

引用格式: SHAO Jiangang, ZHAO Jufeng, CUI Guangmang. Material Recognition Based on Multi-spectral-polarization Information and Physical Constraints[J]. Acta Photonica Sinica, 2026, 55(8):0830003

邵坚钢,赵巨峰,崔光芒. 基于多谱段-偏振信息与物理约束的材质识别[J].光子学报,2026,55(8):0830003

基于多谱段-偏振信息与物理约束的材质识别

邵坚钢^{1,2}, 赵巨峰^{1,2}, 崔光芒^{1,2}

(1 杭州电子科技大学 全省光电智能成像与空天感知重点实验室, 杭州 310018)

(2 杭州电子科技大学 电子信息学院, 杭州 310018)

摘 要:提出一种融合 RGB、光谱与偏振信息的多模态材质识别框架,采用物理先验驱动且结构适中的三支网络,对材质本征特性与表面反射机理进行统一建模,在不同表面工艺条件下实现对材质类别的稳定、准确识别。首先,框架分别提取 RGB 外观特征、光谱强度特征及偏振派生特征,并通过统一的归一化与同步裁剪对齐策略,保持跨模态输入的一致性;其次,模型在特征学习阶段引入注意力融合机制,联合挖掘光谱与偏振信息在材质本征属性与表面态差异表征上的互补信息;同时引入基于马吕斯定律的物理约束,使偏振角与偏振度响应符合成像机理,从而在特征空间有效抑制强镜面高光引起的异常响应并增强跨工艺类别的判别能力。实验结果表明,本文方法在整体材质识别任务中的准确率高达 99.51%,较单一模态及常规融合方法提升了 2.16% 至 22.03% 不等,并展现出跨主流骨干网络的架构普适性。

关键词:偏振成像;光谱成像;材质识别;多模态融合;深度学习

中图分类号: TP391.4

文献标识码: A

doi: 10.3788/gzxb20265508.0830003

0 引言

材质类型的高精度识别在工业检测与智能制造领域至关重要,对目标材质及其表面状态的稳定和准确表征已成为自动化系统视觉感知能力的关键核心。已有大量研究针对具体材质及其表面态开展识别检测工作,例如在零部件表面缺陷检测、腐蚀损伤识别等任务中,相关方法主要依赖目标外观表征所引起的视觉差异,间接区分不同材质类别及其局部退化状态^[1-2]。但在同材质异表面、弱纹理及复杂光照场景中,仅凭颜色或纹理特征难以实现精细感知。因此,如何在复杂表面与多样光照条件下稳定识别材质并联合理解表面态,仍是视觉感知的关键难题。

近年来,学界沿多条路径推进材质识别研究。仅利用光谱信息的方法已从早期通过光谱角映射度量谱形差异,演进为现阶段利用深度学习进行特征自主提取^[3];HE Kaiming^[4]等应用残差学习进行高光谱识别;CHOLLET F 等^[5]通过深度可分离卷积提升表达效率。此类方法参数规模友好、工程实现简洁,但由于忽略了表面态差异及反射机理的物理刻画,对“同材质、异工艺”场景的区分能力有限。研究者开始探索光谱-空间联合建模方法以进一步增强判别能力。LIU Bing 等^[6]提出深度多视角学习框架用以编码跨波段关联与空间上下文;TANG Wei 等^[7]提出三维八度卷积与空间-光谱注意力网络(3DOC-SSAN)以挖掘谱-空特征;ROY S K 等^[8]通过 HybridSN 级联三维与二维卷积。虽然性能有所提升,但此类方法对光照敏感,且因缺乏物理量显式建模,鲁棒性受限。在此基础上,Transformer 结构也被引入以建模长程依赖。MOHAMED S 等^[9]通过谱-空因子化自注意力提升高光谱分类性能;Xu Yichu 等^[10]提出的 DSFormer 则通过双选择融合机

基金项目:浙江省自然科学基金(LMS26F050010, LMS26F050009)

第一作者:邵坚钢,1974610383@qq.com

通讯作者:赵巨峰,dabaozjf@hdu.edu.cn

收稿日期:2026-04-19;录用日期:2026-07-06

<http://www.photon.ac.cn>

制自适应聚合不同感受野与重要 token。然而,此类方法普遍依赖较大的标注数据规模与相对理想的成像条件,对偏振等物理先验利用不足,泛化能力与物理可解释性有待提升。此外,自监督与跨域学习成为重要补充方向。TU Lilin 等^[11]提出谱-空分层掩码建模框架 S2HM2,使网络学习光谱维度间的非线性相关性以及空间结构的语义连续性后再迁移至下游分类任务;KANG X 等^[12]构建自监督谱-空预训练框架以减少标注依赖;Fang Zhuoqun 等^[13]则提出域适应框架 CLDA,在未标注目标域中迭代筛选高置信样本以缓解跨场景条件下的分布偏移问题。虽然提升了泛化能力,但其本质上仍未脱离单一强度观测的范畴,难以解耦高光引起的谱形畸变。部分研究尝试探索偏振维度对识别的增益效果。Zhao Yongqiang 等^[14]通过多波段偏振图像融合,验证了偏振参量在区分典型材质时的有效性,为材质分类提供了基础物理依据。尽管该类偏振成像在材质识别领域的相关工作引入了新维度,但往往视其为数据驱动的增强通道,忽视了相位关联与反射过程的物理约束,在高反光工况下稳定性不足。

针对上述问题,本文认为现有融合方法性能受限的根源在于:多模态信息仅在特征维度进行简单拼接,缺乏对物理成像机理的显式约束。虽然部分研究引入了偏振维度,但其未能充分利用偏振相位关联对表面微结构进行去耦,导致在强高光或弱纹理等极端工况下,模型难以从紊乱的强度响应中提取稳定的材质本征特征。与上述方法相比,本文提出一种融合 RGB、光谱与偏振三种模态的材质识别框架,核心为偏振增强多谱段材质识别网络(Polarization-Boosted Multi-Spectral Network, PBMS-Net)。本文方法分别提取 RGB 外观特征、光谱强度特征与表面微结构引导的偏振特征,在统一的数据策略下进行全局归一化以保留跨波段的相对强度,并对多模态数据执行同步裁剪以保证像素级对齐。在网络设计上,各分支采用中等规模的卷积主干结合浅层物理分支,通过显式引入物理一致性约束,将大量判别信息前移至先验建模阶段,从而减少对超大参数模型的依赖。模型层面通过多分支特征提取与注意力融合机制,从光谱与偏振两个物理维度联合刻画材质与表面属性,并行构建基于马吕斯定律的重建损失,将偏振对不同表面微结构的区分能力与光谱的材质本征特性进行深度融合,并从特征层面抑制强高光引起的异常响应,在保持适中模型规模与推理时延的前提下,提高了复杂工况下的识别稳定性。需要说明的是,现有公开数据集在通用场景覆盖方面已取得进展,而工业检测中“同材质异工艺、弱纹理、强反射”等细粒度任务仍缺乏针对性的多谱段-偏振数据支撑。基于此,本文搭建双相机多谱段-偏振成像系统并构建 MSP-6 数据集,重点覆盖材质-表面联合识别场景,为该类工业任务的模型训练、验证与对比提供实验基础。本文同时在公开数据集上开展对比实验,以评估方法的跨场景泛化能力。

1 原理与方法

1.1 偏振增强多谱段材质识别网络

PBMS-Net 整体框架如图 1 所示,由 RGB 分支、光谱分支、偏振分支以及特征融合与分类模块四部分构成。网络输入源于双相机同步采集的 RGB 图像(RGB.tif)与多谱段-偏振序列(band-pol.tif),数据进入网络后首先进行物理特征预处理与分流:RGB 图像经归一化与同步裁剪后映射为外观切片(RGB slice),多谱段-偏振序列经强度截断与归一化后依据四向偏振观测计算各谱段的斯托克斯参数(Stokes parameters, S_0, S_1, S_2),并将 6 波段的总强度 S_0 在波段维堆叠得到光谱强度切片(S_0 slice)。随后,由斯托克斯参数张量进一步计算线偏振度(Degree of Linear Polarization, DoLP)和线偏振角(Angle of Linear Polarization, AoLP),构造 DoLP、 $\sin(2AoLP)$ 、 $\cos(2AoLP)$ 三类偏振派生特征;再以总强度 S_0 的低强度分位数构建低强度掩膜,在偏振派生通道上抑制暗区噪声,并在通道维度堆叠得到 18 通道的偏振切片(PolarRaw slice)。随后,将这三组模态数据分别交由互不共享参数的特征提取器以端到端方式编码。其中,偏振分支特有的 PolarFuser 模块通过 1×1 卷积将 18 通道偏振张量压缩为 4 通道紧凑子空间,随后各分支均通过 ResNet18 主干网络,获得维度齐整但表征语义异构的高层特征表示,分别刻画材质的纹理外观、本征光谱响应及表面微结构属性。

在特征融合与判别阶段,各模态的高层特征首先通过拼接模块在通道维进行聚合,随即送入级联的通道注意力与空间注意力模块,对异构特征进行深层次交叉融合,自适应地在通道维抑制冗余噪声响应、在空间维增强材质显著区域,输出最终的联合特征表示。融合后的特征经全局平均池化汇聚为紧凑向量,并输入材质分类器得到最终判别结果。此外,为实现物理机理驱动的可靠学习,网络引入物理一致性约束路径:

物理头直接在偏振原特征 PolarRaw 上回归每个波段的斯托克斯参数,并通过 Malus 模块重建四个偏振角下的多角度光强。重建误差与分类交叉熵按权重构成混合损失函数,将物理光学定律显式嵌入特征学习空间,从而在复杂工况下实现对异构感知信息的联合判别,并保持较好的可解释性与稳定性。

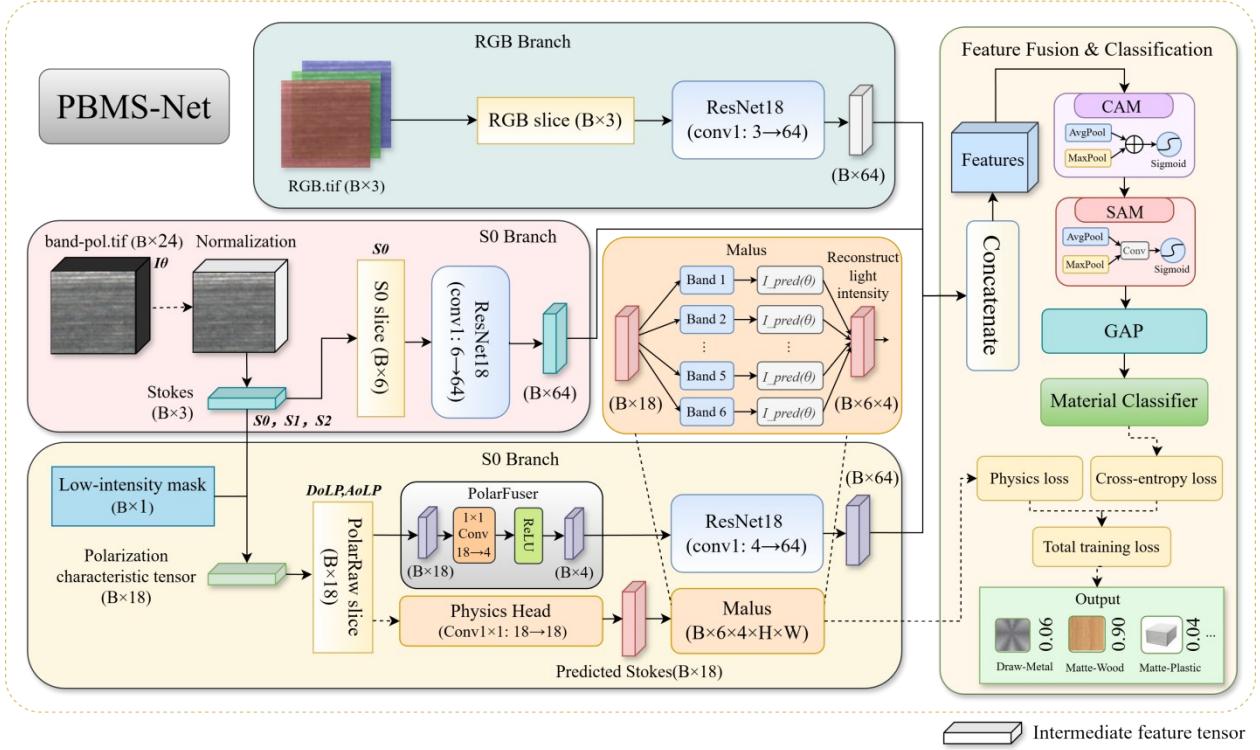


图1 PBMS-Net 整体框架
Fig.1 Overall framework of PBMS-Net

1.2 光谱-偏振数据预处理

针对光谱偏振数据的预处理过程,本文设置了一系列步骤来保证数据的有效性和物理一致性。这些步骤包括:对原始强度进行百分位截断与线性归一化、基于多偏振角图像计算斯托克斯参数并提取偏振特征,以及对计算得到的斯托克斯向量进行物理可行性检查与修正,从而为后续物理一致性约束和网络训练提供稳定、物理合理的输入。

首先,对每个偏振角 θ 与谱段 ϕ 下的原始灰度图像的强度 $I_{\theta, \phi}$ 进行百分位截断,抑制热噪声及异常高亮值。截断范围为每个波段的 0.5% 到 99.5% 的百分位,并将强度值线性归一化到 $[0, 1]$ 区间

$$LB = \text{percentile}(I_{\theta, \phi}, 0.5), UB = \text{percentile}(I_{\theta, \phi}, 99.5) \quad (1)$$

$$I_{\text{clip}} = \text{clip}(I, LB, UB), I_{\text{norm}} = \frac{I_{\text{clip}} - LB}{UB - LB} \quad (2)$$

该操作可有效抑制异常值对后续斯托克斯参数估计的影响。

基于归一化强度,计算每谱段的斯托克斯参数,包括总强度 S_0 、两组正交方向的线偏振分量 Q 和 U

$$S_0 = \frac{I_0 + I_{45} + I_{90} + I_{135}}{4}, Q = I_0 - I_{90}, U = I_{45} - I_{135} \quad (3)$$

通过这些斯托克斯参数进一步计算偏振特征

$$\text{DoLP} = \frac{\sqrt{Q^2 + U^2}}{S_0 + \epsilon}, \text{AoLP} = 0.5 \cdot \text{atan2}(U, Q) \quad (4)$$

式中, ϵ 为除零保护常数。

考虑到实际偏振成像系统的非理想性以及测量噪声,直接由式(3)、(4)计算得到的偏振量可能出现少量物理超界样本。为保证输入数据的物理可行性,在计算斯托克斯参数后引入物理一致性检查与修正:对

每个像素,当 $\sqrt{Q^2 + U^2} > S_0$ 时,将 (Q, U) 投影到以原点为中心、半径为 S_0 的圆盘上

$$(Q', U') = S_0 \cdot \frac{(Q, U)}{\sqrt{Q^2 + U^2}} \quad (5)$$

并用修正后的 (Q', U') 重新计算 DoLP 与 AoLP。对于未触发该条件的像素,则保持原始 (Q, U) 不变。该投影操作等价于将斯托克斯向量约束到满足如下条件的物理可行区域

$$0 \leq \text{DoLP} \leq 1, S_0^2 \geq Q'^2 + U'^2 \quad (6)$$

该操作有效消除了噪声和系统非理想性导致的超界问题。下文物理一致性损失中所使用的偏振监督量均基于上述修正后的斯托克斯向量。

在特征构建阶段,采用 S_0 作为强度通道,并基于修正后的偏振量构造偏振特征表示。具体地,使用 DoLP 表征偏振强度,将 AoLP 映射为 $\cos(2 \cdot \text{AoLP})$ 与 $\sin(2 \cdot \text{AoLP})$ 两个分量,从而得到一个由 $(S_0, \text{DoLP}, \cos(2 \cdot \text{AoLP}), \sin(2 \cdot \text{AoLP}))$ 组成的多通道偏振特征张量。实验中,出于网络结构和与 RGB 图像量纲对齐的考虑,选取其中三个通道作为输入特征,并对各通道进行统一的线性缩放与归一化,使其数值分布与常规强度/颜色通道处于相近范围。

此外,为减弱不同场景间曝光差异带来的影响,对所有样本的强度通道 S_0 进行统一的尺度重标定,构建基于分位数的全局强度尺度 I_{raw} ,使得不同场景的输入在整体亮度上保持可比性。经上述预处理后,输入网络的光谱-偏振数据同时满足基本物理约束并具有良好的数值稳定性。

1.3 光谱-偏振协同的三支并行业务与注意力融合框架

在完成基于偏振特性的高光抑制与特征增强处理后,为进一步提升弱纹理与高反光场景下的判别稳定性与可解释性,需要将四角度偏振观测在统一动态范围下有效融入卷积特征学习流程。为此,本文以“RGB-强度谱-偏振谱”三支并行为基础,通过轻量的极化融合与注意力的协同加权,实现对多维信息的高效聚合;并在语义分类之上引入基于物理一致性与光谱平滑的软约束,使网络在复杂光照与弱纹理条件下依旧保持稳健泛化。

如图1所示,整体管线首先将偏振与强度特征以统一表达对齐到网络可学习的特征空间。随后,偏振谱经由极化融合模块完成轻量线性映射,强度谱与 RGB 分支并行进入残差主干抽取鲁棒特征,三支路的高层语义再通过注意力机制完成协同融合,最终输出材质分类结果;在判别的同时,结合可微物理重建与光谱正则作为软约束,保证特征与物理观测的一致性。

为将偏振谱有效纳入残差主干,本文采用 1×1 线性映射的极化融合模块,将18通道压缩为4通道,既实现降维又保留主要偏振基

$$X_{\text{pol}} = \text{Conv}_{1 \times 1}(X_{\text{pol_raw}}; W_{\text{fuse}}) \quad (7)$$

融合后的偏振谱与 RGB 与强度谱共同输入到三条残差主干

$$F_{\text{rgb}} = \Phi_{\text{rgb}}(X_{\text{rgb}}), F_{s_0} = \Phi_{s_0}(X_{s_0}), F_{\text{pol}} = \Phi_{\text{pol}}(X_{\text{pol}}) \quad (8)$$

并在通道维拼接得到联合特征

$$F = [F_{\text{rgb}}; F_{s_0}; F_{\text{pol}}] \quad (9)$$

式中, Φ 表示 ResNet18 的卷积子网络。

三支的设计动机在于充分利用 RGB、强度谱(宏观反射与照度)以及偏振谱(微表面取向与反射分量)之间的互补性;在弱纹理与强反光条件下,单一模态常难以稳定分辨,而多模态协同能够显著提升特征可分性。

为在弱纹理场景下强化关键响应,本文引入卷积块注意力模块(Convolutional Block Attention Module, CBAM)^[15]对拼接特征进行加权,并参考了 CBAM 的“通道-空间串联”思想,保持其轻量实现形式,以避免在三支框架中引入过高参数开销。在任务适配上,本文未将注意力模块分别置于各单模态分支内部,而是作用于 RGB、强度谱与偏振谱拼接后的联合特征,以显式建模跨模态互补关系并抑制单一模态噪声放大。进一步地,考虑到偏振信息在高光边缘与局部反射热点处呈现更强空间异质性,本文采用先通道重标定、后空间筛选的顺序,以提升弱纹理与强反光区域的响应稳定性。需要注意的是,与通用视觉任务不同,本研究针对光谱与偏振特征的物理异构性,利用该模块实现跨模态信息的自适应筛选。在注意力模块中,采用通

道注意力对拼接特征 $F \in \mathbb{R}^{C \times H' \times W'}$ 的全局统计进行门控

$$a = \sigma(W_2 \phi(W_1 \text{GAP}(F)) + W_2 \phi(W_1 \text{GMP}(F))), F_{\text{ch}} = a \odot F \quad (10)$$

式中, GAP、GMP 分别为全局平均池化与最大池化, $\phi(\cdot)$ 为非线性激活, $\sigma(\cdot)$ 为 Sigmoid 激活函数。进一步地, 空间注意力在 F_{ch} 的空间维进行显著性筛选

$$m = \sigma(\text{Conv}_{7 \times 7}([\text{Avg}(F_{\text{ch}}), \text{Max}(F_{\text{ch}})])), F_{\text{sp}} = m \odot F_{\text{ch}} \quad (11)$$

两类注意力相互协同: 通道维提升判别子空间的响应, 空间维强化结构显著区域, 使融合特征在弱纹理与高反光区域具备更高的可分性与稳定性。相比于直接使用 SE 等单一维度的注意力机制, CBAM 这种串联的空间与通道加权更契合偏振谱在物体边缘与反光点上的空间异质特性。随后, F_{sp} 经全局平均池化得到紧凑表征

$$z = \text{GAP}(F_{\text{sp}}) \quad (12)$$

并分别进入线性判别器, 输出类别后验

$$y_{\text{mat}} = \text{softmax}(W_{\text{mat}} z + b_{\text{mat}}) \quad (13)$$

为保证融合特征与物理观测的一致性, 本文在偏振分支旁路引入极简可微的物理头(Physics Head), 直接在原始偏振谱 $X_{\text{pol_raw}}$ 上回归斯托克斯分量

$$[\hat{S}_0, \hat{S}_1, \hat{S}_2] = \text{split}(\text{Conv}_{1 \times 1}(X_{\text{pol_raw}}; W_{\text{phys}})) \quad (14)$$

并在前向模型层面并按照式(16)执行四角度的强度重建, 构建重建损失, 使得网络在学习判别表征的同时, 亦对可解释的物理量进行显式建模。

在训练阶段, 三支并行的注意力融合保证多模态信息的充分表达, 物理头的可微重建则为特征学习提供了“观测-物理”的一致性约束。为避免物理分支对分类主干造成干扰, 本文采用参数组分离与学习率缩放策略, 在保证稳定性的同时提升软约束的收敛效率。 y_{mat} 则在推理阶段作为最终判别输出, 同时可选择性地给出斯托克斯分量 S_0, S_1, S_2 与四角度强度的物理重建, 可用于后续的物理一致性检验与可视化。

本方法将四角度偏振观测在统一动态范围下表达为可学习的强度谱与偏振谱; 通过“RGB-强度谱-偏振谱”三支并行的注意力融合, 充分利用颜色、宏观强度与微表面偏振的互补性; 在分类任务之上叠加马吕斯一致性与光谱平滑的软约束, 使网络在弱纹理与高反光场景下获得更稳健的判别与更强的物理可解释性, 从而实现“偏振增强-特征融合-物理一致”的端到端优化。

1.4 反射机理与联合损失函数设计

从表面反射机理来看, 材料的成像辐射不仅取决于其光谱反射特性, 还与表面微观几何结构密切相关。经典双向反射分布函数(Bidirectional Reflectance Distribution Function, BRDF)理论^[16]通常采用微表面假设描述粗糙表面的镜面反射行为, 其中最具代表性的 Torrance-Sparrow 微表面模型^[17]可表示为

$$f_r = \frac{D \cdot G \cdot F}{4 \cos \theta_i \cos \theta_r} \quad (15)$$

式中, D, G, F 分别为微表面法向分布函数、几何遮蔽函数与非涅耳反射系数。表面较光滑时, 微表面法向分布更集中, 镜面反射中的偏振特征更易保持; 而当表面粗糙度增大时, 微面元朝向分散会导致不同偏振态叠加平均, 使偏振响应减弱。因此, 偏振信息不仅反映材料反射特性, 也能够一定程度上表征表面粗糙度与表面工艺状态。基于此, 本文将偏振成像物理规律引入网络训练过程, 以增强模型对真实表面反射机理的约束。

在物理重建损失计算方面, 为了在保证偏振参数物理可行性的前提下约束网络输出, 本文采用基于马吕斯定律的前向模型构建预测强度。具体地, 利用经过预处理阶段物理可行性检查与投影修正后的斯托克斯向量(S_0, S_1, S_2)作为物理监督量, 对给定偏振角 θ 处的预测强度表示为

$$I_{\text{pred}}(\theta) = 0.5(S_0 + S_1 \cdot \cos(2\theta) + S_2 \cdot \sin(2\theta)) \quad (16)$$

式中, S_0, S_1, S_2 为由物理头回归得到的斯托克斯分量。由于参与损失计算的斯托克斯向量已在预处理阶段被投影到满足式(6)的物理可行区域, 故物理一致性损失不会驱动网络去拟合由系统非理想性或噪声导致的超界样本, 而是逼近物理上可实现的偏振状态。

在综合物理一致性、光谱连续性以及多模态分支可学习性的基础上, 本文采用联合损失进行端到端

优化

$$L_{\text{total}} = \lambda_{\text{phys}} L_{\text{phys}} + \lambda_{\text{diff}} L_{\text{diff}} + \lambda_{\text{spec}} L_{\text{spec}} + \lambda_{\text{rgb}} L_{\text{rgb}} \quad (17)$$

其中,物理重建项通过加权均方误差约束预测强度与观测强度的一致性

$$L_{\text{phys}} = \text{mean}(W \odot (I_{\text{pred}} - I_{\text{raw}})^2) \quad (18)$$

为刻画材料光谱响应随波长变化的连续性,本文在同一空间位置与同一偏振态下对相邻波段施加差分约束,即固定偏振态索引 p ,沿波段索引 b 对 b 与 $b+1$ 的响应变化趋势进行对齐,使预测的“光谱变化趋势”与观测保持一致

$$L_{\text{diff}} = \text{mean}(|\Delta_b I_{\text{pred}} - \Delta_b I_{\text{raw}}|) \quad (19)$$

此外,为避免融合训练中单一模态表征被削弱,在融合前对光谱分支与RGB分支施加轻量辅助监督,分别定义为

$$L_{\text{spec}} = \text{mean}(|I_{\text{pred}}^{\text{spec}} - I_{\text{raw}}|), L_{\text{rgb}} = \text{mean}(|I_{\text{pred}}^{\text{rgb}} - I_{\text{raw}}|) \quad (20)$$

式中, W 为与 I_{raw} 同维度的权重张量, λ_{phys} 、 λ_{diff} 、 λ_{spec} 、 λ_{rgb} 为各损失项权重系数。在最终的端到端训练配置中,以主任务分类损失为基准(权重默认为1.0),物理重建项权重 λ_{phys} 设为0.3,相邻波段差分项权重 λ_{diff} 设为0.05,光谱与RGB分支辅助监督权重 λ_{spec} 和 λ_{rgb} 均统一设为0.1。上述权重系数的有效性与必要性通过消融实验与敏感性分析进行了验证。

2 系统与数据集

2.1 实验系统与流程

本研究搭建的双相机多谱段-偏振融合成像实验系统整体结构如图2所示。图中,1为高密度可调光强LED面光源,2为被测目标样本,3为分光镜,4为多通道滤光轮,5为偏振旋转片,6为光谱-偏振支路成像物镜,7为灰度成像相机,8为RGB支路成像物镜,9为RGB成像相机。为保证实验数据的物理一致性与可重建性,整个成像系统依据同轴分光式设计,将目标反射光引入分光镜,并将光束分别分配至两条相互垂直的光轴,沿第一光轴依次布置4、5、6、7构成光谱-偏振支路,用于采集多波段、多偏振角度的光谱与偏振数据;沿第二光轴则布置8与9构成RGB支路,用于采集高分辨率彩色图像。两路光学通道均采用同系列、同分辨率的工业相机与一致焦距的成像镜头,以确保空间视场完全重合,从而在后续光谱-偏振特征与RGB纹理信息的融合处理中避免视角误差与时序偏移。

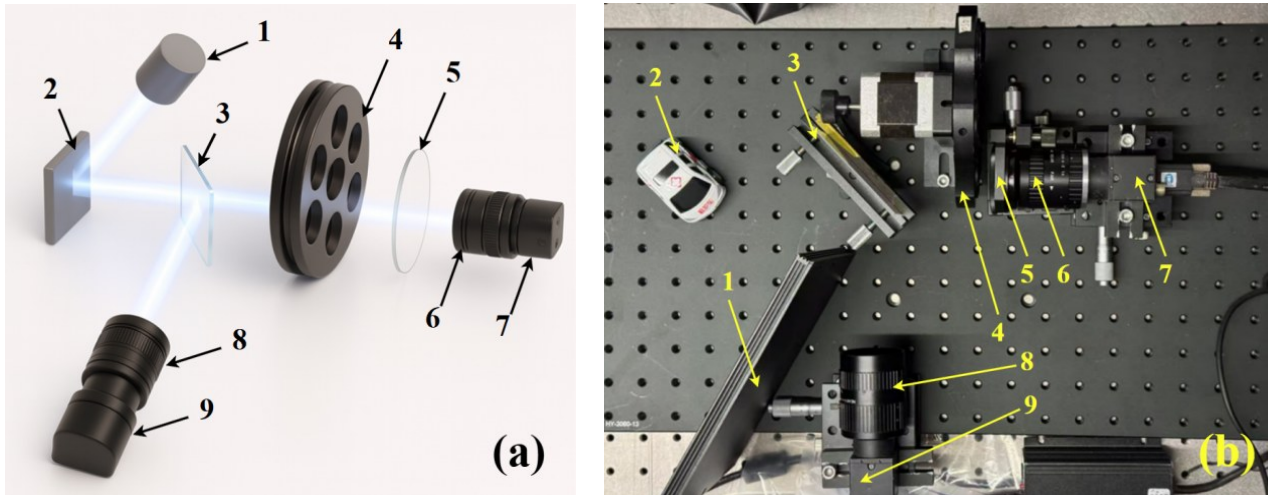


图2 双相机多谱段-偏振融合成像实验系统。(a)装置示意图;(b)装置实物图

Fig.2 Dual-camera multispectral-polarization fusion imaging experimental system. (a) Schematic diagram of the device; (b) Photograph of the device

1—High-density adjustable light intensity LED surface light source; 2—Target sample under test; 3—Beam splitter; 4—Multi-channel filter wheel; 5—Polarization rotator; 6—Spectral-polarization branch imaging objective lens; 7—Grayscale imaging camera; 8—RGB branch imaging objective lens; 9—RGB imaging camera

在光谱-偏振通道中,目标反射光首先经过电机驱动的多通道滤光轮,实现450nm、480nm、510nm、550nm、590nm与640nm六个窄带波段的顺序采集。滤光片带宽控制在10nm,可有效覆盖可见光区具有代表性的材质反射变化段,使系统获得足够的光谱区分能力。在滤光片之后,入射光经角度可控的偏振旋转片实现对目标的偏振态选择,使系统能够在每个波段下获得四幅偏振图像。最终形成包含6种光谱波段与4种偏振角度的多通道数据立方体,既保留了材质在不同波长下的反射特征,又同步捕获了各波段的偏振依赖性。

系统在实验前首先利用棋盘格标定靶对两路相机分别完成内参及畸变参数标定,并基于同一靶标视图求解两相机之间的外参关系,建立两路成像平面之间的初始几何映射;在此基础上,再在标定靶上提取角点、边缘等特征,根据匹配结果估计最优仿射变换参数,对RGB图像进行重采样与几何校正,使其在主视场范围内与光谱-偏振通道实现近似像素级的视角对齐。随后,系统基于四向偏振图像计算斯托克斯参数,进一步求解DoLP与AoLP,建立从灰度测量到偏振物理量的映射关系。与光谱-偏振通道垂直的RGB成像通道经分光镜折射实现光路分离,采用同系列工业相机构建实时成像子系统。两台相机通过计算机软件触发实现精确同步,当光谱-偏振相机完成一轮多通道采集时,RGB相机同时输出空间配准的高分辨率彩色图像。RGB图像既可作为光谱-偏振数据的纹理参照,也用于数据配准标定与深度模型的外观先验补充。

成像系统装置被固定于精密光学平台上,外部使用吸光率大于99%的黑色遮光结构构建半封闭暗室环境,以抑制环境光干扰。实验对象选取金属、塑料、木材等具有典型表面差异的材质样本,表面粗糙度、工艺类型等种类均涵盖多种变化情况。光源采用可调光强的高密度LED阵列发光板,波长覆盖滤光轮所需的450nm-640nm,并可保持固定光强持续输出,以保证系统工作的稳定性。上述实验平台采用双相机、多通道滤光轮与偏振旋转片组成的时间复用式多谱段-偏振采集方案,目的是保证多模态数据的物理一致性与实验可重复性,因此更适用于方法验证与高质量数据采集;在实际部署中,硬件配置可根据检测精度、响应速度与系统集成度要求进行适当简化,例如通过筛选关键波段、减少偏振角度或采用集成式偏振成像器件等方式,在保证主要判别信息的前提下降低系统复杂度。

整个数据采集流程如下:系统首先完成空间与偏振的联合标定,并基于偏振标定靶建立偏振物理量映射模型。随后,在稳定光照条件下,对样本进行多轮扫描式采集:滤光轮在六个波段之间顺序转位,每当定位完成后触发旋转片完成四个偏振角的快速切换并对应完成曝光;RGB相机同步输出其对应的彩色帧。最终形成结构化的多模态数据组,包括RGB图像、六波段光谱灰度图像以及对应的四向偏振图像,且确保了数据间的像素级空间配准与场景一致性,可为后续的光谱-偏振联合特征提取、材质类别识别与表面状态估计提供稳定可靠的实验基础。

2.2 MSP-6数据库构建

依托上述成像平台,课题组自主采集并构建了一套面向材质-表面联合识别的多模态数据库(Multi-Spectral-Polarization with 6 spectral bands, MSP-6)。对每一目标场景,光谱-偏振通道在六个窄带波段下分别获取四偏振角下的灰度图像,共形成24幅光谱-偏振图像;RGB通道同步输出一幅三通道彩色图像,两路图像原始分辨率均为 1600×1200 像素。因此,每一份原始数据均包含24幅光谱-偏振灰度图像与1幅RGB图像。从输入构成看,RGB输入对应1幅三通道彩色图像,主要提供外观与纹理信息;光谱输入对应6个窄带波段,用于表征跨波长反射响应;偏振输入则在各光谱波段下进一步引入4个偏振角观测,用于描述表面微结构及反射特性。相较于RGB输入,光谱与偏振模态在观测维度上更为丰富,信息维度更高。为控制数据规模,本文未采用原始图像堆叠方式,而是先通过斯托克斯参数计算构造偏振派生特征,并利用PolarFuser模块进行压缩表征,以在保留核心物理属性的同时优化计算效率。

采集时,每幅原始图像视场内只包含同质表面样本。在统一坐标系下,于非边缘区域随机生成裁剪窗口,并同时作用于24幅光谱-偏振图像及对应RGB图像,从原始图像中截取若干子图。裁剪尺寸可按应用灵活设定:大尺寸用于覆盖完整表面纹理,小尺寸用于小目标或微区识别;本文实验选用 128×128 像素,以兼顾细节与样本量。得益于这一策略,每个子样本内部各通道均实现像素级严格对齐,为后续光谱-偏振物理特征与RGB纹理信息的深度融合提供了可靠前提。

选取的样本涵盖金属、塑料、木材等典型工程材质及拉丝、哑光、镜面等多种表面形貌,具体包括拉丝金

属、磨砂金属、镜面金属、磨砂塑料、光滑塑料、粗糙木材等类型。对于每一类型,对不同目标材质样本采集多幅尺寸为 1600×1200 像素的原始图像,在每幅图像中预先划定远离边缘和遮挡物的安全裁剪区域,仅在该区域内随机生成多个 128×128 裁剪窗口,共得合计4185份不同种类样本。其中,包含各类表面金属材质样本1860份,各类表面塑料材质样本1395份,木质样本930份,每份由24幅光谱-偏振灰度图和1幅RGB图像组成并附材质标签。

在模型训练与评估阶段,MSP-6数据库的训练集、验证集和测试集划分均以原始样本为最小单位进行。实验设置在原始图像层面随机划分训练、验证、测试三类集合,随后仅在各自集合中图像的安全裁剪区域内生成 128×128 子图。由同一幅原始大图裁剪得到的所有子图被完整划入同一数据子集,确保其不会同时出现在训练集和测试集中,从而避免了样本泄漏问题,保证了识别性能评估的独立性和客观性。

图3展示了采集一典型拉丝金属样本得到的 6×4 通道原始灰度图像序列:随波长增加整体亮度与纹理对比度呈渐变,不同偏振角下高光与暗部灰度分布差异明显,反映了材质光谱反射特性及各向异性偏振响应,为后续斯托克斯参量与线偏振度、线偏振角反演以及材质-表面联合识别提供了丰富物理信息。

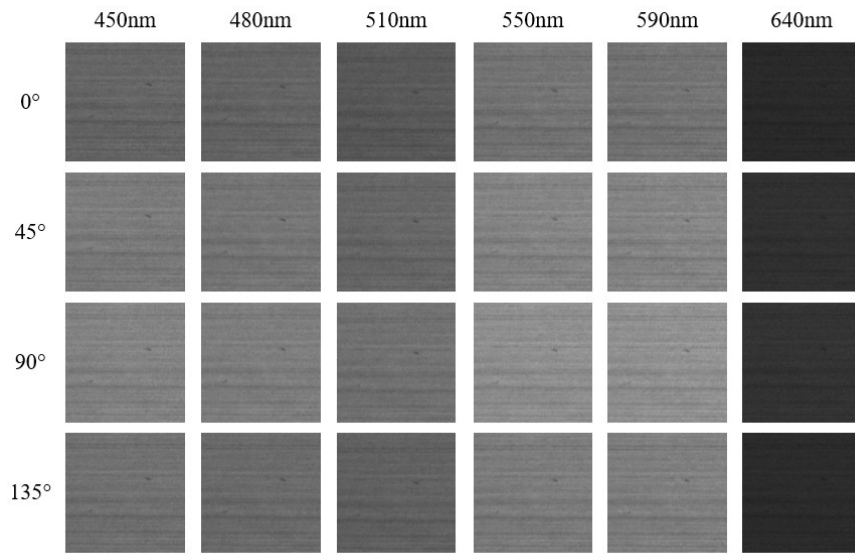


图3 数据库采集的典型目标样本(六波段 $\times$ 四偏振角)成像序列示意

Fig.3 Spectral-polarization image sequence (six bands $\times$ four polarization angles) of a typical target acquired by the proposed system

为进一步体现本文系统对不同材料表面信息的表征能力,图4给出了部分典型材料的样本图。其中,图4(a)~(d)为系统在同一波段和特定偏振角条件下采集的表面局部灰度图,图4(e)~(h)为对应样品的彩色RGB图像,图4(i)~(l)为相应区域的平均光谱强度与平均线偏振度随波长变化的特征曲线,横坐标为波长,左纵坐标为平均光谱强度,右纵坐标为平均线偏振度。每行从左至右分别为光滑塑料、磨砂金属、镜面金属和粗糙木质。不难发现,不同材料主要表现出表面纹理、粗糙度、颗粒度及镜面反射特性的差异,这些信息有助于表征材料的空间形貌特征。而材料可分性的核心在于其跨波段光谱响应差异,即不同材料具有不同的光谱反射率变化规律。图4(i)~(l)进一步表明,四类材料在波长维度上的平均光谱强度变化趋势存在明显差别,同时其线偏振度曲线也呈现出不同的变化特征。因而,表面形貌信息提供了空间纹理层面的辅助判据,而不同材料固有的光谱反射率曲线则是实现材质光谱可分性的根本物理依据;偏振响应则进一步增强了对表面反射机制和材料属性差异的表征能力。上述结果共同说明,本文所构建的光谱-偏振-RGB联合信息能够从光谱特性、偏振特性和空间纹理三个层面提升不同材质表面的区分能力。

2.3 公共数据集处理

MSP-6数据集的设计目标是服务于工业检测中的同材质异工艺识别,不追求大规模通用类别覆盖。为增强对真实场景的外推验证,课题组引入Princeton University和POSTECH联合发布的CVPR 2024数据集Spectro-polarimetric Real-world Dataset (SPRD)作为公共数据集进行对比实验^[18]。该数据集于自然环境中

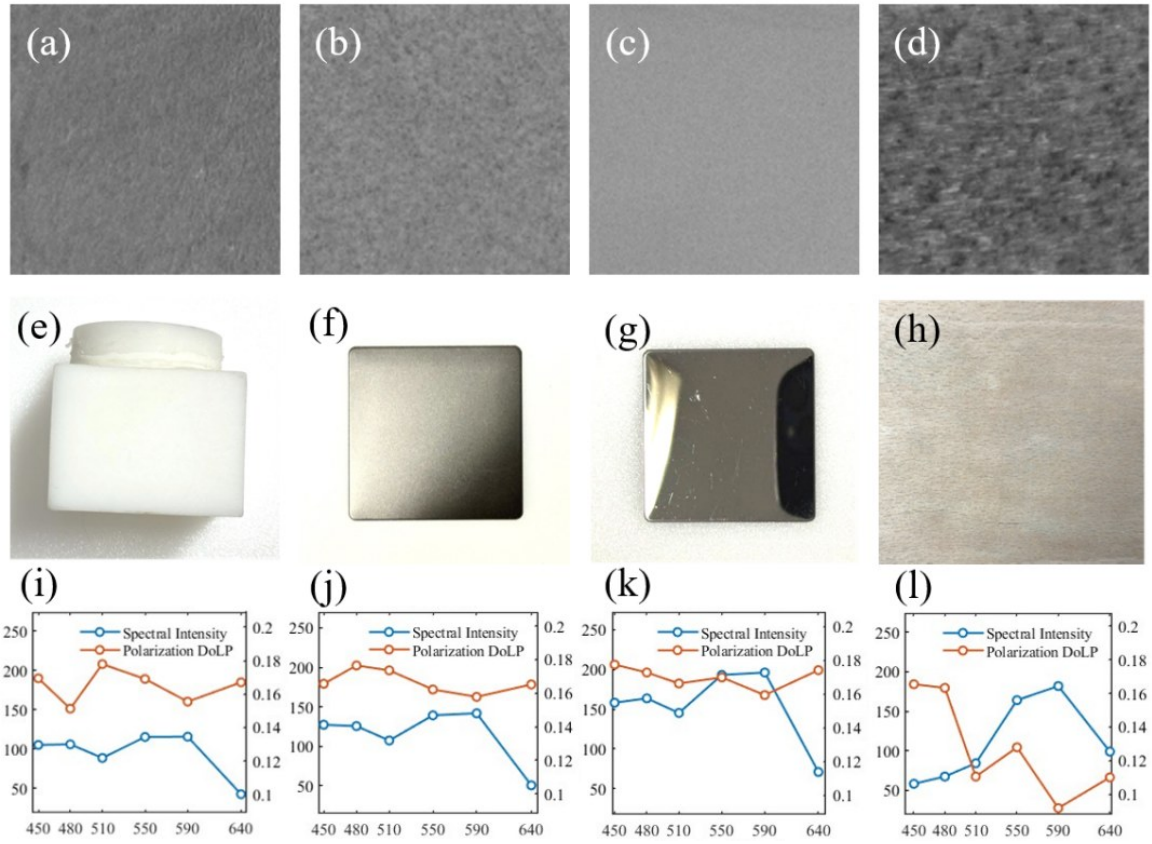


图4 同波段下四种典型材料的局部灰度图(a-d)、对应彩色RGB图像(e-h)及其光谱-偏振特征曲线(i-l)。(a,e,i)光滑塑料;(b,f,j)粗糙金属;(c,g,k)镜面金属;(d,h,l)粗糙木质

Fig.4 Local grayscale images (a-d), corresponding color RGB images (e-h), and spectral-polarization characteristic curves (i-l) of four typical materials in the same waveband. (a, e, i) smooth plastic; (b, f, j) matte metal; (c, g, k) mirror metal; (d, h, l) matte wood

采集,覆盖多材质与多表面形貌,提供多波段与多偏振视角的成像序列,具有更丰富的照明条件与姿态变化,更接近工程应用需求。本文使用该数据集公开的部分样本开展评估,并不改变其采集语义的前提下,将其整理为面向“材质-表面联合识别”的可用资源。该数据在复杂背景、真实光照与多样外观的综合作用下,能够更全面检验光谱-偏振融合策略的稳健性与普适性。需要说明的是,SPRD与本文自建MSP-6的定位不同:SPRD面向自然环境下的通用场景,具有更丰富的背景干扰、光照变化与姿态差异;MSP-6则聚焦工业检测中的“同材质异工艺、弱纹理、高反光”细粒度识别任务,强调可控采集与物理一致性约束,二者互为补充而非替代关系。

为保证材质样本的准确标记与与MSP-6协议的一致性,本文首先依据数据集提供的标注文件对样本类型进行筛选,选取数据集中115份标记为“对象级拍摄”目标的样本作为不同材质的单一来源集合,在统一坐标系下于非边缘安全区域随机生成裁剪窗口,并同步作用于所选光谱-偏振序列及对应RGB图像,得到共计5750份尺寸为 128×128 像素的子样本。标记策略采用无监督实例级伪标签,同一材质样本裁剪所得的所有子样本统一赋予该样本编号对应的材质类别,从而以“样本来源”刻画材质类别;同时结合光谱-偏振通道的低强阈掩蔽与百分位归一化,弱化极化派生量在低信噪区域的数值不稳定。上述筛选与标记使SPRD数据集在单材质来源、像素级对齐、输入模态与训练评估协议上与课题组自建数据集保持一致,从而保证跨数据域的对比公平性与结论的可复用性。

3 实验结果与分析

3.1 方法性能分析

为验证多谱段-偏振融合模型在材质识别任务中的有效性,本文在MSP-6数据集上将所提出方法与多

种主流深度学习网络进行了系统对比,结果如表1所示。对比模型涵盖仅使用单模态输入的经典卷积网络、多模态CNN结构,以及两种针对高光谱数据提出的Transformer模型FactoFormer和DSFormer。为确保对比的公平性,所有模型均在统一实验协议下重新训练,保持数据划分、输入分辨率及预处理流程高度一致。各方法均采用Adam优化器,初始学习率设为0.001,Batch size为32,并辅以StepLR调度策略。针对跨领域对比模型,本文在保留其核心架构的基础上完成了输入适配,相关超参数及损失权重均经验证集敏感性分析确定。值得注意的是,受限于MSP-6偏振-多谱段数据与传统高光谱数据在成像机理及维度上的固有差异,部分对比方法在本数据集上的性能与其在原始任务中的表现可能存在一定偏差。本文中,“多谱段”(Multispectral,MS)指包含多个光谱波段的数据,“偏振”(Polarization,Pol)指不同偏振角度的成像信息。

表1 基于MSP-6数据库的方法性能对比
Table 1 Performance comparison of methods based on the MSP-6 database

Method	Year	Input	Material Acc/% ↑	Surface aware- ness	Time /ms ↓	Parameters (Params/M) ↓
Ours	2025	MS+Pol	99.51	✓	6.596	27.749
MultiSpectralResNet	2025	MS+Pol	81.39	✓	6.605	33.848
RegNetY-8GF	2020	MS+Pol	97.35	✓	5.944	39.392
EfficientNetV2	2021	RGB	86.02	×	13.267	21.500
DSFormer	2025	MS	88.56	×	8.868	59.091
FactoFormer	2024	MS	86.74	×	4.540	85.061
ConvNeXt-Tiny	2022	MS	83.41	×	3.585	28.600
RegNetY-8GF	2020	MS	77.48	×	5.830	39.381

结果表明,本文方法在整体识别准确率上显著优于所有对比方法。相比同样使用多谱段+偏振的RegNetY-8GF,准确率提升2.16%;相比仅依赖RGB的EfficientNetV2,提升超过13个百分点;相对于仅使用多谱段、未引入偏振信息的ConvNeXt-Tiny和RegNetY-8GF,提升幅度分别达到16.10%和22.03%。其中,MultiSpectralResNet是课题组前期提出的一种多谱段-偏振基线网络,将RGB及各偏振相关谱段在输入层简单堆叠后直接送入统一ResNet主干进行特征提取,其识别准确率仅为81.39%,表明通道的简单堆叠难以充分挖掘偏振信息的判别能力,需要结合专门设计的多分支结构和物理先验约束,对不同模态进行有针对性的特征提取与交互融合。

作为强基线的FactoFormer与DSFormer属于典型谱-空Transformer,其原始设计更适用于高光谱场景下较高维的光谱立方体输入。在本文的MSP-6数据集上,受限于仅有6个波段,其可利用的谱维冗余与长程谱间依赖显著弱于常规高光谱任务。为此,本文在实验中采用了适配低波段输入的轻量化实现,统一输入分辨率与训练协议后以6波段多谱段数据进行训练和评估。然而,这类方法仍主要依赖自注意力机制进行数据驱动的谱-空特征建模,并未显式编码偏振物理规律或表面反射先验。在MSP-6数据集上,其综合准确率仍仅为86.74%和88.56%,明显低于本文方法。这表明,在存在强镜面反射、高亮区域以及不同材质间颜色和纹理高度相似等复杂场景时,仅依赖数据驱动的特征自适应难以充分区分材质差异与光照差异;相比之下,基于马吕斯定律和表面反射先验的物理约束能够提供额外可分性线索,从而显著提升多谱段与偏振联合建模效果。

为客观评价算法的运算性能,本文及对比方法均在统一环境下进行测试。实验环境配置为: Intel(R) Xeon(R) Platinum 8470Q CPU,显卡采用NVIDIA vGPU-32GB,软件环境为Python 3.8、PyTorch 1.11.0及CUDA 11.3。计时流程如下:加载各模型官方预训练权重并设为评估模式,采用与MSP-6数据集一致的输入分辨率。为消除系统波动,各方法在正式计前先进行50次GPU预热,并通过显式调用硬件同步指令来确保计时的准确性。最终取测试集上连续500幅图像推理时延的平均值作为对比结果。全流程未启用自动混合精度加速,确保了不同方法在相同计算精度下的可比性。结合表1的参数量与推理时延分析,本文方法在保持高精度识别的同时,维持了较低的计算开销。这表明该模型能够有效部署于具备常规GPU加速能力的工业视觉平台,而不依赖超大规模计算资源。针对实时性或成本敏感型场景,后续可通过主干网络轻量

化、光谱波段筛选及偏振维度压缩等策略,进一步优化推理效率与资源占用。

表1还给出了各方法的表面感知能力、模型参数量及单幅图像推理时间。本方法在实现最高材质识别准确率并保留对表面变化敏感性的同时,单幅推理时间为6.596 ms,仍处于毫秒级。从模型规模来看,本文方法的参数量与ConvNeXt-Tiny和EfficientNetV2等经典轻量化基线处于同一数量级,且远低于FactorFormer等模型。可以看出,在与经典CNN和Transformer基线处于可比参数规模的前提下,本文方法在准确率、表面感知能力和推理效率三方面取得了较好平衡,从实验上验证了多谱段光谱信息与偏振信息联合建模以及物理一致性约束的有效性。

表2 基于SPRD数据库的方法性能对比
Table 2 Performance comparison of methods based on the SPRD database

Method	Year	Input	Material Acc/% ↑	Surface awareness
Ours	2025	MS+Pol	86.83	✓
RegNetY-8GF	2020	MS+Pol	85.56	✓
EfficientNetV2	2021	RGB	70.51	×
ConvNeXt-Tiny	2022	MS	52.78	×
RegNetY-8GF	2020	MS	83.89	×

基于SPRD公共数据集开展的定量评估结果见表2。需要指出,SPRD的采集规则与本文系统在通道组织、标注语义与预处理流程上均存在差异,本节为实现可比性而进行一致化改造与强制兼容评估,因而整体精度预期低于原生MSP-6。即便如此,本文方法在该公共数据上仍取得了86.83%的材质识别准确率,依然优于RegNetY-8GF在多模态与单模态输入下的性能表现;与仅依赖RGB图像的EfficientNetV2及仅使用多谱段输入的ConvNeXt-Tiny相比,准确率分别提升了16.32%和34.05%。

相较于MSP-6,本文方法相比于传统方法呈现的优势有所收敛。分析其原因主要有以下三点:首先,SPRD数据采集自复杂真实场景,其极化角度与强度存在一定的标定误差,且在低辐照区域引入噪声,对基于偏振的派生特征形成扰动,从而削弱了偏振通道在短周期训练及统一归一化设定下的增益效果;其次,该数据集的标签生成方式采用基于“对象级拍摄”的实例级无监督聚类,其任务界定更偏向于样本来源的区分,而非严格意义上的材质学分类。在此设定下,光谱维度上的稳定差异对分类的贡献相对更大;此外,为保障不同数据集之间的可比性,本文在实验中未针对各模型进行差异化的融合权重调优或正则化参数调整。尽管如此,光谱-偏振融合机制在复杂光照与表面形貌变化条件下,仍持续提升材质识别的准确性。这一结果表明,本文方法在材质判别任务中具备良好的可复用性与补充价值。

3.2 分类结果分析

表3 多谱段-偏振特征在不同表面条件下的材质分类准确率
Table 3 Accuracy of material classification using multispectral-polarization features under different surface conditions

Material/Surface	Draw/%	Matte/%	Mirror/%	Smooth/%	Macro Avg /%
Metal	99.80	99.65	99.95	—	99.80
Plastic	—	95.25	—	99.50	97.38
Wood	—	98.50	—	—	98.50
Overall Macro Avg.	99.80	97.80	99.95	99.50	98.56

本节进一步从本文方法自身的分类结果分析其具体表现。为同时考察模型在不同表面状态下的稳定性,实验将拉丝、哑光、镜面等表面条件视为同一材质类别内部的自然扰动因素,对各材质在不同表面组合下的识别结果进行统计,其材质维度与表面条件维度的综合结果如表3所示,需要说明的是表中“Macro Avg”为等权平均准确率,对每种材质,先在其出现的表面条件上取算术平均,再对三类材质取平均;与全文“99.51%”的按样本加权总体准确率(Overall Accuracy)统计口径不同。考虑到MSP-6数据集中金属、塑料与木材样本的类别分布存在一定差异,为避免总体准确率受多数类样本影响,本文除在表1报告总体准确率

(Overall Accuracy)为99.51%的同时,在表3给出了按类别等权统计的宏平均结果(Macro Average),并在对应表注中明确两类指标的计算方式。分析结果表明,尽管两类指标因统计口径不同存在数值差异,但其整体水平与趋势保持一致,说明当前类别分布差异未改变本文方法在材质识别任务中的基本结论。

从整体性能来看,在不同种类的材质识别场景中,模型的综合分类准确率达到98.56%。在没有显式建模复杂几何形状和大尺度场景先验的情况下,仅依靠多谱段-偏振数据构成的特征空间,即可在多类样本之间建立清晰的判别边界,表明多谱段-偏振成像为材质识别提供了足够丰富且稳定的物理表征。

从材质维度的统计结果来看,所有种类材质的平均识别率均超过97%。其中,金属材质在各类表面条件下的平均识别准确率达到99.80%,表现最为突出;木材材质的平均识别准确率为98.50%,同样保持在较高水平;塑料材质的平均识别准确率为97.38%,略低于前两类,但仍处于整体较高区间。这说明,即使在表面粗糙度和纹理形态存在明显差异的情况下,多谱段与偏振特征依然能够稳定刻画材质本征反射特性与偏振响应差异,使得材质类别在特征空间中呈现类内紧致、类间分离的聚类分布。

在各类表面条件中,镜面场景尤具代表性。实验结果表明,在镜面反射条件下的整体材质识别准确率达到99.95%,且该类样本几乎没有出现误判。需要强调的是,镜面高光往往会导致光谱强度出现饱和或明显的非线性畸变,使得仅依赖光谱或RGB信息的模型极易在强反射区域失效;而在本方法中,引入偏振信息之后,系统一方面利用偏振对镜面高光的天然抑制能力,在观测层面有效分离镜面分量与漫反射分量,削弱了高光对特征表达的“遮盖效应”;另一方面,通过马吕斯定律构建物理重建约束,使网络在训练阶段被迫同时满足“分类正确”和“偏振物理响应一致”这两个条件,在高反射强度、复杂光照的极端场景下仍能维持稳定、可解释的特征分布。因此,优秀的识别准确率印证了本文所提出的“偏振高光抑制+马吕斯重构一致性”协同机制在镜面反射与复杂光照场景中的有效性,表明其在工业检测与智能制造等高反光应用场景中的实用价值。

在此基础上,对原始多谱段-偏振数据中的物理超界样本进行了统计分析。以本实验训练集为例,偏振度或斯托克斯向量明显超出理论物理范围的像素比例始终保持在0.5%以下,不同场景大多落在0.1% - 0.3%的范围内,仅占总体样本的极小一部分。通过预处理阶段的裁剪与投影操作,这些异常样本被统一映射到满足公式(6)的物理可行区域,其分布与主体数据分布保持一致。对比采用和不采用该物理可行性修正的实验结果可以发现,各材质在不同表面条件下的分类准确率变化极小,综合准确率仍保持在98%以上,说明该处理不会引入显著偏差;同时,结合马吕斯定律的物理一致性约束,网络输出的斯托克斯向量中不再出现明显的超界异常,从而有效抑制了非理想偏振系统和测量噪声对识别结果的潜在负面影响。

综上所述,所提出的多谱段-偏振融合模型在拉丝、哑光、镜面等多种表面扰动条件下仍能保持稳定的材质判别性能。多波段光谱强度为区分材质成分和折射率差异提供了“固有指纹”,偏振信息则通过高光抑制与物理一致性约束,有效保证了在镜面反射、复杂表面等条件下的识别准确率。两者协同使得模型在高相似度材质之间依然能够构建清晰的决策边界,从而验证了其在不同识别场景下的有效性与可推广性。

需要指出的是,本文实验结果主要是在受控采集条件下获得的。MSP-6数据集在半封闭暗室、固定LED光源以及稳定成像几何关系下完成采集,这种设置有利于保证多谱段-偏振观测的物理一致性、空间配准精度与实验可重复性,因此能够较为准确地验证本文方法在典型材质-表面联合识别任务中的有效性。与此同时,本文结论目前主要适用于受控条件下的工业材质识别场景,对于更复杂现场环境中可能存在的环境光扰动、背景变化、姿态偏移及局部遮挡等因素,其泛化性能仍有待后续进一步验证。

3.3 消融实验

在完成整体模型对比实验之后,为进一步分析不同输入模态对材质识别性能的具体贡献,本文在MSP-6数据集上进行了多模态消融实验。实验基于RGB、光谱强度与偏振信息构建了不同的输入模式,并在完全相同的训练与测试划分下,对各模式的分类性能进行系统评估。表4给出了各模态组合在总体准确率(Accuracy)、平衡准确率(BAL_ACC)、多类别F1指标(MACRO_F1)、相关系数(MCC)、Kappa系数以及若干不确定性与校准指标(ECE)上的综合表现,用于全面刻画不同信息通道对材质识别的影响。

分析表4可见,单一RGB模态的整体准确率仅为0.5110,BAL_ACC与MACRO_F1接近随机,MCC与Kappa也几乎为零,说明在自然照明下不同材质的RGB反射差异有限,不同表面状态带来的纹理和高光特

表 4 多模态性能评估
Table 4 Multimodal performance evaluation

Input Mode	Accuracy ↑	BAL_ACC	MACRO_F1	MCC ↑	Kappa ↑	ECE ↓
		↑	↑			
rgb-only	0.5110	0.3350	0.2312	0.0152	0.0033	0.3139
s0-only	0.6717	0.5050	0.4776	0.4804	0.3789	0.2374
rgb+s0	0.8183	0.6517	0.6022	0.7157	0.6738	0.1189
rgb+polar	0.7426	0.5841	0.5430	0.5982	0.5215	0.1852
rgb+s0+polar	0.9951	0.9900	0.9927	0.9918	0.9918	0.0186

征难以被稳定区分。引入光谱强度后,“仅光谱”模式准确率提升至0.6717,性能指标明显改善,表明多谱段反射在区分金属、塑料、木材等材质成分方面起到基础性作用,但从BAL_ACC与MACRO_F1仍可看到不同表面条件下的混淆。进一步融合多模态时,“RGB+光谱”模式准确率提升至0.8183,性能指标同步提高,说明光谱提供材质的主要“固有指纹”,RGB补充局部纹理与颜色细节。值得注意的是,对比“RGB+偏振”与“RGB+光谱”模式可以发现,虽然引入偏振信息能将准确率提升至0.7426,但其性能仍低于包含光谱信息的输入组合。这一对比有力地论证了光谱信息在材质分类中提供了独特的、不可替代的鉴别能力,是刻画材质理化属性的核心维度。而在三种模态全部引入的情形下,准确率显著跃升至0.9951。

综合来看,RGB单模态难以支撑材质识别,多谱段光谱是刻画材质成分与吸收特性的基础信息通道,而偏振信息在抵抗镜面高光、表面粗糙度等扰动方面发挥关键作用;只有在光谱与偏振信息同时被引入并进行深度融合时,模型才能在识别准确率和不确定性控制两方面同时逼近饱和,多谱段与偏振信息的联合建模由此为“通过融入偏振提升材质识别能力”这一目标提供了清晰的实验支撑。

为进一步验证各核心模块的贡献,我们进行了模块独立消融实验,结果如表5所示。

表 5 模块级消融实验与综合性能评估
Table 5 Ablation study and comprehensive performance evaluation of core modules

Ablation	Module			Evaluation Metrics				
	PolarFuser	Attention	Physical Constraint	Accuracy ↑	BAL_ACC ↑	MACRO_F1 ↑	MCC ↑	ECE ↓
(a)	×	✓	✓	97.15	0.966	0.968	0.955	0.026
(b)	✓	×	✓	98.24	0.978	0.980	0.971	0.020
(c)	✓	✓	×	96.52	0.958	0.961	0.943	0.033
Full	✓	✓	✓	99.51	0.990	0.993	0.992	0.019

分析表5可见,各模块的移除均会导致模型性能出现不同程度下降,说明PolarFuser、注意力融合和物理约束三者均对最终识别结果具有积极贡献。相比完整模型,移除PolarFuser后,模型的Accuracy由99.51%下降至97.15%,BAL_ACC、MACRO_F1与MCC也分别下降至0.966、0.968和0.955,同时ECE由0.019上升至0.026。这表明,PolarFuser能够对高维偏振原始特征进行有效压缩与重组,在降低通道冗余的同时保留关键偏振判别信息,从而提升模型对表面微结构差异的表征能力。移除注意力融合模块后,模型Accuracy降至98.24%,BAL_ACC、MACRO_F1和MCC分别为0.978、0.980和0.971,整体性能虽仍保持较高水平,但相较完整模型仍存在稳定退化。这说明注意力机制能够在多模态特征融合过程中自适应突出更具判别性的通道与空间区域,抑制冗余响应与噪声干扰,从而进一步增强融合特征的有效性与鲁棒性。

值得注意的是,移除物理约束模块后,模型性能下降最为显著,Accuracy由99.51%下降至96.52%,BAL_ACC、MACRO_F1和MCC分别降至0.958、0.961和0.943,且ECE从0.019明显上升至0.033。这一结果表明,物理约束不仅直接提升了分类精度,也显著改善了模型预测结果的校准能力与稳定性。其原因在于,基于马吕斯定律构建的物理一致性约束能够将偏振成像机理显式嵌入特征学习过程,使网络在拟合判别边界的同时受到真实物理规律的约束,从而有效抑制由高光、噪声及表面反射变化引起的伪判别信息。综合来看,三项核心模块在本文框架中具有明显的协同作用,其中物理约束模块贡献最为关键,PolarFuser

为偏振特征提供了紧凑而有效的表示,注意力融合则进一步提升了多模态信息交互的质量,三者共同保证了模型在识别精度、类别均衡性与置信度校准等方面的优异表现。

3.4 泛化能力验证

为进一步验证所提出的三支多模态融合结构及物理约束机制的通用性与泛化能力,本研究在保持所有实验设置不变的前提下,仅替换主干网络,对比不同网络结构在相同输入和相同训练策略下的分类性能。实验结果如表6所示,测试范围涵盖从轻量级网络(MnasNet-A1、MobileNetV3-Large)、中等规模网络(EfficientNet系列)、到较高容量的RegNetY系列,以及具备良好偏振表征能力的ConvNeXt-Tiny等多种网络框架。在这些多样化的网络结构中,本研究提出的整体框架均展现出一致而显著的性能提升趋势,充分表明所设计的三支融合结构并不依赖某一特定backbone,而是能够与多种网络结构稳定适配并发挥优势。

表6 模型替换骨干网络性能汇总
Table 6 Summary of performance of model-replaced backbone network

Method	Year	Material Acc/% $\uparrow$	Material F1 (Macro/Weighted) $\uparrow$
EfficientNetV2-S	2021	84.11	0.85 / 0.84
MobileNetV3-Large	2019	95.37	0.95 / 0.95
RegNetY-8GF	2020	94.01	0.95 / 0.94
MnasNet-A1	2019	84.72	0.80 / 0.84
ConvNeXt-Tiny	2022	78.68	0.78 / 0.79

从表6可以观察到,无论主干网络的规模大小以及其特征提取机制如何差异,结合三支多模态融合与物理约束后,各种替换的骨干网络在材质分类任务上均取得了较高且相对稳定的识别精度。以轻量级模型为例,MobileNetV3-Large在本任务下的材质分类准确率达到95.37%,对应的F1分数也均为0.95;结构更紧凑的MnasNet-A1亦能达到84.72%的准确率和较高的F1水平,说明即使在参数量和计算量受限的条件下,所提出的方法仍能够显著提升轻量网络的材质判别能力,为未来在资源受限平台上的部署提供了可能。

在中等规模网络中,EfficientNetV2-S的材料分类准确率为84.11%,宏平均和加权平均F1分数分别为0.85和0.84;相比其在传统RGB任务中的表现,本研究在保持相对较小网络规模的同时,通过引入多谱段-偏振三模态输入和三支融合结构,使其在材料识别任务上仍保持着较为可观的性能。

在更高容量的RegNetY-8GF中,模型的材料分类准确率进一步提升至94.01%,对应的宏平均和加权平均F1分数达到0.95和0.94,在所有替换骨干网络中表现优异。这表明随着骨干网络表征能力的增强,所提出的多模态融合与物理约束机制能够充分利用更强的语义特征,对材料光谱-偏振响应进行更细致的建模,从而进一步提高识别精度。

总体来看,本研究提出的三支融合结构与物理约束机制在多种骨干网络上均取得了稳定且优秀的材料识别性能,跨越轻量级、中量级与大容量网络,性能波动幅度小,且整体水平均处于较高区间。这一现象说明,本方法的核心优势并不依赖某一特定主干网络,而是源于有效的多模态融合策略与偏振物理先验驱动的特征建模。无论是面向移动端推理的轻量模型,还是用于工业检测、智能制造的高容量主干网络,本方法均能充分挖掘光谱-偏振多模态数据的联合特征结构,体现出良好的架构独立性与任务泛化能力,从而为在不同资源约束和应用场景下推广基于偏振增强的材料识别提供了有力支撑。

4 结论

本文面向多材质目标在复杂表面与复杂光照条件下的识别需求,提出了一种具有物理一致性约束的多谱段-偏振三支融合方法。该方法将RGB、六波段光谱强度及偏振派生特征作为三路输入,在特征层与损失层显式引入马吕斯定律约束,充分发挥光谱对材质本征特性的表征能力以及偏振对镜面高光和表面微结构的敏感性,从而显著提升材料识别的精度与鲁棒性。实验结果显示,在本文数据协议与评测设置下,该方法在MSP-6数据集上取得了较高分类准确率,并在不同骨干网络配置下表现出较好的结构适配性与跨模型一致性。

总体而言,本文工作强调面向工业检测场景的针对性设计与工程可用性验证。结果表明,多谱段与偏

振信息的协同建模结合物理一致性约束能有效提升复杂工况下的材质识别性能。尽管 MSP-6 数据集支撑了本文方法在典型材质上的有效性,但受限与数据集的类别规模与环境多样性,后续仍需在更广泛的材质类型及复杂工业现场中验证其泛化性能。未来工作将重点围绕以下几个方面展开:1)系统分析不同小像素裁剪尺度与识别性能之间的关系,寻求在计算效率与噪声抑制能力之间取得平衡的最优窗口尺寸;2)研究在现有六波段、四偏振角采集策略下各波段-偏振组合的判别贡献,筛选出在观测数量受限时仍能保持高精度的代表性组合,构建简化而高效的成像方案;3)进一步扩展 MSP-6 数据集,采集更多材质类别及多种表面加工工艺样本,并在更贴近真实工业现场的条件开展验证,从而持续增强所提方法的泛化能力、实用性与推广价值。

参考文献

- [1] LI Hui, XUE Qingsheng, BAI Haoxuan, et al. Polarization spectrum measurement technology based on intensity modulation[J]. *Acta Photonica Sinica*, 2023, 52(4):0430001
李辉,薛庆生,白皓轩,等. 基于强度调制的偏振光谱测量技术[J]. *光子学报*, 2023, 52(4):0430001
- [2] ZHAO Langyue, WU Yiquan. Research progress of surface defect detection methods based on machine vision[J]. *Chinese Journal of Scientific Instrument*, 2022, 43(1): 198-219.
赵朗月,吴一全. 基于机器视觉的表面缺陷检测方法研究进展[J]. *仪器仪表学报*, 2022, 43(1):198-219.
- [3] LI Shutao, SONG Weiwei, FANG Leyuan, et al. Deep learning for hyperspectral image classification: An overview [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57 (9): 6690-6709.
- [4] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [5] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017:1251-1258.
- [6] LIU Bing, YU Anzhu, YU Xuchu, et al. Deep multiview learning for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7758-7772.
- [7] TANG Xu, MENG Fan, ZHANG Xiang, et al. Hyperspectral image classification based on 3-D octave convolution with spatial-spectral attention network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(3):2430-2447.
- [8] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: Exploring 3D-2D CNN for hyperspectral image classification [J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(2):279-283.
- [9] MOHAMED S, HANGHIGHAT M, FERNANDO T, et al. FactoFormer: Factorized hyperspectral transformers with self-supervised pretraining[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 62:1-14.
- [10] XU Yichu, WANG Di, ZHANG Lefei, et al. DSFormer: A dual-selective transformer network for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62:5502617.
- [11] TU Lilin, LIN Zhaohui, LI Wei, et al. S2HM2: A spectral - spatial hierarchical masked modeling framework for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62:1-18.
- [12] KANG Xudong, DENG Bin, DUAN Puhong, et al. Self-supervised spectral-spatial transformer network for hyperspectral oil spill mapping[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61:1-10.
- [13] FANG Zhuoqun, YANG Yuexin, LI Zhaokui, et al. Confident learning-based domain adaptation for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60:5527116.
- [14] ZHAO Yongqiang, PAN Quan, ZHANG Hongcai. Material classification based on multi-band polarimetric images fusion [C]. *SPIE*, 2006, 6240:62400G.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [J]. *ECCV*, 2018: 3-19.
- [16] NICODEMUS F E, RICHMOND J C, HSIA J J, et al. Geometrical considerations and nomenclature for reflectance [R]. *National Bureau of Standards (US)*, 1977.
- [17] TORRANCE K E, SPARROW E M. Theory for off-specular reflection from roughened surfaces [J]. *Journal of the Optical Society of America*, 1967, 57(9): 1105-1114.
- [18] JEON Y, CHOI E, KIM Y, et al. Spectral and polarization vision: Spectro-polarimetric real-world dataset[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 22098-22108.

Material Recognition Based on Multi-spectral-polarization Information and Physical Constraints

SHAO Jiangang^{1,2}, ZHAO Jufeng^{1,2}, CUI Guangmang^{1,2}

(1 Zhejiang Key Laboratory of Optoelectronic Intelligent Imaging and Aerospace Sensing, Hangzhou Dianzi University, Hangzhou 310018, China)

(2 School of Electronic and Information Engineering, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: Material recognition for weakly textured surfaces that share the same base material but differ in surface processing remains a challenging task in intelligent inspection and industrial quality control. Traditional methods mainly relying on spectral imaging often suffer from insufficient observational cues, which makes it difficult to extract features with both strong discriminability and good robustness under varying surface conditions. To address this issue, a multimodal material recognition framework is proposed, in which RGB appearance, spectral intensity, and polarization cues are jointly exploited under the guidance of physical priors. The objective of this framework is to achieve stable and accurate material classification for same-substrate surfaces with different processing-induced states, while maintaining a balanced model complexity suitable for practical deployment. The proposed framework adopts a three-branch network architecture to integrate RGB, spectral, and polarization information in a unified manner. Firstly, RGB images, spectral cubes, and polarization measurements are preprocessed using a common normalization and synchronized cropping strategy, so that geometric and radiometric consistency across different modalities can be ensured and pixel-wise or region-wise feature alignment can be achieved. Then, the RGB branch is used to learn appearance and mesoscopic texture cues, the spectral branch is employed to mine wavelength-dependent intensity patterns related to intrinsic material properties, and the polarization branch is designed to extract polarization-derived features, such as the degree of polarization and angle of polarization, which are sensitive to surface micro-geometry and reflection mechanisms. Subsequently, an attention-based fusion module is introduced to adaptively reweight and aggregate multimodal features, so that the complementary roles of spectral and polarization information in characterizing intrinsic material attributes and surface-state differences can be fully explored. Furthermore, a Malus-law-based physical constraint is incorporated into the polarization branch to regularize the predicted polarization responses. By enforcing the relationship among polarization angle, degree of polarization, and incident-reflection geometry to be consistent with imaging physics, the proposed framework can effectively suppress abnormal activations caused by strong specular highlights and guide the learned feature space toward physically plausible solutions. Extensive experiments are conducted on a material dataset comprising samples with identical substrates but diverse surface processing procedures and weak texture characteristics. The quantitative results demonstrate that the proposed multimodal framework achieves an overall recognition accuracy of 98.56% on the material classification task. Compared with single-modality baselines based on RGB, spectral, or polarization information alone, as well as with conventional fusion methods, the proposed approach yields accuracy improvements ranging from 2.03% to 22.03%, which verifies the effectiveness of jointly utilizing complementary modalities under physical guidance. Ablation studies further show that removing the attention-based fusion module leads to a significant degradation in recognition performance, confirming its important role in selectively emphasizing informative channels and modalities. Similarly, when the Malus-law-based constraint is removed, the model becomes more sensitive to specular highlight artifacts and less robust under varying surface processes, which further demonstrates the necessity of embedding physical priors into the learning process. In addition, experiments conducted with different mainstream backbone networks indicate that the proposed fusion and constraint mechanisms can be conveniently integrated into various architectures and can consistently improve recognition performance, showing good structural generality and scalability. Finally, by jointly leveraging RGB appearance, spectral intensity, and physically constrained polarization cues within a three-branch attention-fusion architecture, this work provides a robust and accurate solution for material recognition of weakly textured surfaces with the same base material under diverse processing conditions. The introduction of the Malus-law-based physical constraint effectively alleviates the adverse influence of strong specular highlights and enhances the discriminability of feature representations across different surface processes. Owing to its superior recognition performance, effective physical guidance, and good adaptability to different backbone networks, the proposed framework shows strong potential for practical applications in complex industrial material inspection and quality control scenarios.

Key words: Polarization imaging; Spectral imaging; Material recognition; Multimodal fusion; Deep

Learning

OCIS Codes: 110.5405; 110.4235; 100.5010

CSTR: 32255.14.gzxb20265508.0830003