

多模态融合知识蒸馏的无人机路桥检测^①



梁 巧¹, 杨德刚², 王 杰³

¹(重庆对外经贸学院 数学与计算机科学学院, 重庆 401520)

²(重庆师范大学 信息技术中心, 重庆 401331)

³(重庆对外经贸学院 教育学院, 重庆 401520)

通信作者: 梁 巧, E-mail: liangqiao1011@126.com

摘 要: 针对传统路桥检测技术存在的效率低、成本高及安全风险问题, 以及当前无人机检测中多模态模型参数量大、难以在机载平台实时部署的挑战, 本文提出一种基于交叉蒸馏的多模态特征融合路桥检测模型. 该模型采用双分支教师网络与单分支学生网络架构, 通过教师网络间的特征交互与协同蒸馏机制, 实现多模态特异性知识的高效迁移; 同时引入基于注意力机制的动态特征融合模块, 强化对路桥缺陷关键特征的感知能力. 实验表明: 在保持检测精度 $mAP@0.5$ 为 89.6% 的同时, 该模型参数量降至 8.2M, 推理速度达 32.6 f/s, 性能显著优于传统多模态融合及轻量化方法. 相比特征拼接、单模态蒸馏后融合等策略, 其检测精度与计算效率均具明显优势. 消融实验证实了交叉蒸馏机制与注意力融合模块的有效性. 该模型成功实现了路桥缺陷的高精度轻量化检测, 为无人机路桥检测工程应用提供了技术基础.

关键词: 交叉蒸馏; 多模态特征融合; 双分支教师网络; 注意力机制

引用格式: 梁巧, 杨德刚, 王杰. 多模态融合知识蒸馏的无人机路桥检测. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10086.html>

UAV-based Road and Bridge Inspection Combining Multimodal Fusion and Knowledge Distillation

LIANG Qiao¹, YANG De-Gang², WANG Jie³

¹(School of Mathematics and Computer Science, Chongqing College of International Business and Economics, Chongqing 401520, China)

²(Information Technology Center, Chongqing Normal University, Chongqing 401331, China)

³(School of Education, Chongqing College of International Business and Economics, Chongqing 401520, China)

Abstract: To address the inefficiencies, high costs, and safety risks inherent in traditional road and bridge inspection techniques, as well as the challenges posed by the large parameter volumes of current multimodal detection models and the difficulty in achieving real-time deployment on unmanned aerial vehicle (UAV) platforms, this study proposes a multimodal feature fusion road and bridge detection model based on cross distillation. The model employs a dual-branch teacher network and a single-branch student network architecture. Efficient knowledge transfer of modality-specific features is achieved through feature interaction and collaborative distillation mechanisms between the teacher networks. Concurrently, a dynamic feature fusion module, utilizing attention mechanisms, is introduced to enhance the perception of critical features associated with road and bridge defects. Experimental results demonstrate that, while maintaining a detection precision of 89.6% $mAP@0.5$, the proposed model reduces its parameter size to 8.2M and achieves an inference speed of 32.6 f/s. These results significantly outperform traditional multimodal fusion and lightweight methods. Compared to strategies utilizing feature concatenation or post-distillation unimodal fusion, the proposed model shows clear advantages in both detection accuracy and computational efficiency. Ablation studies confirm the effectiveness of

① 基金项目: 重庆市教育委员会科学技术研究计划青年项目 (KJQN202402002); 2024 年重庆对外经贸学院科研项目 (KYZK2024005); 重庆对外经贸学院教育教学改革研究项目 (JG2025034)

收稿时间: 2025-08-15; 修改时间: 2025-09-16; 采用时间: 2025-09-26; csa 在线出版时间: 2026-01-08

the cross-distillation mechanism and the attention-based fusion module. The model successfully enables high-precision, lightweight detection of road and bridge defects, thus providing a technical foundation for the engineering application of UAV-based road and bridge inspection.

Key words: cross-distillation; multimodal feature fusion; dual-branch teacher network; attention mechanism

随着无人机技术的快速发展,其在道路桥梁检测领域的应用日益广泛.多模态融合知识蒸馏技术通过整合多种传感数据,显著提升了检测精度与效率^[1].道路桥梁作为交通基础设施的核心构件,其结构健康状况直接关乎交通运行效率及人民生命财产安全,因此实施定期且精准的检测至关重要.目前国内外无人机路桥检测方面仍存在诸多问题:缙永涛等^[2]提出的无人机路桥病害检测系统,虽实现了自动化数据采集功能,但尚未解决多模态模型的轻量化设计问题.徐倩等^[3]在无人机红外目标检测技术综述中指出,红外模态在隐蔽缺陷检测中具有应用潜力,然而现有研究对多模态融合策略的优化仍显不足.国外研究方面,Gou等^[4]的知识蒸馏综述虽验证了该技术在模型压缩中的有效性,但未涉及多模态异构数据的知识迁移机制.Himeur等^[5]研究了遥感领域知识蒸馏的应用,虽阐释了多模态技术在遥感检测中的优势,但针对路桥场景的适配性研究仍较为匮乏.本研究提出的多模态特征融合知识蒸馏方法,可有效提升结构缺陷识别的准确性并实现模型轻量化,为解决深度学习模型在无人机平台部署的难题提供了有效途径.

1 无人机路桥检测

当前无人机路桥检测研究已取得一定进展:数据采集自动化技术趋于成熟^[6],基于卷积神经网络的算法已能对典型结构缺陷进行初步识别^[7].但仍存在不足之处:数据处理环节仍依赖人工干预^[8],在复杂环境下的识别任务鲁棒性欠佳^[9].多模态融合研究主要集中在3个层面:数据层融合易受模态间差异干扰^[10];特征层融合难以充分挖掘多模态间的互补关系^[11];决策层融合虽稳定性高,但易受单一模态误判影响^[12].单一模态数据存在固有局限,包括可见光图像易受光照、阴影干扰,难以识别细微裂缝;红外热成像虽能捕捉结构内部温度差异,却缺乏空间纹理细节;激光雷达点云可提供深度信息以定位坑洞、沉降,但对裂缝等平面缺陷表征不足.多模态融合可通过互补不同模态的优势,解

决单一模态在复杂路桥场景下的检测盲区,显著提升缺陷识别准确率与鲁棒性.

当前的主流融合策略对多模态特征间的内在关联性建模不足^[13],知识蒸馏技术在多模态场景下的适配性亦有待优化^[14],异构数据特性导致知识迁移困难^[15];现有方法缺乏专门针对多模态融合任务设计的迁移机制,该领域将知识蒸馏与多模态融合相结合的研究尚处于起步阶段^[16].本研究采用的多模态数据具体包括3类:第1类分辨率1920×1080的可见光图像,用于捕捉表面纹理、裂缝形态;第2类分辨率640×512的红外热成像数据,反映结构内部温度异常,识别剥落、渗漏等隐蔽缺陷;第3类点密度200点/m²的激光雷达点云数据,提供三维深度信息,定位坑洞、沉降、桥面断裂等立体缺陷;旨在构建一种交叉蒸馏机制,通过促进教师模型间的特征交互并实现向学生模型的有效知识迁移,达成多模态特征的深度融合与模型的高效压缩,从而提升复杂环境下的检测精度与实时性,为实际工程应用提供技术支撑.

2 交叉蒸馏特征融合

2.1 实验模型整体框架设计

基于交叉蒸馏的多模态特征融合路桥检测模型框架以可见光图像、红外热图像及激光雷达点云等多模态路桥数据作为输入.数据经预处理后分别输入至模型的特征提取模块.该框架主要由4个核心模块构成:(1)特征提取模块:针对不同模态数据设计专用特征提取子网络,深度挖掘各模态的独特判别信息;(2)交叉蒸馏模块:本文的核心工作,通过教师网络间的特征交互以及教师-学生网络间的知识迁移,实现多模态知识的有效传递与整合,其多特征融合网络算法结构如图1所示;(3)特征融合模块:对经蒸馏处理的多模态特征进行自适应融合,生成蕴含全面信息的综合特征;(4)检测头模块:基于融合特征执行路桥缺陷的定位与分类,输出最终检测结果.

该框架的优势体现在:其一,通过多模态特征融合,

充分利用不同模态数据的互补性, 显著提升对复杂路桥缺陷的识别能力; 其二, 借助交叉蒸馏机制, 在确保检测精度的同时, 有效降低学生模型的参数量与计算复杂度, 使其满足无人机平台的部署需求.

2.2 交叉蒸馏机制设计

2.2.1 教师-学生网络架构

模型采用“双分支教师网络-单分支学生网络”架构, 如图 2 所示.

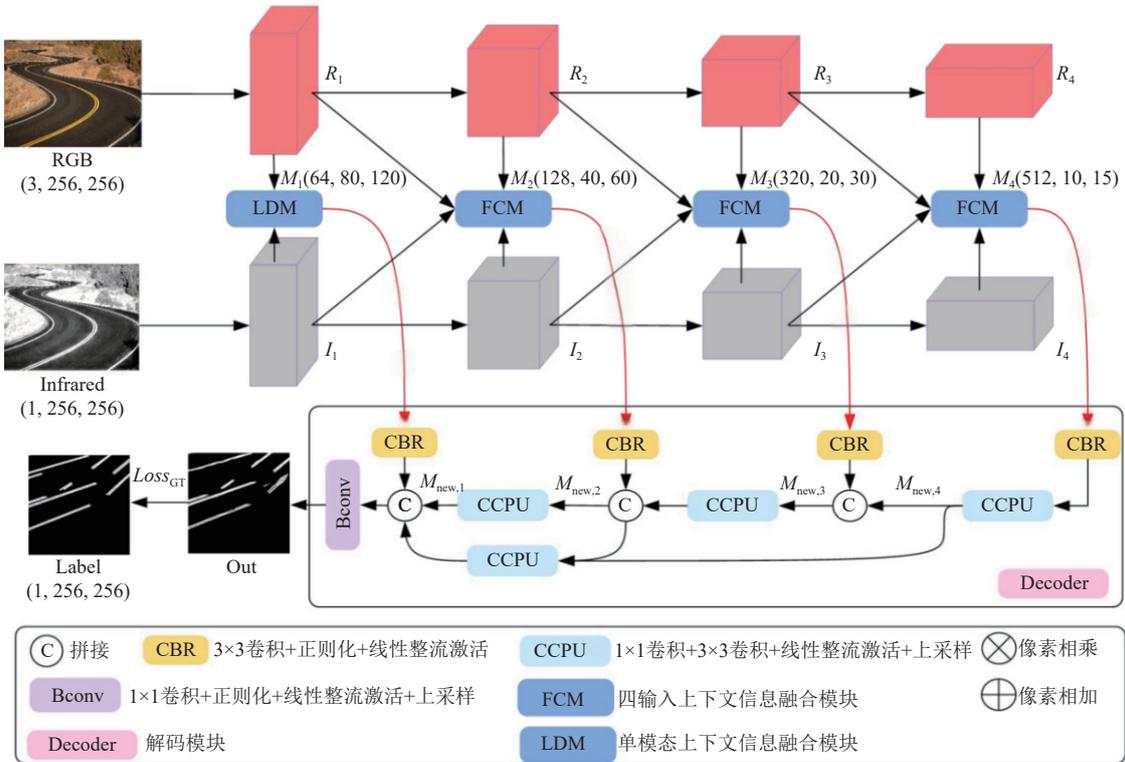


图 1 多特征融合网络算法结构图

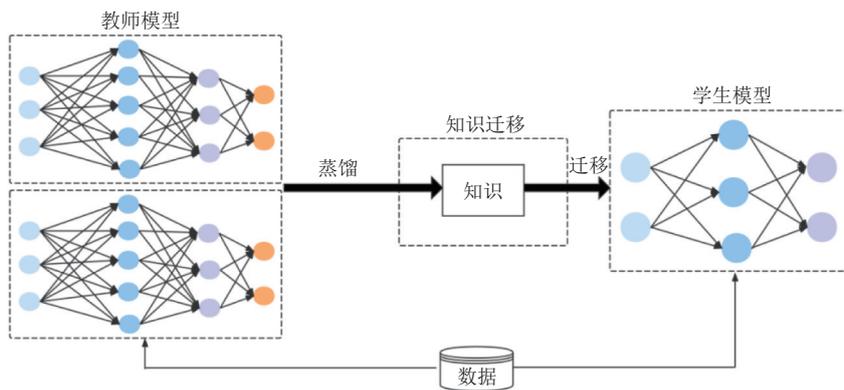


图 2 教师-学生网络架构

双分支教师网络分别处理异构多模态数据: 教师模型 A (Teacher-A) 基于改进的 Faster R-CNN 架构^[17], 专注于学习可见光与红外图像的纹理及温度特征, 输出缺陷的二维位置坐标与类别概率分布. 教师模型 B (Teacher-B) 同样基于改进的 Faster R-CNN 架构, 聚焦于激光雷达点云数据的深度特征提取, 输出缺陷的三

维形态参数. 两者通过特征交互模块实现协同: 在特征提取中间层引入特征对齐模块, 将 Teacher-A 的二维特征映射至三维空间, 并与 Teacher-B 的点云特征进行空间对齐与特征交互. 该机制促使 Teacher-A 学习深度关联信息, 同时使 Teacher-B 融合纹理特征, 从而实现跨模态特征互补.

学生网络采用轻量化 YOLO11n 架构^[18]改进, 通过知识蒸馏融合双教师网络知识, 实现高效路桥缺陷检测. 选择 YOLO11n 作为基础架构, 系因其在保证检测精度前提下具备模型体积小、推理速度快的特性, 更适于计算资源受限的无人机平台部署.

2.2.2 交叉蒸馏策略

为了在较小的网络上获得更好的性能, 我们利用响应蒸馏函数中的 Kullback-Leibler 散度 (KLD) 损失, 将更多样化的知识从多特征融合教师网络 (multi-feature fusion network-teacher, MFNet-T) 转移到多特征融合学生网络 (multi-feature fusion network-student, MFNet-S), 为此, 本文设计交叉蒸馏方法如图 3 所示. 交叉蒸馏策略主要通过两个层面实现知识的高效迁移: 第 1 阶段为教师间交叉蒸馏, 采用特征匹配损失函数, 如式 (1) 所示. 其目的是约束两个教师网络的中间层特征分布一致, 实现多模态特异性知识的共享. 第 2 阶段为教师-学生协同蒸馏, 双教师网络共同指导基于轻量化 YOLO11n 的学生网络进行知识传递, 即将 Teacher-A 与 Teacher-B 输出的融合特征, 通过特征匹配损失, 引导学生网络学习跨模态特征关联. 为了学习更多的特征信息, 去除无用的深层信息, 在最后一层 (图 3 中 Layer4) 使用骰子损失进行特征蒸馏, 因篇幅有限, 不再详述. 双教师网络输出预测概率分布, 并通过 KLD 损失引导学生网络学习其决策逻辑, 即软标签损失, 如式 (2) 所示. 学生网络自身的硬标签监督损失如式 (3) 所示, 其通过结合双教师监督 (如式 (4)) 和真实标签约束 (如式 (5)) 而形成优化机制.

$$L_{\text{feature}} = L_{\text{srm}}(S_1, T_2) + L_{\text{srm}}(S_2, T_1) + L_{\text{srm}} + L_{\text{srm}}(S_3, T_4) + L_{\text{srm}}(S_4, T_3) \quad (1)$$

$$L_{\text{kld}} = \text{KLD}(Out_S, MFS_{\text{label}}) \quad (2)$$

其中, Out_S 表示 MFNet-S 的预测图, S_i ($i=1, 2, 3, 4$) 是 MFNet-S 的 M_i ($i=1, 2, 3, 4$) 通过上下采样和 CBR 调整通道数得到的特征输出, 分别与 MFNet-T 的 M_i ($i=1, 2, 3, 4$) 通过 T_i ($i=1, 2, 3, 4$) 来表示进行对齐. 硬标签监督损失 L_k 以及 L_{svm} 、 L_{srm} 的计算方法如式 (3)–式 (5).

$$L_k = L_{\text{svm}}(MFT_{\text{out}}, MFS_{\text{label}}) + L_{\text{srm}}(MFT_{\text{out}} + MFS_{\text{label}}) \quad (3)$$

$$L_{\text{svm}} = \frac{\sum_{i=1}^N y_i \log p_i}{N} + (1 - y_i) \log(1 - p_i) \quad (4)$$

$$L_{\text{srm}} = 1 - \frac{2 \sum_{i=1}^N s_i t_i}{\sum_{i=1}^N s_i + \sum_{i=1}^N t_i} \quad (5)$$

其中, MFT_{out} 和 MFS_{label} 分别表示 MFNet-T 和 MFNet-S 的输出和真实标签. s_i 和 t_i 分别表示双教师监督中预测和标签每个像素值. y_i 和 p_i 分别表示真实标签约束中预测和标签的每个像素的值. L_{svm} 和 L_{srm} 损失用于监督具有 GT 信息的输出, 本文在 L_k 中为两个损失函数分配相同的权重.

最后, 获得整体损失函数如下:

$$L_{\text{total}} = L_{\text{kld}} + L_{\text{feature}} + L_k \quad (6)$$

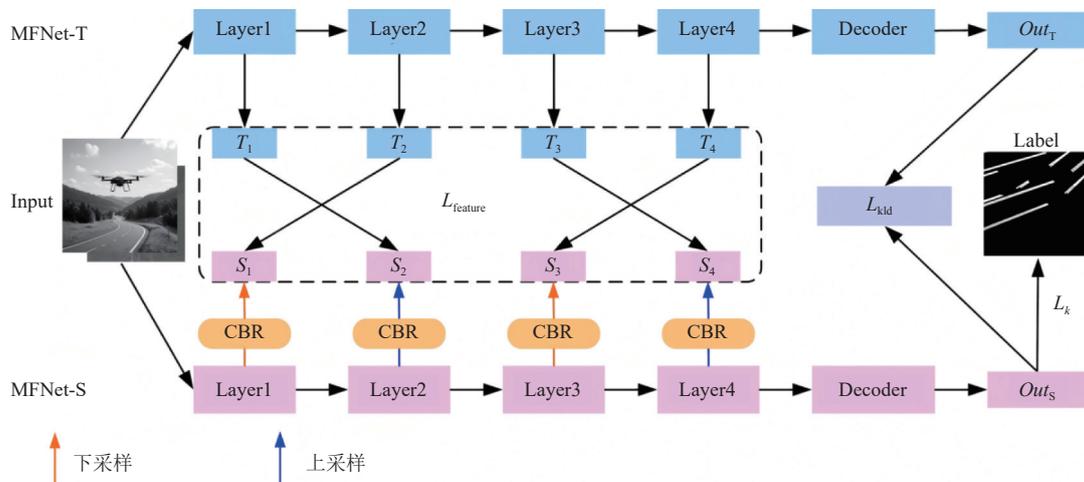


图 3 交叉蒸馏知识方法

2.3 多模态特征融合模块

基于注意力机制的多模态特征融合分4步实现.

第1步: 特征预处理. 接收来自 Teacher-A 网络的可见光-红外融合特征、Teacher-B 网络的激光雷达点云特征, 以及学生网络自身提取的单模态特征. 利用 1×1 卷积层统一各特征图的通道维度并执行降维操作, 以减少后续计算复杂度.

第2步: 注意力权重分配. 引入空间注意力模块与通道注意力模块. 空间注意力模块通过 3×3 卷积层与 Sigmoid 激活函数生成空间权重图, 聚焦于裂缝、坑洞等缺陷区域, 增强目标区域特征响应强度. 通道注意力模块则通过全局平均池化与全连接层生成通道权重向量, 依据任务需求自适应调整通道重要性. 例如, 为裂缝检测分配更高权重至可见光纹理通道, 而为坑洞检测则分配更高权重至激光雷达深度通道.

第3步: 双向跨尺度融合. 采用“自下而上”与“自上而下”相结合的双向路径进行特征融合. 低层特征经上采样传递丰富的细节信息, 强化高层特征的缺陷边缘表征能力; 高层特征则通过下采样传递高级语义信息, 优化低层特征的类别判别性能. 融合过程中, 通过动态权重系数实现不同层级特征的加权融合.

第4步: 融合特征输出. 将跨尺度融合后的特征输入 Conv Fusion 层, 经 3×3 卷积与批量归一化 (Batch-Norm) 处理后, 输出最终的多模态融合特征图, 用于后续检测头执行缺陷定位与分类任务.

基于注意力机制的多模态特征融合模块设计了多路径特征融合与自适应加权机制, 显著提升了不同尺度特征图间的交互能力. 通过双向融合机制, 低层特征经由自下而上的路径增强高层特征的细节信息, 同时高层特征通过自上而下的路径传递丰富的语义信息. 尤其在跨尺度融合过程中, 通过动态权重分配策略, 确保了低层特征与高层特征的高效交互, 有效缓解了传统上下采样过程中的细节信息丢失问题, 结构如图4所示.

2.4 模型训练过程

模型训练过程分为3个阶段: 第1阶段为教师网络预训练, 如图5所示. 在此阶段, 分别于大规模标注的多模态路桥数据集上独立训练 Teacher-A 与 Teacher-B 网络. 优化器采用随机梯度下降 (SGD), 初始学习率设定为 0.001, 批大小 (batch size) 为 16, 训练持续 50

个轮次 (epochs), 直至模型收敛. 此阶段旨在使教师网络获得较强的单模态或双模态缺陷检测能力.

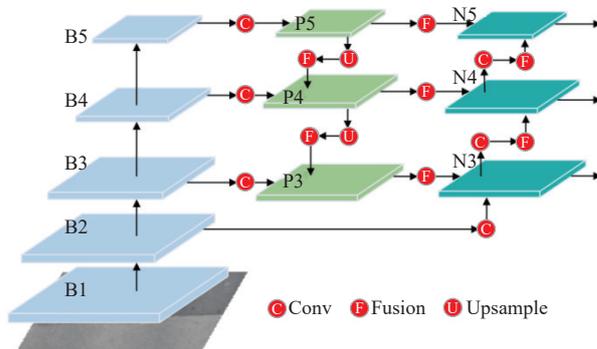


图4 基于注意力机制的多模态特征融合模块

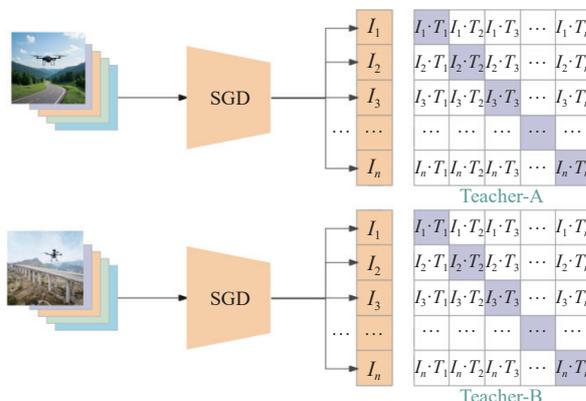


图5 教师网络训练

第2阶段为学生网络蒸馏训练, 如图6所示. 固定经预训练的双分支教师网络参数, 以其特征输出与决策输出作为监督信号, 指导学生网络的训练. 此阶段采用 Adam 优化器, 初始学习率为 0.0001, batch size 为 32, 并引入设计的混合损失函数, 训练 30 个 epochs. 该过程旨在促进学生网络逐步习得教师网络的知识.

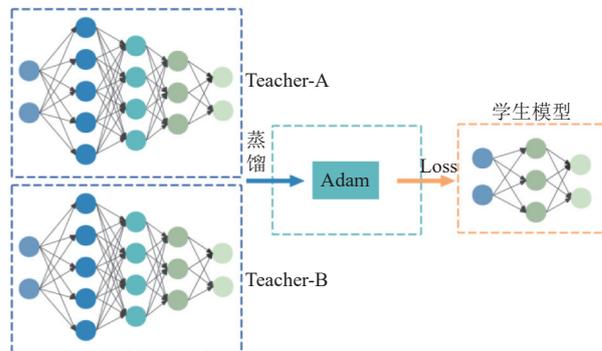


图6 学生网络蒸馏训练

第3阶段为联合微调, 如图7所示. 解除教师网络参数的固定状态, 将教师网络与学生网络视为整体进

行训练. 学习率采用较小的 0.000 01, batch size 为 16, 继续训练 20 个 epochs. 通过反向传播算法同步优化教师网络与学生网络的参数, 以进一步提升模型的整体性能. 整个训练过程中, 学习率采用余弦退火策略进行动态调整, 并引入早停机制: 若验证集上的平均精度均值 (*mAP*) 连续 10 个 epochs 未显著提升, 则终止训练, 以避免模型过拟合.

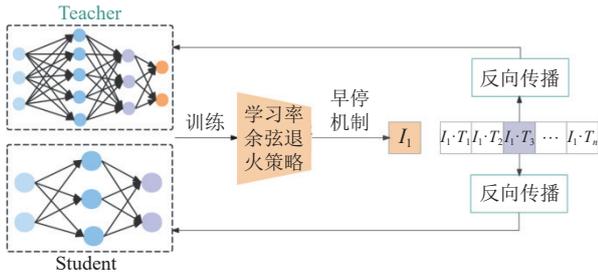


图 7 教师-学生网络联合微调

3 实验结果与分析

3.1 实验平台与超参数设置

实验硬件环境配置如下: 中央处理器 (CPU) 采用 Intel Core i9-12900K, 16 核 24 线程, 基础频率 3.2 GHz; 图形处理器 (GPU) 采用 NVIDIA GeForce RTX 3090, 24 GB GDDR6X 显存; 系统内存为 64 GB DDR4-3200. 软件环境基于 Ubuntu 20.04 LTS 操作系统, 深度学习框架选用 PyTorch 1.10.0, 并集成 CUDA 11.3 与 CuDNN 8.2.0 以加速计算过程. 模型训练超参数设置如下: 初始学习率设置为 0.001, 采用余弦退火策略动态调整学习率, 衰减系数为 0.9. Batch size 固定为 16, 训练总轮次 epochs 为 100. 知识蒸馏温度系数 (temperature parameter) 设定为 3.0, 用于软化教师网络输出的概率分布;

特征匹配损失、软标签损失与硬标签损失的权重系数分别为 0.5、0.3 及 0.2, 通过加权组合构成混合损失函数.

3.2 实验数据集简介

本研究所用数据集由两部分构成: 一部分为实地采集数据, 桥梁数据包括沙溪嘉陵江大桥 (主跨 180 m)、重庆城市快速路立交桥 (3 层互通) 以及 3 座县级公路简支梁桥; 城市主干道数据包括重庆市合川区中心主干道 (沥青路面, 通车 5 年)、工业园区主干道 (水泥路面, 通车 8 年) 以及产业园区主干道 (复合路面, 通车 3 年). 采集时间覆盖 2025 年 3-6 月, 每处场景采集多模态数据各 200-300 组. 另一部分为公开数据集, 选用 Bridge Damage Dataset^[19] 与 Road Defect Dataset^[20] 数据集. Bridge Damage Dataset 来自美国加州大学伯克利分校土木与环境工程系 (2018 年发布), 包含 2000 个桥梁缺陷样本. Road Defect Dataset 来自英国运输研究实验室 (2023 年发布), 包含 3000 个道路缺陷样本. 该数据集严重程度等级分为 3 级: 轻微 (缺陷面积小于 0.1 m² 或宽度小于 1 mm)、中度 (缺陷面积为 0.1-1 m² 或宽度为 1-3 mm)、重度 (缺陷面积大于 1 m² 或宽度大于 3 mm).

本文使用标注工具为 Label Studio, 标注精度控制在 ±1 像素. 所有数据的不同道路桥梁缺陷类型分布情况详见表 1, 具体示例如图 8 所示.

表 1 各种路桥数据标签及数量

缺陷类型	标签	数量	缺陷类型	标签	数量
路纵裂缝	D00	3132	桥纵裂缝	D10	1104
路横裂缝	D01	2687	桥横裂缝	D20	955
路复杂裂缝	D02	3275	桥复杂裂缝	D30	2241
道路坑洞	D03	780	桥梁松耦	D40	456
道路补丁	Repair	356	桥面断裂	Fracture	14

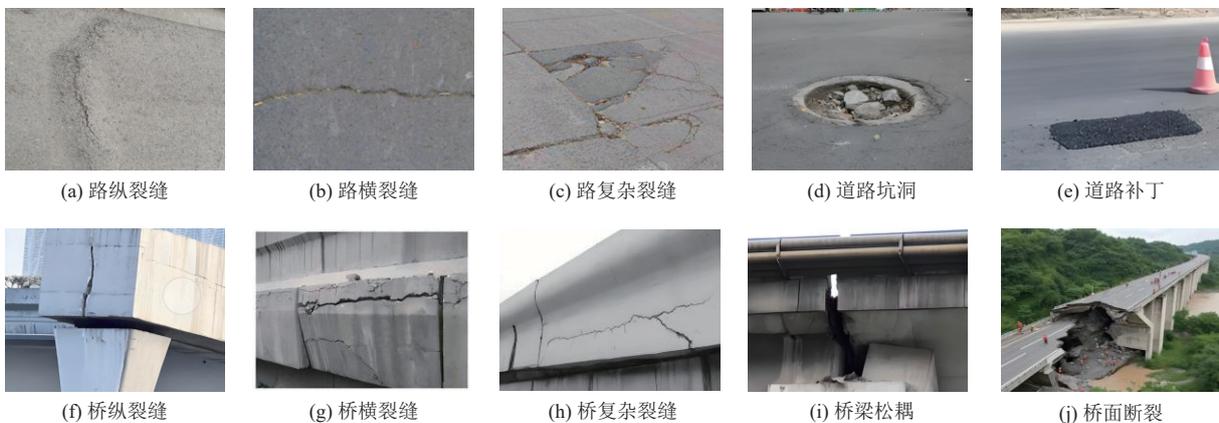


图 8 不同路桥缺陷示例图

3.3 评估指标

实验采用目标检测领域6项核心指标评估模型性能: 精确度 (*Precision*) 表征模型预测为正样本中实际正样本的比例. 召回率 (*Recall*) 描述模型识别实际正样本的能力. *F1* 值 (*F1-score*) 为精确度与召回率的调和平均数, 用于综合评价模型的性能. 平均精度均值 (mean average precision, *mAP*) 通过计算平均精度 (average precision, *AP*) 的均值获得, 本文采用各缺陷类别在 IoU 阈值为 0.5 时的 *mAP*, 即 *mAP@0.5*. 模型参数量 (params) 以百万 (M) 为单位, 衡量模型的存储开销. 浮点运算量 (floating-point operations, FLOPs) 以十亿次运算 (G) 为单位, 表征模型的计算复杂度. 推理速度以每秒处理帧数 (frames per second, FPS) 表征, 用于评估模型的实时处理能力. 评价指标的具体计算如下:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

$$mAP@0.5 = \frac{1}{C} \sum_{c=1}^C AP_c \quad (10)$$

其中, C 是缺陷类别的总数量, AP_c 是第 c 类在 $IoU \geq 0.5$ 时的平均精度.

3.4 不同融合策略效果对比

如表 2 所示, 本研究对交叉蒸馏融合策略与两种传统融合策略 (特征拼接策略: 通过通道维度直接拼接多模态特征; 单模态蒸馏后融合策略: 对各模态模型单独蒸馏后再融合检测结果) 进行了性能对比实验.

表 2 不同融合策略性能对比

融合策略	<i>mAP@0.5</i> (%)	参数量 (M)	FLOPs (G)	推理速度 (f/s)
特征拼接策略	82.3	12.5	28.7	24.1
单模态蒸馏融合策略	85.1	10.8	22.5	27.8
交叉蒸馏融合策略 (本文)	89.6	8.2	15.3	32.6

由表 2 实验结果可得以下结论.

(1) 检测精度: 交叉蒸馏融合策略的 *mAP@0.5* 达 89.6%, 分别超出 82.3% 的特征拼接策略和 85.1% 的单模态蒸馏后融合策略 7.3 个百分点和 4.5 个百分点.

(2) 模型效率: 交叉蒸馏融合策略的模型参数量为 8.2M、FLOPs 为 15.3G, 明显低于特征拼接策略 (12.5M,

28.7G) 和单模态蒸馏后融合策略 (10.8M, 22.5G).

(3) 推理速度: 交叉蒸馏融合策略的推理速度为 32.6 f/s, 优于 24.1 f/s 的特征拼接策略和 27.8 f/s 的单模态蒸馏后融合策略.

上述结果证明, 交叉蒸馏融合策略能够高效整合多模态知识, 在显著提升检测精度的同时, 有效优化模型轻量化性能.

3.5 消融实验

为验证各核心模块的有效性, 设计了 3 组消融实验方案, 实验结果如表 3 所示: 仅保留多模态融合的移除交叉蒸馏机制、采用特征拼接替代的移除注意力融合模块、同时移除交叉蒸馏机制与注意力融合模块. 实验结果表明, 完整模型的 *mAP@0.5* 为 89.6%. 移除交叉蒸馏机制后, *mAP@0.5* 下降 6.4 个百分点, 至 83.2%, 模型参数量增至 11.5M, FLOPs 升至 23.7G, 表明该机制在维持模型精度与轻量化特性方面发挥关键作用; 移除注意力融合模块后, *mAP@0.5* 下降 3.8 个百分点, 至 85.8%, 说明注意力机制能有效增强多模态特征的判别能力; 同时移除两个模块后, *mAP@0.5* 显著下降 11.1 个百分点, 至 78.5%, 实验结果进一步验证了各模块在模型中的协同作用.

表 3 消融实验

实验配置	<i>mAP@0.5</i> (%)	基准模型偏移 (百分点)	Params (M)	FLOPs (G)
完整模型 (基准)	89.6	0	8.2	15.3
移除交叉蒸馏机制	83.2	-6.4	11.5	23.7
移除注意力融合模块	85.8	-3.8	8.9	17.6
同时移除两个模块	78.5	-11.1	12.3	26.5

为验证模型中交叉蒸馏机制与注意力融合模块的有效性进行了核心模块消融实验如表 4 所示, 包含 3 种实验配置: 完整模型 (基准)、仅移除交叉蒸馏机制、替换特征提取子网络 (用 ResNet50 替代)、移除混合损失中的 KLD 损失. 实验结果显示, 替换特征提取子网络后, 模型 *mAP@0.5* 下降 3.3 个百分点, 至 86.3%, 参数量与 FLOPs 分别增至 9.8M、19.2G; 移除 KLD 损失后, *mAP@0.5* 下降 2.5 个百分点, 至 87.1%. 数据证实了交叉蒸馏机制对模型轻量化与精度的关键支撑作用, 以及注意力融合模块对多模态特征判别能力的提升效果, 且两模块存在协同增强模型性能的作用.

3.6 不同模型性能对比

本文提出的交叉蒸馏模型与 6 类主流模型 (可见光单模态模型 YOLO11n、多模态特征拼接模型 Faster

R-CNN、无蒸馏多模态融合模型、EMF-YOLO、SCE-YOLO、多模态 Transformer 模型) 进行性能对比, 结果如表 5 所示, 评估指标包括检测精度 ($mAP@0.5$)、模型参数量 (Params) 与推理速度 (FPS).

表 4 核心模块消融实验结果

实验配置	$mAP@0.5$ (%)	基准模型偏差 (百分点)	Params (M)	FLOPs (G)
完整模型 (基准)	89.6	0	8.2	15.3
替换特征提取子网络 (用ResNet50替代原架构)	86.3	-3.3	9.8	19.2
移除混合损失中的 KLD损失	87.1	-2.5	8.2	15.3

表 5 交叉蒸馏模型与主流模型性能对比

模型	$mAP@0.5$ (%)	Params (M)	推理速度 (f/s)
YOLO11n	76.3	7.8	35.2
Faster R-CNN	82.5	42.8	15.3
无蒸馏多模态融合模型	90.2	18.6	21.4
EMF-YOLO ^[21]	84.2	9.5	29.8
SCE-YOLO ^[22]	85.7	10.2	27.5
多模态Transformer模型 ^[23]	88.9	15.7	18.3
交叉蒸馏模型 (本文)	89.6	8.2	32.6

表 5 结果显示, 本文模型 $mAP@0.5$ 达 89.6%, 仅比无蒸馏多模态模型低 0.6 个百分点, 却较 YOLO11n (76.3%)、Faster R-CNN (82.5%)、EMF-YOLO (84.2%)、SCE-YOLO (85.7%)、多模态 Transformer 模型 (88.9%) 分别提升 13.3、7.1、5.4、3.9、0.7 个百分点; 参数量 (8.2M) 与 YOLO11n (7.8M) 相近, 但显著低于 EMF-YOLO (9.5M)、SCE-YOLO (10.2M)、多模态 Transformer 模型 (15.7M)、无蒸馏多模态融合模型 (18.6M) 及 Faster R-CNN (42.8M); 推理速度 (32.6 f/s) 接近 YOLO11n (35.2 f/s) 且显著优于 EMF-YOLO (29.8 f/s)、SCE-YOLO (27.5 f/s)、无蒸馏多模态融合模型 (21.4 f/s)、多模态 Transformer 模型 (18.3 f/s) 与 Faster R-CNN (15.3 f/s), 实时性能卓越; 充分证明其在精度、轻量化与实时性间实现了更优平衡.

如表 6 所示, 将交叉蒸馏模型与主流模型进行对比: 包括可见光输入单模态模型 YOLO11n、多模态特征拼接模型 Faster R-CNN 及无蒸馏多模态融合模型.

表 6 不同模型性能对比

模型	$mAP@0.5$ (%)	参数量 (M)	推理速度 (f/s)
YOLO11n	76.3	7.8	35.2
Faster R-CNN	82.5	42.8	15.3
无蒸馏多模态融合模型	90.2	18.6	21.4
交叉蒸馏模型 (本文)	89.6	8.2	32.6

图 9 所示结果表明, 交叉蒸馏模型的 $mAP@0.5$ 达到 89.6%, 较 76.3% 的 YOLO11n 提升了 13.3 个百分点, 较 82.5% 的 Faster R-CNN 提升了 7.1 个百分点, 虽略低于 90.2% 的无蒸馏多模态模型, 但差距仅为 0.6 个百分点; 在模型规模方面, 交叉蒸馏模型的参数量为 8.2M, 显著低于 42.8M 的 Faster R-CNN 和 18.6M 的无蒸馏多模态模型; 在推理速度方面, 交叉蒸馏模型为 32.6 f/s, 优于 Faster R-CNN 的 15.3 f/s 和无蒸馏多模态模型的 21.4 f/s, 并接近 YOLO11n 的 35.2 f/s. 综上, 交叉蒸馏模型在精度、模型规模和推理速度之间实现了更优的平衡.

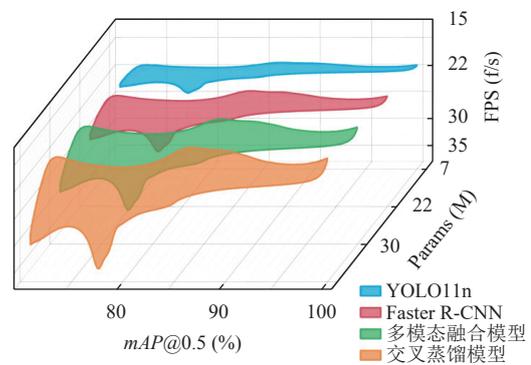


图 9 不同模型 $mAP@0.5$ 对比

3.7 路桥缺陷检测可视化及分析

路桥缺陷可视化对比结果详见图 10.

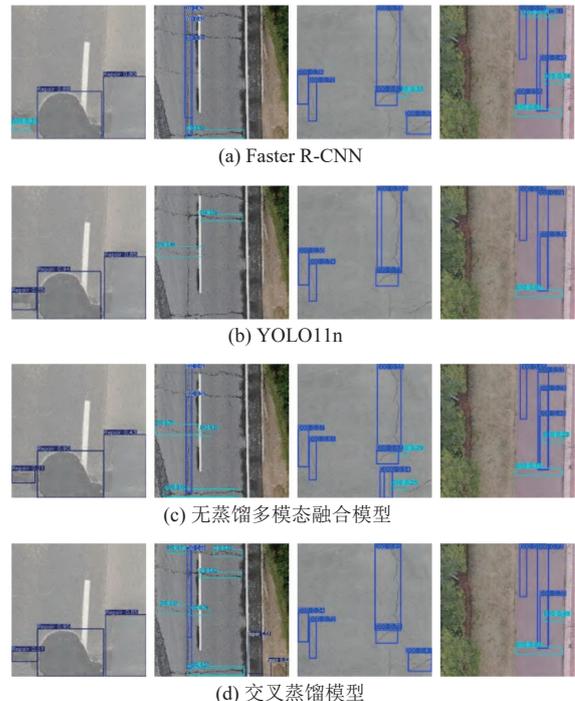


图 10 路桥缺陷可视化对比

对典型缺陷案例进行可视化分析表明:在宽度小于 2 mm 的细微裂缝检测中,交叉蒸馏模型能精准定位裂缝端点,而 YOLO11n 存在漏检现象, Faster R-CNN 则出现定位偏移.针对大面积剥落区域,交叉蒸馏模型对边缘轮廓的识别完整性显著提升,此优势归因于激光雷达深度特征的融合机制.在复杂场景下,交叉蒸馏模型展现出较强鲁棒性:逆光环境中通过红外数据辅助可将裂缝检出率维持在 85%,相较之下单模态模型降至 62%;植被遮挡区域通过注意力机制抑制干扰特征,其误检率较 Faster R-CNN 降低 18 个百分点.漏检现象主要集中于宽度小于 0.5 mm 的极细微裂缝及强噪声干扰区域,误检多由阴影与裂缝的纹理相似性导致.

3.8 热力图可视化分析

通过 Grad-CAM 算法生成注意力热力图的结果如图 11 所示.分析图 11 可知,完整模型生成的注意力热力图在裂缝、坑洞等缺陷区域呈现出显著的激活响应,且能完整捕捉缺陷边缘的形态学特征;移除交叉蒸馏机制后,热力图对多模态特征的注意力分布呈现发散态势,局部区域出现响应缺失现象;学生网络与教师网络热力图的结构相似性指数 (SSIM) 达到 89%,证实交叉蒸馏机制有效传递了特征关注模式.在裂缝检测任务中,热力图显示模型主要聚焦于可见光纹理特征与激光雷达深度突变区域的叠加响应,有效佐证了多模态知识的协同融合机制,从而增强了模型决策过程的可解释性.

4 结论与展望

4.1 研究结论

针对无人机路桥检测任务中精度与轻量化的双重需求,本研究提出一种基于交叉蒸馏的多模态特征融合模型.所提模型构建了双分支教师网络与单分支学生网络的架构框架,通过教师网络间的特征交互与联合蒸馏机制,实现了多模态特异性知识的高效迁移;此外,引入基于注意力机制的动态特征融合模块,有效增强了模型对路桥缺陷关键特征的捕捉能力.实验结果表明,该模型在保持 89.6% $mAP@0.5$ 检测精度的前提下,将参数量压缩至 8.2M,推理速度提升至 32.6 f/s,其性能显著优于传统多模态融合及模型轻量化方法,成功实现了路桥缺陷的高精度轻量化检测,验证了所提交叉蒸馏策略与融合机制的有效性与优越性.

4.2 研究展望

本研究存在如下局限性:首先,数据集场景覆盖范

围有限,主要集中于中东部地区常见路桥类型,对高原、严寒等特殊环境的路桥数据收录不足;其次,极端气象,如暴雨、浓雾等条件下的检测鲁棒性有待进一步提高,多模态数据在恶劣环境中的特征退化问题尚未完全解决;最后,模型对微观尺度缺陷的检出率仍具提升空间.

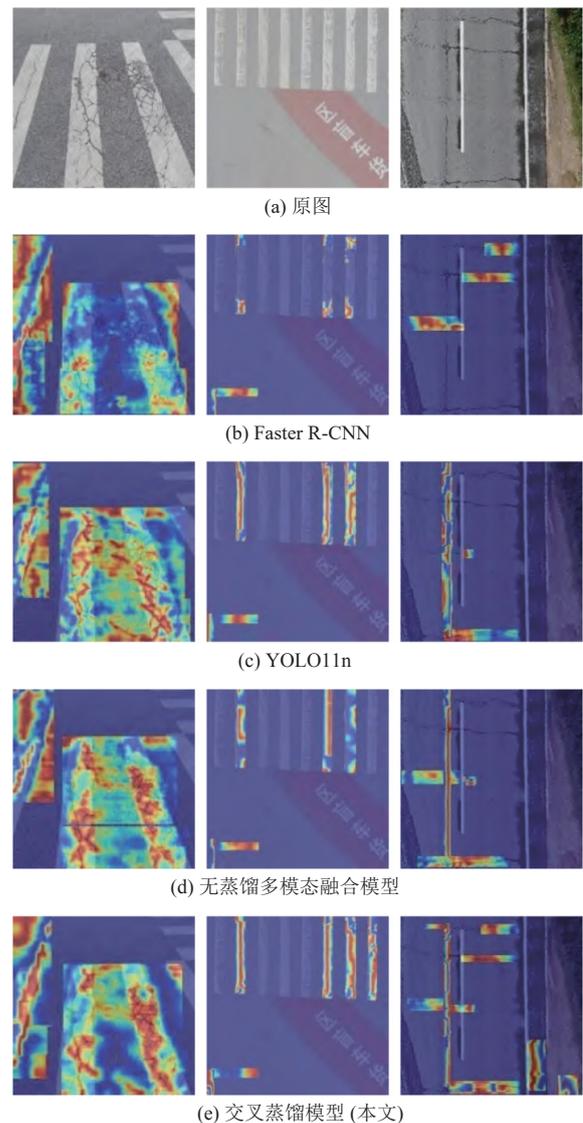


图 11 不同模型热力图结果

未来研究可从以下 4 个方向推进:其一,扩充数据集的场景覆盖范围,补充特殊环境与极端气象条件下的路桥数据,以增强模型泛化能力;其二,优化交叉蒸馏策略,通过引入动态温度调节机制与模态自适应权重分配,提升跨模态知识迁移效率;其三,丰富多模态数据类型,融合应力波、声学等物理信号,深入探究多源数据与结构健康状态的关联机制;其四,将模型扩展

至路桥变形监测领域, 结合时序多模态数据实现对结构形变的长期动态监测与预警, 从而提升检测系统的工程实用价值。

参考文献

- 1 毛克彪, 代旺, 郭中华, 等. AI 知识蒸馏技术演进与应用综述. 农业大数据学报, 2025, 7(2): 144–154.
- 2 缙永涛, 李文博, 段学锋. 基于无人机的路桥病害检测系统. 测绘通报, 2022(12): 141–146, 169.
- 3 徐倩, 韩楠, 董文安, 等. 无人机红外目标检测技术: 现状与展望. 无线电工程, 2025, 55(9): 1764–1774.
- 4 Gou JP, Yu BS, Maybank SJ, *et al.* Knowledge distillation: A survey. *International Journal of Computer Vision*, 2021, 129(6): 1789–1819. [doi: [10.1007/s11263-021-01453-z](https://doi.org/10.1007/s11263-021-01453-z)]
- 5 Himeur Y, Aburaed N, Elharrouss O, *et al.* Applications of knowledge distillation in remote sensing: A survey. *Information Fusion*, 2025, 115: 102742. [doi: [10.1016/j.inffus.2024.102742](https://doi.org/10.1016/j.inffus.2024.102742)]
- 6 田煜. 基于人工智能的路桥工程施工关键技术与管理研究. 新型城镇化, 2024(7): 59–63.
- 7 易文. 基于多尺度特征融合与 Transformer 的轻量化钢桥结构缺陷检测研究 [硕士学位论文]. 上海: 东华大学, 2025.
- 8 王云升. 智能交通系统在路桥设计中的应用与优化. 运输经理世界, 2024, (22): 77–79.
- 9 赵生军. 复杂环境条件下路桥隧道工程施工安全风险预测. 江苏航运职业技术学院学报, 2023, 22(2): 37–43.
- 10 Ding ZZ, Yang CS, Hu B, *et al.* Lightweight CNN combined with knowledge distillation for the accurate determination of black tea fermentation degree. *Food Research International*, 2024, 194: 114929. [doi: [10.1016/j.foodres.2024.114929](https://doi.org/10.1016/j.foodres.2024.114929)]
- 11 Mungoli N. Adaptive feature fusion: Enhancing generalization in deep learning models. arXiv:2304.03290, 2023.
- 12 陈佳慧, 高树辉, 袁红敏, 等. 单模态至多模态生物特征身份识别研究现状与应用展望. 西安交通大学学报 (医学版), 2025, 46(4): 551–561.
- 13 戴鹏飞, 邹京汕, 杨柳, 等. 基于 YOLOv5 的无人机桥面病害检测算法研究. 现代电子技术, 2025, 48(16): 7–12.
- 14 于子朔. 基于无人机影像处理的建筑外墙维护检查系统. 信息记录材料, 2025, 26(8): 232–234, 248.
- 15 王蕊蕊, 王慧琴, 王可, 等. 基于改进 YOLOv8 的无人机视角下高速公路异常目标检测方法. 计算机工程与应用, 2025, 61(11): 105–118.
- 16 王楠, 王淇, 欧阳丹彤. 基于知识蒸馏与动态调整机制的多模态情感分析模型. 计算机学报, 2025, 48(8): 1923–1942.
- 17 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述. 计算机工程与应用, 2021, 57(8): 10–25.
- 18 李彬, 李生林. 改进 YOLOv11n 的无人机小目标检测算法. 计算机工程与应用, 2025, 61(7): 96–104.
- 19 Zhang CB, Chang CC, Jamshidi M. Bridge damage detection using a single-stage detector and field inspection images. arXiv:1812.10590, 2018.
- 20 Rathee M, Bačić B, Dobrajeh M. Automated road defect and anomaly detection for traffic safety: A systematic review. *Sensors*, 2023, 23(12): 5656. [doi: [10.3390/s23125656](https://doi.org/10.3390/s23125656)]
- 21 秦乐, 谭泽富, 雷国平, 等. EMF-YOLO: 轻量化多尺度特征提取路面缺陷检测算法. 计算机工程与应用, 2025, 61(14): 101–111.
- 22 张帅, 王波涛, 涂嘉怡, 等. SCE-YOLO: 改进 YOLOv8 的轻量级无人机视觉检测算法. 计算机工程与应用, 2025, 61(13): 100–112.
- 23 Cui X, Zhu M, Qin YL, *et al.* Multi-level optimal transport for universal cross-tokenizer knowledge distillation on language models. *Proceedings of the 39th AAAI Conference on Artificial Intelligence*. Philadelphia: AAAI Press, 2025. 23724–23732.

(校对责编: 张重毅)