

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Yang Xiuhong, Zhang Yuxuan, Shi Zhenghao, Liu Hantao, Karakus Oktay, Dai Jiakun. Causal-inspired deconfounding for pear rust detection in UAV imagery[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260468. (杨秀红, 张聿璇, 石争浩, 刘晗涛, Oktay Karakus, 戴佳钡. 因果启发去混杂的无人机梨锈病检测[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.260468.) [DOI: 10.11834/jig.260468]

因果启发去混杂的无人机梨锈病检测

杨秀红^{1,2*}, 张聿璇¹, 石争浩^{1,2}, 刘晗涛³, Oktay Karakus³, 戴佳钡⁴

1. 西安理工大学计算机科学与工程学院, 陕西西安 710048; 2. 陕西省网络计算与安全技术重点实验室, 陕西西安 710048; 3. 英国卡迪夫大学计算机科学与信息学院, 英国威尔士 CF24 3AA; 4. 陕西省生物农业研究所, 陕西西安 710043

摘要: 目的 梨锈病斑检测是梨树病害监测和精准防控的重要环节。在无人机自然采集条件下, 锈病叶片常伴随阴影、遮挡、局部曝光、亮点等混杂信息, 传统目标检测模型在学习病斑特征时易将这些非病害区域与锈病标签之间的虚假统计关联作为判别依据, 导致大量误检。针对上述问题, 本文提出一种因果启发去混杂的无人机梨锈病检测方法。方法 其核心是提出因果混杂判别筛选(Causal Confounder Discriminative Filtering, CCDF)模块, 旨在显式建模预测查询中的混杂风险, 并对检测置信度进行后门调整和稳定校准, 无需依赖像素级混杂掩码或域标签。首先, 构建查询级目标代理和混杂代理。从实时检测Transformer(Real-Time Detection Transformer, RT-DETR)解码器分类头输入端提取256维查询特征, 将与真实标注(Ground Truth, GT)有效匹配的查询聚合为目标原型集合, 作为真实病斑证据在查询特征空间中的目标代理(正例), 将高置信度、低交并比(Intersection over Union, IoU)的误检查询聚合为混杂原型集合, 作为混杂因素在查询特征空间中的可观测混杂代理(负例); 然后, 设计轻量多层感知机(Multilayer Perceptron, MLP)作为判别器, 充当后门路径的显式阻断器。利用正负查询对学习真实病斑与混杂模式间的非线性判别边界, 为每个预测查询输出独立的混杂概率, 为后续可微分的连续化后门调整作准备; 最后, 采用连续且单调的混杂感知重评分函数, 在不改变边界框位置和类别预测的前提下, 对高风险预测进行差异化分数抑制, 使预测置信度逐步剥离混杂因素的虚假影响, 趋近于仅由真实病斑证据驱动。结果 实验表明, 在GYMNSA测试集上, 本文方法将精确率(Precision, P)提升至0.781, IoU=0.50时平均精度均值(mean Average Precision at IoU=0.50, mAP50)提升至0.723, 假阳性(False Positive, FP)数量降至348, 召回率(Recall, R)为0.641。不同抑制强度的消融实验进一步表明, 随着抑制增强, 精确率持续提高, 呈现出明确的Precision-Recall (P-R)权衡关系。结论 CCDF模块能够针对性地压低由阴影、高光、叶脉和相似纹理等混杂因素引起的误检分数, 在不重新训练检测器、不改变框回归结果的前提下提升检测结果的可信度, 尤其适用于误检代价较高的梨锈病早期筛查与无人机巡检场景。

关键词: 梨锈病检测; 因果推理; 去混杂; 后门调整; 无人机影像

Causal-inspired deconfounding for pear rust detection in UAV imagery

Yang Xiuhong^{1,2*}, Zhang Yuxuan¹, Shi Zhenghao^{1,2}, Liu Hantao³, Karakus Oktay³, Dai Jiakun⁴

1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; 2. Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China; 3. School of Computer Science & Informatics, Cardiff University, CF24 3AA, Wales United Kingdom; 4. Shaanxi Institute of Biological Agriculture, Xi'an 710043, China

收稿日期: 2026-07-31; 修回日期: 2026-09-02

* 通信作者: 杨秀红 yangxiuhong@xaut.edu.cn

基金项目: 陕西省自然科学基金项目(2023-JC-YB-586); 陕西省重点研发计划项目(2024NC-YBXM-216)

Supported by: Shaanxi Provincial Natural Science Project (2023-JC-YB-586); Shaanxi Provincial Key Development Program (2024NC-YBXM-216)

Key words: pear rust disease detection; causal inference; deconfounding; backdoor adjustment; UAV imagery

论文引用格式: DOI:10.11834/jig.260648

Abstract Objective Pear rust is a common foliar disease that can reduce photosynthetic capacity, accelerate premature leaf fall, and ultimately affect fruit quality and yield. Unmanned aerial vehicle (UAV) imagery provides an efficient means of surveying orchard canopies over large areas, but reliable automatic detection remains difficult under natural acquisition conditions. Rust lesions are often small and locally similar to background patterns, while infected leaves frequently co-occur with shadows, highlights, white veins, occlusion boundaries, and branch textures. These factors may become spuriously correlated with disease labels during detector training and can produce high-confidence false positives. Existing agricultural detectors mainly improve feature extraction, multi-scale fusion, attention, or detection heads, but seldom model the confounding risk of individual prediction queries. This study therefore proposes a causal-inspired deconfounding method for pear rust detection in UAV imagery, aiming to reduce the influence of background-associated spurious cues on detection confidence while preserving the original localization capability of the detector. **Method** Real-Time Detection Transformer (RT-DETR) is adopted as the baseline detector because its Transformer decoder represents each candidate prediction with an explicit object query that is naturally aligned with the final classification and box-regression outputs. The proposed framework introduces a lightweight causal confounder discriminative filtering module, referred to as CCDF, after the baseline detector. First, query-level target and confounding proxies are constructed using a frozen RT-DETR. For each image, the 256-dimensional hidden query representation output by the last Transformer decoder layer and fed into the final classification prediction head is extracted before classification logits are produced. This level is selected because it has undergone the complete decoder updates and corresponds directly to the final candidate predictions. Earlier decoder-

layer representations could be used, whereas encoder or backbone features would require additional candidate-region alignment and pooling. Changing the extraction layer would require reconstructing the proxy sets and retraining the discriminator; cross-layer performance differences are not quantified here. Candidate queries are divided according to their original confidence and maximum intersection-over-union (IoU) with ground-truth boxes. Queries with confidence not lower than 0.25 and IoU not lower than 0.50 are used as target-evidence proxies, whereas predictions with confidence not lower than 0.50 and maximum IoU lower than 0.10 are used as confounding proxies. The interval between the two IoU criteria is excluded to reduce label ambiguity caused by localization errors around small lesions. The thresholds are fixed to balance proxy purity and sample availability; adaptive thresholds remain feasible but require validation. A lightweight 256-128-2 multilayer perceptron (MLP) is then trained on the two proxy sets to learn a nonlinear boundary between target-evidence and confounding patterns and to output a query-level confounding probability. Finally, the estimated probability is converted into a continuous risk weight and used in a monotonic multiplicative re-scoring function. The mechanism suppresses candidates with higher estimated confounding risk more strongly while relatively preserving lower-risk candidates. It changes only confidence values and their ranking, without modifying the predicted category, bounding-box coordinates, or RT-DETR parameters. Three suppression configurations are evaluated to characterize the precision-recall trade-off, with strategy A used as the main configuration for stronger false-positive control. To avoid information leakage, proxy mining and discriminator training use only the training split; the validation split is used for discriminator evaluation and re-scoring configuration selection, while the test set is reserved for final performance evaluation and visualization. Ground-truth annotations from the test set are used only after inference to com-

pute evaluation metrics and are never provided to CCDF during prediction. **Results** Experiments are conducted on the public GYMNSA pear rust detection dataset using a unified test set containing 140 UAV images and 2,135 annotated rust instances. The baseline RT-DETR is trained for 100 epochs under the same evaluation protocol. On the held-out validation data, the MLP confounding discriminator achieves an area under the receiver operating characteristic curve (AUROC) of 0.813, an area under the precision-recall curve (AUPRC) of 0.971, and an average score-separation value of 0.363, indicating that target-evidence and confounding queries are distinguishable in the selected query representation space and that the discriminator output can serve as a meaningful risk variable for subsequent re-scoring. On the GYMNSA test set, the RT-DETR baseline obtains a Precision of 0.755, a Recall of 0.664, a mean average precision at IoU=0.50 (mAP@50) of 0.722, and 460 false-positive detections. With the proposed CCDF and the main suppression strategy, Precision increases to 0.781 and mAP@50 reaches 0.723, while the number of false positives decreases to 348. This corresponds to 112 fewer false positives, or a reduction of 24.35%, whereas Recall decreases to 0.641. This is consistent with the intended role of CCDF: it does not generate boxes or alter regression, but preferentially down-weights predictions resembling the confounding proxy distribution. Ablation experiments with three suppression intensities show a systematic precision-recall trade-off. Stronger suppression produces higher Precision and fewer false positives, whereas weaker suppression retains more difficult low-confidence lesions. Candidate-level score analysis further shows that score reduction increases with estimated confounding probability. For high-confidence confounding candidates, the median score reduction is 0.172; 66.73% of such candidates are moved below the decision threshold of 0.55, compared with 23.54% of the corresponding target candidates. Qualitative analysis also shows that many baseline errors occur around reflections, leaf veins, shadows, and occlusion boundaries, and that

CCDF effectively removes a substantial portion of these high-confidence predictions from the final detection set. **Conclusions** The proposed method introduces lightweight causal-inspired deconfounding into an end-to-end Transformer detector without pixel-level confounder masks, domain labels, detector retraining, or changes to bounding-box regression. By treating confident background false positives as observable query-level proxies for confounding effects and well-matched lesion queries as proxies for target evidence, CCDF module converts otherwise difficult-to-observe environmental interference into a supervised risk estimation problem. The learned confounding probability then supports continuous confidence calibration that preferentially attenuates predictions dominated by spurious background patterns. Results demonstrate that this strategy substantially reduces false alarms while maintaining nearly unchanged mAP@50, thereby improving the reliability of UAV-based pear rust screening in scenarios where false positives are costly. The decrease in Recall also indicates that aggressive deconfounding may suppress a subset of difficult true lesions, so CCDF should be understood as an explicit precision-recall control mechanism rather than a universal performance enhancer. Future work will investigate adaptive proxy-selection thresholds, cross-layer feature representations, and end-to-end integration of deconfounding constraints to improve difficult-lesion recall and robustness across orchards, illumination conditions, flight configurations, and disease stages.

0 引言

梨锈病是梨树生产中常见的病害之一,主要危害叶片,也可影响叶柄和幼果。该病害发生后,叶片表面通常形成黄色至橙黄色病斑,病斑扩展会削弱叶片光合作用能力,影响树体养分积累和果实正常发育,严重时可能造成早期落叶、果实品质下降和产量损失。因此,梨锈病防控的关键在于发病初期的及时发现、准确定位和提前干预。梨锈病通常在梨树枝叶的生长期,成龄梨树树冠高度通常可达数米,叶片逐渐展开并形成较复杂的冠层结构,病斑可能分

布在不同高度、不同朝向和不同遮挡程度的叶片上。传统人工巡田依赖人工逐行、逐株观察,不仅劳动强度大、效率低,而且容易受到树体高度、枝叶遮挡、观察角度和主观经验影响,难以及时覆盖大面积果园。因此,利用无人机获取果园冠层图像,并结合深度学习实现梨锈病自动检测,对于提高病害巡查效率、降低人工成本和支撑精准植保具有重要意义。

随着低空遥感平台和深度学习检测模型的发展,无人机已成为农业病虫害监测的重要数据获取手段。与人工巡田相比,无人机能够在较短时间内获取大面积农田或果园冠层图像,并为病害早期识别、空间定位和精准防控提供连续、客观的数据支持。近年来,相关研究主要集中于作物病害识别、病斑定位和病虫害区域化监测等方向。CropdocNet模型(Shi等,2022)利用无人机高光谱影像提取马铃薯晚疫病的光谱-空间层级特征,缓解了田间病害早期识别困难和人工调查效率低的问题。Kurihara等人(2022)利用无人机高光谱成像与随机森林算法识别油棕基腐病植株,实现了健康、早期感染、晚期感染和死亡植株的分类。Xiao等人(2022)基于无人机多光谱影像构建苹果火疫病检测模型,提高了果园尺度病害植株快速识别能力。Prasad等人(2022)提出结合生成对抗网络和两阶段机器学习的作物病害检测框架,用于缓解低分辨率无人机图像与高质量病害图像之间的数据差异问题。Amarasingam等人(2022)利用无人机RGB图像对甘蔗白叶病进行检测,并比较YOLOv5、YOLOv7、DETR和Faster R-CNN等目标检测模型,验证了通用检测网络在无人机作物病害定位中的适用性。Bao等人(2023)针对无人机茶叶叶枯病图像中病斑尺度小、背景复杂和光照变化明显等问题,提出DDMA-YOLO模型,通过改进特征表达和检测结构提升了茶叶叶枯病检测精度。MCD-Yolov5(Li等,2023)结合多层特征融合、注意力机制和检测Transformer模块,实现了无人机图像中多类作物病虫害的实时识别。Liu等人(2024)利用无人机光学与热红外图像提取光谱、纹理和温度特征,建立小麦白粉病严重度估计模型,提高了病害定量监测能力。Zhang等人(2024)针对复杂田间环境下白菜黄萎病轻度感染样本不足和病害特征不明显的问题,提出DSConv-GAN模型,将CycleGAN样本生成与改进YOLOv8检测网络结合,提高了无人机图像中轻度病株检测效果。YOLOv5_DWMix

(Zhao等,2025)将深度可分离卷积、MixConv、注意力机制和改进损失函数引入YOLOv5,实现了无人机图像中水稻冠层病虫害目标识别。冷佳旭等人(2023)系统总结了无人机视角目标检测中的尺度、空间、类别、语义和优化目标不均衡问题;潘晓英等人(2023)从数据增强、多尺度特征融合、上下文语义和注意力机制等方面梳理了小目标检测方法;冯琪涵等人(2025)进一步通过前景细化与多维归纳偏置自注意力提升了复杂背景下的无人机小目标检测性能。

上述研究表明,无人机与深度学习结合能够有效提升农业病虫害监测效率,但复杂树冠环境下的梨锈病检测仍面临明显挑战。无人机采集的树冠叶片图像通常包含复杂光照、枝叶重叠、叶片遮挡、局部曝光、叶面亮点和背景阴影等复杂干扰因素。这些区域往往与梨锈病叶片共同出现,容易使目标检测模型在训练过程中学习到非病害区域与病害标签之间的虚假关联关系。例如,局部强光产生的亮点、叶片边缘形成的阴影、被遮挡叶片上的纹理以及白色线状叶脉,可能在检测结果中形成高置信度预测框,但其与真实标注框的匹配程度较低,最终导致误检。现有无人机农业病虫害检测方法多从网络结构、特征融合、注意力机制、数据增强和检测头改进等方面提升模型表达能力,对模型为何将阴影、遮挡、曝光和亮点等非目标因素误认为病害目标的机制分析不足,也缺少面向误检区域的显式去混杂约束。因此,梨锈病目标检测不仅需要提高模型对小尺度病斑的表达能力,还需要抑制模型对混杂信息的依赖,使检测器更加关注梨锈病病斑本身的稳定特征。

针对无人机梨锈病检测中复杂干扰导致的误检问题,本文提出一种基于因果启发去混杂的无人机梨锈病检测方法。该方法不依赖像素级混杂掩码或域标签,直接从基线检测器的预测结果中提取正确匹配查询和高置信度误检查询,分别构建目标证据和混杂因素在查询特征空间中的可观测代理。在统一的256维查询空间内,设计轻量MLP学习真实病斑与混杂模式间的非线性判别边界,并将其输出解释为预测框的混杂风险。最后,设计一个连续非线性重评分函数,基于预测框的混杂概率对原始类别置信度进行连续化后门调整和稳定校准,对高风险误检施加显著抑制,对低风险真实病斑预测予以相

对保留,从而有效抑制阴影、高光、叶脉及类病斑纹理等混杂因素所诱发的误检响应。

本方法的核心在于设计了一个因果混杂判别筛选模块(Causal Confounder Discriminative Filtering, CCDF),其主要贡献如下:

1)提出查询级混杂代理构建策略:本文不依赖像素级混杂掩码标注,而是利用RT-DETR解码器输出的最终层查询特征,将高置信度、低IoU的误检查询定义为混杂代理(负例),将与GT匹配的查询定义为目标证据代理(正例),在同一特征空间中进行统一建模。把分散于检测结果中的典型误检转化为可监督的查询级混杂代理,有效解决现有因果推理方法在无人机梨锈病检测任务中缺乏可监督误检样本、难以直接迁移的关键难题。

2)设计轻量MLP混杂判别器:对后门路径进行显式阻断。利用正负查询对学习离线误检证据,输出可解释的逐查询混杂概率。该模块在查询特征空间刻画真实病斑与各类混杂干扰之间的非线性判别边界,将离散的误检证据转化为连续的可微分风险变量,为实现连续化后门调整作准备。

3)稳定混杂感知重评分机制:设计连续单调的分数抑制策略,对每个候选预测的置信度施加一个负向依赖于其混杂概率的后验权重,使预测置信度逐步剥离混杂因素的虚假影响,趋近于仅由真实病斑证据驱动;进一步设置A、B、C三种抑制强度,其中策略A在Recall(R)基本保持的同时显著提升Precision(P)并降低虚警,其余策略可灵活适配不同P-R需求的实际场景。

1 相关工作

1.1 农业目标检测的相关研究

近年来,深度学习目标检测技术已广泛应用于作物病害识别、害虫监测和农田智能巡检等领域。与通用自然图像相比,农业目标通常尺度小、类间相似度高、分布密集且背景复杂,并易受光照变化、枝叶遮挡、拍摄距离及成像质量等因素干扰。

一方面,大量研究基于YOLO(You Only Look Once)系列模型(如YOLOv5、YOLOv8)进行改进,主要围绕小目标特征增强、多尺度信息融合、注意力机制和轻量化部署等方向展开。例如,Improved Pest-YOLO(Tang等,2023)引入高效通道注意力和跨阶

段特征融合,强化细小害虫的通道响应;DCF-Yolov8(Zhang等,2023)借鉴DenseBlock思想改进C2f模块,增强低层边缘与纹理细节的复用;Xu等人(2023)提出的SSV2-YOLO轻量化骨干网络以适应端侧部署;Yang等人(2023)提出的Maize-YOLO优化多尺度检测结构;FSM-YOLO(Yan和Yang,2024)设计自适应特征增强和空间上下文感知模块;Insect-YOLO(Wang等,2025)和Pest-YOLOv8(Su等,2025)分别嵌入CBAM和三重注意力机制,结合改进的损失函数提升低质量图像下的定位精度。然而,这类方法本质上仍依赖卷积操作的局部感受野和预设锚框,对全局上下文和远距离依赖关系的建模能力有限,因此在背景高度杂乱、目标被严重遮挡或光照剧烈变化时,容易将背景噪声误判为病害或害虫,产生较多假阳性。同时,大多数YOLO改进模型侧重于特征提取层的优化,而在预测输出层缺乏对每个候选框置信度的混杂风险评估,难以主动抑制由错误统计关联引发的误检,导致在复杂农田场景下的可靠性仍有待提升。

另一方面,Transformer架构因其全局上下文建模和长程依赖捕获能力,也被引入农业目标检测领域。该类方法能够更好地理解目标与周围场景的关系,且摒弃了锚框和非极大值抑制等人工设计组件,将检测重构为集合预测问题。RT-DETR(Real-Time DEtection TRansformer)(Zhao等,2024)通过解耦多尺度特征交互与解码器迭代优化,在保持全局建模能力的同时显著提升推理速度,成为首个真正适用于实时场景的端到端Transformer检测器。Xu等人(2024)将Cascade R-CNN与Swin Transformer结合,利用局部-全局特征提取和自适应多尺度增强,有效应对光照、反光和尺度差异;Wang等人(2024)提出端到端PD-TR模型,基于Transformer查询机制直接输出病害类别与位置,减少对人工候选框的依赖,并借助自注意力权重提供一定可解释性。总体而言,这类方法更注重场景关系理解和端到端学习,能够缓解局部特征歧义,兼顾了Transformer在复杂场景下的鲁棒表征能力和工业部署所要求的效率,为农业目标检测中高精度与实时性的平衡提供了可行方案。尽管这些研究工作通过注意力融合、多尺度特征增强等手段改善了漏检问题,但现有工作大多侧重于提升特征表达和检测精度,对于模型产生误检的内在原因关注不足,缺少从预测查询层面对真

实病害证据与混杂误检模式的显式区分,也缺乏针对每个候选框的混杂风险估计与差异化抑制机制。

1.2 因果推理在目标检测中的相关研究

为削弱深度学习模型对非因果相关特征的依赖,因果推理近年来逐渐被引入图像识别、细粒度分类、目标检测和跨域泛化等视觉任务。Wang 等人(2021)提出因果注意力模块,通过无监督方式标注混杂因素并进行因果干预,使视觉识别模型在分布外场景下更多关注稳定的目标因果特征。Rao 等人(2021)提出反事实注意力学习方法,利用反事实干预衡量注意力区域对预测结果的因果贡献,改善了细粒度分类和重识别任务中注意力不稳定的问题。Huang 等人(2021)针对目标检测中上下文信息容易导致虚假相关的问题,提出面向目标检测的因果干预框架,通过结构因果模型分析目标、上下文和预测之间的关系,削弱上下文混杂因素对分类结果的影响。Liu 等人(2022)提出因果上下文去偏模块,通过后门调整和反事实推理消除上下文共现关系造成的预测偏置,使多标签视觉识别模型获得更接近目标直接因果效应的预测结果。Li 等人(2023)从条件因果视角分析少样本目标检测中的知识蒸馏过程,提出 Disentangle and Remerge 方法,通过后门调整缓解有限样本条件下学生检测器学习错误语义关联的问题。Xu 等人(2023)面向未知域目标检测提出多视角对抗判别器,从因果机制出发挖掘并移除域不变特征中残留的非因果因素,提升了检测模型在未见过场景中的泛化性能。Unbiased Faster R-CNN (Liu 等, 2024)将单源域泛化目标检测建模为受场景混杂和目标属性混杂共同影响的因果问题,并通过因果注意力学习和因果原型学习降低非因果偏置特征对检测结果的影响。

因果推理方法已被证实能够为视觉模型引入去偏能力和泛化约束。然而,现有工作大多聚焦于通用图像识别、细粒度分类或域泛化目标检测等相对受控的任务,且通常需要显式构建结构因果模型 (Structural Causal Model, SCM) 或执行反事实推理,计算开销较大。更为关键的是,在无人机采集的梨锈病图像中,阴影、遮挡、曝光亮点以及类病斑纹理等混杂因素在图像层面上难以被独立分割或精确标注,因而那些需要像素级混杂掩码或域标签作为输入的既有因果方法难以直接迁移至本任务,其应用及部署存在巨大障碍。

2 基线检测与混杂问题定义

2.1 RT-DETR 基线框架

本文以 RT-DETR (Zhao 等, 2024) 作为基础检测框架,该框架由 HGNetV2 骨干网络、混合编码器和 Transformer 解码器组成。其中,骨干网络提取多尺度图像特征;混合编码器完成尺度内特征交互与跨尺度融合;解码器利用目标查询输出类别置信度和边界框。与基于锚框的检测器相比,RT-DETR 采用端到端集合预测方式,可避免额外的非极大值抑制 (Non-Maximum Suppression, NMS),并能够在实时性与检测精度之间取得良好的平衡。

在 GYMNSA 测试集上的可视化结果表明,漏检主要集中于尺度较小、对比度较低或被叶片遮挡的病斑;误检则多出现在强反光、白色叶脉、阴影边缘、枝条节点及相似叶片纹理区域。本文重点处理后两者,即在不改变基线框回归的前提下识别并抑制高混杂风险预测。

2.2 混杂因素定义与误检抑制策略

设 X 为输入图像, Z 为病斑本身的视觉证据 (包括:病斑颜色、局部纹理和形态证据), C 为混杂因素集合 (包括:阴影、高光、叶脉、遮挡和枝条纹理等), Q 为 RT-DETR 解码器输出的查询特征, Y 为模型输出。他们之间的因果关系可用图 1 所示的 SCM 表示:

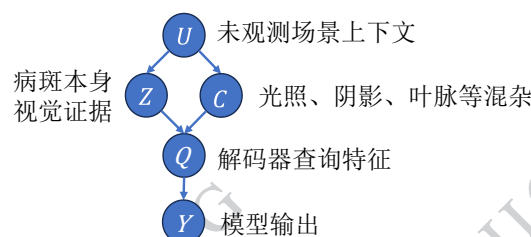


图1 梨锈病检测中的结构因果模型 SCM

Fig. 1 Structural Causal Model (SCM) in pear rust disease detection

自然采集过程中, Z 与 C 可能同时出现在病叶附近, $Z \leftarrow U \rightarrow C$ 构成 Z 与 C 之间的关联, 导致模型在有限样本上学习到 $C \rightarrow Q \rightarrow Y$ 虚假统计关联, 而 $Z \rightarrow Q \rightarrow Y$ 表示因果路径。当测试场景中的光照、叶片姿态或背景组成发生变化时, 虚假关联会诱发高置信度误检。

根据后门准则(Pearl, 2009),若要估计 Z 对 Y 的因果效应 $P(Y|do(Z))$,需要对混杂因素 C 进行阻断或调整。其后门调整的形式化公式为:

$$P(Y|do(Z)) = \sum_{c \in C} P(Y|Z, C=c) \cdot P(C=c) \quad (1)$$

C 在无人机采集的梨锈病图像中难以直接观测和显式建模。因此,本文提出一种**可观测代理调整策略**:利用基线 RT-DETR 在训练集上产生的高置信度误检作为 C 在查询特征空间中的投影 $\tilde{C}$,而以与真实框匹配的查询作为目标证据 Z 的代表。通过对 $\tilde{C}$ 进行后门调整的近似实现,削弱 $C \rightarrow Q \rightarrow Y$ 路径对预测 Y 的干扰。

3 基于因果启发去混杂的无人机梨锈病检测方法

本方法在 RT-DETR 端到端检测框架的基础上,通过从 Transformer Decoder 最终层查询特征中引出一条轻量混杂判别支路,构建 CCDF 模块,实现对候选预测中混杂风险的显式估计与原始置信度的差异化校准。RT-DETR 主干网络仍负责多尺度特征提取、特征融合、目标查询解码以及边界框与类别置信度预测,CCDF 则以与各候选框一一对应的 256 维

decoder query 为输入,估计其属于混杂误检模式的概率,并将该概率作为风险变量对原始类别置信度进行连续重评分。该过程不改变 RT-DETR 的边界框回归结果、类别索引及主体网络参数,仅调整候选框的置信度及其排序,从而降低阴影、高光、叶脉、枝条和类病斑纹理等非病害因素对最终检测结果的干扰,总体流程如图 2 所示。

核心模块 CCDF 由查询级混杂代理构建、MLP 混杂判别和稳定混杂感知重评分三部分组成。首先,在冻结 RT-DETR 参数的条件下,根据候选框原始置信度及其与真实标注框的最大 IoU,从最终层 decoder query 中筛选与 GT 有效匹配的目标证据查询,以及高置信度、低 IoU 的混杂误检查询,分别构建可观测目标代理和混杂代理。其次,利用两类带标签查询训练结构为 256-128-2 的轻量 MLP 判别器,使其在统一查询特征空间中学习真实病斑证据与混杂误检模式之间的非线性判别边界,并在推理阶段为每个候选查询输出独立的混杂概率。最后,将混杂概率映射为连续风险权重,并通过乘性稳定重评分函数对相应候选框的原始置信度进行校准,使混杂风险较高的预测获得更强抑制,而低风险目标预测仅受到较小影响,最终实现对高置信度误检的自适应筛选。

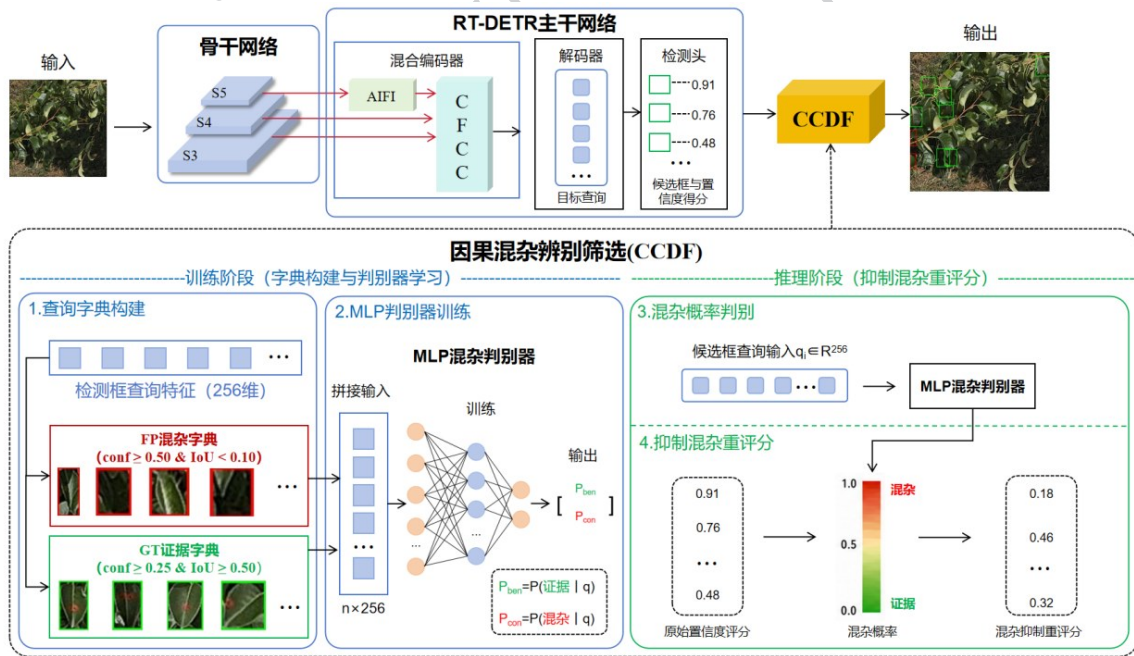


图2 CCDF-RT-DETR 无人机影像梨锈病检测总体流程

Fig. 2 Overall pipeline of CCDF-RT-DETR for pear rust detection in UAV imagery

3.1 查询级混杂代理构建

字典构建以冻结的 RT-DETR 基线为特征提取器。对输入图像完成前向推理后,本文选取 Transformer Decoder 最后一个解码层输出的隐藏查询表示作为 CCDF 特征,对应最终分类预测头 dec_score_head[-1] 的输入,而非分类头输出的类别 logits。选择该层是因为其已经历完整的 decoder 迭代更新,并与最终候选预测保持一一对应。设一幅图像包含 N 个候选查询,第 i 个查询特征为 q_i ,其维度为 256,并与候选框 b_i 和原始类别置信度 s_i 保持索引一一对应。查询张量表示为:

$$Q = [q_1, q_2, \dots, q_N]^T \in \mathbb{R}^{N \times 256} \quad (2)$$

对候选框 b_i , 计算其与当前图像全部 GT 框的最大交并比,以衡量候选预测与真实病斑的空间匹配程度:

$$\text{IoU}_i^{\max} = \max_j \text{IoU}(b_i, g_j), g_j \text{ 为第 } j \text{ 个 GT 框} \quad (3)$$

结合原始置信度与最大 IoU 构建两类查询集合。GT 目标证据查询要求预测分数不低于 0.25 且与至少一个 GT 框的 IoU 不低于 0.50; 混杂查询要求预测分数不低于 0.50 且与任意 GT 框的最大 IoU 低于 0.10。相应筛选规则为:

$$Q_{\text{GT}} = \{q_i | s_i \geq 0.25, \text{IoU}_i^{\max} \geq 0.50\} \quad (4)$$

$$Q_{\text{C}} = \{q_i | s_i \geq 0.50, \text{IoU}_i^{\max} < 0.10\} \quad (5)$$

式(4)中的目标证据查询主要表示与真实病斑有效匹配的目标级表征;式(5)中的混杂查询具有较高预测置信度,但与任何 GT 框几乎不重叠,主要覆盖阴影、高光、白色叶脉、枝条节点、遮挡边缘和类病斑纹理等纯背景误检。采用 $\text{IoU} < 0.10$ 的严格条件能够尽量排除仅由小目标框偏移造成的定位困难样本,降低混杂字典的标签噪声。

上述 IoU 阈值依据目标匹配定义与降低代理样本标签噪声的原则预先设定。目标证据查询采用 $\text{IoU} \geq 0.50$, 混杂查询采用 $\text{IoU} < 0.10$, 并将 $0.10 \leq \text{IoU} < 0.50$ 作为不参与字典构建的模糊区间,以减少定位偏差造成的标签混淆。

从因果推断视角看,混杂因素 C 在高维图像空间中难以直接观测和量化,但其对解码器查询特征 q 的影响会在模型预测中留下可观测的痕迹。如式(5)所示,较高预测置信度但与真实目标几乎不重叠的误检框所对应的查询特征视为 C 在查询特征空间中的可观测代理 $\tilde{C}$ 。

可观测代理的有效性依赖于如下因果假设:若某查询特征 q_i 由混杂因素 C 引起,而与真实病斑证据 Z 无关,其对应的预测框 b_i 与任何真实目标框的空间重叠度应接近于 0,又由于 $C \rightarrow Q \rightarrow Y$ 的虚假关联,模型对其预测的置信度 s_i 仍然较高。通过式(5)的规则可筛选出由 C 引起的查询样本作为 $\tilde{C}$ 的实例。

进一步地,本文将 GT 匹配查询作为由真实病斑证据 Z 驱动的查询样本,作为 Z 在查询特征空间中的投影 $\tilde{Z}$ 。通过构建 $\tilde{Z}$ 和 $\tilde{C}$ 两组原型集合,本文将不可直接观测的目标证据和混杂风险转化为可监督学习的查询级标签,为后续基于后门调整思想的置信度校准提供数据基础。

目标证据查询与混杂查询分别堆叠为 $Q_{\text{GT}} \in \mathbb{R}^{N_{\text{GT}} \times 256}$ 和 $Q_{\text{C}} \in \mathbb{R}^{N_{\text{C}} \times 256}$ 。实验中两类查询最多各采样 20000 个,并在训练集与验证集上分别构建字典。字典同时保存图像索引、候选框坐标、原始置信度、最大 IoU 和 query 编号,便于对高分混杂样本进行回溯和人工核查。

为排除测试集信息泄漏,本文对查询字典构建、MLP 判别器训练、参数选择和最终测试进行数据隔离。RT-DETR 基线的参数更新使用训练集完成,基线确定后冻结其参数;训练集用于查询字典构建和 MLP 训练,验证集用于判别器评价及重评分参数选择,测试集仅在模型和参数固定后用于最终性能评价与可视化,测试集 GT 不参与 CCDF 推理,仅在推理完成后用于计算 P、R、mAP50 和 FP 等评价指标。

3.2 MLP 混杂判别器——混杂路径的阻断器

将目标证据查询标记为 $y_i = 0$, 将混杂查询标记为 $y_i = 1$, 构成监督数据集 $D = \{(q_i, y_i)\}_{i=1}^M$ 。混杂判别器由两层全连接网络构成,输入为 RT-DETR 解码器输出的 256 维查询特征,首先通过线性映射将特征维度压缩至 128 维,并依次经过高斯误差线性单元 (Gaussian Error Linear Unit, GELU) 激活和 Dropout 正则化,查询向量经线性映射、GELU 激活与 Dropout 后输出两维 logits $z_i = g_\theta(q_i)$ 。其中第 1 维对应混杂类别,混杂概率定义为:

$$p_i^c = [\text{Softmax}(g_\theta(q_i))]_1 = \frac{\exp(z_{i,1})}{\exp(z_{i,0}) + \exp(z_{i,1})} \quad (6)$$

判别器使用二元交叉熵损失进行优化。设训练样本数为 M , 则分类损失为:

$$L_{cls} = \frac{1}{M} \sum_{i=1}^M [y_i \ln p_i^c + (1 - y_i) \ln (1 - p_i^c)] \quad (7)$$

训练过程中采用类别均衡采样,并在独立验证集上以 AUROC、AUPRC 和平均分数间隔评价判别能力,避免仅依赖单一阈值准确率。AUROC 衡量 GT 查询与混杂查询的整体排序可分性, AUPRC 更关注高混杂概率区域的精确识别,平均分数间隔则反映两类样本输出分布的分离程度。上述指标共同用于判断 MLP 输出能否作为稳定重评分的风险变量。训练完成后冻结判别器参数。推理时,将第 i 个候选框对应的 256 维最终层查询 q_i 输入 MLP,得到 $p_i^c \in [0, 1]$ 。

从因果推断视角看,MLP 判别器的功能定位是阻断后门路径 $C \rightarrow Q \rightarrow Y$,即通过学习 $\tilde{Z}$ 和 $\tilde{C}$ 在查询特征空间中非线性边界,估计每个候选查询 q_i 受混杂因素 C 驱动的概率 p_i^c 。由式(6)可知, p_i^c 越大,表示该查询越接近混杂字典中的高置信度误检模式,即 q_i 受 C 影响越强, $Z \rightarrow Q \rightarrow Y$ 的因果效应越弱;反之, p_i^c 越小,表示其表征更接近 GT 匹配查询, q_i 更接近真实因果效应。由于 $p_i^c \in [0, 1]$ 的连续变量,该后门调整策略不再是传统硬截断离散操作,而是实现了可微分的连续化后门调整,更加适应深度学习端到端优化的工程需求。

从额外计算开销来看,CCDF 仅在 RT-DETR 输出最终查询特征后增加一个 $256 - 128 - 2$ 的轻量两层全连接 MLP,参数量约为 0.033 M,按 300 个目标查询计算,每幅图像额外增加约 9.91M MACs (约 0.01 GMAC);在 RTX 4070 SUPER 平台上额外推理时延约为 0.2 ms/幅,处于亚毫秒量级,对 RT-DETR 整体推理速度影响较小。因此,CCDF 在实现针对性混杂抑制的同时,仅引入较低的参数量、计算量和推理时延,可视为轻量级后筛选模块。

3.3 稳定混杂感知重评分——连续后门调整的近似实现

从因果推断视角看,后门调整通过式(1)的形式对混杂因素 C 的边缘分布进行加权平均,从而消除后门路径的偏置。在本文的查询特征空间中,可将 p_i^c 视为 q_i 属于混杂分布 $P(C)$ 的后验概率,而对预测置信度的调整实质相当于从联合分布 $P(Y|Q)$ 中剥离 C 的贡献、保留 Z 的因果效应的过程。然而,直接计算 $P(Y|do(Z))$ 在高维连续空间中不可行,本文提

出如下连续后门调整的近似策略。

设 RT-DETR 输出的原始类别置信度为 s_i 。为避免直接删除候选框或采用线性相减造成分数突变,首先利用中心参数 c 和温度参数 T 将混杂概率 p_i^c 映射为连续风险权重 ω_i 为:

$$\omega_i = [\text{Sigmoid}(\frac{p_i^c - c}{T})]^\gamma \in (0, 1), \quad (8)$$

上式中, c 决定抑制开始显著生效的风险中心, T 控制过渡区间的平滑程度, γ 控制高混杂概率样本的聚焦强度。

在此基础上,将原始置信度与风险权重进行乘性稳定校准:

$$s'_i = (1 - \lambda \omega_i) s_i \quad (9)$$

式中, s'_i 表示经 CCDF 重评分后的校准置信度, $\lambda \in [0, 1]$ 控制最大抑制幅度。

从式(8)和(9)可以看出, ω_i 随 p_i^c 单调增加,混杂风险越高,校准分数 s'_i 下降越明显;低风险候选的分数变化相对较小。该后门调整策略等价于对每个候选预测的置信度施加一个负向依赖于其混杂风险的后验权重,使得最终校准分数 s'_i 趋近于 $P(Y|do(Z))$ 的近似估计,即剥离了 C 的虚假影响后,仅由 Z 驱动的预测置信度。

本文将策略 A 设为主配置: $\lambda = 0.30$ 、 $c = 0.55$ 、 $T = 0.20$ 、 $\gamma = 1.0$ 。较低的风险中心使中高风险候选较早进入抑制区间,用于强调 Precision 和 FP 控制。策略 B 采用 $\lambda = 0.30$ 、 $c = 0.70$ 、 $T = 0.20$ 、 $\gamma = 2.0$,形成中等抑制;策略 C 采用 $\lambda = 0.30$ 、 $c = 0.75$ 、 $T = 0.15$ 、 $\gamma = 1.5$,具有较高触发门槛和较弱整体抑制。三种策略共享同一 RT-DETR 与 MLP 判别器,仅调整重评分参数,用于分析误检抑制强度与 Recall 保持之间的权衡。

4 实验结果与分析

4.1 GYMNSA 数据集

本文实验采用公开的 GYMNSA 梨锈病目标检测数据集。该数据集由 Maß 等人联合德国朱利叶斯·屈恩研究所果树育种研究所 (Julius Kühn-Institut, Institute for Breeding Research on Fruit Crops) 与莱布尼茨农业工程与生物经济研究所构建,使用计算机视觉标注工具 (Computer Vision

Annotation Tool, CVAT) 进行人工标注,共标注 16251 个病斑实例。原始 RGB 影像由 DJI Phantom 4 Pro V2.0(P4P)和 DJI Matrice 300 RTK 两种无人机平台采集,其中 Matrice 300 RTK 搭载 Zenmuse P1 相机,公开数据所对应的 10 次飞行任务相对飞行高度约为 4.8~12.0 m;两套相机系统的原始影像分辨率分别为 5472×3648 像素和 8192×5460 像素,并在数据集制作阶段裁剪为 768×768 像素图像块用于病斑标注。数据集按 8:2 的比例划分训练集和测试集,其中训练集包含 531 幅图像,统一测试集包含 140 幅图像、共计 2135 个梨锈病实例。不同飞行高度、拍摄时段、光照和树冠条件使图像具有较丰富的场景变化,同时病斑尺度较小、分布较密集且局部症状差异细微,因此适合用于评价目标检测模型在复杂果园环境下的检测性能(Maß 等,2025)。

4.2 实验环境与评价指标

实验环境为 Windows 系统,深度学习框架为 Ultralytics 8.4.10、Python 3.12.3 和 PyTorch 2.5.1,使用 NVIDIA GeForce RTX 4070 SUPER 12 GB GPU。默认输入尺寸为 768,RT-DETR baseline 训练 100 epoch,batch size 为 4,关闭自动混合精度(Automatic Mixed Precision, AMP)。

评价指标包括精确率(P)、召回率(R)、mAP50 和假阳性数量(FP)。其中,P 反映预测为病斑的候选中正确预测的比例,R 反映全部真实病斑中被检出的比例,mAP50 表示 IoU 阈值为 0.50 时 PR 曲线下面积。由于本文方法 CCDF 主要通过重评分调整预测置信度排序,而非改变检测框位置或数量,因此 FP 数量的变化能够直接衡量其对误检的抑制效果。所有模型均在相同的 GYMNSA 测试集上进行评估。

表 1 实验环境与训练设置
Table 1 Experimental environment and training settings

项目	设置
框架	Ultralytics 8.4.10
Python/PyTorch	Python 3.12.3 / torch 2.5.1
GPU	NVIDIA GeForce RTX 4070 SUPER 12 GB
默认输入尺寸	768
RT-DETRbaseline 训练轮数	100
batchsize	4
混合精度训练 AMP	False

4.3 混杂判别器有效性验证

为验证 MLP 判别器能否有效区分目标证据查询与混杂查询,本文在固定查询字典上训练 MLP,并在独立验证集上评估其判别性能。结果如表 2 所示。较高的 AUROC 和 AUPRC 表明两类查询在特征空间中具有稳定可分性,平均分数间隔(即正负样本判别器输出均值的绝对差)进一步说明判别器的区分能力并非仅在单一阈值附近取得优势,而是具有一定的阈值鲁棒性。该实验验证了将 MLP 输出作为后续重评分风险变量的合理性。

4.4 对比实验

为验证本文方法在梨锈病检测任务中的性能,本文设置 Transformer 检测器、YOLO 检测器以及 RT-DETR 基线(Zhu Y A 等,2024)作为对比模型。

表 2 MLP 混杂判别器验证结果

Table 2 Evaluation results of the MLP confounder discriminator on the held-out validation set

指标	AUROC	AUPRC	平均分数间隔
结果	0.813	0.971	0.363

Transformer 检测器包括 Deformable DETR(Zhu X Z 等,2021)、DAB-DETR(Yang 等,2025)和 DINO-DETR(Zhang 等,2023);YOLO 检测器包括 YOLOv8s、YOLOv8m 和 YOLOv8n。

为保证对比公平性,表 3 中所有对比模型均采用相同的 GYMNSA 数据划分和 768×768 模型输入尺寸进行训练与测试,并在统一测试集和一致的评价协议下计算 P、R、mAP50 和 FP;YOLO 系列与 Trans-

former 系列模型均未单独使用更高输入分辨率或额外测试增强。

我们的方法将 P 由 0.755 提高至 0.781, 提升 2.6 个百分点; R 由 0.664 下降至 0.641。FP 误检框数量减少 112 个, 说明本方法在减少误检上有较大提升。该结果符合本文方法定位: CCDF 通过压低高混杂风险框减少误检, 而不生成新的检测框, 因此不会直接提高小病斑召回。

表 3 不同检测模型在 GYMNSA 测试集上的对比结果

Table 3 Comparison of different detection models on the GYMNSA test set

Models	P ↑	R ↑	mAP50 ↑	FP ↓
YOLOv8s	0.741	0.650	0.717	485
YOLOv8m	0.744	0.663	0.723	487
YOLOv8n	0.738	0.599	0.684	454
Deformable DETR	0.735	0.652	0.691	502
DAB-DETR	0.713	0.607	0.658	522
DINO-DETR	0.765	0.683	0.702	448
RT-DETR baseline	0.755	0.664	0.722	460
Ours	0.781	0.641	0.723	348

4.5 消融实验

在保持 RT-DETR、查询字典和 MLP 判别器完全相同的条件下, 仅改变稳定重评分策略中的风险阈值参数, 设计消融变体 CCDF-A、CCDF-B、CCDF-C, 其中 A、B、C 分别代表强、中、弱三种抑制强度。CCDF-A 最早触发并实施较强抑制, CCDF-B 和 CCDF-C 逐步提高风险门槛, 以保留更多低分真实病斑。结果如表 4 所示。随着抑制强度由 C、B 增强至 A, P 由 0.753 提高至 0.781, 而 R 由 0.664 下降至 0.641, 呈现明确的 P-R 权衡。策略 A 的 FP 降至 348, 误检抑制最为明显, 因此选为主实验方案; CCDF-B 和 CCDF-C 分别适用于更重视召回的平衡场景, 可作为实际部署中的备选参数配置。

4.6 可视化分析

为进一步分析 CCDF 对复杂果园场景中混杂误检的作用机制, 本文从基线误检来源、典型场景检测结果、候选框分数变化及混杂判别响应区域四个维度验证所提因果调整的有效性。误检模式统计用于揭示基线模型容易受到哪些非病害因素干扰; 检测结果对比用于展示 CCDF 对典型高置信度误检的实

表 4 CCDF 不同抑制强度消融实验结果

Table 4 Ablation results of CCDF under varying suppression intensities

Models	P ↑	R ↑	mAP50 ↑	FP ↓
RT-DETR baseline	0.755	0.664	0.722	460
CCDF-A	0.781	0.641	0.723	348
CCDF-B	0.756	0.660	0.718	454
CCDF-C	0.753	0.664	0.717	465

际抑制效果; 分数分布分析用于刻画重评分前后不同候选框的置信度变化; 混杂响应热力图则用于观察 MLP 判别结果与图像局部区域之间的对应关系。

4.6.1 误检模式分布

为分析基线模型在自然果园环境中的主要误检来源, 本文在统一测试集上对固定置信度阈值下的误检框进行统计。统计条件设置为置信度阈值 0.55、匹配 IoU 阈值 0.50, 基线 RT-DETR 在该条件下共产生 460 个误检框。根据误检区域中的主要视觉特征, 将其划分为反光干扰、叶脉白色纹理干扰、阴影与暗色斑点干扰、叶片遮挡与边缘干扰等四类。

图 3 给出了不同误检类型的典型样本。基线模型的误检主要集中于反光、阴影、遮挡和相似叶脉等干扰类型, 说明模型容易将与梨锈病病斑在颜色、局部对比度或纹理形态上相似的非病害区域误识别为目标。其中, 强光和高光区域具有较高亮度及明显的局部对比, 叶脉和枝条则呈现细长或斑块状纹理; 遮挡边缘和阴影区域会形成不规则的深浅变化。这些干扰因素与真实病斑在有限样本中形成较强统计关联, 是产生高置信度误检的重要原因, 也为后续构建混杂字典和混杂判别器提供了直接问题动机。

4.6.2 检测结果可视化

为直观比较基线模型与 CCDF 在复杂果园场景中的检测差异, 本文选取叶片遮挡且阴影覆盖较为密集的测试图像进行可视化。所有结果均采用相同的置信度阈值 0.55 和 IoU 匹配阈值 0.50。图 4 中绿色框、红色框和橙色框分别表示真阳性 (True Positive, TP)、假阳性 (False Positive, FP) 和假阴性 (False Negative, FN)。

在遮挡场景中, 重叠叶片形成的不规则边缘和局部暗色区域 (对应图 3 中的 (a) 与 (b) 场景) 同样会引发基线模型的误检。经过 CCDF 重评分后, 部分



图3 容易误检的叶片模式

Fig. 3 Distribution of false-positive patterns produced by the baseline model on the GYMNSA test set ((a) Shadow; (b) Occlusion; (c) Reflection; (d) Similar leaf veins)

非病害预测被有效过滤,本文所提模型能够在保留大部分真实病斑的同时,减少病斑周围叶脉、阴影及重复纹理产生的冗余误检。

在固定置信度阈值 0.55 下,基线模型的 FP 数量为 460,策略 A 将其降低至 348,减少 112 个,降幅为 24.35%;P 由 0.755 提高至 0.781。可视化结果与定量统计一致,说明 CCDF 对高置信度混杂误检具有较强的选择性抑制作用。与此同时,策略 A 的 R 由 0.664 下降至 0.641,表明较强抑制也可能使少量困难真实病斑降至阈值以下,因此需要结合 P-R 权衡评价其实际效果。

为区分 CCDF 与简单提高全局置信度阈值的作用,本文进一步进行了阈值对照。对基线 RT-DETR 仅提高固定置信度阈值,使 FP 数量降至与 CCDF-A 接近约 350 个时,结果显示其 Recall 降至 0.596;相比之下,CCDF-A 在 FP=348 时仍保持 R=0.641。二者虽然均能减少误检,但提高全局阈值会对全部候选实施统一的硬截断,而 CCDF 依据 MLP 输出的混杂概率对候选实施差异化重评分。因而在相近 FP 水平下,CCDF 能够相对保留更多真实样本,说明其误检抑制效果并非简单等价于提高检测置信度阈值。

4.6.3 CCDF 对候选框置信度分布的影响

为分析 CCDF 对候选框置信度分布的影响,本文统计了策略 A 重评分前后的原始置信度、混杂概

率(即 MLP 判别器输出的 softmax 归一化结果)及分数变化量。由于本文采用乘法抑制形式,校准后分数不高于原始分数。因此,分数分布分析的重点不是正样本分数是否得到提高,而是低混杂风险目标候选的分数是否得到相对保持,以及高混杂风险候选是否受到更显著的抑制。

图 5 分别给出了 GT 匹配候选与高置信度混杂候选在重评分前后的置信度分布。低混杂风险目标候选的分数分布整体变化较小,而高置信度混杂候选的分布明显向低分区域移动,且更多候选被压低至 0.55 决策阈值以下。这表明判别器输出能够对候选施加抑制,即对真实病斑候选相对保持分数,而对混杂候选实施更强压制。图 6 进一步绘制了混杂概率与分数变化量之间的关系。可以看到,随着混杂概率升高,候选框的分数下降幅度逐渐增大,分箱中位数曲线整体呈单调上升趋势。这说明稳定混杂感知重评分函数能够根据混杂风险自适应调节抑制强度,对高风险候选施加更显著的分数量压制。

进一步统计显示,高置信度混杂候选的中位变化量为 0.172;被压低至 0.55 阈值以下的混杂候选占 66.73%,而相应目标候选占 23.54%。上述结果说明 CCDF 并非对所有候选框实施相同幅度的缩放,而是根据混杂风险对高风险预测进行更强抑制。此外,少量 GT 匹配候选也被不同程度压低,表明较强的抑制策略在降低误检的同时会伴随一定的召回代价,这与 4.5 节消融实验中 P-R 权衡的结论一致。

4.6.4 引入 CCDF 前后的目标响应热力图对比

为进一步分析 CCDF 对检测结果的影响,本文选取典型测试样本,对比引入 CCDF 前后的目标响应热力图。具体而言,以候选框对应的目标证据得分 z 作为反向传播目标,计算其对 RT-DETR 空间特征图的梯度,并依据梯度贡献对各通道特征进行加权融合;将所得响应图归一化并上采样至原图尺寸,生成目标证据热力图,颜色越强表示该区域对当前候选框目标判断的贡献越大。需要指出的是,热力图反映的是检测器对梨锈病目标特征的响应,而非对混杂因素本身的直接定位。因此,该实验重点表明:CCDF 在抑制高混杂风险候选后,最终保留的预测框特征响应明确集中于病斑区域,从特征响应层面验证了 CCDF 能够有效抑制非病害干扰对检测决策的影响。

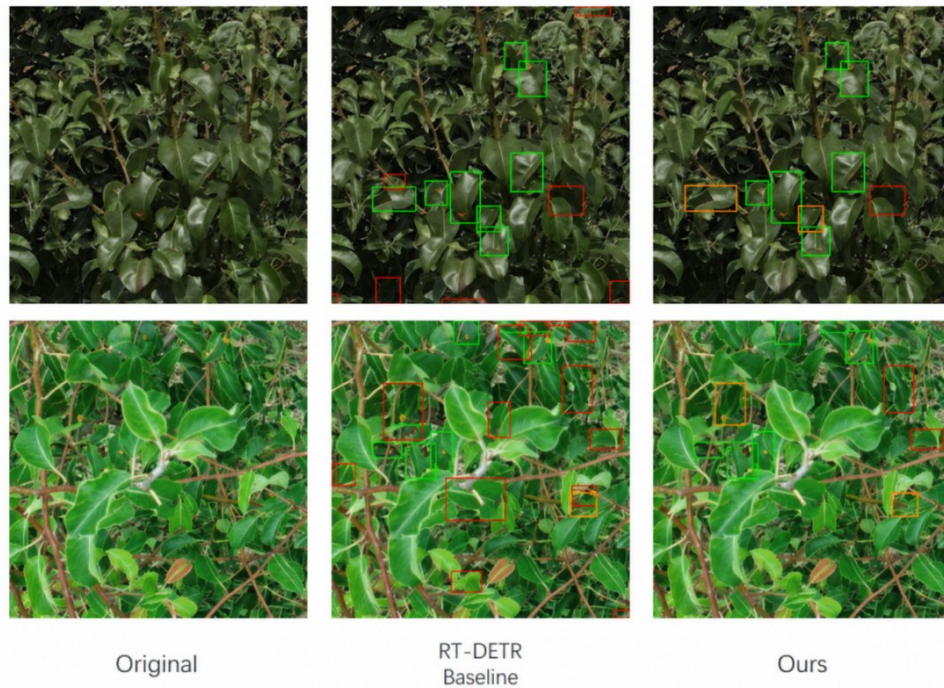


图4 检测结果可视化

Fig. 4 Visualization of detection results

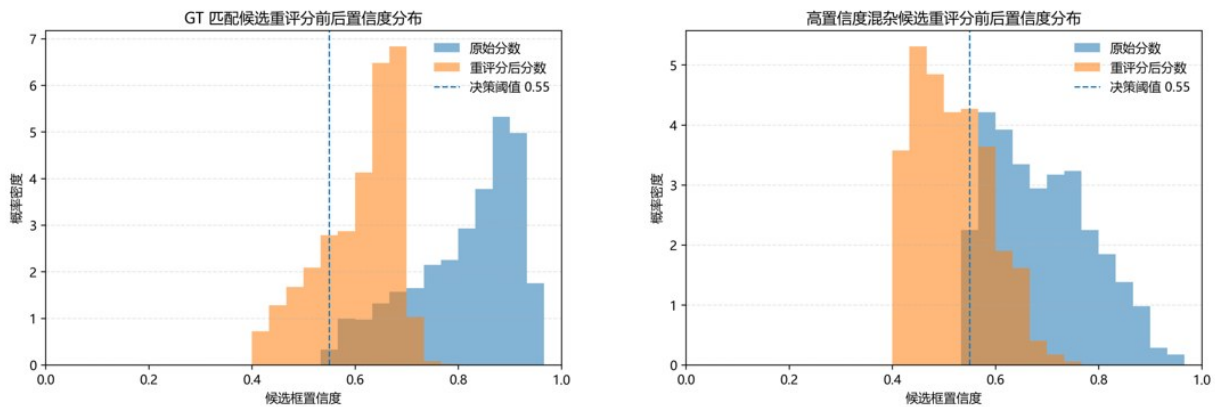


图5 CCDF对候选框置信度分布的影响

Fig. 5 Effect of CCDF on the confidence distribution of candidate boxes

从图7可以看出,未引入CCDF时,部分候选框的响应区域扩散至叶脉、叶片边缘等非病害区域;引入CCDF后,高混杂概率候选框的预测分数被显著压低,最终保留的预测响应与真实病斑区域具有更高的一致性,高响应区域集中于病斑本体,非病害背景区域的响应强度明显减弱。

对于同一候选框,由于CCDF不修改边界框坐标和主干特征,其空间响应结构通常不发生根本性变化;图中响应强度或集中程度的变化主要源于混

杂感知分数校准以及最终候选框筛选结果的差异。结合前述FP数量和置信度分布分析,可以认为CCDF能够在保留主要病斑响应的同时,有效抑制由背景混杂模式支撑的高置信度误检,从而提升检测结果的可信度。

5 结论

本文针对无人机梨锈病检测中阴影、高光、叶
© 中国图象图形学报版权所有

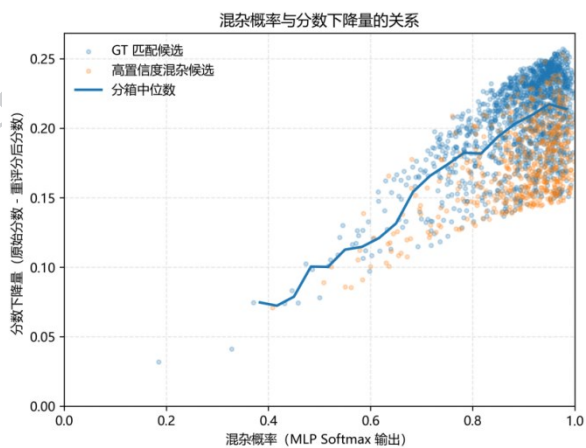


图6 混杂概率与分数下降的关系

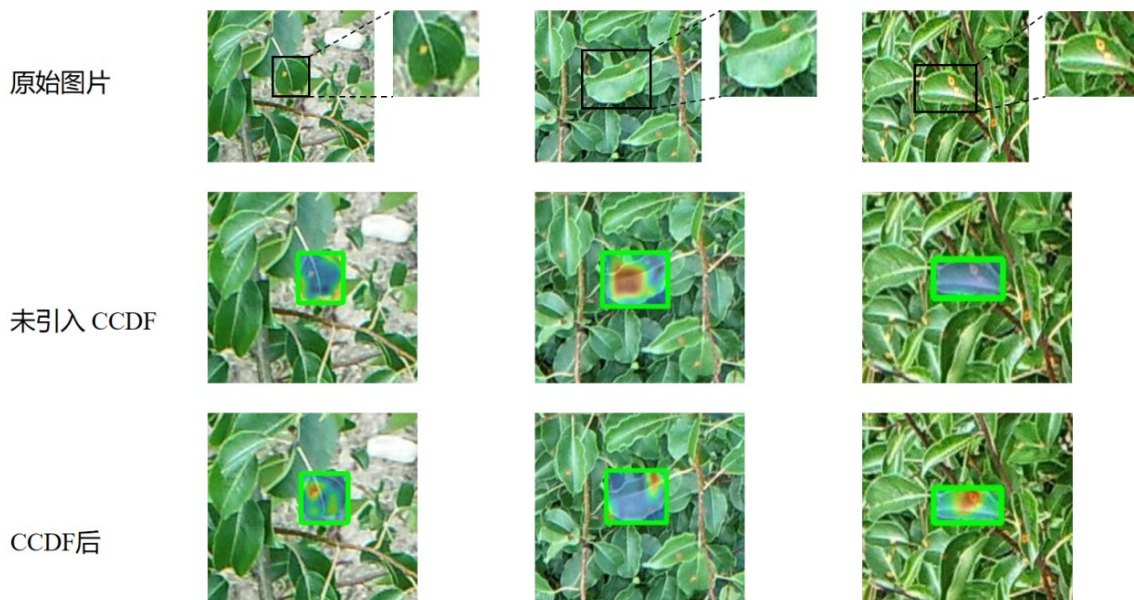
Fig. 6 Relationship between confounding probability and score reduction

脉、遮挡及类病斑纹理等混杂因素引发的高置信度误检问题,提出了一种因果启发式去混杂的梨锈病检测方法,核心为CCDF模块。该方法在不依赖像素级混杂标注的前提下,将与真实框有效匹配的解码器查询视为目标证据,将高置信度、低IoU的误检查询视为混杂因素的可观测代理,并在同一特征空间中进行统一建模;进而利用轻量MLP估计各候选查询的混杂风险,并通过连续单调的重评分函数对高风险预测施以后验抑制,从因果干预的角度切断

了非病害背景与检测决策之间的虚假关联($C \rightarrow Q \rightarrow Y$)。

实验结果表明,目标证据查询与混杂查询具有较好的特征可分性。引入CCDF后,模型精确率P由0.755提高至0.781,假阳性数量FP由460个降低至348个,mAP50有一定提高;同时召回率R由0.664下降至0.641,体现出误检抑制与目标保留之间的P-R权衡。

值得指出的是,本文的方法框架不依赖于特定检测架构或特定的混杂形态先验,具有一定的通用参考价值。后续研究将进一步探索自适应混杂抑制策略,并将查询级去混杂约束融入端到端训练过程,以兼顾误检抑制、困难目标召回和跨场景泛化能力。



((a) Original image; (b) Without CCDF; (c) After CCDF re-scoring)

图7 CCDF重评分前后的目标证据响应热力图

Fig. 7 Target-evidence response heatmaps before and after CCDF re-scoring

参考文献(References)

- Shi Y, Han L X, Kleerekoper A, Chang S and Hu T. 2022. Novel Crop-docNet model for automated potato late blight disease detection from unmanned aerial vehicle-based hyperspectral imagery. *Remote Sensing*, 14(2): 396 [DOI: 10.3390/rs14020396]
- Kurihara J, Koo V C, Guey C W, Lee Y P and Abidin H. 2022. Early detection of basal stem rot disease in oil palm tree using unmanned aerial vehicle-based hyperspectral imaging. *Remote Sensing*, 14(3): 799 [DOI: 10.3390/rs14030799]
- Xiao D Q, Pan Y Q, Feng J Z, Yin J J, Liu Y F and He L. 2022. Remote sensing detection algorithm for apple fire blight based on UAV multispectral image. *Computers and Electronics in Agriculture*, 199: 107137 [DOI: 10.1016/j.compag.2022.107137]
- Prasad A, Mehta N, Horak M and Bae W D. 2022. A two-step machine learning approach for crop disease detection using GAN and UAV technology. *Remote Sensing*, 14(19): 4765 [DOI: 10.3390/rs14194765]
- Amarasingam N, Gonzalez F, Salgadoe A S A, Sandino J and Powell K. 2022. Detection of white leaf disease in sugarcane crops using UAV-derived RGB imagery with existing deep learning models. *Remote Sensing*, 14(23): 6137 [DOI: 10.3390/rs14236137]
- Bao W X, Zhu Z Q, Hu G S, Zhou X G, Zhang D Y and Yang X J. 2023. UAV remote sensing detection of tea leaf blight based on DDMA-YOLO. *Computers and Electronics in Agriculture*, 205: 107637 [DOI: 10.1016/j.compag.2023.107637]
- Li L P, Zhao H and Liu N. 2023. MCD-Yolov5: accurate, real-time crop disease and pest identification approach using UAVs. *Electronics*, 12(20): 4365 [DOI: 10.3390/electronics12204365]
- Liu Y, Liu G H, Sun H, An L L, Zhao R M, Liu M J, Tang W J, Li M Z, Yan X J, Ma Y T and Zhao F K. 2024. Exploring multi-features in UAV-based optical and thermal infrared images to estimate disease severity of wheat powdery mildew. *Computers and Electronics in Agriculture*, 225: 109285 [DOI: 10.1016/j.compag.2024.109285]
- Zhang J, Zhang D F, Liu J Y, Zhou Y H, Cui X S and Fan X F. 2024. DSCONV-GAN: a UAV-based model for Verticillium wilt disease detection in Chinese cabbage in complex growing environments. *Plant Methods*, 20: 186 [DOI: 10.1186/s13007-024-01303-2]
- Zhao G Y, Lan Y B, Zhang Y L and Deng J Z. 2025. Rice canopy disease and pest identification based on improved YOLOv5 and UAV images. *Sensors*, 25(13): 4072 [DOI: 10.3390/s25134072]
- Maß V, Alirezazadeh P, Seidl-Schulz J, Leipnitz M, Fritzsche E, Ibrahim R A A, et al. 2025. Annotated image dataset with different stages of European pear rust for UAV-based automated symptom detection in orchards. *Data in Brief*, 58: 111271 [DOI: 10.1016/j.dib.2025.111271]
- Leng J X, Mo M J C, Zhou Y H, Ye Y M, Gao C Q and Gao X B. 2023. Recent advances in drone-view object detection. *Journal of Image and Graphics*, 28(9): 2563-2586 (冷佳旭, 莫梦竟成, 周应华, 叶永明, 高陈强, 高新波. 2023. 无人机视角下的目标检测研究进展. *中国图象图形学报*, 28(9): 2563-2586) [DOI: 10.11834/jig.220836]
- Pan X Y, Jia N X, Mu Y Z and Gao X R. 2023. Survey of small object detection. *Journal of Image and Graphics*, 28(9): 2587-2615 (潘晓英, 贾凝心, 穆元震, 高炫蓉. 2023. 小目标检测研究综述. *中国图象图形学报*, 28(9): 2587-2615) [DOI: 10.11834/jig.220455]
- Feng Q H, Wang Z X, Sun C C and Shao Z W. 2025. Small object detection in drone images via foreground refinement and multidimensional inductive bias self-attention. *Journal of Image and Graphics*, 30(11): 3547-3563 (冯琪涵, 王志晓, 孙成成, 邵志文. 2025. 融合前景细化和多维归纳偏置自注意力的无人机图像小目标检测. *中国图象图形学报*, 30(11): 3547-3563) [DOI: 10.11834/jig.250017]
- Tang Z, Lu J, Chen Z, Qi F and Zhang L. 2023. Improved Pest-YOLO: real-time pest detection based on efficient channel attention mechanism and transformer encoder. *Ecological Informatics*, 78: 102340 [DOI: 10.1016/j.ecoinf.2023.102340]
- Zhang L, Ding G, Li C and Li D. 2023. DCF-Yolov8: an improved algorithm for aggregating low-level features to detect agricultural pests and diseases. *Agronomy*, 13(8): 2012 [DOI: 10.3390/agronomy13082012]
- Xu W Y, Xu T, Thomasson J A, Chen W, Karthikeyan R, Tian G Z, Shi Y Y, Ji C Y and Su Q. 2023. A lightweight SSV2-YOLO based model for detection of sugarcane aphids in unstructured natural environments. *Computers and Electronics in Agriculture*, 211: 107961 [DOI: 10.1016/j.compag.2023.107961]
- Yang S, Xing Z, Wang H B, Dong X R, Gao X, Liu Z, Zhang X D, Li S M and Zhao Y Y. 2023. Maize-YOLO: a new high-precision and real-time method for maize pest detection. *Insects*, 14(3): 278 [DOI: 10.3390/insects14030278]
- Yan C M and Yang K Y. 2024. FSM-YOLO: apple leaf disease detection network based on adaptive feature capture and spatial context awareness. *Digital Signal Processing*, 155: 104770 [DOI: 10.1016/j.dsp.2024.104770]
- Wang N, Fu S W, Rao Q, Zhang G Y and Ding M Q. 2025. Insect-YOLO: a new method of crop insect detection. *Computers and Electronics in Agriculture*, 232: 110085 [DOI: 10.1016/j.compag.2025.110085]
- Su B, Zhu Y Y and Lin Y F. 2025. Pest-YOLOv8: enhanced detection for small and complex agricultural pests using triple attention and Wise-IoU. *Journal of Plant Diseases and Protection*, 132: 135 [DOI: 10.1007/s41348-025-01134-w]
- Zhao Y A, Lv W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, et al. 2024. DETRs beat YOLOs on real-time object detection//Proceed-

- ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]
- Xu R K, Yu J J, Ai L B, Yu H F and Wei Z C. 2024. Farmland pest recognition based on Cascade RCNN combined with Swin-Transformer. *PLoS ONE*, 19(6): e0304284 [DOI: 10.1371/journal.pone.0304284]
- Wang H X, Nguyen T H, Nguyen T N and Dang M. 2024. PD-TR: end-to-end plant diseases detection using a transformer. *Computers and Electronics in Agriculture*, 224: 109123 [DOI: 10.1016/j.compag.2024.109123]
- Wang T, Zhou C, Sun Q R and Zhang H W. 2021. Causal attention for unbiased visual recognition//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3091-3100 [DOI: 10.1109/ICCV48922.2021.00308]
- Rao Y M, Chen G Y, Lu J W and Zhou J. 2021. Counterfactual attention learning for fine-grained visual categorization and re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 1005-1014 [DOI: 10.1109/ICCV48922.2021.00106]
- Huang W Q, Jiang M, Li M, Meng B, Ren J X, Zhao S X, Bai R W and Yang Y. 2021. Causal intervention for object detection//Proceedings of 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence. Washington, DC, USA: IEEE: 770-774 [DOI: 10.1109/ICTAI52525.2021.00122]
- Liu R Y, Liu H, Li G, Hou H D, Yu T H and Yang T. 2022. Contextual debiasing for visual recognition with causal mechanisms//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12755-12765 [DOI: 10.1109/CVPR52688.2022.01242]
- Li J M, Zhang Y N, Qiang W W, Si L Y, Jiao C B, Hu X H, Zheng C W and Sun F C. 2023. Disentangle and remerge: interventional knowledge distillation for few-shot object detection from a conditional causal perspective. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1): 1323-1333 [DOI: 10.1609/aaai.v37i1.25216]
- Xu M J, Qin L Y, Chen W J, Pu S L and Zhang L. 2023. Multi-view adversarial discriminator: mine the non-causal factors for object detection in unseen domains//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 8103-8112 [DOI: 10.1109/CVPR52729.2023.00783]
- Liu Y J, Zhou S J, Liu X Y, Hao C H, Fan B J and Tian J D. 2024. Unbiased Faster R-CNN for single-source domain generalized object detection//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 28838-28847 [DOI: 10.1109/CVPR52733.2024.02724]
- Zhu X Z, Su W J, Lu L W, Li B, Wang X G and Dai J F. 2021. Deformable DETR: deformable Transformers for end-to-end object detection//Proceedings of the 9th International Conference on Learning Representations. Virtual Event [DOI: 10.48550/arXiv.2010.04159]
- Yang J, Zhang H, Zhou Y, et al. 2025. Improved DAB-DETR model for irregular traffic obstacles detection in vision-based driving environment perception scenario. *Applied Intelligence*, 55(7): 541 [DOI: 10.1007/s10489-025-06440-2]
- Zhang H, Li F, Liu S L, Zhang L, Su H, Zhu J, et al. 2023. DINO: DETR with improved denoising anchor boxes for end-to-end object detection//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda [arXiv: 2203.03605]

作者简介

张聿璇,男,西安理工大学计算机学院硕士研究生,主要研究方向为数字图像处理、目标检测和深度学习。E-mail: 2645442077@qq.com。

石争浩,男,西安理工大学计算机学院教授,博士生导师,主要研究方向为图像处理、计算机视觉、医学影像计算机辅助诊断。E-mail: yishi@xaut.edu.cn。

刘晗涛,男,英国卡迪夫大学计算机科学与信息学院教授,主要研究方向为图像/视频质量评估、多媒体信号处理与视觉感知。Email: hantao.liu@cs.cardiff.ac.uk。

Oktay Karakus,男,英国卡迪夫大学计算机科学与信息学院讲师(助理教授)。主要研究方向为统计与机器学习方法在复杂成像、传感器及环境数据分析中的应用。Email: karakuso@cardiff.ac.uk。

戴佳锟,男,陕西省生物农业研究所研究员,主要研究方向为作物表型智能解析与无人机遥感检测在育种方面的应用。Email: djkxa@163.com