

文章编号 1004-924X(2024)20-2987-12

基于稀疏自注意力的偏振表面法线估计

于智超¹, 万振华², 赵开春^{1*}

(1. 清华大学 精密仪器系, 北京 100084;
2. 广西大学 机械工程学院, 广西 南宁 530004)

摘要: 对于物体表面法线的准确估计在各种计算机视觉任务中扮演着重要角色。采用基于物理的偏振表面法线估计方法存在诸多限制, 限制了其实用性。相反, 基于深度学习的偏振表面法线估计方法在精度和适用性上均超越了基于物理的方法, 为了进一步提高基于深度学习的偏振表面法线估计精度, 以使其能适用于更多的实际任务, 本文提出了一种新的方法。首先, 结合斯托克斯矢量提出了一种新的偏振信息表示方法, 旨在提高模型对偏振物理先验信息的提取能力。随后结合基于双层路由的稀疏自注意力机制, 以改进模型对全局上下文信息的感知能力, 来更好地对局部偏振信息消歧。在 DeepSP 数据集和自建测试数据进行测试。实验结果表明: 提出的方法在 DeepSP 数据集上平均角度误差为 13.37°, 并在精度和角度误差等所有测试指标上均优于现有方法, 证明了该方法在估计法线效果方面的显著改进。通过引入新的偏振信息表示方法和稀疏自注意力机制, 我们的方法提高了偏振表面法线估计的精度和适用性, 为实际任务的应用提供了更强的支持。

关键词: 偏振信息表达; 稀疏自注意力机制; 偏振表面法线估计

中图分类号: TP394.1; TH691.9 **文献标识码:** A **doi:** 10.37188/OPE.20243220.2987

Shape from polarization based on sparse self-attention

YU Zhichao¹, WAN Zhenhua², ZHAO Kaichun^{1*}

(1. Department of Precision Instrument, Tsinghua University, Beijing 100084, China;

2. Faculty of Mechanical Engineering, Guangxi University, Nanning 530004; China)

* Corresponding author, E-mail: kaichunz@mail.tsinghua.edu.cn

Abstract: Accurate estimation of surface normal plays a vital role in various computer vision tasks. Physically-based shape from polarization methods have limitations restricting their applications. Conversely, learning-based shape from polarization methods outperform physical methods in both accuracy and applicability. To further improve the accuracy of shape from polarization and make it applicable to a broader range of practical tasks, we proposed a novel method. First, we introduced a new polarization information representation combining Stokes vectors, enhancing the model's ability to extract polarization physical prior information. Then, we integrated a bi-level routing sparse self-attention mechanism to improve the model's perception of global contextual information, enabling better disambiguation of local polarization information. Testing on the DeepSP dataset and out test data, experimental results demonstrate our proposed method achieves an average angular error of 13.37° on the DeepSP dataset, outperforming existing meth-

收稿日期: 2024-05-09; 修订日期: 2024-07-22.

基金项目: 教育部联合基金资助项目 (No. 6141A02022606)

ods in all tested metrics including accuracy and angular error. This indicates a significant improvement in normal estimation effectiveness with our proposed method. By introducing a novel polarization information representation and sparse self-attention mechanism, our approach enhances the accuracy and applicability of polar surface normal estimation, providing stronger support for practical task applications.

Key words: polarization information representation; sparse self-attention mechanism; shape from polarization

1 引 言

准确估计物体表面法向量对于表面感知和恢复物体几何结构至关重要。这一技术在诸如三维重建、工业产品检测、机器人导航与环境感知、增强现实/虚拟现实、文物保护以及月面三维重建等领域的实际应用中发挥着重要作用^[1-4]。然而,实现准确的物体表面法线估计是实现这些应用的前提。目前主要的单目法向量估计方法是光度立体法,它通过在不同光照条件下获取目标的图像来估计物体表面的法向量。这种被动式的重建方式具有快速准确提取信息、适用性强的特点^[9],在特定条件下能够获得相对高精度的法向量,并且已经得到了广泛的研究和发展^[5-8]。然而,基于光度的方法通常需要多个不同光照下的目标图像,而且重建通常在受控光源下进行,因此对光源的要求较高。在一般场景中不受控制的照明条件下,其实际重建效果会受到影响。相比之下,激光雷达等主动式的测量手段虽然每个激光测量点的精度较高,但通常实际使用时物体表面的测量数据点较为稀疏,并且受到噪声的影响^[10]。因此,通过这种方式重建的结构可能不够准确,恢复的表面结构也不够可靠。

环境中的光除了携带强度信息外,还包含重要的偏振信息,这种偏振态反映了光所具有的方向性信息,而在各种视觉任务中结合偏振信息能改善性能提升效果,比如去雾^[17],图像超分辨^[12],目标姿态估计^[26]等。事实上即便无偏光照环境下,物体反射的光往往会呈现出一定的偏振性。这种偏振态与光线在反射点上物体表面的法线方向密切相关。因此,通过获取物体表面反射光的偏振信息,可以有效地帮助恢复物体表面的结构。相比于传统方法,这种基于偏振信息的方法对光源的要求更为宽松。而随着偏振 CMOS 技术的不断成熟,获取偏振信息也变得越来越

便捷。

结合偏振信息来恢复物体表面结构的方法自提出以来越来越受到学术界的关注。早在 1979 年, Koshikawa 等人就提出了通过偏振信息来估计物体形状的方法^[11]。Mahmoud 等人提出了结合偏振和阴影信息来恢复形状,以解决偏振估计中的法线歧义问题^[25]。Vage 等人提出了考虑漫反射和镜面特性的真实世界表面模型,并将这两种反射类型分离以应用适当的 SFP (Shape from Polarization) 方法^[13]。Yu 等人提出了一种通过最小化在各个偏振器角度处的预测强度与所有像素的观测强度之间的平方残差之和来解决表面高度的方法,并采用了高斯-牛顿类型的优化方法^[14]。随着深度学习的发展,将偏振恢复形状与深度学习相结合成为一种趋势。Ba 等人提出了第一个结合深度学习的偏振恢复形状的方法,并创建了一个高质量的偏振重建数据集用于研究^[15]。相比于传统方法,该方法在精度上取得了较大的提升。Zou 等人提出了一种利用偏振推导的几何线索并采用深度学习来从单一偏振图像中估计人体姿态和形状的方法^[16]。Valentin 等人结合仿真和真实数据集用于训练深度学习模型,并估计物体三维形状与表面双向反射分布函数的变化^[24]。Lei 等人提出了一种结合偏振的深度学习用于场景级法线估计,实现了对大场景的法线估计效果^[10]。

尽管目前结合偏振的表面形状恢复方法已经取得了一定的进展,但在实际应用中,精度仍然需要进一步提高。因此,本文提出了一种基于深度学习的法线估计方法,旨在用于三维重建等任务,并通过采用新的偏振信息编码方式来更有效地利用偏振信息进行法线回归。该方法采用深度学习模型,将新的偏振信息编码为模型的输入,以实现更高的精度。具体而言,本文采用卷积神经网络作为模型骨架,并结合双层路由的稀

疏自注意力机制,以改善卷积网络对全局信息的利用不足问题。通过这种方式,可以更好地处理局部偏振信息存在的歧义性,从而提高模型的性能。本文通过定量评估和定性分析证明了本文方法的有效性。

2 偏振测量原理

光的偏振态可以用四参数的斯托克斯矢量表达:

$$\mathbf{S} = [S_0, S_1, S_2, S_3]^T, \quad (1)$$

其中: S_0 表示光的强度, S_1 是电磁波中平行于参考平面和垂直于参考平面的偏振分量强度差异, S_2 是相对于参考平面夹角分别为 45° 和 -45° 偏振分量的强度差异, S_3 表达的是圆偏振分量强度。光与介质的交互包括反射、折射等,可以用穆勒矩阵建模。假设入射光的斯托克斯矢量为 $\mathbf{S}_i$,经介质某一处反射后光的偏振态为 $\mathbf{S}_R = [S_{0R}, S_{1R}, S_{2R}, S_{3R}]^T$,假设该反射点处穆勒矩阵为 $\mathbf{M}_R$,则有:

$$\mathbf{S}_R = \mathbf{M}_R \mathbf{S}_i, \quad (2)$$

其中: $\mathbf{M}_R$ 与介质表面的方向及材质等相关。对于普通相机而言,只能测量到反射光中的 S_{0R} 分量,即光的强度。要测量偏振信息,需要使用偏振片来进行测量。对于偏振角为 θ 的偏振片,其透射行为可以用穆勒矩阵 $\mathbf{M}_\theta$ 建模。反射光经偏振角为 θ 的偏振片后得到的光偏振态为:

$$\mathbf{S}' = \mathbf{M}_\theta \mathbf{S}_R = \mathbf{M}_\theta \mathbf{M}_R \mathbf{S}_i, \quad (3)$$

其中, $\mathbf{M}_\theta$ 表示偏振片的穆勒矩阵,定义如下:

$$\mathbf{M}_\theta = \frac{1}{2} \begin{bmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \sin 2\theta \cos 2\theta & 0 \\ \sin 2\theta & \sin 2\theta \cos 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4)$$

经偏振片后相机测量的 S_1 的强度项为:

$$I_\theta = \frac{1}{2} (S_{0R} + S_{1R} \cos 2\theta + S_{2R} \sin 2\theta). \quad (5)$$

因此经偏振片后光的强度与经偏振片之前光的偏振分量有关,因此可以通过至少三次不同角度偏振片的光强测量可以恢复光的偏振分量 S_{0R}, S_{1R}, S_{2R} 。

本文采用的是分焦平面相机,该相机在CMOS上集成了四个不同方向的偏振片,如图2所示。

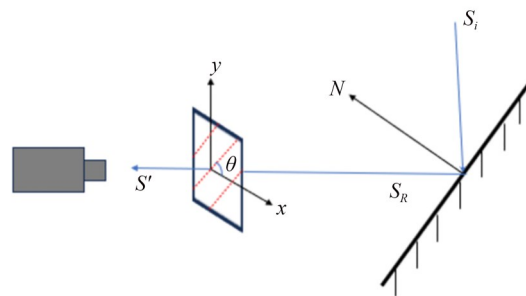


图1 反射光偏振测量原理示意图

Fig. 1 Schematic diagram of reflective light polarization measurement

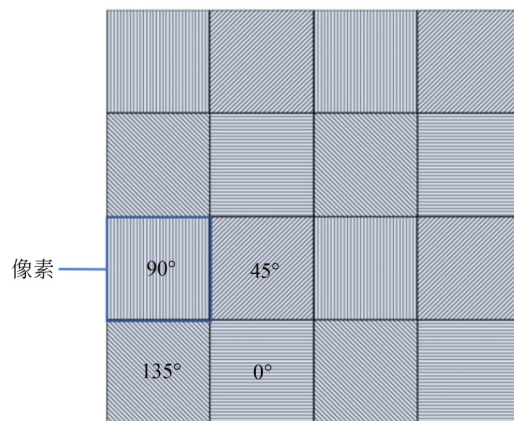


图2 偏振相机CMOS

Fig. 2 CMOS of polarization camera

因此可以通过一次拍摄获取四个不同偏振片角度下的测量结果。在反射过程中,以入射面为参考,以CMOS水平方向为 0° ,代入方程(5)可以计算反射光的斯托克斯矢量:

$$\begin{cases} S_{0R} = I_0 + I_{90} \\ S_{1R} = I_0 - I_{90} \\ S_{2R} = I_{45} - I_{135} \end{cases} \quad (6)$$

由于线偏振片只能测量光的线偏振情况,忽略光的圆偏振分量,则线偏振度 ρ 和线偏振角 φ 可以按式(7)计算:

$$\rho = \frac{\sqrt{S_{1R}^2 + S_{2R}^2}}{S_{0R}}, \quad (7)$$

$$\varphi = \frac{1}{2} \arctan 2(S_{2R}, S_{1R}), \quad (8)$$

其中:偏振角其物理含义为光的偏振方向与入射面夹角,偏振度则体现光的偏振程度,其范围在 $[0, 1]$ 之间。

3 双层路由稀疏自注意力

自注意力机制能够捕捉长距离上下文间的依赖关系,实现对特征全局信息的关注,经典的注意力按公式(9)计算^[20]:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{C}}\right)V, \quad (9)$$

其中: $Q \in R^{N_q \times C}$ 为查询, $K \in R^{N_k \times C}$, $V \in R^{N_v \times C}$ 为键值对。

对于多头自注意力而言 QKV 都来自输入 $X \in R^{N \times C}$ 的线性投影,对于图像特征而言, $N = H \times W$,其中 H, W 为图像的高和宽。多头自注意力计算的复杂度为 $O((HW)^2)^{[21]}$,其原理图如图 3 所示。

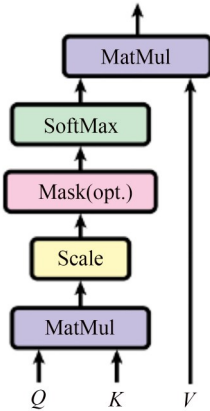


图 3 经典注意力计算原理图^[20]

Fig. 3 Schematic diagram of attention calculation

双层路由的自注意力与一般的自注意力计算有所不同,其计算原理如图 4 所示。

首先对特征划分为不重叠的 $S \times S$ 个特征

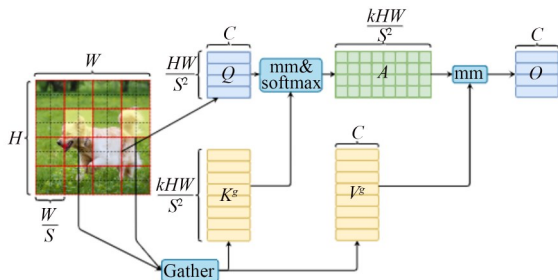


图 4 双层路由稀疏注意力计算原理图^[21]

Fig. 4 Schematic diagram of bi-level routing sparse attention calculation

块,将输入 $X \in R^{H \times W \times C}$ 重构为 $X^r \in R^{S^2 \times \frac{HW}{S^2} \times C}$,在对其分别做线性映射,得到查询 Q 、键 K 和值 V :

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v. \quad (10)$$

然后,通过构造有向图来确定不同区域之间的参与关系,并计算每个区域应该参与的区域。具体来说,通过逐区域计算查询和键的内积得到邻接矩阵 A^r :

$$A^r = Q^r (K^r)^T, \quad (11)$$

其中: $Q^r, K^r \in R^{S^2 \times C}$, $A^r \in R^{S^2 \times S^2}$, A^r 中元素表示不同区域之间大的语义关联性大小。随后通过对 A^r 中每行元素排序,保留每行的前 k 个最关联的区域,得到路由索引矩阵:

$$I^r = \text{topkIndex}(A^r) \in N^{S^2 \times k}. \quad (12)$$

也就是 I^r 的第 i 行包含 k 个对应最关联区域的下标。随后依据 I^r 中的下标重新构建键值对:

$$K^s = \text{gather}(K, I^r), V^s = \text{gather}(V, I^r), \quad (13)$$

其中, $K^s, V^s \in R^{S^2 \times \frac{kHW}{S^2} \times C}$ 。

再对查询 Q 应用筛选过的 K^s, V^s 来计算注意力输出:

$$O = Attention(Q, K^s, V^s) + LCE(V), \quad (14)$$

其中, LCE 为逐深度卷积,用作局部上下文增强项。这种注意力计算的复杂度为 $O((HW)^{\frac{4}{3}})^{[21]}$,由于在计算注意力时对键值对进行了筛选,每次查询并非对所有键值对都进行计算,而是只对最相关的区域进行注意力计算。因此,这是一种稀疏的注意力计算方案,能够避免一些无效的计算,筛除冗余信息,从而提高了计算效率。

4 结合偏振的法线估计方法

为了实现更优的偏振法线估计精度,本文采用了基于深度学习的偏振法线估计方法。通过改进偏振信息的表达方法,使得深度学习模型能够更好地从偏振信息中提取物理先验知识。同时,采用了双层路由的稀疏自注意力方案来增强模型对于全局上下文信息的利用与关联能力,从而提高了模型的性能。整体网络结构如图 5 所示。

对于从分焦平面式偏振相机拍摄的偏振图像,其包含 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 四个不同偏振角度下

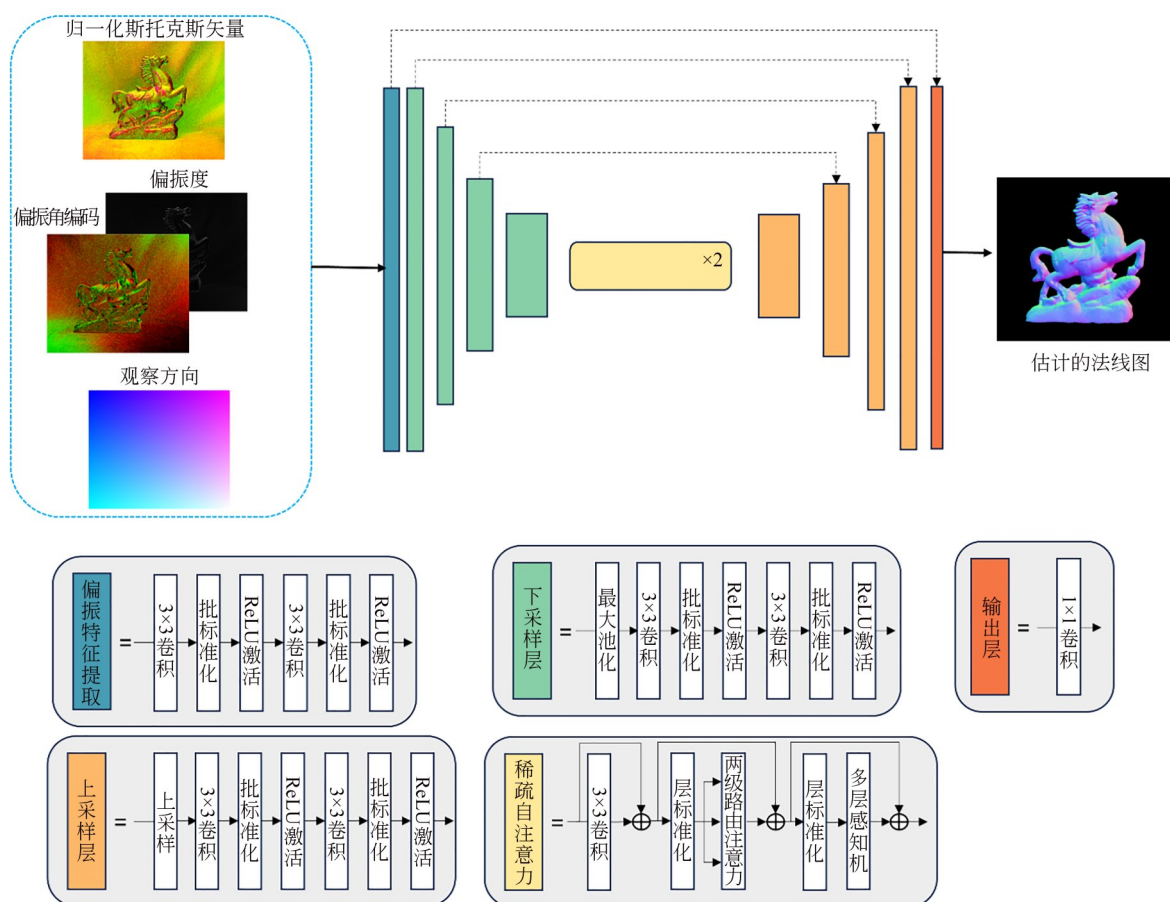


图 5 提出的方法整体结构示意图

Fig. 5 Overview of the proposed model

的强度图,在用于输入深度学习模型前,对其进行了一些处理以突出其中的物理先验信息,使得模型能够更好地提取其中的特征。首先按公式(6)计算斯托克斯矢量的前三个分量,并对 S_1, S_2 做归一化处理,对斯托克斯矢量中 S_1, S_2 归一化处理后得到的矢量称之为归一化的斯托克斯矢量 $\hat{S}$ 。除了计算归一化的斯托克斯矢量外,还计算了偏振度 ρ 和偏振角 φ ,并对偏振角按文献[18]中的方式计算偏振角编码 $\varphi_e = [\cos 2\varphi, \sin 2\varphi]$,这种偏振角表达方式相较于原始的偏振角好处在于在偏振角 π 歧义的存在并不对这种偏振角编码方式的结果产生影响。另外相机投影模型不同时,偏振角的计算方式实际存在不同^[19],两种不同方式计算的偏振角存在一定程度的误差,为了减轻这种影响,按文献[18]中计算方法计算各个位置像素点的观察方向 v_d 并将其作为模型输入,使得模型能够结合观察方向

最终估计出更准确的法线结果,因此,采用了以下形式的偏振信息编码方式作为模型的输入:

$$X = \text{cat}([S_0, \frac{S_1}{\sqrt{S_1^2 + S_2^2}}, \frac{S_2}{\sqrt{S_1^2 + S_2^2}}, \rho, \cos 2\varphi, \sin 2\varphi, v_d]), \quad (15)$$

其中:cat为拼接操作, v_d 的通道数为3,因此输入通道数为9。估计的法线可以表达为:

$$\tilde{n} = f(X), \quad (16)$$

其中: $f(\cdot)$ 表示提出的模型, $\tilde{n}$ 为估计的法线结果。

本文模型采用编码-解码架构实现,模型首先对输入偏振信息进行初步的特征提取,随后经4个下采样模块进行高级特征提取,经每个下采样模块后特征的宽高减小为一半同时特征的通道数增加为原来的2倍。随后利用2个双层路由的稀疏自注意力模块来获取全局上下文之间的关联信息,使得模型能够利用全局上下文间的信

息对局部消歧,对提取的特征进行增强。最后通过 4 个上采样模块对特征进行上采样恢复特征尺度,每个上采样模块将特征的宽高分别扩大为 2 倍,同时降低通道维度至原通道维度的一半,同时由于经多个下采样后特征分辨率下降容易丢失高频细节信息,因此为了改善最终恢复的结果,输入采样的特征与下采样后对应分辨率的特征通过跳跃连接拼接,以保持高频特征,减缓下采样过程高频特征逐渐丢失导致最后估计结果过于平滑的问题。

5 实验与分析

5.1 实验设置

本文模型采用 PyTorch 实现,并在一块 10 GB 显存的 Nvidia RTX 3080 上以批样本大小为 4 训练 500 个 epoch。本文使用 Adam 优化器^[22]更新模型参数,初始学习率设置为 0.000 1,并采用余弦退火策略来自动调整学习率,并采用余弦相似度作为损失函数训练。

采用 DeepSfP 数据集^[15]来训练和测试效果,它包含用分焦平面式偏振相机拍摄的多个不同目标偏振图像,以及通过高精度扫描仪得到的目标表面法线真值,该数据集法线真值质量高,光照条件包含室内室外常见的照明情况,按照数据集原始的训练测试集划分方式进行训练与结果测试,训练集和测试集分别包含 236 64 幅偏振图像,并在训练时将图像裁剪成 512×512 的图像块来进行数据增强。另外由于 DeepSfP 数据集并未提供相机参数,本文用像素坐标代替观察方向作为输入^[18]。此外为了测试模型的泛化性能,实

际拍摄了同一物体的不同光照和角度的五组数据,来定量测试算法性能。利用桌面 3D 扫描设备获取物体的 3D 形状,然后做图像与形状的对齐来估计每次拍摄时相机相对于物体的位姿^[6],然后使用 Mitsuba3^[18]进行渲染得到法线图真值。

对结果进行评估时,采用了多个指标从不同维度来评估结果,分别是按测试集中每张图像计算平均法线角度误差以及结构相似度,并对测试集所有图像估计结果分别计算平均值,用于衡量整体的精度以及从图像质量角度评估结果,结构相似度指标范围在 $[0, 1]$,越高说明结果越好。另外对测试集所有图像的有效像素的估计结果角度误差计算了六个指标,前三个是均值,中值和均方根误差,这三个指标越小结果越好,另外计算了 11.25° , 22.5° , 30.0° 指标,代表对应像素估计的角度误差分别小于对应角度所占的百分比,这一指标越高越好。

5.2 方法对比

将本文方法与现有的三个方法进行比较,包含一个由 Smith 等人提出的基于物理的方法^[23]和两个基于深度学习的方法,DeepSfP^[15]和 sfp-wild^[18]。其中,Smith 等人的方法按照其提供的代码在测试集上测试效果,而 DeepSfP 和 sfp-wild 方法对其按照同样的训练集训练以复现结果,使结果与原作者给出的效果接近。如表 1 所示,本文方法在所有指标上均超过了对比的方法,取得了最优的结果,相比于第二的结果,本文方法在平均角度误差上提升了约 1.2° ,中值和均方根误差均提升约 1.3° ,角度误差大约提升 10%,并且在 11.25° 误差以下像素占比提升了 5.2%。

表 1 DeepSfP 数据集量化评估结果

Tab. 1 Quantitative evaluation results of the DeepSfP dataset

方法	按图像		按像素					
	平均误差↓	SSIM↑	均值↓	中值↓	均方根↓	11.25° ↑	22.5° ↑	30.0° ↑
Smith et al.	36.24	0.874	36.29	31.81	43.08	12.0	32.2	46.7
DeepSfP	19.16	0.900	18.57	14.84	23.33	35.8	71.9	83.3
sfp-wild	14.56	0.924	14.02	10.03	19.22	55.3	83.1	90.3
本文方法	13.37	0.930	12.79	8.73	17.98	60.5	84.8	91.3

如图 6 为部分图像在不同方法下估计的法线图,可以看到本文方法能得到更准确的结果,同

时能保持较好的高频细节。基于物理的方法在比较的方法中表现最差,因为该方法包含许多假

设约束,比如光源为远距离点光源,这实际上在现实世界中并不都能满足。基于深度学习的方法效果好于基于物理的方法,并且本文方法结果上超过了其他基于深度学习的方法,这也说明了本文方法的有效性。

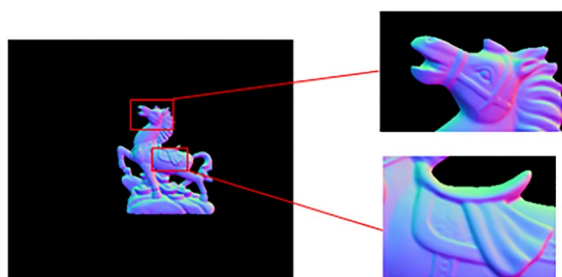
另外为了验证模型的泛化性,在自建的测试数据上进行效果测试,将本文方法与 sfp-wild 方法和 Smith 等人方法对比,定量测试结果如表 2 所示。可以看到本文方法相比于 sfp-wild 方法在角度误差上有超过 4° 的提升,在精度上各个指标



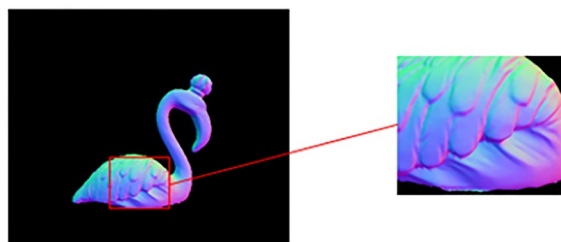
(a) 样本1灰度图
(a) Grey scale image(test sample 1)



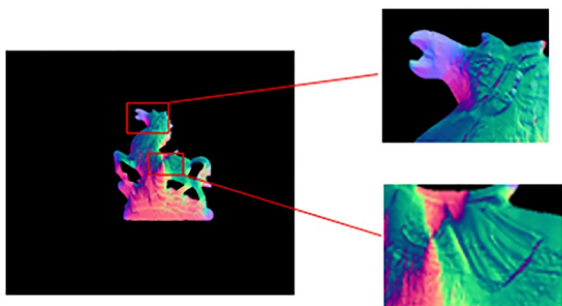
(b) 样本2灰度图
(b) Grey scale image(test sample 2)



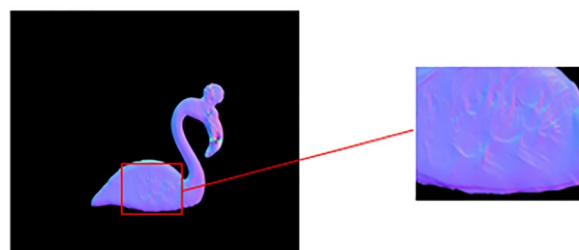
(c) 样本1法线图真值
(c) Ground truth normal map(test sample 1)



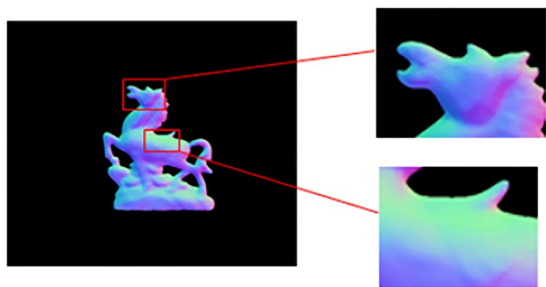
(d) 样本2法线图真值
(d) Ground truth normal map(test sample 2)



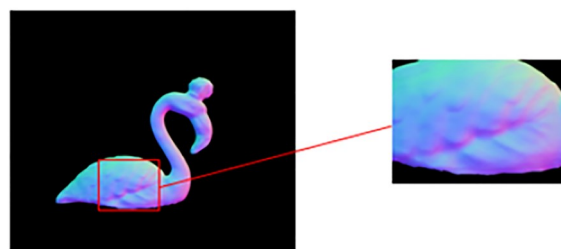
(e) 样本1 Smith *et al.*方法
(e) Smith *et al.*(test sample 1)



(f) 样本2 Smith *et al.*方法
(f) Smith *et al.*(test sample 2)



(g) 样本1 DeepSfp方法
(g) Deepsfp(test sample 1)



(h) 样本2 DeepSfp方法
(h) Deepsfp(test sample 2)

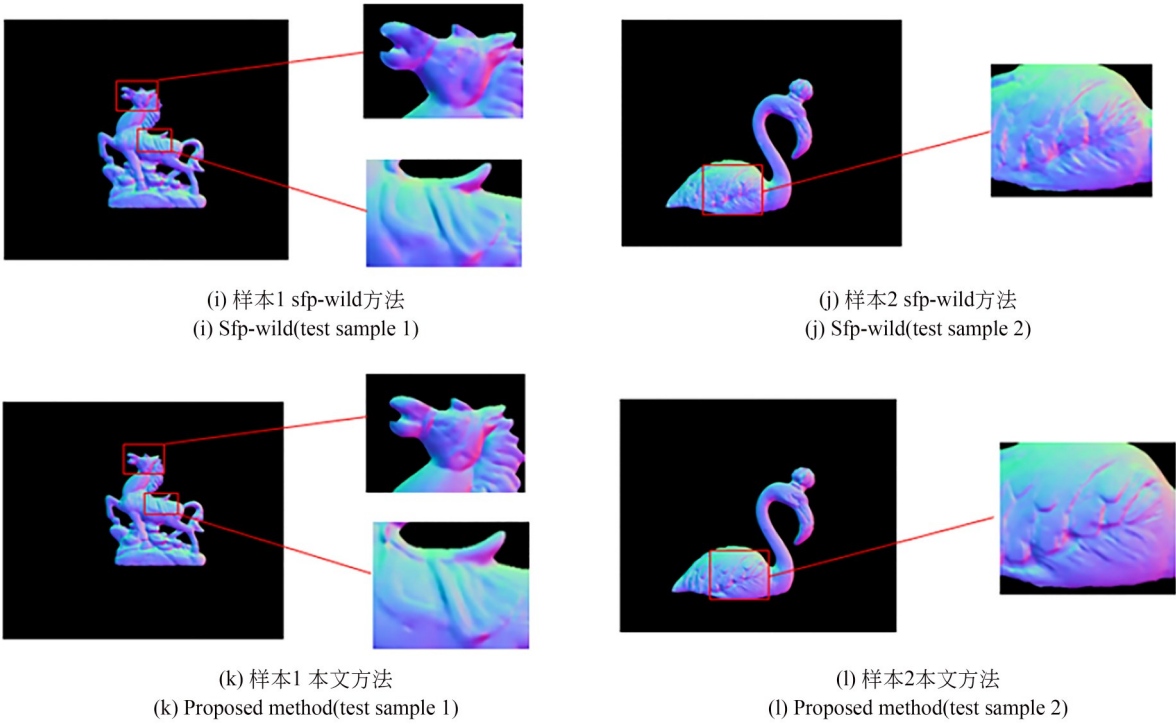


图 6 DeepSfP 数据集中部分测试样本结果

Fig. 6 Results of selected test samples from the DeepSfP dataset

表 2 自建测试数据不同方法定量测试效果

Tab. 2 Quantitative evaluation of different methods on our test data

方法	角度误差			精度		
	均值↓	中值↓	均方根误差↓	11. 25° ↑	22. 5° ↑	30. 0° ↑
Smith et al.	34. 49	33. 24	38. 19	6. 4	25. 0	42. 4
sfp-wild	22. 85	19. 49	27. 21	21. 7	58. 6	74. 7
本文方法	18. 26	15. 15	22. 53	34. 8	72. 4	85. 7

百分比都有超过 10% 的提升,其中 11. 25° 以内误差占比提升超过 13%,说明本文方法具有良好的泛化性,同时精度依旧对比比方法有较大提升。测试数据测试结果图如图 7 所示。

5.3 消融实验

在本文方法中,采用了连续两个自注意力模块,并对不同数量的自注意力模块进行了性

能测试。测试结果如表 3 所示,可以观察到增加自注意力模块的数量会导致模型性能有一定程度的下降,同时计算量也会增加。然而,在自注意力模块数量为 2 时,观察到模型性能达到最佳。因此,选择了自注意力模块数量为 2,以在取得最优性能的同时保持模型相对较小,计算代价更低。

表 3 自注意力模块数量对性能影响

Tab. 3 Results of the number of self-attention modules on performance

自注意力模块数量	按图像		按像素					
	平均误差↓	SSIM↑	均值↓	中值↓	均方根↓	11. 25° ↑	22. 5° ↑	30. 0° ↑
2	13. 37	0. 930	12. 79	8. 73	17. 98	60. 5	84. 8	91. 3
4	13. 53	0. 930	12. 92	8. 72	18. 20	60. 5	84. 4	91. 0
8	13. 83	0. 929	13. 19	9. 06	18. 34	59. 2	83. 8	90. 7

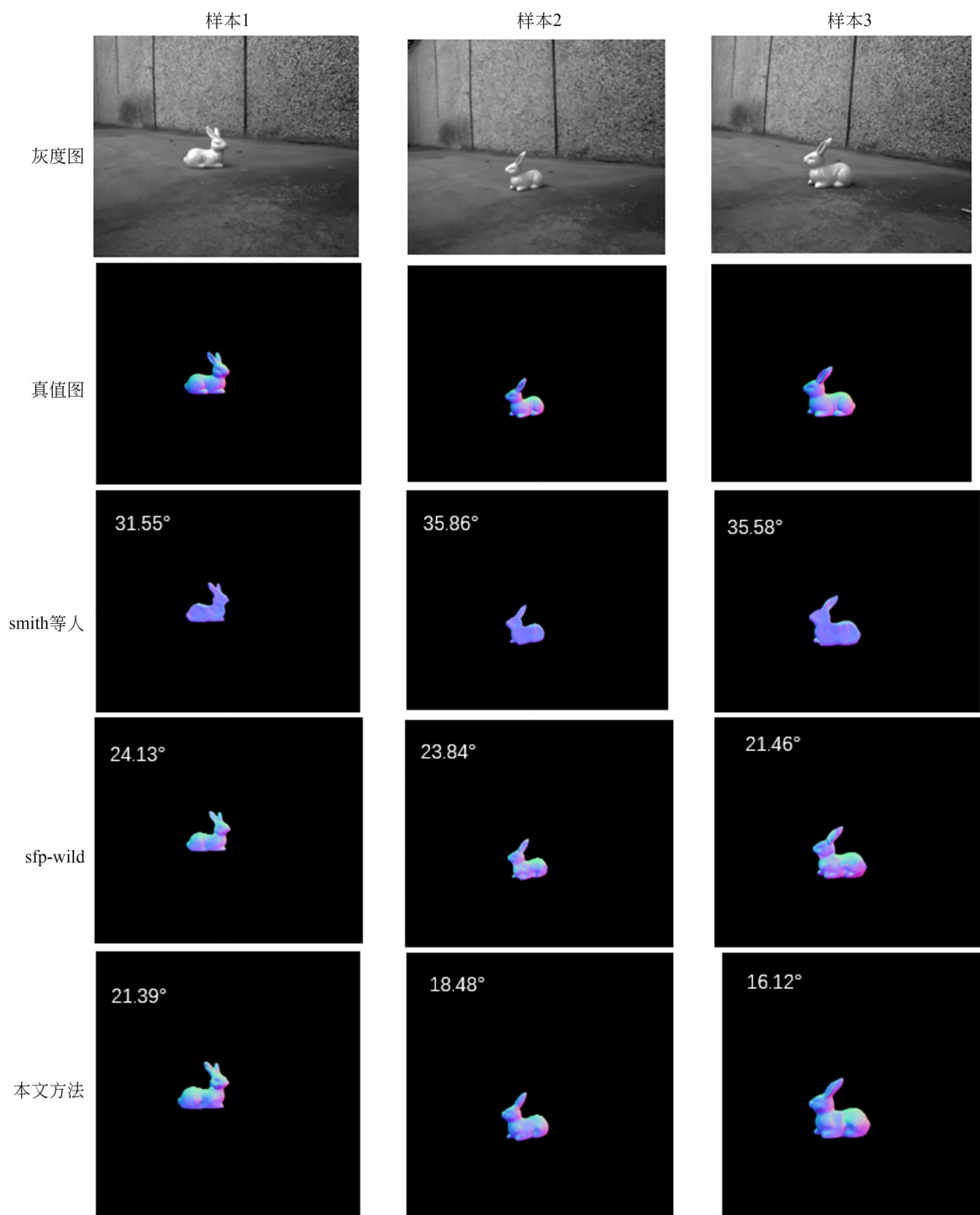


图7 不同方法在自建测试数据测试效果图

Fig. 7 Results of different methods on our test data

为了验证提出的偏振信息表示方式的有效性,将偏振信息表示方法替换为文献[18]中的偏振表示方法重新训练测试效果,结果如表4,可以观察到,提出的偏振表示方式相比于文献[18]中提出的方式在所有指标上都表现更佳,这也说明

了本文提出的偏振表示方式更加有效。

对双层路由的稀疏自注意力机制的有效性同样进行了验证,将模型中的自注意力替换为多头自注意力,重新训练模型测试效果,对比了误差以及每张图像平均处理时间来评估效率,采用

的自注意力相比多头自注意力,在误差和效率上 均有提升,结果如表 5 所示。

表 4 不同偏振表示方法对性能影响

Tab. 4 Results of different polarization representation methods on performance

偏振表示方式	按图像		按像素					
	平均误差↓	SSIM↑	均值↓	中值↓	均方根↓	11. 25° ↑	22. 5° ↑	30. 0° ↑
本文方式	13. 37	0. 930	12. 79	8. 73	17. 98	60. 5	84. 8	91. 3
Sfp-wild	13. 89	0. 928	13. 19	9. 01	18. 45	59. 4	83. 7	90. 3

表 5 不同注意力模块效果

Tab. 5 Results of different attention modules

注意力类型	按图像		按像素					
	平均误差↓	平均处理时间↓	均值↓	中值↓	均方根↓	11. 25° ↑	22. 5° ↑	30. 0° ↑
本文方式	13. 37	0. 548	12. 79	8. 73	17. 98	60. 5	84. 8	91. 3
多头注意力	13. 54	0. 562	12. 88	8. 82	18. 14	60. 5	84. 5	90. 9

6 结 论

本文提出了一种基于深度学习的偏振表面法线估计方法。通过构建高效的偏振表示方式,本文模型能够更好地利用偏振物理先验信息来估计表面法线。同时,采用基于双层路由的稀疏自注意力机制,加强了模型对全局上下文信息的提取能力,有助于模型更好地利用全局信息处理局部歧义。通过在高质量的真实物体数据集以及自建测试数据集上测试验证了方法的有效性。

相比于之前文献中在该数据集表现最佳的方法,本文提出的方法误差相比其降低了约 10%,平均角度误差达到 13. 37°,在 DeepSfp 公开数据集上取得了最佳的性能,并且在本文测试数据上平均误差 18. 26°,相比于 sfp-wild 方法提升了约 4°,在所有测试方法中取得最佳精度。此外,提出的偏振表示方式并未对物体材质做特别的限制,因此本文方法有望应用在更复杂的材质,例如包含金属材料等的物体表面法线估计任务中,从而拓宽了偏振重建方法的适用范围。

参考文献:

[1] CAO Y L, DING B J, CHEN J X, *et al.* Photo-metric-stereo-based defect detection system for metal parts[J]. *Sensors*, 2022, 22(21): 8374.

[2] ZHANG Y F, YE M X, XIANG S W, *et al.* Scene-level surface normal estimation from encoded polarization representation [C]. *International Conference on Intelligent Robotics and Applications. Singapore: Springer*, 2023: 3-14.

[3] YE H C K, MATSUDA N, HUANG X, *et al.* A streamlined photometric stereo framework for cultural heritage [C]. *European Conference on Computer Vision. Cham: Springer*, 2016: 738-752.

[4] 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.

[5] JOHNSON M K, ADELSON E H. Shape estimation in natural illumination [C]. *Proceedings of the IEEE/CVF IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[6] 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.

[7] YAO Z, LI K, FU Y, *et al.* GPS-Net: Graph-

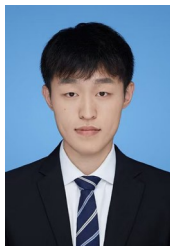
- based photometric stereo network[C]. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [8] CHEN G Y, HAN K, SHI B X, *et al.* Deep photometric stereo for non-lambertian surfaces[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 129-142.
- [9] 王国琿, 卢彦汀. 深度学习技术在光度立体三维重建中的应用[J]. *激光与光电子学进展*, 2023, 60(8): 3788/LOP230431.
- WANG G H, LU Y T. Application of deep learning technology to photometric stereo three-dimensional reconstruction [J]. *Laser & Optoelectronics Progress*, 2023, 60(8): 3788/LOP230431. (in Chinese)
- [10] LEI C Y, QI C Y, XIE J X, *et al.* Shape from polarization for complex scenes in the wild[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA. IEEE, 2022: 12622-12631.
- [11] KOSHIKAWA K. A polarimetric approach to shape understanding of glossy objects[J]. *Advances in Robotics*, 1979, 2(2): 190.
- [12] 王杰, 徐国明, 马健, 等. 轻量级注意力级联网络的偏振计算成像超分辨率重建[J]. *光学精密工程*, 2022, 30(19): 2404-2419.
- WANG J, XU G M, MA J, *et al.* Polarization computational imaging super-resolution reconstruction with lightweight attention cascading network [J]. *Opt. Precision Eng.*, 2022, 30(19): 2404-2419. (in Chinese)
- [13] TAAMAZYAN V, KADAMBI A, RASKAR R. Shape from mixed polarization [EB/OL]. 2016: *arXiv*: 1605.02066. <http://arxiv.org/abs/1605.02066>
- [14] YU Y, ZHU D Z, SMITH W A P. Shape-from-polarisation: a nonlinear least squares approach [C]. 2017 *IEEE International Conference on Computer Vision Workshops (ICCVW)*. Venice, Italy. IEEE, 2017: 2969-2976.
- [15] BA Y H, GILBERT A, WANG F, *et al.* Deep shape from polarization[C]. *European Conference on Computer Vision*. Cham: Springer, 2020: 554-571.
- [16] ZOU S H, ZUO X X, QIAN Y M, *et al.* 3D Human shape reconstruction from a polarization image [C]. *European Conference on Computer Vision*. Cham: Springer, 2020: 351-368.
- [17] 王楚宁, 赵开春. 基于暗通道先验原理的偏振图像去雾[J]. *光学精密工程*, 2023, 31(23): 3474-3481.
- WANG C N, ZHAO K C. Polarization image de-fogging based on dark channel prior principle [J]. *Opt. Precision Eng.*, 2023, 31(23): 3474-3481. (in Chinese)
- [18] JAKOB W, SPEIERER S, ROUSSEL N, *et al.* Mitsuba 3 renderer [EB]. 2022: <http://github.com/mitsuba-renderer/mitsuba3>
- [19] CHEN G C, HE L, GUAN Y S, *et al.* Perspective phase angle model for polarimetric 3d reconstruction [C]. *European Conference on Computer Vision*. Cham: Springer, 2022: 398-414.
- [20] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[EB/OL]. 2017: *arXiv*: 1706.03762. <http://arxiv.org/abs/1706.03762>
- [21] ZHU L, WANG X J, KE Z H, *et al.* BiFormer: vision transformer with bi-level routing attention [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 10323-10333.
- [22] KINGMA D P, BA J. Adam: a method for stochastic optimization[C]. *The International Conference on Learning Representations (ICLR)*, 2015.
- [23] SMITH W A P, RAMAMOORTHY R, TOZZA S. Height-from-polarisation with unknown lighting or albedo[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(12): 2875-2888.
- [24] DESCHAINTE V, LIN Y M, GHOSH A. Deep polarization imaging for 3D SHAPE and SV-BRDF ACQUISITION[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 15562-15571.
- [25] MAHMOUD A H, EL-MELEGY M T, FARAG A A. Direct method for shape recovery from POLARIZATION and shading[C]. 2012 *19th IEEE International Conference on Image Processing*. Oran-

do, FL, USA. IEEE, 2012: 1769-1772.

- [26] 张然, 桂心远, 成昊远, 等. 基于偏振成像的低光照强背景噪声下的目标位姿估计[J]. 光学精密工程, 2021, 29(4): 647-655.

ZHANG R, GUI X Y, CHENG H Y, *et al.* Target pose estimation based on polarization imaging in low light and strong background noise[J]. *Opt. Precision Eng.*, 2021, 29(4): 647-655. (in Chinese)

作者简介:



于智超(1998—),男,江西抚州人,硕士研究生,2021年于清华大学获得学士学位,主要从事偏振三维重建算法相关的研究。E-mail: yuzc21@mails.tsinghua.edu.cn

通讯作者:



赵开春(1973—),男,辽宁大连人,博士,副研究员,清华大学精密仪器系教师,研究领域为仿生微纳光栅阵列器件的设计、制备与测试,仿生导航传感器系统的设计、构建与环境试验,微小卫星的姿态测量与控制技术。E-mail: kaichunz@mail.tsinghua.edu.cn