.60
ST14 ^[43] (2013)	1.74	6.12	6.51	6.12	8.78	10.60	10.09	13.63	13.93	25.44	10.30
SPLINE-Net ^[44] (2019)	4.51	6.49	8.29	5.28	10.89	10.36	9.62	15.50	7.44	17.93	9.63
EW20 ^[45] (2020)	1.58	6.30	6.67	6.38	7.26	13.69	11.42	15.49	7.80	18.74	9.53
DPSN ^[22] (2017)	2.02	6.54	7.05	6.31	7.86	12.68	11.28	15.51	8.01	16.86	9.41
CK18 ^[46] (2018)	1.50	5.74	6.24	4.97	8.64	8.86	10.00	11.44	11.33	21.90	9.06
IRPS ^[47] (2018)	1.47	5.44	6.09	5.79	7.76	10.36	11.47	11.03	6.32	22.59	8.83
LMPS ^[48] (2019)	2.40	6.11	6.54	5.23	7.48	9.89	8.61	13.68	7.98	16.18	8.41
PS-FCN ^[26] (2018)	2.82	6.16	7.13	7.55	7.25	7.91	8.60	13.33	7.33	15.85	8.39
SDPS-Net ^[49] (2019)	2.77	8.06	8.14	6.89	7.50	8.97	11.91	14.90	8.48	17.43	9.51
Attention-PSN ^[50] (2020)	2.93	6.14	6.92	4.86	6.97	7.75	8.42	12.90	6.86	15.44	7.92
DR-PSN ^[51] (2021)	2.27	5.42	7.08	5.46	7.21	7.84	8.49	12.74	7.01	15.40	7.90
GPS-Net ^[52] (2020)	2.92	5.42	6.04	5.07	7.01	7.77	9.00	13.58	6.14	15.14	7.81
CHR-PSN ^[53] (2022)	2.26	5.97	7.04	6.35	6.76	7.15	8.32	12.52	6.05	15.32	7.77
CNN-PS ^[28] (2018)	2.20	4.60	5.40	8.30	6.00	7.90	7.30	12.60	8.00	14.00	7.60
PS-FCN(Norm) ^[54] (2020)	2.67	4.76	6.17	7.72	7.15	7.53	7.84	10.92	6.72	12.39	7.39
MASR-PSN ^[55] (2023)	2.43	4.74	6.32	6.99	6.70	7.24	8.39	10.69	6.92	12.70	7.31
SDPS-MIX ^[29] (2024)	3.40	9.60	9.10	4.50	7.20	9.50	9.90	18.60	6.70	19.70	9.80
FS24 ^[30] (2024)	2.32	6.87	7.44	5.23	7.18	8.24	9.77	15.04	7.26	15.65	8.50
MF-PSN ^[32] (2022)	2.07	5.00	7.20	5.83	6.81	6.88	7.46	12.2	5.90	13.38	7.27
FUPS-Net ^[33] (2025)	2.30	4.71	6.03	5.27	6.58	7.31	7.54	11.29	5.74	13.75	7.05
MMF-Net(Ours)	2.22	4.55	6.19	7.27	6.15	7.35	7.56	10.04	6.09	12.02	6.95

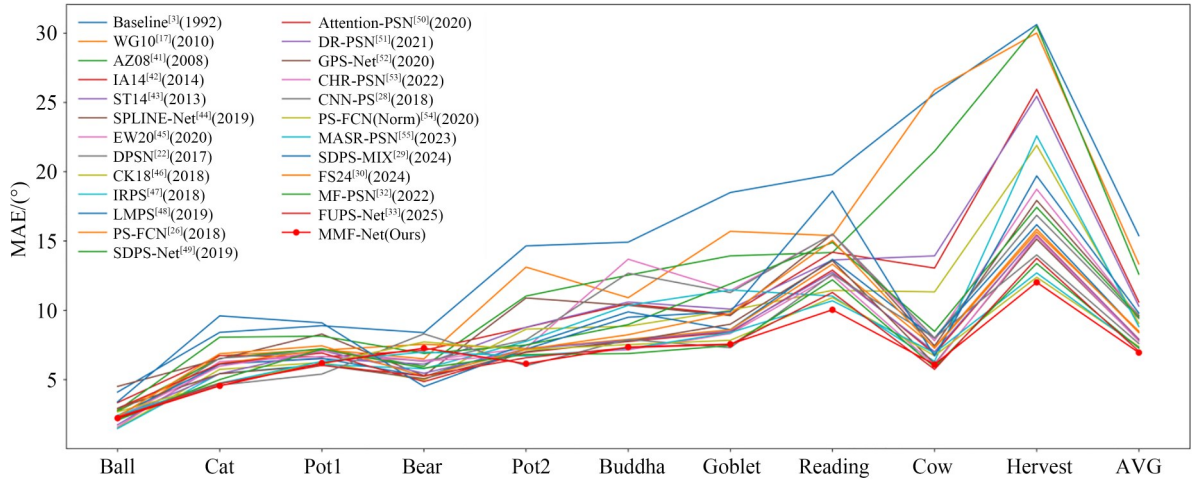


图 6 对比实验数据可视化

Fig. 6 Comparison of experimental data visualisation

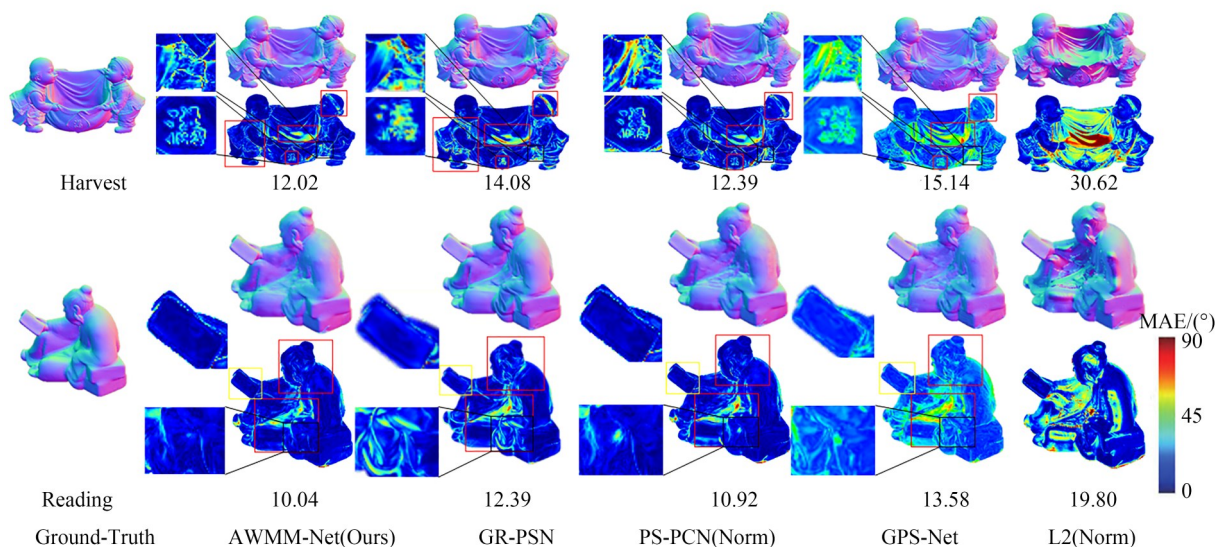


图 7 两幅含有高频信息物体预测误差可视化对比

Fig. 7 Visualisation of experimental error for comparison of two complex pictures

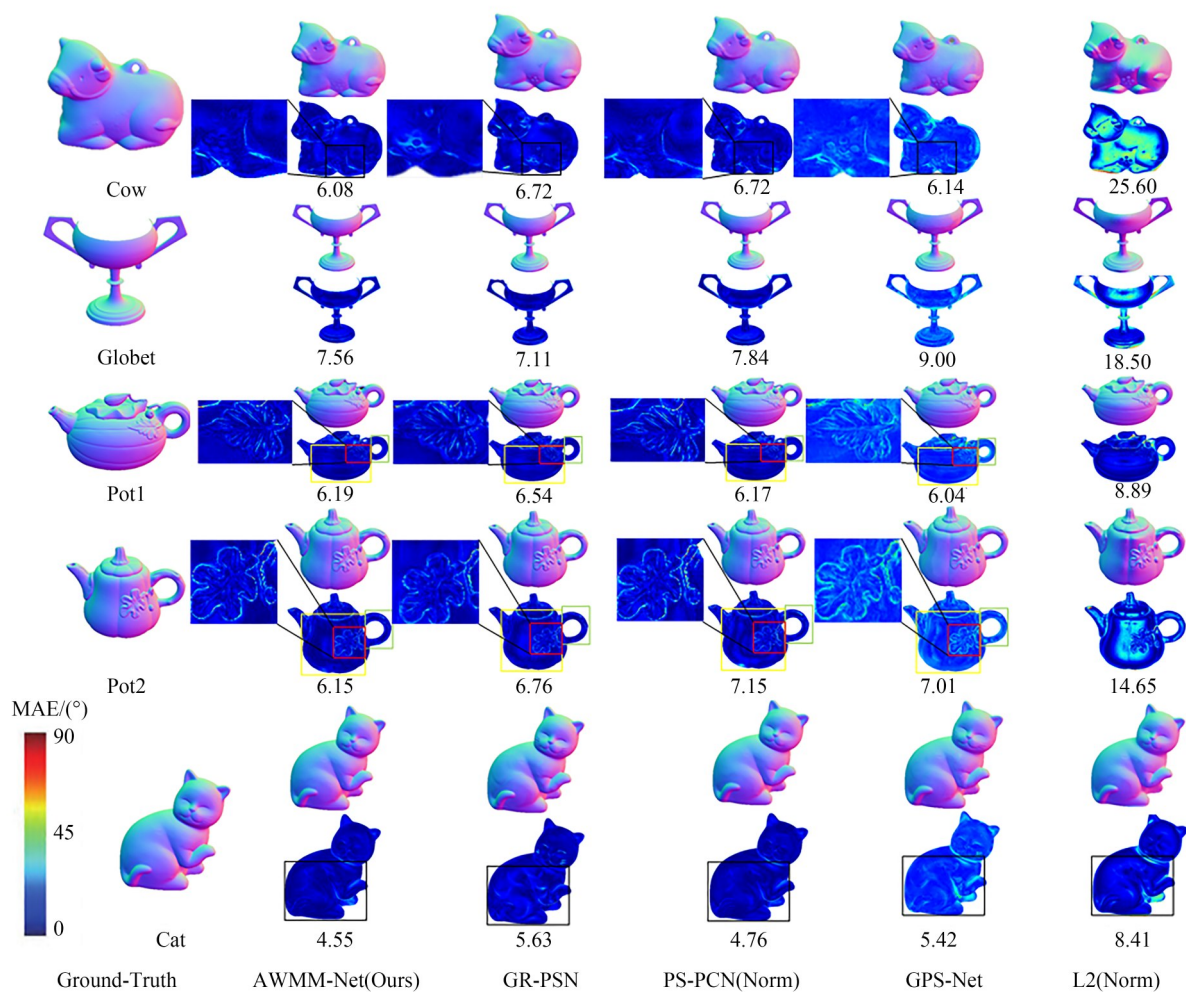


图 8 剩余图片对比实验误差可视化

Fig. 8 Visualisation of residual picture comparison test errors

现,跳跃连接结构的引入使得不同层级的特征能够协同工作,这对于处理图像的纹理以及一些低频区域具有显著的帮助。

3.4 公开数据集泛化性实验及分析

为了进一步说明本文提出的网络的有效性和泛化性,本文额外在 DiLiGenT-MV 公开数据集上进行了实验并分别进行定量以及定性分析。该数据集包含了 reading, buddha, pot2, cow, bear 五个物体在 20 个不同的角度下的不同光照方向的图像(每个角度各有 96 个不同的光照方向)。表 3 为对提出的网络对该数据集进行的定量实验分析。同样,该数据集进行定

性分析,如图 9 所示,本文所提出的网络可以准确地重建物体表面的法向量。实验证明,对于物体表面含有纹理较为丰富复杂的高频信息(红框,如 buddha 手捧元宝区域和面部细节,reading 老人的衣袍褶皱区域),网络获得了较小的表面法线重建误差,在物体表面变化较为剧烈的区域(黄框,如 pot2 的壶嘴,cow 的眼睛与拉环,bear 存钱罐的投币口等),网络也能够很精准地将其表面法向量预测出来。此外,在表面较为平坦的区域(绿色框),网络也表现出优秀的性能,这进一步说明了本文方法对真实世界的泛化性和稳定性。

表 3 DiLiGenT-MV 数据集对比实验
Tab. 3 Comparison experiments on DiLiGenT-MV dataset (°)

Methods	MAE					
	Bread	Pot2	Buddha	Reading	Cow	AVG
PS-FCN	6.80	7.60	9.51	12.56	7.53	8.80
MMF-Net(Ours)	6.74	6.85	9.16	10.08	6.21	7.81

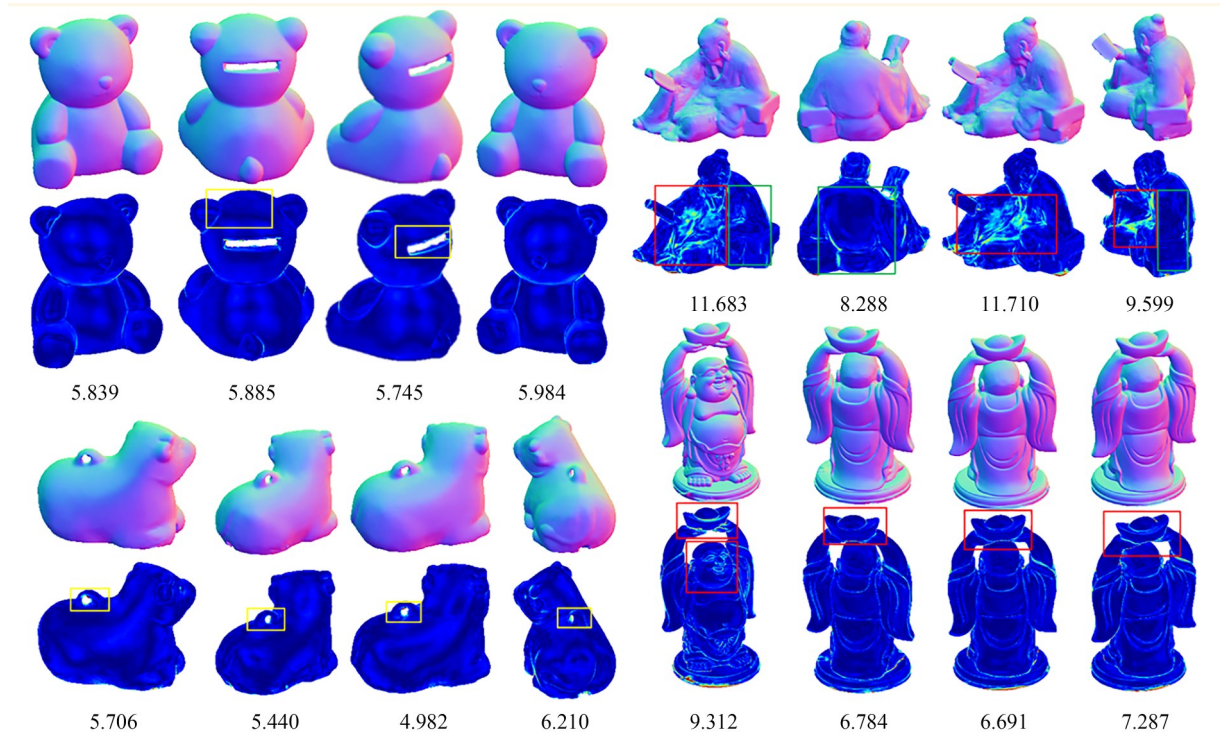


图 9 DiLiGenT-MV 数据集部分物体预测误差可视化

Fig. 9 Visualisation of prediction errors for some objects in the DiLiGenT-MV dataset

3.5 自建数据集泛化性实验及分析

为了更加有力地证明 MMF-Net 的泛化性,

本文自建了光度立体数据集,然后在自建数据集上进行与基准模型 PS-FCN 的对比实验及误差

可视化对比。自建光度立体数据集所用建模软件为 Blender,是一款免费开源的全功能三维创作软件,其功能满足实验对相机、光源、物体材质选择及法向量图获取的要求。

3.5.1 自建光度立体数据集的获取过程

为了构建高质量的光度立体数据集并确保与现有标准的可比性,本节效仿 DiLiGenT benchmark 数据集的布光策略,在 Blender 场景中构建了光源布局(以待测物体为中心在其环绕球面均匀布置 96 个近场点光源),之后进一步对摄像机位置参数设置以及视角进行了设置,以确保能够满足光度立体算法对图像采集的要求,并为后续的渲染和数据集生成提供高质量的图像样本。

在完成摄像机相关位置参数的设置后,还需要获取待测物体表面法向量的真值信息,以便为网络的训练提供准确的参考依据,利用 Blender 的法线贴图功能实现对物体表面法向量真值的获取。为了将法线贴图上的像素值转换为物体表面法向量的真值,需要进行相应的数学变换,具体公式如下:

$$nx_{ij} = \frac{CR_{ij} - 127.5}{128}, \tag{7}$$

$$ny_{ij} = \frac{CG_{ij} - 127.5}{128}, \tag{8}$$

$$nz_{ij} = \frac{CB_{ij} - 127.5}{128}, \tag{9}$$

其中: nx, ny, nz 分别代表该像素点空间向量的 xyz 方向分量, CR, CG, CB 分别代表图像对应的

RGB通道像素值, ij 为图片的第 i 行第 j 列像素。

为了构建一个具有广泛代表性和说服力的光度立体数据集,本文选取多种具有不同材质和复杂度的物体进行渲染输出,涵盖了从低频的均匀表面到含有剧烈变化的高频表面,以及从具有高光反射的陶瓷材质到漫反射较为强烈的木纹材质等多种覆盖材质。分别是面包、马、侏儒、桶、雕像、贝壳、石头 1、石头 2、石头 3、石头 4、大象和黄鸭共计 12 个物体模型。此外,渲染前将场景的背景设置为黑色模拟真实遮光环境。

在完成物体模型的选择后,对每个物体模型的光照图片进行渲染输出。为了更好地模拟物体在真实世界环境中的反射表现,提高数据集的实用性和可靠性,本文为图片中包含物体的区域施加了随机噪声扰动。经过上述处理流程,最终得到能够用于验证光度立体模型性能的自建光度立体数据集。

3.5.2 实验及分析

本文首先进行了与基准数据集 PS-FCN 的定量对比实验。如表 4~表 5 所示,针对含有高频信息物体表面法向量重建优化设计的网络结构在自建数据集上的平均 MAE 相较于基准网络模型 PS-FCN 提升了 14%,并且在数据集包含的 12 个物体的 MAE 指标相较于 PS-FCN 均有所改善,且对于表面含有高频信息的面包、马、贝壳、石头 1~4,网络结构性能提升更为明显,n 再次验证了 MMF-Net 网络结构对于局部高频信息重建方面的有效性。

表 4 自建数据集对比实验 1
Tab. 4 Self-constructed dataset comparison experiment 1 (°)

Methods	MAE					
	Bread	Horse	Dwarf	Barrel	Statue	Shell
PS-FCN	5.72	3.61	6.62	7.21	4.83	8.61
MMF-Net(Ours)	3.91	1.86	6.22	6.74	4.53	7.37

表 5 自建数据集对比实验 2
Tab. 5 Self-constructed dataset comparison experiment 2 (°)

Methods	MAE						
	Stone1	Stone2	Stone3	Stone4	Elephant	Duck	AVG
PS-FCN	4.59	6.49	7.01	5.85	9.71	7.49	6.48
MMF-Net(Ours)	4.27	5.46	6.81	5.58	8.67	5.39	5.57

其次,进行了表面法向量预测误差可视化的定性分析实验,图 10 为数据集中网络性能区别较明显的物体表面法向量重建对比。图 11 为数据集中网络性能区别较明显的误差可视化对比。首先,对于局部高频信息(红框),MMF-Net 相较于基准模型 PS-FCN 对其表面法向量的预测误差有着显著降低,如物体马的上半部分鬃毛、物体侏儒的眼睛和持灯手部区域、物体桶的两个把手、物体贝壳的整体区域、物体石头 3 中间区域以及物体黄鸭的眼部区域。这进一步证明了 MMF-Net 针对物体局部高频区域预测设计的多层多元特征提取融合层确实能够有效提升模型对于含高频信息物体表面法向量预测的性能。其次,对于低中频区域(黄框)的表面法向量预测,例如物体马的下半部分鬃毛,腹部,腿、物体侏儒的腰带区域,也能够发现 MMF-Net 的性能明显优于基准模型 PS-FCN。证明 MMF-Net 设

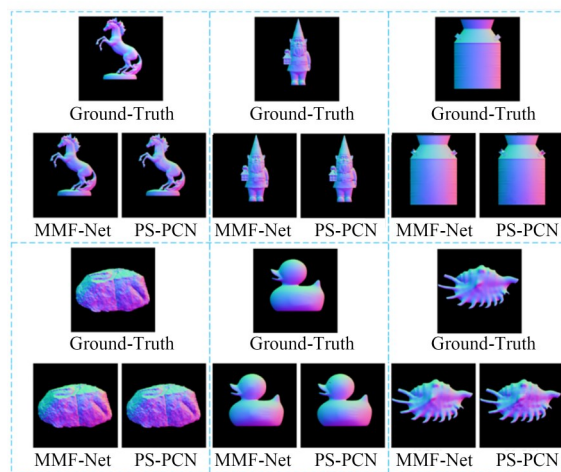


图 10 物体表面法向量重建

Fig. 10 Reconstruction of normal vectors of object surface

计的跳跃连接结构对于同时保留图像高频和低频信息的性能有着正面影响。

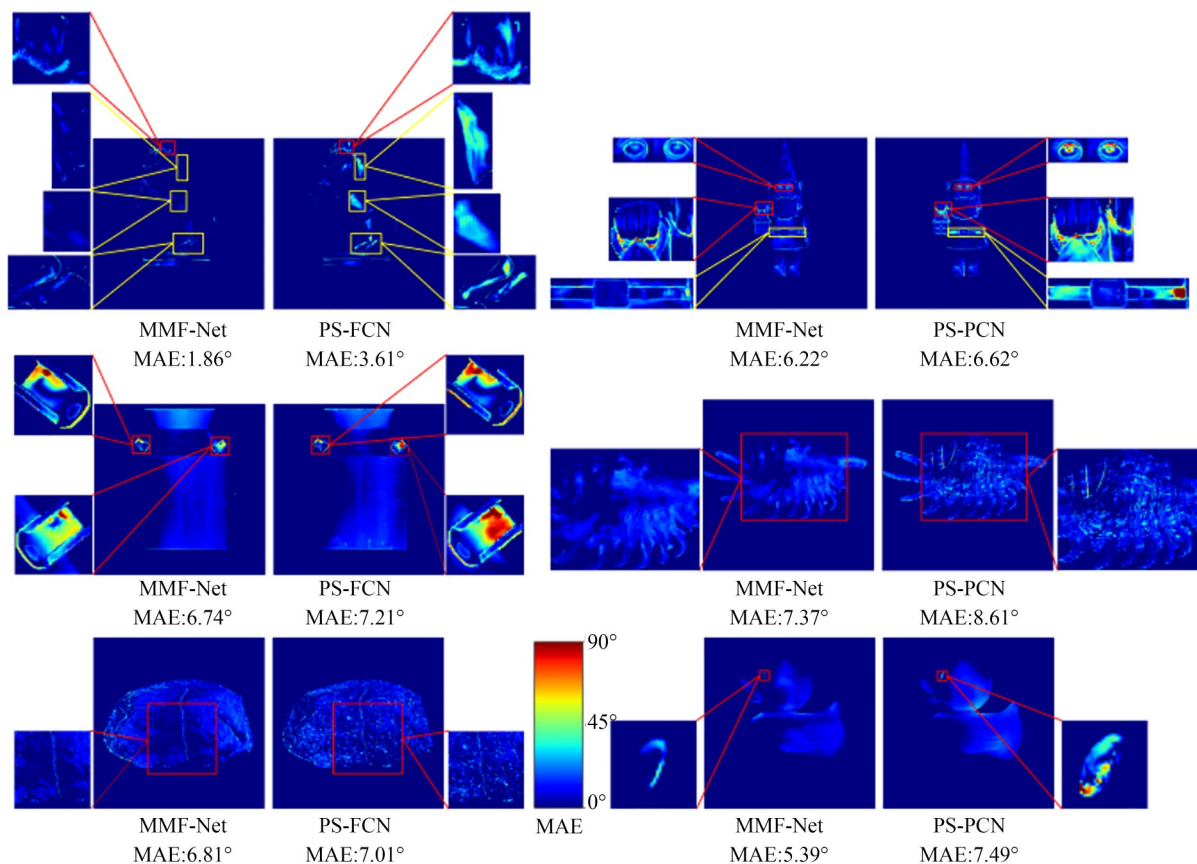


图 11 物体表面法向量预测误差可视化对比

Fig. 11 Comparison of prediction error visualisation of normal vectors on object surface

4 结 论

本文基于PS-FCN网络架构,优化设计了多层多元特征权重自适应融合网络MMF-Net,在基于学习的光度立体方法中取得了比其他方法更为准确的结果。首先,提出一种多层多元特征提取结合层,分别从各个尺度的特征中提取出对网络有益的部分,从而让模型能够更好地学习输入图像的邻域特征以及每像素细节。其次,为了更好地平衡不同尺度特征在本层所占比例,提出了自适应权重机制,让难以人为界定的尺度特征比例能够在网络的训练中达到自主学习的目的,获得更好效果。而后,采用特征提取层与对称式解码重建层的类对称结构,使模型能够提升对不同特征和模式的感知能力。最后,本文提出的网络模型充分利用了输入图像的深浅层特征的结合,使得两者在各自优势的区域互补。实验结果表明,本文提出的网络架构在对物体的三维表面进行重建方面表现出显著的有效性。与其他网络相比,在相同的输入数据条件下,网络能够得到更高准确性的高频图像重建模型,尤其在处理褶皱较多的复杂表面区域时表现优秀,为复杂表面的精准还原提供了有效支持。在DiLiGenT benchmark数据集上的测试结果实现平均MAE达到6.95°,对比先前的优秀网络架构FUPS-Net的7.05°提升了1%的性能,对比基准网络架构PS-FCN的7.39°提升了6%的性能,此外,在其中两幅复杂程度较高的harvest和reading的重建结果上,重建误差均值为11.03°,对比FUPS-Net的12.52°提升了12%,对比基准网络模型PS-FCN的11.66°提升了5%。在这两幅图片的对

比实验中,MMF-Net均取得最佳MAE,证明了MMF-Net在处理高频图像模型时的有效性。最后,为了证明MMF-Net的泛化性和稳定性,在公开数据集DiLiGenT-MV和自建光度立体数据集上进行了广泛的泛化性实验,实验结果表明MMF-Net具有优秀的泛化性能。在DiLiGenT-MV数据集上平均重建误差达到7.8°,在自建数据集上平均重建误差达到5.57°,对比基准网络模型PS-FCN的6.48°提升了14%,证明了MMF-Net在面对不同情况具有优秀的稳定性和泛化性。

但本文所提出的方法仍有改进空间。例如多尺度特征融合方面,可以通过增加不同尺度特征的数量来获得更佳优势互补效果。输入图像方面,可通过增加不同颜色/模态的光源来对模型进一步约束,以及受渲染软件制约,自建数据集渲染时,存在真实场景中的噪声、反光干扰等因素考虑不足等缺点。因此,如何充分利用不同种类的信息进行更有效的融合互补,考虑真实数据的环境影响因素,进而提升图像表面法向量的重建性能以及保证数据集与真实数据的一致性,仍是未来需要进一步研究和解决的问题。

作者贡献声明:

王标:研究方法的提出,论文构思和项目支持;

李影:测量实验数据分析与论文撰写;

融百川:测量实验设计;

刘璟:实验数据分析;

王永红:论文审核与指导;

张进:论文修改。

参考文献:

- [1] 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.
- [2] HORN B K P. *Robot Vision* [M]. Wang L, Jiang X L, Transl. Beijing: China Youth Publishing Group, 2014.
- [3] WOODHAM R J. Photometric method for determining surface orientation from multiple images[J]. *Optical Engineering*, 1980, 19(1): 139.
- [4] 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.
- [5] 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. IEEE, 2021: 3803-3813.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] BOYD S. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foundations and Trends® in Machine Learning*, 2010, 3(1): 1-122.
- [10] 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.
- [11] 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. IEEE, 2017: 350-359.
- [12] ZHANG Y F, YANG Y H. Illuminant direction determination for multiple light sources[C]. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000*. June 15-15, 2000, Hilton Head Island, SC. IEEE, 2000: 269-276.
- [13] POWELL M W, SARKAR S, GOLDFELD D. Calibration of light sources[C]. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000*. June 15-15, 2000, Hilton Head, SC, USA. IEEE, 2000: 263-269.
- [14] HARA K, NISHINO K, LKEUCHI 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.
- [15] 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.
- [16] 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.
- [17] WU L, GANESH A, SHI B X, et al. *Robust Photometric Stereo Via Low-rank Matrix Completion and Recovery*[M]. Computer Vision-ACCV 2010. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 703-717.
- [18] WU T P, TANG C K. Photometric stereo via expectation maximization[C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. January 23, 2009, IEEE, 2010: 546-560.
- [19] 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. IEEE, 2012: 318-325.
- [20] RUITERS R, KLEIN R. Heightfield and spatially varying BRDF reconstruction for materials with interreflections [J]. *Computer Graphics Forum*, 2009, 28(2): 513-522.
- [21] HERTZMANN A, SEITZ S M. Example-based photometric stereo: shape reconstruction with general, varying BRDFs[C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. August 31, 2005, IEEE, 2005: 1254-1264.
- [22] 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. IEEE, 2017: 501-509.
- [23] NIELSEN J B, JENSEN H W, RAMAMOORTHY R. On optimal, minimal BRDF sampling for reflectance acquisition[J]. *ACM Transactions on Graphics*, 2015, 34(6): 1-11.
- [24] 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.
- [25] 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.
- [26] CHEN G Y, HAN K, WONG K Y K. *PS-FCN: a Flexible Learning Framework for Photometric Stereo* [M]. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11213: 3-19.
- [27] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. May 24, 2016, IEEE, 2017: 640-651.
- [28] IKEHATA S. CNN-PS: CNN-based photometric stereo for general non-convex surfaces[C]. *Computer Vision -ECCV 2018*. Cham: Springer, 2018: 3-19.
- [29] WANG R S, ZHEN C, DONG N. SDPS-M2CAN: improving predictive performance of photometric stereo networks on broad spectrum reflective surfaces with minor indentations or bumps [C]. *2024 China Automation Congress (CAC)*. November 1-3, 2024, Qingdao, China. IEEE, 2024: 3347-3352.
- [30] 方明权, 宋滢. 基于深度学习和自注意力的光度立体方法[J]. *软件工程*, 2024, 27(8): 46-50. FANG M Q, SONG Y. Photometric stereo method based on deep learning and self-attention mechanism [J]. *Software Engineering*, 2024, 27(8): 46-50. (in Chinese)
- [31] 姚卓坤. 基于非结构化神经网络的光度立体方法 [D]. 天津: 天津大学, 2021. YAO Z K. *Photometric Stereo Method Based on the Unstructured Neural Network* [D]. Tianjin: Tianjin University, 2021. (in Chinese)
- [32] LIU Y R, JU Y K, JIAN M W, *et al.* A deep-shallow and global-local multi-feature fusion network for photometric stereo[J]. *Image and Vision Computing*, 2022, 118: 104368.
- [33] JU Y K, SHI B X, WEN B H, *et al.* Revisiting one-stage deep uncalibrated photometric stereo via Fourier embedding[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 47(8): 6185-6199.
- [34] 张蕾, 石岩, 卢文雍, 等. 基于 SURF-OKG 特征匹配的三维重建技术[J]. *光学精密工程*, 2024, 32(6): 915-929. ZHANG L, SHI Y, LU W Y, *et al.* 3D reconstruction technique based on SURF-OKG feature matching[J]. *Opt. Precision Eng.*, 2024, 32(6): 915-929. (in Chinese)
- [35] 邓高旭, 刘梓豪, 柯俐颖, 等. 相位测量的不连续位置自适应分割[J]. *光学精密工程*, 2024, 32(8): 1252-1260. DENG G X, LIU Z H, KE L Y, *et al.* Adaptive discontinuity segmentation in phase measurement [J]. *Opt. Precision Eng.*, 2024, 32(8): 1252-1260. (in Chinese)
- [36] 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.
- [37] JOHNSON M K, ADELSON E H. Shape estimation in natural illumination[C]. *CVPR 2011*. June 20-25, 2011, Colorado Springs, CO, USA. IEEE, 2011: 2553-2560.
- [38] WILES O, ZISSERMAN A. SilNet: single- and multi-view reconstruction by learning from silhouettes[C]. *Proceedings of the British Machine Vision Conference 2017*. London, UK. British Machine Vision Association, 2017.
- [39] 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. IEEE, 2016: 3707-3716.
- [40] 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.
- [41] 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. IEEE, 2008: 1-8.
- [42] IKEHATA S, AIZAWA K. Photometric stereo using constrained bivariate regression for general isotropic surfaces[C]. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 23-28, 2014, Columbus, OH, USA. IEEE, 2014: 2187-2194.
- [43] SHI B X, 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.
- [44] 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. IEEE, 2019: 8548-8557.
- [45] ENOMOTO K, WAECHTER M, KUTULAKOS K N, et al. Photometric stereo via discrete hypothesis-and-test search [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020. Seattle, WA, USA. IEEE, 2020: 2308-2316.
- [46] CHENG K H M, KUMAR A. Revisiting outlier rejection approach for non-lambertian photometric stereo [J]. *IEEE Transactions on Image Processing*, 2018.
- [47] TANIAI T, MAEHARA T. Neural inverse rendering for general reflectance photometric stereo [C]. *International Conference on Machine Learning*, 2024.
- [48] 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. IEEE, 2019: 7560-7568.
- [49] 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. IEEE, 2019: 8731-8739.
- [50] JU Y K, LAM K M, CHEN Y, et al. Pay attention to Devils: a photometric stereo network for better details [C]. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. July 11-17, 2020. Yokohama, Japan. *International Joint Conferences on Artificial Intelligence Organization*, 2020: 694-700.
- [51] 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.
- [52] 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*, 33, 2020, pp. 10306-10316.
- [53] JU Y K, PENG Y X, JIAN M W, et al. Learning conditional photometric stereo with high-resolution features [J]. *Computational Visual Media*, 2022, 8(1): 105-118.
- [54] 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.
- [55] JU Y K, JIAN M W, RAO Y, et al. MASR-PSN: a low-resolution photometric stereo images-relevant deep learning model for high-resolution surface normal reconstruction [J]. *Journal of Image and Graphics*, 2023, 28(7): 2120-2134.

作者简介:



王 标(1980—),男,博士,副教授,硕士生导师,2009年于合肥工业大学获得博士学位,2019年至2020年在新加坡南洋理工大学进行访问学者研究,主要研究包括精度检测与在线测量、微元件材料特性测试、机器视觉以及图像处理。E-mail: wangbiao@hfut.edu.cn

通讯作者:



王永红(1972—),男,博士,教授,博士生导师,2004年获得博士学位,2007-2008年在美国 Oakland University 做博士后,主要研究包括光学精密测试技术、激光散斑干涉测量、机器视觉与图像处理、材料变形应变机械性能测量和复合材料无损检测。E-mail: yhwang@hfut.edu.cn