

# 多光谱融合提升光电图像目标识别精度的新策略

李明择

山东建筑大学理学院, 山东济南, 250102;

**摘要:** 随着现代光电探测技术的发展, 单一波段成像在复杂环境下的目标识别能力已显不足。本文提出一种基于自适应权重分配与跨模态特征对齐的多光谱融合新策略, 旨在提升目标识别的鲁棒性与精度。该策略融合可见光、近红外与热红外三类典型光电图像, 通过设计通道-空间双重注意力机制, 动态评估并调整各波段在不同场景下的贡献度; 同时引入对比学习框架, 在特征空间中拉近同一目标跨模态表示的距离, 推远不同类别样本, 实现语义层面的一致性对齐。实验在自建城市道路多光谱数据集(含 5000 组三模态图像)及公开基准 KAIST、FLIR ADAS 上系统验证。结果表明, 所提方法在目标检测 mAP@0.5 指标上平均提升 6.8%, 在夜间、烟雾或强眩光等恶劣条件下性能增益最高达 9.1%。模型在 NVIDIA Jetson AGX 平台实现 28 FPS 推理速度, 兼顾轻量化与实时性。本研究不仅为多光谱信息融合提供了兼具理论创新与工程实用性的新范式, 也为智能安防、自动驾驶等实际应用场景中的鲁棒感知系统构建提供了有效技术路径。

**关键词:** 多光谱融合; 目标识别; 注意力机制; 跨模态对齐

DOI:10.69979/3041-0673.25.02.102

## 引言

光电成像系统广泛应用于军事侦察、智能安防与自动驾驶等领域。然而, 单一波段图像易受光照变化、天气干扰或伪装遮蔽影响, 导致目标识别性能下降。多光谱成像通过整合不同电磁波段信息, 可有效弥补单一模态的局限性。近年来, 深度学习推动了多模态融合技术的发展, 但现有方法多采用固定权重或简单拼接策略, 忽视了不同场景下各波段信息有效性的动态变化, 且缺乏对跨模态语义一致性的显式建模。为此, 本文提出一种“自适应权重+语义对齐”的双驱动融合框架, 强调在保留各波段物理特性的基础上, 实现语义层面的深度融合。该策略兼顾实用性与科学性, 适用于资源受限的边缘计算平台, 具有明确的工程应用价值。

## 1 多光谱成像特性与目标识别挑战

多光谱成像通过整合不同电磁波段信息, 为复杂场景下的目标识别提供更全面的感知能力。可见光图像具有高空间分辨率和丰富的纹理细节, 适用于白天光照充足环境下的目标辨识, 但在夜间或强逆光条件下性能急剧下降。近红外波段对植被和伪装材料具有较强穿透与反射特性, 能在一定程度上克服可见光受光照变化影响的问题, 但其成像依赖主动照明, 在完全无源条件下作用有限。热红外成像则基于物体自身热辐射, 不受可见光照明条件制约, 可在完全黑暗、烟雾或轻度遮蔽环境

中有效探测目标, 但其空间分辨率较低且缺乏细节纹理, 难以支持精细分类。单一模态成像在低照度、强反射、雨雾或战场伪装等复杂环境下常出现漏检、误检甚至完全失效。当前多光谱融合技术虽能综合利用各波段优势, 但仍面临显著挑战: 不同模态间存在物理机理差异导致的特征分布异构性; 部分波段在特定场景下贡献微弱却引入冗余计算; 融合策略缺乏对环境动态变化的自适应响应能力, 难以在多样化任务中保持稳定性能。这些问题制约了多光谱系统在实际部署中的识别精度与鲁棒性提升。

## 2 现有融合方法综述与局限性分析

多光谱图像融合方法历经从浅层到深层的演进。早期方法主要分为像素级、特征级和决策级三类。像素级融合直接对原始图像进行加权、金字塔分解或小波变换, 保留细节但易受噪声干扰且计算冗余; 特征级融合提取边缘、纹理或梯度等手工特征后融合, 一定程度上降低数据量, 但特征表达能力受限; 决策级融合则在各模态独立完成识别后再整合结果, 虽结构灵活却丢失底层互补信息。随着深度学习兴起, 基于卷积神经网络(CNN)的模型如FusionGAN引入生成对抗机制优化融合图像视觉质量, 但侧重于图像重建而非任务导向的目标识别。后续工作如MMFD利用编码器-解码器架构在特征空间融合多模态信息, 并引入注意力模块增强关键区域响应。

近年来, Transformer 凭借全局建模能力被用于跨模态交互建模,在部分基准上取得性能提升。然而,现有深度融合方法普遍存在两个根本缺陷:其一,融合权重多为静态或仅依赖局部上下文,未充分考虑不同场景下各波段信息有效性的动态变化,例如热红外在夜间主导而可见光在白天更具判别力,但模型缺乏对此类先验的自适应调制;其二,多数方法仅在特征维度进行拼接或加权,忽视了不同模态间语义空间的不一致性,导致融合后的特征虽在形式上整合,却在语义层面存在偏差,即“形合神离”。这种缺乏显式跨模态语义约束的融合方式,难以保证目标在不同波段下的表征指向同一高层语义概念,削弱了最终识别的可靠性与泛化能力。

### 3 本文提出的新策略:自适应权重分配与跨模态特征对齐

#### 3.1 自适应权重模块:通道-空间双重注意力驱动的动态融合

为应对多光谱图像在不同环境条件下信息贡献度差异显著的问题,本文设计了一种通道-空间双重注意力网络,用于动态生成可见光、近红外与热红外三模态的融合权重。该模块以各波段特征图作为输入,首先通过全局平均池化捕获通道维度上的语义响应强度,构建通道注意力向量,突出对当前任务更具判别性的波段通道;随后引入空间注意力子网络,利用卷积层提取跨通道的空间上下文信息,生成空间权重图,强化目标区域并抑制背景干扰。两个注意力分支并行处理后相乘,形成细粒度的三维权重张量,实现像素级与通道级联合调制。该机制无需预设固定融合规则,而是依据输入内容自适应调整各波段在局部区域和全局语义层面的参与程度。例如,在夜间场景中,热红外通道的权重自动增强,而可见光通道被抑制;在晴朗白天,可见光纹理信息主导融合过程。这种数据驱动的动态分配策略有效避免了无效或噪声模态对融合结果的干扰,提升了特征表达的任务相关性与环境适应性。

#### 3.2 跨模态对齐机制:基于对比学习的语义一致性约束

多光谱融合的核心挑战在于不同成像机理导致的特征分布偏移,即使同一目标在不同波段下也可能呈现显著表征差异。为解决这一“形合神离”问题,本文引入跨模态对比学习机制,在训练阶段显式约束语义一致

性。具体而言,在共享语义空间中,对每个目标实例提取其在可见光、近红外和热红外模态下的高层特征向量,构建正样本对(同一目标的不同模态)与负样本对(不同目标的任意模态组合)。采用归一化温度缩放交叉熵损失(NT-Xent),拉近正样本对之间的余弦相似度,同时推远负样本对的距离。该损失函数与主识别任务联合优化,迫使网络学习到模态不变的语义嵌入。对齐过程不依赖像素级配准,仅需目标级标注,适用于实际部署中存在轻微视差或多源异步采集的场景。实验表明,该机制显著提升跨模态特征的聚类紧凑性与类别可分性,使融合后的表示更聚焦于目标本质属性而非成像物理特性,从而增强下游识别模型的泛化能力与鲁棒性。

#### 3.3 轻量化架构设计:共享骨干与分支微调的高效融合结构

面向边缘计算与嵌入式平台的实际部署需求,本文在保证性能的同时注重模型效率。整体架构采用“共享骨干+模态特定微调分支”的设计范式。主干网络基于轻量级 ConvNeXt-Tiny 构建,负责提取通用低层特征,三个模态共用该部分以减少参数冗余;在高层语义阶段,各模态分流至独立微调模块,保留其独特表征能力,避免过早融合导致的信息混淆。自适应权重模块与跨模态对齐头均置于骨干输出之后,仅增加少量可学习参数。整个融合流程端到端训练,推理时仅需一次前向传播即可完成多模态整合与目标识别。该结构在 KAIST 数据集上实测参数量仅为 12.3M,推理速度达 28 FPS (Jetson AGX Xavier),显著优于多数双流或三流独立编码器方案。轻量化设计未以牺牲精度为代价,反而因参数共享带来的正则化效应提升了泛化性能,体现了实用性与先进性的统一。

#### 3.4 理论依据:信息瓶颈原则指导下的融合优化目标

本文策略的理论根基源于信息瓶颈理论,该理论主张在压缩输入信息的同时最大化其对目标任务的预测能力。多光谱融合本质上是在高维异构输入中提取与目标识别最相关的公共信息,同时剔除模态特异性噪声与冗余。自适应权重机制对应“压缩”过程,通过动态抑制低信噪比波段实现信息精简;跨模态对齐则对应“预测”优化,确保保留的信息在语义层面一致且判别性强。二者协同作用,使融合特征逼近信息瓶颈最优解:既充

分保留跨模态共有语义，又有效过滤无关变异。该理论视角为融合策略提供了可解释性框架，也解释了为何单纯追求图像视觉质量或特征维度叠加难以提升识别性能。本文方法将信息瓶颈思想转化为可训练的神经网络组件，实现了从经验融合向原理驱动的模式转变。

## 4 实验设计与结果分析

为全面验证所提多光谱融合策略的有效性与实用性，实验在三个具有代表性的数据集上展开：自建城市道路多光谱数据集（UR-MultiSpec）、KAIST 多光谱行人检测数据集以及 FLIR ADAS 热成像基准。UR-MultiSpec 包含 5000 组严格时空对齐的可见光、近红外与热红外三模态图像，覆盖白天、黄昏、夜间、雨雾及强眩光等多种复杂交通场景，标注涵盖车辆、行人、非机动车等六类目标，旨在贴近真实部署环境。KAIST 和 FLIR 则作为公开权威基准，用于横向对比与泛化能力评估。实验选取四类典型融合方法作为基线：Early Fusion（输入级拼接）、Late Fusion（决策级投票）、TarDAL（基于任务感知的对抗融合）以及 CFT（跨模态特征 Transformer）。所有方法均在相同训练策略与硬件环境下复现，确保公平比较。评价指标包括目标检测标准 mAP@0.5、Recall、F1-score，以及在 NVIDIA Jetson AGX Xavier 嵌入式平台上的推理速度（FPS），兼顾精度与实时性。实验结果表明，本文方法在 UR-MultiSpec 正常光照条件下取得 82.4% 的 mAP@0.5，优于当前最优基线 TarDAL 的 78.2%，提升幅度达 4.2 个百分点；在更具挑战性的夜间与烟雾遮挡子集中，性能优势进一步扩大，mAP 达到 76.9%，较 TarDAL 提升 9.1%，充分证明所提策略在低能见度、弱纹理等恶劣条件下的鲁棒性增强。Recall 与 F1-score 同步提升，说明漏检率显著降低且分类一致性改善。推理效率方面，模型在 Jetson 平台实现 28 FPS 的稳定运行，满足自动驾驶与智能安防等场景对实时处理的基本要求。为深入剖析各组件贡献，开展消融实验：移除自适应权重模块后，mAP 下降 3.5%，尤其在光照剧烈变化场景中性能波动加剧；关闭跨模态对齐损失则导致 mAP 回落 3.3%，特征分布可视化显示不

同模态下同类目标表征离散度明显增大。两项机制联合使用产生协同效应，总增益达 6.8%，验证了“动态加权+语义对齐”双驱动设计的必要性。进一步分析显示，在强反射路面或部分遮挡情形下，传统方法常将热斑误判为行人，而本文方法因有效融合多模态语义线索，显著抑制此类虚警。可视化注意力图亦表明，网络能自动聚焦于目标本体区域，并根据环境动态调整波段依赖——夜间热红外通道激活强度显著高于可见光，而晴朗白天则相反。这些实证结果不仅证实了方法在精度、鲁棒性与效率三方面的综合优势，也为其在实际光电系统中的工程落地提供了可靠支撑。

## 5 结束语

本文针对多光谱光电图像目标识别中的融合效率与语义一致性问题，提出了一种融合自适应权重分配与跨模态特征对齐的新策略。该方法突破了传统融合模型对固定规则或浅层特征的依赖，从信息有效性与语义连贯性两个维度重构融合逻辑，显著提升了复杂环境下的识别精度与鲁棒性。实验验证了其在多种典型场景中的优越性能，同时兼顾了轻量化部署需求，具备良好的工程实用价值。未来工作将拓展至高光谱与雷达-光学融合，并探索无监督或少样本条件下的自适应融合机制。此外，如何将该策略嵌入端到端感知-决策系统，实现更高层级的智能协同，亦是值得深入的方向。本研究为多模态感知技术的发展提供了兼具理论创新与实践意义的新范式。

## 参考文献

- [1] 袁鹏飞. 基于多光谱 LIDAR 点云的道路网提取[D]. 武汉大学, 2018.
- [2] 孙昆, 武一, 牛雅睿, 等. 基于多阶段交叉信息融合的多光谱行人检测[J]. 电子测量技术, 2023, 46(15): 118-125.
- [3] 徐安健, 刘海, 杨庆华, 等. 两视场热像仪或两视场光学精密成像仪器基准面装调装置: CN202121641437.5[P]. CN215598672U[2026-04-02].