

基于改进 YOLO 11n 的轻量级青贮玉米田间收获 人体检测方法

李 君¹ 黄 奕¹ 吕昊瞰¹ 刘 娟¹ 韦强威¹ 许 玲²

(1. 中国农业大学工学院, 北京 100083; 2. 烟台市土地储备和利用中心, 烟台 264003)

摘要: 针对青贮玉米收获机作业过程中田间人员常被密集作物遮挡时可见光的目标检测效果受限, 红外成像可感知田间环境中人体与周围作物温度的显著差异但分辨率较低, 采用传统单一模态方法存在的精度不足及漏检率高的问题, 提出了一种基于改进 YOLO 11n 的红外与可见光图像融合的青贮玉米田间人体检测方法, 通过融合红外与可见光来提升田间被遮挡人体检测的鲁棒性和准确性。在主干网络中以轻量化的 LWGA (Lightweight global-aware) 模块替代原有的 SPPF (Spatial pyramid pooling fast) 结构, 增强多尺度特征提取能力并降低模型复杂度; 结合多层次的 C3k2_TBFE 模块与深层的 C2PSA_biformer 注意力机制, 强化对复杂背景及遮挡目标的特征表达能力; 在颈部网络中引入 C3k2_TAB 模块, 提升多尺度语义融合效果及小目标检测性能。构建青贮玉米收获田间人体被遮挡图像的数据集, 利用该数据集对改进后的模型进行训练和验证。实验结果显示, 改进模型在红外与可见光数据集上 mAP@50 分别达到 91.68% 和 88.32%, 较原始模型分别提升 4.48、3.13 个百分点, 其模型参数量和浮点运算量分别降低 4.26% 与 7.9%。改进后的模型可满足在田间人体实时检测需求且可长期适应户外工作, 并构建双摄像头图像智能识别及多模态感知系统, 保证高效、高精度实时检测, 可为青贮玉米收获机作业过程中复杂环境下田间人体的实时检测提供理论技术支撑。

关键词: 青贮玉米; 田间人体检测; 红外与可见光融合; YOLO 11n; 轻量化网络; 注意力机制

中图分类号: TP391

文献标识码: A

文章编号: 1000-1298(2026)21-0324-12

OSID:



Lightweight Silage Corn Field Personnel Detection with Improved YOLO 11n

Li Jun¹ Huang Yi¹ Lü Haotun¹ Liu Juan¹ Wei Qiangwei¹ Xu Ling²

(1. College of Engineering, China Agricultural University, Beijing 100083, China

2. Yantai Land Reserve and Utilization Center, Yantai 264003, China)

Abstract: Aiming to address the limitations of visible-light-based object detection during silage corn harvesting, where field workers were often obscured by dense crops, and compensate for the low resolution of infrared imaging despite its ability to distinguish temperature differences between humans and surrounding vegetation, a detection method based on the fusion of infrared and visible images was proposed by using an improved YOLO 11n architecture. The original spatial pyramid pooling fast (SPPF) structure in the backbone network was replaced with a lightweight lightweight global-aware (LWGA) module to enhance multi-scale feature extraction and reduce model complexity. Multi-level C3k2_TBFE modules and a deep C2PSA_biformer attention mechanism were incorporated to strengthen the representation of complex backgrounds and occluded targets, while a C3k2_TAB module was introduced in the neck network to improve multi-scale semantic fusion and small-object detection performance. A dataset containing images of occluded field personnel in cornfields was constructed, and the improved model was trained and validated using this dataset. Experimental results showed that detection accuracies of 91.68% and 88.32% were achieved on infrared and visible datasets, respectively, with improvements of 4.48 and 3.13 percentage points compared with that of the original model. The model's parameter count and computational complexity were reduced by 4.26% and 7.9%, respectively. The improved model was capable of meeting real-time detection requirements in field environments and supporting long-

收稿日期: 2025-08-04 修回日期: 2025-08-22

基金项目: 北京市乡村振兴农业科技课题 (NY2502290125)

作者简介: 李君 (2002—), 男, 硕士生, 主要从事智能农业装备与技术研究, E-mail: sy20243071706@cau.edu.cn

通信作者: 吕昊瞰 (1983—), 男, 讲师, 博士, 主要从事传感检测与伺服控制研究, E-mail: lvhaotun@cau.edu.cn

term outdoor operation, and a dual-camera intelligent recognition and multi-modal perception system was constructed to ensure efficient, accurate, and real-time detection, providing theoretical and technical support for personnel detection under complex conditions during silage corn harvesting.

Key words: silage corn; field personnel detection; infrared and visible image fusion; YOLO 11n; lightweight network; attention mechanism

0 引言

在青贮玉米机作业过程中,人员误入作业区域所引发的伤亡事故频发,其主要原因在于田间环境复杂,人体目标难以及时、准确识别^[1]。目前,作业区域人体的检测仍以人工监控和可见光视觉识别为主,但由于田间背景杂乱、青贮玉米植株遮挡频繁以及光照条件变化剧烈,单一视觉通道难以实现稳定、可靠的人体识别^[2]。尤其是在作业视野受限、天气昏暗或粉尘较多等不利条件下,人体目标边缘模糊、纹理信息弱,极易造成漏检与误报,限制了智能预警系统的实际应用效果^[3]。因此,亟需一种适应复杂环境、具备高鲁棒性和实时性的人体检测方法,以提升农机作业的主动安全水平。

红外热成像技术凭借其其对热辐射的敏感性,具有全天候、非接触和抗干扰等优势,能够有效提取人体热特征,是提升农业作业区域人体检测能力的重要手段^[4]。研究表明,在低光照、背光或植被遮挡等条件下,红外图像可有效弥补可见光图像的不足,展现出更强的通用性与稳定性^[5]。同时,深度学习目标检测算法近年来在计算机视觉领域取得显著进展,为红外与可见光图像的智能融合提供了可行的技术路径。

红外与可见光图像的融合在智能安防、自动驾驶等复杂场景中已获得良好应用^[6-8]。现有研究充分验证了红外与可见光融合在低光照、遮挡和复杂背景条件下的互补性与鲁棒性。

然而,红外与可见光融合在农业场景中的应用尚处于起步阶段。现有工业应用研究多采用浅层融合策略,如像素级加权平均或特征级通道拼接,但未能充分挖掘多模态数据的深层语义信息与互补特性。此外,传统模型如 YOLO v8n 虽在煤矿等场景中取得较高检测精度,但其计算复杂度较高,难以在资源受限的农机边缘设备上高效部署。更为关键的是,以青贮玉米田为代表的高秆作物环境下,目标检测面临光照剧烈变化、植株密集遮挡、前后景难以分辨等困难,使得现有注意力机制(如 SE/CBAM)对小目标的特征提取能力不足,导致漏检率居高不下,严重影响检测模型的稳定性与实用性^[9]。因此,在青贮玉米田间高复杂度环境中,如何实现具备深层

语义感知能力的红外与可见光特征融合,是当前亟待解决的关键技术问题。

为此,本文提出一种基于 YOLO 11n 架构和轻量级青贮玉米田间人体检测方法。与现有方案相比,该方法采用决策级动态融合机制,通过双通道图像采集,结合可见光与红外检测结果,实现对人体目标的互补识别,有效提升检测的准确率、鲁棒性与系统实时性。同时,对模型结构进行轻量化与模块化设计,增强其多尺度特征提取能力和在资源受限平台上的部署适应性,以期为农机作业安全提供可靠的技术支撑。

1 数据采集与多模态融合方法

1.1 数据采集

本研究实验图像数据采集于北京市延庆区沈家营镇香村营村青贮玉米田(40.497550° N, 116.076631° E)。该区域青贮玉米种植行距为 53 cm,株间距为 23 cm,植株高度约 2.7 m。结合农业生产的经济性考量,图像核心采集设备选用鑫思特 HT-S320 型红外热成像摄像头^[10]。为保障采集操作的安全性,并最大限度贴近实际作业环境,摄像头采用人工架设方式布置于高处,安装高度设定为 3.8 m,与实际青贮玉米收获机的摄像头安装高度保持一致。采集期间,采集对象在青贮玉米田间随机行走,以获取不同遮挡条件下的图像数据。红外热成像摄像头通过连接线与搭载 Windows 11 操作系统的计算机相连,借助自主编写的 Python 程序实现摄像头调用与视频录制功能,同步完成实时画面传输与控制。

为模拟多种光照与温度条件下的实际应用场景,提升模型的泛化能力,数据采集覆盖多个时间段,每日分 3 次进行图像采集,具体时段为 07:00—09:00、13:00—15:00 及 19:00—21:00。数据采集的实地工作场景如图 1 所示。

1.2 图像融合

图像融合技术按融合层级可划分为像素级融合、特征级融合及决策级融合 3 类^[11]。其中,像素级融合通过对源图像像素点的直接运算实现信息整合,其核心优势在于能够最大限度保留原始图像的细节特征,但存在数据处理量大、计算复杂度高的固

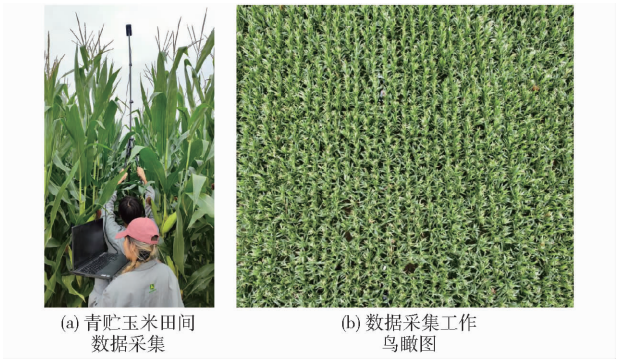


图 1 数据采集的实地工作场景

Fig. 1 On-site data collection scenarios in corn field

有局限;特征级融合需先经边缘检测、纹理提取等算法完成关键特征提取,再对特征层进行融合处理,该方式在保证处理效率的同时,可显著增强系统的抗干扰能力;决策级融合则是在各源图像分别完成目标分析之后,依据特定的决策规则对分析结果进行整合,具有较强的鲁棒性^[12]。

在青贮玉米田间人体检测的初步研究中,对像素级、特征级和决策级 3 种融合策略进行了分析与尝试。像素级融合虽能保留细节,但因模态差异大,易出现边缘模糊与信息混淆,检测性能提升有限且计算开销增加。特征级融合在 CNN 中间层拼接或加权特征,但存在特征对齐困难、信息冗余和训练收敛慢等问题,实验结果显示其 mAP@ 0.5 比决策级方案低 2~3 个百分点。相比之下,决策级融合通过逻辑规则整合检测结果,有效利用红外与可见光的互补优势,避免信息混淆和特征异构性,计算开销低且鲁棒性更高。

在青贮玉米田间人员检测场景中,实践研究揭示出两类关键技术瓶颈:①红外图像与可见光图像的模态特性存在本质差异,红外图像以目标与环境的温度对比度为核心特征,可见光图像则依赖场景的光学反射特性^[13],这种跨模态特征空间的异构性会导致特征匹配过程中频繁出现误匹配与特征丢失现象,显著增加了特征级融合的实施难度。②在实际检测过程中,当青贮玉米田间人体与摄像头的距离小于 2.8 m 时,红外摄像头所获取的人体图像因红外辐射特征与周围环境趋同,使得目标区分度大幅下降,检测效能显著衰减;而可见光摄像头在此近距离场景下,由于遮挡干扰较少,能够精准识别田间作业人员的形态特征,展现出独特优势。反之,可见光摄像头在强光直射、逆光阴影等复杂光照条件及高密度植被遮挡环境中,易受光学噪声干扰而导致检测精度降低^[14],此时红外摄像头凭借其对于复杂环境的穿透性优势可有效弥补这一缺陷。

基于上述研究发现,鉴于红外与可见光摄像头

在不同场景下呈现出的互补性优势,且特征级融合在跨模态场景中实施难度较大,采用决策级融合策略,构建双摄像头多模态感知系统,通过二者的功能互补突破单模态检测的性能瓶颈。

系统采用双通道检测架构,为实现多模态信息的有效聚合,采用逻辑或运算作为决策融合规则,即

$$P(x) = I(x) \cup V(x) \tag{1}$$

式中 x ——目标区域

$I(x)$ ——红外摄像头检测输出

$V(x)$ ——可见光摄像头检测输出

$P(x)$ ——最终检测输出

$\cup$ 为逻辑或运算符,其运算逻辑为:当 $I(x) = 1$ 或 $V(x) = 1$ 时,融合结果 $P(x) = 1$,最终判定为存在人体;仅当 $I(x) = 0$ 且 $V(x) = 0$ 时, $P(x) = 0$,最终判定为无人体。公式(1)的决策融合需通过空间对齐实现:基于双目相机标定参数,将红外图像像素坐标通过投影变换映射至可见光图像坐标系,确保同一空间位置的目标被关联。该融合规则的核心是对同一空间位置在 2 种成像模态上的检测结果进行逻辑运算。然而,由于双摄像头安装位姿的差异,同一目标场景在两幅图像中的投影存在非刚性形变,导致目标像素位置不一致。因此,实现跨模态检测结果的可靠融合,必须对红外与可见光图像进行严格的空

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \tag{2}$$

式中 m_{11} 、 m_{12} 、 m_{13} 、 m_{21} 、 m_{22} 、 m_{23} 为参数。

设 $P'(x',y')$ 为可见光图像某点, $P(x,y)$ 为红外图像上对应的同名点,将点 $P(x,y)$ 变换到可见光图像坐标系的点 $P'(x',y')$ 的公式为

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

仿射矩阵 M 有 6 个参数,理论上只需在红外图像和可见光图像之间找到 3 对精确对应的特征点即可求解。为了减小特征点定位误差和误匹配的影响,提升配准鲁棒性,采用 RANSAC 算法。该算法能够从包含噪声的观测数据中估计出最优模型参数。具体操作是:在红外和可见光图像上手动标记多对高置信度的特征点。将这些特征点坐标输入 RANSAC 算法,最终求解出最优的仿射矩阵 M 。

利用求得的仿射矩阵 M 对可见光图像进行变

换,即可得到与可见光图像空间对齐的配准后红外图像。配准前后图像对比如图 2a 和图 2b 所示。红外图像与配准后可见光图像叠加得到的融合示意图如图 2c 所示。从图 2c 可以观察到,配准后 2 类图

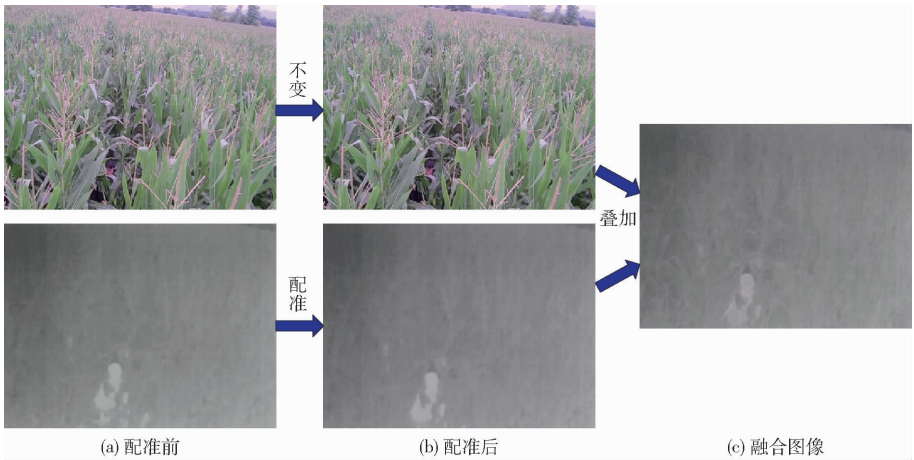


图 2 热红外和可见光图像融合示例

Fig. 2 Example of thermal infrared and visible light images fusion

这种“存在即判定”的融合逻辑本质上是通过构建冗余检测机制实现多模态信息的互补,在红外摄像头因距离过近导致检测失效的场景中,可见光摄像头可凭借清晰的光学成像完成目标补充检测;而在可见光摄像头受作物遮挡或光照干扰时,红外摄像头能通过温度特征实现目标精准定位。该跨模态冗余设计不仅有效降低了单一传感器失效引发的漏检风险,更在统计层面提升了田间人体检测的空间覆盖率与时间连续性。

1.3 数据集构建

采用决策级融合策略,需分别构建红外与可见光 2 个独立数据集。所采用的数据集包含 2 部分数据集 A 和 B:数据集 A 由近距离采集的可见光图像组成,主要服务于基于深度学习模型的人体自动检测任务;数据集 B 由青贮玉米田间较远距离且存在严重遮挡条件下采集的热红外图像构成,用于提升复杂环境中人体检测的鲁棒性与准确率。其中,可见光图像为 3 通道 RGB 格式,保留作物纹理与人体色彩特征;红外图像为单通道灰度图,像素值映射温度梯度。

使用目标检测模型分别对可见光图像与热红外图像中的人体进行定位。在模型训练前,需对图像数据进行标注,并剔除遮挡严重及模糊不清的图像样本,最终得到 1 420 幅可见光图像和 2 435 幅热红外图像。其中,可见光图像划分为训练集 852 幅、验证集 284 幅、测试集 284 幅;热红外图像划分为训练集 1 461 幅、验证集 487 幅、测试集 487 幅。

考虑到样本多样性对模型泛化能力和鲁棒性的显著影响,研究从训练集选取部分图像进行数

据增强。从可见光图像中选取 200 幅样本,从热红外图像中选取 400 幅样本,分别采用图像旋转、添加高斯噪声、锐化、亮度增强及亮度减弱 5 种方式进行增强,共生成 1 420 幅可见光增强图像和 2 435 幅热红外增强图像,并按原始划分比例分配至各子集中。两数据集的数据增强示例如图 3 和图 4 所示。上述数据增强方法能够使训练后的青贮玉米田间人体检测模型更好地适应不同视角、光照条件及噪声干扰,有效提升模型的泛化能力与鲁棒性,同时缓解因部分场景样本数量不足可能导致的过拟合问题。

最终数据集 A(可见光)共计 2 840 幅图像,数据集 B(热红外)共计 4 870 幅图像,均包含增强后样本。所有图像均使用 Labelme 软件^[15]进行“person”类别的手动标注,生成 JSON 格式标注文件,随后进一步转换为适用于深度学习模型训练的 TXT 格式文件,用于后续目标检测模型的训练与评估。模型通过双输入分支分别处理:可见光分支采用 3×3 卷积提取颜色与边缘特征,红外分支通过 1×1 卷积将单通道特征升维至与可见光一致的通道数,再进入主干网络。

1.4 实验平台

训练模型所用的实验平台参数如表 1 所示。训练阶段采用构建的红外与可见光融合数据集,设置批量大小为 8,迭代次数为 200,初始学习率为 0.001。为进一步提升模型的泛化能力,训练过程中同步采用数据增强与正则化技术。整个训练流程在指定硬件环境下执行,保障训练过程的稳定性与高效性。通过合理的参数配置及优化策略,模型在保

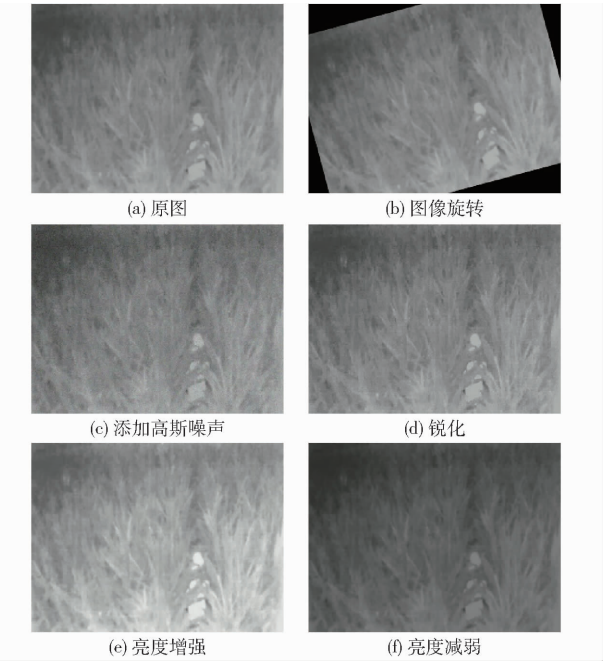


图3 红外数据集图像增强示例

Fig.3 Examples of infrared dataset image augmentation

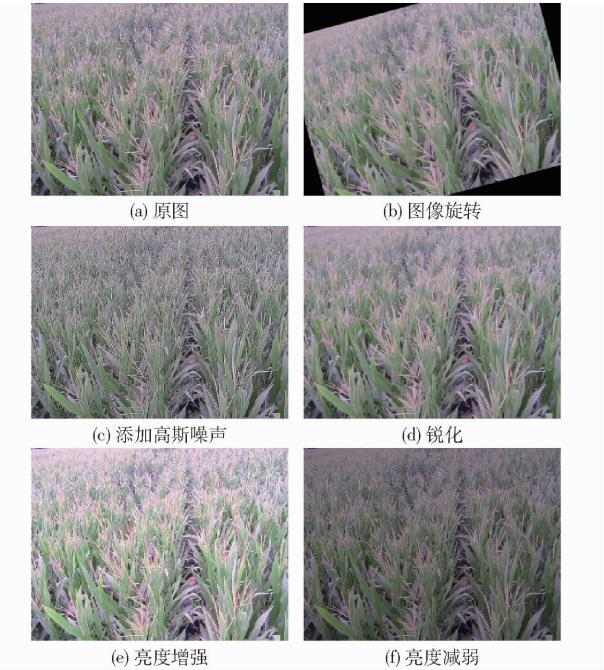


图4 可见光数据集图像增强示例

Fig.4 Examples of visible light dataset image augmentation

表1 实验平台参数

Tab.1 Experimental platform parameters	
参数	配置
CPU	Intel(R) Core(TM) i9-13900HX@ 2.20GHz
内存容量/GB	16
GPU	NVIDIA GeForce RTX 4060
GPU 加速库	CUDA 11.8、CUDNN 9.1.0
操作系统	Windows 11(64 位)
环境	Python 3.10、Pytorch 2.7.1+cu118

持轻量化特性的同时,实现了较好的收敛效果。

2 青贮玉米田间人体检测方法

2.1 改进 YOLO 11n 目标检测模型

2.1.1 轻量级青贮玉米田间人体目标检测模型

青贮玉米田间人体检测模型的核心目标是识别田间作业人体位置并定位目标区域,从而实现对人体的精准监测。选择 YOLO 11n 作为基线模型,针对青贮玉米田间人体目标开展训练,以确保模型在复杂环境下具备优异的检测性能与定位精度。该模型由主干模块、颈部模块和头部模块 3 部分构成:主干模块负责提取图像多尺度特征,颈部模块承担不同层级特征的融合任务,头部模块则输出最终检测结果,包括目标位置坐标与人体类别置信度。输入图像首先经主干网络处理,通过标准卷积层与改进的 C3k2 结构提取多层次特征;随后进入 SPPF (Spatial pyramid pooling fast) 模块完成多尺度上下文信息整合,其后续的 C2PSA 模块引入注意力机制以增强对显著区域的关注能力,并将高质量特征传递至后续网络;特征融合阶段采用 PAN 与 FPN 相结合的架构,通过上采样与下采样操作实现多分辨率特征的双向传递与融合^[16],强化模型对复杂场景中目标的识别能力;最终,融合后的特征输入检测头部结构,在不同尺度特征图上完成目标边框与类别信息的预测,输出位置坐标、置信度及类别得分等信息,并经非极大值抑制 (NMS) 模块去除重叠冗余框,生成最终目标检测结果。

YOLO 11n 在常规场景下表现出良好的目标检测性能,但在青贮玉米田间复杂环境中仍存在局限性:模型参数量较大且计算开销较高,难以在边缘计算设备上实现高效部署;受青贮玉米田密集植被、强遮挡及复杂光照等因素干扰,模型对小尺度人体目标及部分遮挡目标的检测稳定性不足,存在定位精度下降与鲁棒性欠缺的问题。为此,提出一种改进的 YOLO 11n 目标检测模型,结构如图 5 所示。主要改进策略包括:采用轻量化 LWGA (Lightweight global-aware) 模块替代 SPPF 结构,在增强多尺度目标特征提取能力的同时保证模型轻量化特性,适用于资源受限环境部署;在主干网络中引入多层次 C3k2_TBFE 模块,并结合深层 C2PSA_biformer 注意力机制,显著提升局部与全局特征表达能力,以优化复杂背景及遮挡条件下的识别性能;颈部网络采用 C3k2_TAB 模块构建特征融合路径,有效促进不同尺度语义信息交互,进一步提高多尺度小目标的检测精度与鲁棒性。

2.1.2 LWGA 模块

在实际应用场景中,人体检测模型需适应多样

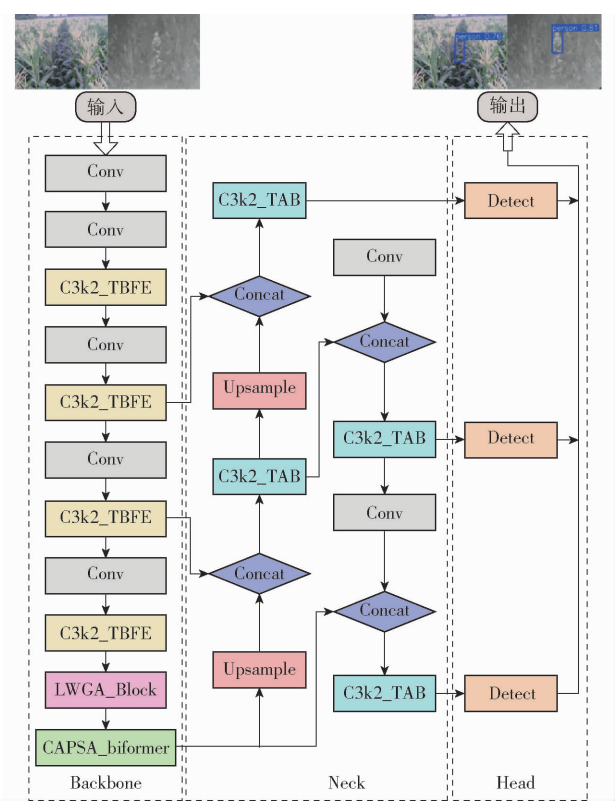


图 5 改进 YOLO 11n 网络结构图

Fig. 5 Improved YOLO 11n network architecture

化复杂环境,特别是在资源受限的边缘设备中部署运行时,因此需在检测精度与模型复杂度之间建立良好平衡。YOLO 11n 网络虽具备较强的特征提取能力,但其包含的大量标准卷积层会导致参数量与计算复杂度偏高,使其在低功耗设备上的实时运行受到限制。传统主干网络(如 ResNet^[17] 和 VGGNet^[18])虽性能优异,但因结构庞大、推理耗时较长,难以满足嵌入式平台的应用需求;而 MobileNet^[19] 和 ShuffleNet^[20] 等轻量化网络通过深度可分离卷积显著减少参数与计算开销,更适配资源受限场景,但其在青贮玉米田间复杂背景及小目标遮挡等条件下,检测稳定性与精度表现欠佳。

针对这一矛盾,在 YOLO 11n 网络中采用 LWGA^[21] 模块替代原有 SPPF 卷积结构。该模块融合采用轻量卷积的局部路径,与通过通道重标定捕获全局上下文信息的全局路径,两者并行处理后进行加权融合,生成兼具局部细节与全局感知能力的高质量特征图,有效提升模型对不同尺度目标的识别能力。LWGA 模块的引入不仅能保证检测性能,同时可显著降低模型的内存占用与运行延迟,提升在边缘设备上的部署适应性与实用价值,其结构如图 6 所示。

2.1.3 C3k2_TBFE 模块和 C2PSA_biformer 模块

在青贮玉米田间这类遮挡严重的环境中,人体目标常被密集植株部分遮挡,从采集数据集中可观

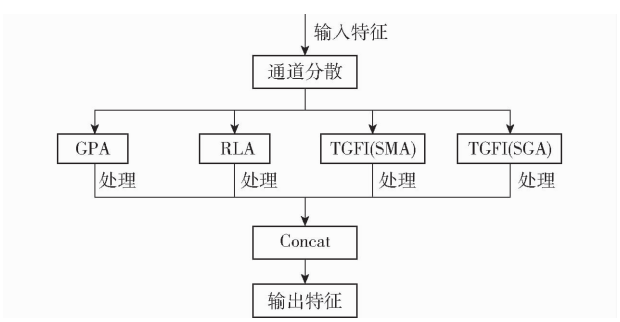


图 6 LWGA 模块内部结构图

Fig. 6 LWGA module architecture diagram

察到,人体在图像中所占像素比例较小,属于典型的小目标检测问题。在此背景下,传统 YOLO 11n 目标检测模型易出现漏检与误检,难以满足高精度检测需求。为提升模型在该类场景中的检测性能,对 YOLO 11n 网络结构进行改进,提出将原模型中的 C3k2 模块替换为 TBFE (Two-branch feature extractor) 模块,构建新的 C3k2_TBFE 组合结构。

TBFE 模块的设计充分利用 CNN 在局部特征提取方面的优势,通过并行引入 3D 卷积与 2D 卷积,分别从光谱维度与空间维度进行特征提取,在特征提取早期实现多维度信息的有效融合^[22]。其中,低通道数的逐点卷积层用于调整图像的通道维度,既降低后续网络的计算复杂度,又保留关键特征信息;3D 卷积结合扩展与挤压机制,增强模型对光谱特征的捕获能力;2D 卷积则专注于提取光谱-空间联合特征。两者互为补充、协同作用,有效提升模型对复杂场景下小目标的感知与表征能力,从而提高青贮玉米田间人体检测的准确性与鲁棒性,其具体结构如图 7a 所示。

为进一步提升模型检测准确率,在改进的 YOLO 11n 目标检测模型中,将原有的 C2PSA 模块与 biformer 结构相结合,构建出融合注意力机制的 C2PSA_biformer 模块。biformer^[23] 作为一种新型视觉 Transformer 架构,采用双层路由注意力机制(Bi-level routing attention, BRA),能够实现动态、查询感知的稀疏注意力分配,在显著提升模型建模长距离依赖能力的同时,保持较高计算效率。结合青贮玉米田间人体检测的实际场景,环境复杂、植株密集、光照变化大,人体目标常呈现尺寸小、遮挡严重等特点,传统卷积结构难以充分建模远距离上下文信息,导致漏检率较高。C2PSA_biformer 模块通过引入全局感知机制,增强模型对长距离特征的捕获与聚合能力,提升在复杂背景中对小目标或被遮挡目标的识别性能,从而更好地适应青贮玉米田间等复杂农业场景下的人体检测任务,其具体结构如图 7b 所示。

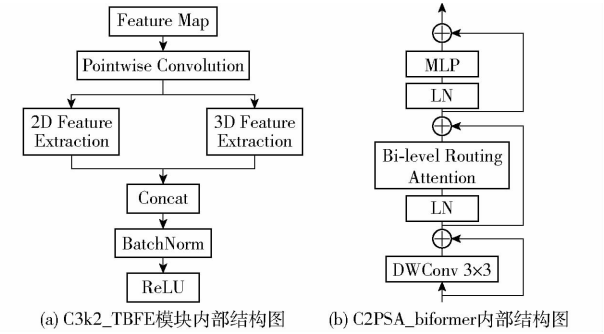


图 7 C3k2_TBFE 模块和 C2PSA_biformer 模块的内部结构图

Fig. 7 Internal structures of C3k2_TBFE and C2PSA_biformer modules

2.1.4 C3k2_TAB 模块

传统基于卷积神经网络 (CNN) 的方法由于卷积操作的局部感受野限制,在捕捉全局依赖关系方面存在不足。虽然采用复杂结构能够提升性能,但全局自注意力机制的计算复杂度较高,影响推理速度。现有的改进方法如分块处理存在内容不可知的缺陷,而多注意力机制如 ATD 则增加了计算负担。针对这些问题,在 YOLO 11n 网络的 C3k2 结构中引入了基于内容感知令牌聚合的 TAB (Token aggregation block) 模块^[24],以实现高效的长距离信息交互和快速推理。TAB 利用聚类思想,将图像令牌根据内容相似度划分为多个组,借鉴 Routing Transformer,通过共享全局令牌中心及训练阶段的指数移动平均 (EMA) 机制,精确聚合内容相似的令牌,从而实现更有效的长距离信息传播。

TAB 包含 4 个主要部分: Content-aware token aggregation (CATA) 模块负责按相似度分组并通过子组划分解决组内令牌数量不均衡问题,以提升并行效率; Intra-group self-attention (IASA) 允许子组内查询令牌关注相邻子组的键,缓解内容相似令牌拆分带来的影响; Inter-group cross-attention (IRCA) 借助训练阶段学习的全局令牌中心,在子组与令牌中心间进行交叉注意力计算,增强全局信息交互; 卷积操作用于融合前端特征,优化表达。通过引入 TAB 模块, YOLO 11n 能更高效捕捉长距离依赖,整合全局信息,提升对青贮玉米田间人体这一小目标且遮挡严重目标的检测能力。同时, TAB 优化了计算流程,降低冗余,显著提升模型推理速度,使改进后的 YOLO 11n 在保持高检测性能的基础上,更好地满足青贮玉米田间复杂环境下的实时检测需求。其内部结构如图 8 所示。

2.2 评价指标

选用精确率 (Precision, P)、召回率 (Recall, R)、平均精度均值 (mean Average Precision, mAP)、模型

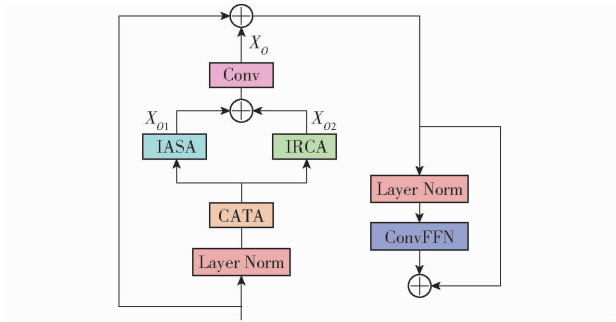


图 8 C3k2_TAB 模块内部结构图
Fig. 8 Internal structure of C3k2_TAB module

参数量及浮点运算量 (FLOPs) 作为模型评价指标。

精确率表示检测结果中被正确识别为目标的本占所有检测为目标样本的比例; 召回率表示检测出的目标样本占所有实际目标样本的比例; 平均精度均值反映模型在不同交并比阈值下的检测精度, 一般考查交并比阈值为 0.5 时的检测精度, 即 mAP@50^[25]。除此之外, 模型参数量指模型中所有可训练参数的总数, 用于反映模型的轻量化程度; 浮点运算量指模型推理过程中执行的浮点计算次数, 用于衡量计算复杂度和推理效率。综合以上指标, 可全面评估模型的检测性能及其在实际部署中的计算与存储效率。

3 实验结果与分析

3.1 消融实验

为全面评估各改进模块对模型性能的提升效果, 以 YOLO 11n 为基准模型, 逐步引入各项优化策略, 设计了 8 组消融实验, 从模块独立应用与组合效果 2 个维度, 系统分析其对检测精度与推理效率的影响。具体改进措施包括: 采用轻量化的 LWGA 模块替代原有 SPPF 结构, 以提升多尺度目标特征提取能力并降低模型复杂度, 更适配资源受限的实际部署场景; 在主干网络中引入多层级的 C3k2_TBFE 模块, 并结合深层 C2PSA_biformer 注意力机制, 增强局部与全局特征的表达能力, 提升模型在复杂背景与遮挡环境下的识别鲁棒性; 同时在颈部网络中引入 C3k2_TAB 模块, 构建更高效的特征融合路径, 强化不同尺度语义信息间的交互能力。各组改进模块对模型检测精度与推理速度的具体影响如表 2 所示。

实验结果表明, LWGA 模块在 2 个数据集上精确率、召回率和 mAP@0.5 等精度指标略有下降, 原因是其轻量化设计简化了多尺度特征聚合过程, 对密集玉米植株遮挡下的小目标高频细节捕捉能力减弱, 但其在参数量和浮点运算量方面表现突出, 有效降低了模型复杂度, 此外可将局部路径卷积核调整

表 2 消融实验结果

Tab.2 Ablation study results

模型	数据集	P/%	R/%	mAP@ 0. 5/%	参数量	浮点运算量
YOLO 11n	红外	85. 83	84. 67	87. 20	2. 58 × 10 ⁶	6. 3 × 10 ⁹
	可见光	83. 85	82. 84	85. 19		
YOLO 11n + LWGA	红外	84. 21	83. 93	86. 05	2. 06 × 10 ⁶	5. 6 × 10 ⁹
	可见光	82. 98	81. 65	84. 03		
YOLO 11n + C3k2_TBFE + C2PSA_biformer	红外	87. 85	86. 35	90. 36	2. 78 × 10 ⁶	6. 7 × 10 ⁹
	可见光	84. 62	84. 36	86. 32		
YOLO 11n + C3k2_TAB	红外	86. 32	85. 39	89. 68	2. 65 × 10 ⁶	6. 5 × 10 ⁹
	可见光	84. 48	83. 64	85. 72		
YOLO 11n + LWGA + C3k2_TBFE + C2PSA_biformer	红外	86. 30	85. 12	89. 15	2. 42 × 10 ⁶	5. 8 × 10 ⁹
	可见光	83. 95	83. 85	85. 86		
YOLO 11n + LWGA + C3k2_TAB	红外	85. 96	84. 85	89. 06	2. 36 × 10 ⁶	5. 7 × 10 ⁹
	可见光	84. 27	82. 93	85. 65		
YOLO 11n + C3k2_TBFE + C2PSA_biformer + C3k2_TAB	红外	90. 65	89. 32	92. 65	2. 89 × 10 ⁶	6. 8 × 10 ⁹
	可见光	87. 87	87. 68	89. 69		
YOLO 11n + LWGA + C3k2_TBFE + C2PSA_biformer + C3k2_TAB	红外	89. 84	88. 39	91. 68	2. 47 × 10 ⁶	5. 8 × 10 ⁹
	可见光	86. 96	86. 35	88. 32		

为 5 × 5,增强对小目标高频细节的提取能力;相比之下,由于 C3k2_TBFE 模块是通过 3D 卷积与 2D 卷积进行的并行特征提取,结合引入的 C2PSA_biformer 注意力机制显著提升了检测性能,使 3 项精度指标均有明显提升。这表明其在特征提取与表达方面具有较强能力,对作物遮挡下的局部特征敏感,然而这也带来了较高的计算开销和推理延迟;此外,C3k2_TAB 模块的加入在进一步提升检测精度的同时,也不可避免地增加了一定的计算负担。

最终,融合上述 3 种模块后,提出的改进模型在红外数据集上的精确率、召回率和 mAP@ 0. 5 分别达到 89. 84%、88. 39%和 91. 68% ,较基线模型分别提升 4. 01、3. 72、4. 48 个百分点;在可见光数据集上,精确率、召回率和 mAP@ 0. 5 分别达到 86. 96%、86. 35%和 88. 32% ,较基线模型分别提升 3. 11、3. 51、3. 13 个百分点,参数量和浮点运算量分别降低 4. 26%与 7. 9%。

综合来看,提出的基于改进 YOLO 11n 的轻量

级青贮玉米田间人体检测模型在显著提升检测精度的同时,有效控制了模型复杂度,兼顾了检测性能与资源效率,具备良好的实际部署价值。

3. 2 对比实验

为进一步评估改进后模型在青贮玉米田间人体目标检测任务中的性能优势,将所提模型与当前工业领域广泛应用的 YOLO 系列模型进行对比分析。选取的对比模型包括 YOLO v5s、YOLO v8n 以及作为基线模型的 YOLO 11n。为更直观地展现各模型在多维性能指标下的优劣,采用雷达图对比精确率、召回率、mAP@ 0. 5、参数量与浮点运算量(FLOPs)等指标,其中精确率、召回率和 mAP@ 0. 5 为越大越优的指标,进行了正向归一化处理;参数量与浮点运算量为越小越优的指标,采用反向归一化处理。各项指标的具体数值如表 3 所示,由表 3 可见,YOLO 11n 模型在红外数据集上分别比上述模型提升了 7. 21、8. 49、4. 01 个百分点,在可见光数据集上分别提升了 5. 60、6. 27、3. 11 个百分点。

表 3 不同模型对比实验结果

Tab.3 Comparison of experimental results for different models

模型	数据集	P/%	R/%	mAP@ 0. 5/%	参数量	浮点运算量
YOLO v5s	红外	82. 63	81. 36	86. 30	7. 10 × 10 ⁶	2. 18 × 10 ¹¹
	可见光	81. 36	82. 54	84. 65		
YOLO v8n	红外	81. 35	80. 68	85. 69	3. 00 × 10 ⁶	8. 10 × 10 ⁹
	可见光	80. 69	81. 78	83. 26		
YOLO 11n	红外	85. 83	84. 67	87. 20	2. 58 × 10 ⁶	6. 30 × 10 ⁹
	可见光	83. 85	82. 84	85. 19		
本文模型	红外	89. 84	88. 39	91. 68	2. 47 × 10 ⁶	5. 80 × 10 ⁹
	可见光	86. 96	86. 35	88. 32		

相比于 YOLO v5s 的简单 SPPF 结构和 YOLO v8n 的基础 C2f 模块,改进模型的 C3k2_TBFE 模块更适配红外图像的温度特征分布,而 C2PSA_biformer 注意力机制可有效过滤玉米冠层的热辐射干扰,所以从图 9 可见,所提出的改进模型在 2 个数据集上均展现出优越的检测性能,精确率、召回率及

mAP@0.5 等关键指标均处于领先水平。同时,由于 LWGA 模块以“通道重标定”替代传统池化,在保留全局感知能力的同时减少 37% 的卷积操作,更适配农机边缘设备的算力限制,因此模型在轻量化设计与计算效率方面亦表现出良好的控制能力,充分实现了检测精度与实时性之间的有效平衡。

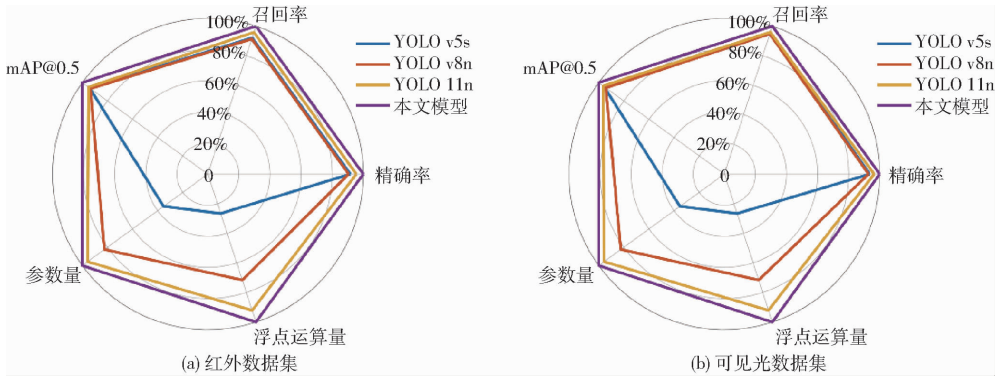


图 9 各模型性能指标雷达图

Fig.9 Radar chart of performance metrics for each model

为直观对比不同模型在青贮玉米田间人体目标检测任务中的性能,图 10 展示了模型与其他对比模型在典型人体目标场景中的检测效果,表 4 给出了各模型对待检图像进行推理后的置信度对照情况。由表 4 可见,模型在对典型人体目标的检测中取得

了最高的置信度且没有出现漏检和框重叠,进一步体现了所提改进模型在青贮玉米田间复杂环境下人体目标检测任务中的优越性能。

通过对比分析可知,模型在青贮玉米田间人体目标检测中表现出色,具备较强的鲁棒性与检测精

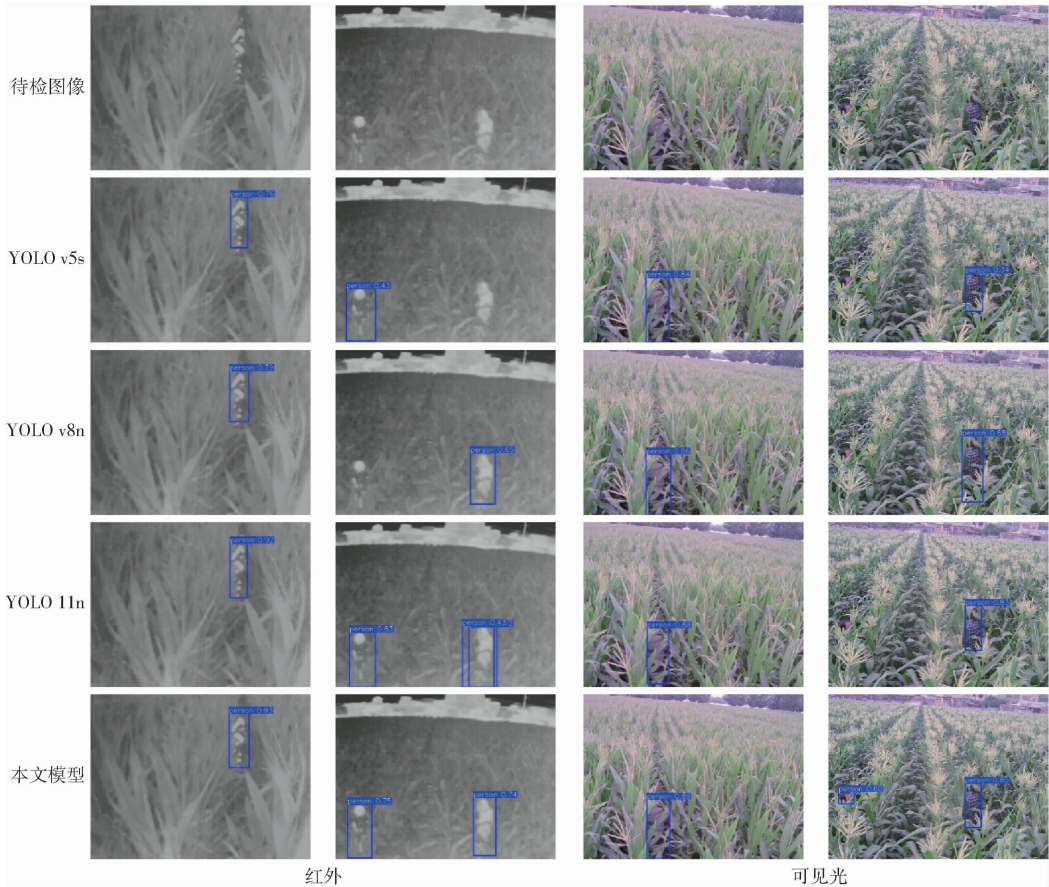


图 10 各模型检测效果

Fig.10 Detection results of different models

表 4 各模型对青贮玉米田间人体识别置信度

Tab.4 Confidence comparison of different models for cornfield personnel detection

模型	红外置信度			可见光置信度		
	单目标	多目标		单目标	多目标	
		左侧	右侧		左侧	右侧
		目标	目标		目标	目标
YOLO v5s	0.76	0.43	(漏检)	0.84	(漏检)	0.34
YOLO v8n	0.79	(漏检)	0.65	0.86	(漏检)	0.68
YOLO 11n	0.92	0.83	(框重叠)	0.88	(漏检)	0.83
本文模型	0.93	0.75	0.74	0.89	0.60	0.85

度。未来结合边缘设备部署与实地作业需求,模型具有良好的实际应用前景,能够为农业自动化作业中的人体识别与安全保障提供有力支持。

4 检测系统设计

4.1 检测系统组成

为保障青贮玉米机作业过程中田间作业人员的安全,降低因作物遮挡或视野盲区引发的碰撞风险,设计并搭建了一套基于红外与可见光图像融

合的青贮玉米田间人体视觉检测与自动报警系统。系统主要包括数据采集单元、数据处理单元与数据响应单元。其中,数据采集单元采用鑫思特 HT-S320 型红外热成像摄像头,该摄像头集成了红外热成像摄像头与可见光摄像头,能够同时获取 2 种模态图像信息;数据处理单元选用 NVIDIA Jetson Orin Nano(8 GB)嵌入式计算平台,用于执行图像处理、目标检测与距离估算等任务;数据响应单元则采用声光报警器,用于在识别出人体目标后发出警示信号。

4.2 系统工作原理

系统具备图像采集、目标检测与距离估算功能。红外与可见光图像分别输入改进型 YOLO 11n 模型进行人体目标识别,并采用逻辑或机制融合判定结果:只要任一通道检测到人体,即视为存在潜在风险。结合摄像头安装高度、俯视角及目标在图像中的位置,系统基于几何模型估算人体与农机之间的水平距离。仅当识别到人体目标且其距离处于预设范围内时,系统会立即触发声光报警,其工作原理如图 11 所示。

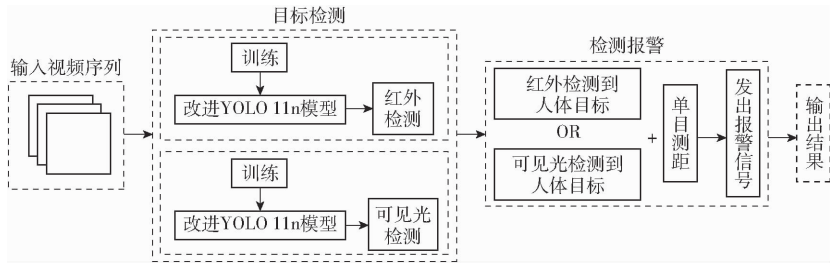


图 11 系统工作原理图

Fig. 11 System working principle diagram

4.3 田间应用效果

为验证基于红外与可见光融合的青贮玉米田间人体检测系统在实际收获作业中的可行性与稳定性,于 2025 年 9 月 16 日在北京市怀柔区梨园庄村(40.331244°N, 116.735240°E)开展田间应用实验,如图 12 所示。

实验田种植青贮玉米,行距约 53 cm,株距约 23 cm,植株高度约 2.7 m,田间种植密度与植株形态与前期数据采集场景基本一致。实验当天天气晴朗,环境温度约 25 ℃,光照条件良好,实验时段覆盖上午和下午两个作业时段,以模拟实际收获作业中不同光照条件。

实验将红外热成像摄像头与可见光摄像头集成安装于青贮玉米收获机驾驶舱顶部,摄像头离地高度约 3.8 m,靠近驾驶员视野中心线,声光报警器安装于驾驶舱左上方。数据处理单元采用 NVIDIA Jetson Orin Nano(8 GB)嵌入式平台,负责图像处理、目标检测与距离估算。实验前完成双摄像头空间配

准,基于仿射变换模型和 RANSAC 算法求解红外与可见光图像之间的几何映射关系,确保同一空间位置的目标在两幅图像中实现对齐。

实验过程中,收获机以低速沿田间作业路线行进,测试人员在作业区域内以站立、行走、下蹲、正面朝向和背向等多种姿态随机出现,模拟实际田间测试人员误入作业区域的情形。系统分别对红外图像和可见光图像进行独立检测,并采用逻辑或决策融合规则判定是否存在人体目标。当任一通道检测到人体目标且其距离处于预设风险范围时,系统立即触发声光报警。试验同步记录检测结果、报警响应时间及系统运行状态,以评估系统在实际田间环境中的检测精度、鲁棒性与实时性。

田间应用结果表明,该系统在复杂田间环境下能够稳定运行并有效识别作业区域内的人体目标。改进型 YOLO 11n 模型在红外和可见光数据集上的 mAP@ 50 分别达到 91.68% 和 88.32%,较原始 YOLO 11n 模型分别提升 4.48、3.13 个百分点;精确

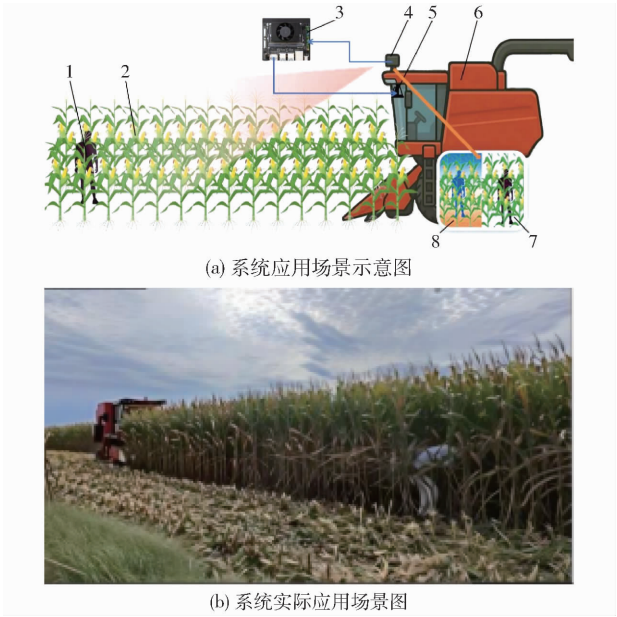


图 12 系统实际应用场景示意图

Fig. 12 System actual application scenario diagram

1. 待检测人体目标 2. 青贮玉米植株 3. NVIDIA Jetson Orin Nano 4. 鑫思特 HT-S320 型红外热成像摄像头 5. 声光报警器 6. 青贮玉米收获机 7. 可见光摄像头所获取的视频数据 8. 红外摄像头所获取的视频数据

率分别达到 89.84% 和 86.96%，召回率分别达到 88.39% 和 86.35%。与 YOLO v5s、YOLO v8n 和 YOLO 11n 相比，本文模型在红外数据集上的精确率分别提升 7.21、8.49、4.01 个百分点，在可见光数据集上分别提升 5.60、6.27、3.11 个百分点。同时，模型参数量和浮点运算量较原始 YOLO 11n 分别降低 4.26% 和 7.9%，在保证检测精度的同时具备良好的轻量化特性。

5 结论

- (1) 针对当前青贮玉米田间人体检测存在效率低下、误检漏检率高及自动化程度有限等问题，提出一种融合计算机视觉技术的智能检测方法，提升了检测的准确性与作业的自动化水平，提高了青贮玉米田间作业人员的安全性。
- (2) 提出了一种基于红外与可见光融合图像的青贮玉米田间人体识别方法。该方法充分利用可见光图像纹理细节丰富、人体特征显著的优势，同时结

合红外图像对温度差异的敏感感知能力，构建多模态特征互补机制。在青贮玉米田间复杂遮挡背景下，通过融合两种图像的关键信息，有效克服单一模态图像在光照变化、植株遮挡等场景下的识别局限，显著提升对田间人体的识别性能。

(3) 提出了一种改进的 YOLO 11n 模型用于青贮玉米田间人体识别。该模型在基线结构的基础上，采用轻量化的 LWGA 模块替代原有 SPPF 结构，以提升多尺度特征提取效率，同时减少模型复杂度；在主干网络中引入 C3k2_TBFE 结构与 C2PSA_biformer 注意力机制，增强模型对复杂背景下小目标人员的识别能力；在颈部网络中集成 C3k2_TAB 模块，进一步优化特征融合与信息传递。通过 8 组消融实验验证了各改进模块的有效性及其协同增强效果。与原始 YOLO 11n 模型相比，模型参数量和浮点运算量分别减少了 4.26% 和 7.9%，mAP@50 分别达到 91.68% 和 88.32%，较原始模型分别提升 4.48、3.13 个百分点。

(4) 与主流轻量级目标检测模型 YOLO v5s、YOLO v8n 和 YOLO 11n 相比，所提出的改进 YOLO 11n 模型在青贮玉米田间人体识别任务中具有更高的检测精度。在人体检测精确率方面，本文模型在红外数据集上分别比上述模型提升了 7.21、8.49、4.01 个百分点，在可见光数据集上分别提升了 5.60、6.27、3.11 个百分点。同时，模型较好地平衡了检测性能、实时性与资源效率，具有更强的部署适应性，适用于资源受限的农业作业环境。

(5) 设计了一种基于红外和可见光融合的青贮玉米田间人体视觉检测与自动报警系统。系统结合两类图像的互补特性，构建双通道目标检测结构，采用改进型 YOLO 11n 模型实现对人体目标的独立识别，并通过逻辑融合机制增强检测鲁棒性。同时，集成基于几何模型的单目测距模块，能够在识别人体目标的同时估算其与农机之间的相对距离，并在安全阈值范围内触发声光报警，提升田间作业的安全保障能力。系统结构完整、功能明确，为青贮玉米收获机作业中人员安全监测提供了一种可行的设计方案，具有良好的工程应用潜力。

参 考 文 献

[1] 李庆慧. 玉米收获机事故原因分析及监管工作要点[J]. 农机质量与监督, 2024(3):29,41.
Li Qinghui. Analysis of the causes of corn harvester accidents and key points for regulatory work[J]. Quality and Supervision of Agricultural Machinery, 2024(3):29,41. (in Chinese)

[2] 付瀚祥. 基于深度学习的红外与可见光图像融合研究[D]. 广汉:中国民用航空飞行学院,2025.
Fu Hanxiang. Research on infrared and visible image fusion based on deep learning[D]. Guanghan: Civil Aviation Flight University of China,2025. (in Chinese)

[3] 王琦, 吕海涛. 浅析玉米收割机伤人事故及防范措施[J]. 农业开发与装备, 2015(1):95-96.
Wang Qi, Lü Haitao. A brief analysis of corn harvester injuries and preventive measures[J]. Agricultural Development &

- Equipments, 2015(1):95–96. (in Chinese)
- [4] 宋利杰. 基于热像仪的玉米收割机前方人员入侵检测系统设计[D]. 杨凌:西北农林科技大学,2024.
Song Lijie. Design of intrusion detection system based on thermal camera for people in front of corn harvesters[D]. Yangling: Northwest A&F University,2024. (in Chinese)
- [5] 孙婧, 王志社, 杨帆, 等. 红外与可见光图像多层感知机交互融合方法[J]. 红外技术,2025, 47(5):619–627.
Sun Jing, Wang Zhishe, Yang Fan, et al. Multi-layer perceptron interactive fusion method for infrared and visible images[J]. Infrared Technology,2025,47(5):619–627. (in Chinese)
- [6] 李钰瑾, 邢卉, 石英琦, 等. 基于红外与可见光图像融合的电缆火情检测研究[J]. 家电维修,2025(6):134–136.
Li Yujing, Xing Hui, Shi Yingqi, et al. Study on cable fire detection using fusion of infrared and visible images[J]. Home Appliance Maintenance,2025(6):134–136. (in Chinese)
- [7] 南晶晶, 潘红光, 蒋泽, 等. 基于低光照增强的红外和可见光图像融合的井下人员识别[J]. 工矿自动化,2025,51(4):107–113,145.
Nan Jingjing, Pan Hongguang, Jiang Ze, et al. Underground personnel recognition based on low-light enhancement of infrared and visible light image fusion[J]. Industry and Automation, 2025, 51(4):107–113,145. (in Chinese)
- [8] 官敬超, 刘家辉, 赵庆辉, 等. RFC-YOLO v8:可见光和红外图像自适应融合的目标检测[J]. 电光与控制,2025, 32(9):1–9.
Guan Jingchao, Liu Jiahui, Zhao Qinghui, et al. RFC-YOLO v8: adaptive fusion of visible and infrared images for object detection[J]. Electro-Optics and Control,2025,32(9):1–9. (in Chinese)
- [9] 张天宇, 韩静, 廖洪晖, 等. 基于优化 YOLO v5 算法的玉米苗间杂草检测研究[J]. 中国农机化学报,2025, 46(7):220–225.
Zhang Tianyu, Han Jing, Liao Honghui, et al. Research on weed detection among maize seedlings based on optimized YOLO v5 algorithm[J]. Journal of Chinese Agricultural Mechanization,2025,46(7):220–225. (in Chinese)
- [10] 田文斌, 姚渝, 吕昊曦, 等. 基于改进 YOLO v11n-seg 的奶牛乳房炎检测方法[J]. 农业机械学报,2025, 56(6):237–246.
Tian Wenbin, Yao Yu, Lü Haotun, et al. Detection method for cow mastitis based on improved YOLO v11n-seg[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(6):237–246. (in Chinese)
- [11] 朱自文, 宋晓鸥, 崔巍, 等. 可见光-红外图像融合的目标检测综述[J]. 计算机工程与应用,2025, 61(17):1–26.
Zhu Ziwen, Song Xiaou, Cui Wei, et al. Visible and infrared image fusion for intelligent object detection: a review[J]. Computer Engineering and Applications, 2025, 61(17):1–26. (in Chinese)
- [12] 杨华, 林卉. 数据融合的研究综述[J]. 矿山测量,2005(3):24–28.
Yang Hua, Lin Hui. A research survey on data fusion[J]. Mine Surveying, 2005(3):24–28. (in Chinese)
- [13] 李霖, 王红梅, 李辰凯. 红外与可见光图像深度学习融合方法综述[J]. 红外与激光工程,2022,51(12):337–356.
Li Lin, Wang Hongmei, Li Chenkai. A review of deep learning fusion methods for infrared and visible images[J]. Infrared and Laser Engineering,2022,51(12):337–356. (in Chinese)
- [14] 高训东, 陈辉, 姚亚宁, 等. 基于双分支特征分解的红外与可见光融合网络[J]. 激光与光电子学进展,2025(14):482–492.
Gao Xundong, Chen Hui, Yao Yaning, et al. Infrared-visible image fusion network based on dual-branch feature decomposition [J]. Laser & Optoelectronics Progress,2025(14):482–492. (in Chinese)
- [15] Torralba, Russell B C, YUEN J. LabelMe: online image annotation and applications[J]. Proceedings of the IEEE,2010, 98(8):1467–1484.
- [16] Lin T Y, Dollár P, Girshick R B, et al. Feature pyramid networks for object detection[J/OL]. CoRR,2016–12–09. <http://arxiv.org/abs/1612.03144>.
- [17] 李炳臻, 刘克, 顾佼佼, 等. 卷积神经网络研究综述[J]. 计算机时代,2021(4):8–12,17.
Li Bingzhen, Liu Ke, Gu Jiaojiao, et al. Review of the researches on convolutional neural networks[J]. Computer Era, 2021(4):8–12,17. (in Chinese)
- [18] 倪天宇. 晶圆表面缺陷自动检测技术的研究[D]. 上海:东华大学,2022.
Ni Tianyu. Resear on automatic detection technology of wafer surface dffects[D]. Shanghai: Donghua University,2022. (in Chinese)
- [19] Howard A, Sandler M, Chu G, et al. Searching for MobileNetV3[J/OL]. CoRR,2019–05–06. <http://arxiv.org/abs/1905.02244>.
- [20] Zhang X, Zhou X, Lin M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[J/OL]. CoRR,2017–07–04. <https://arxiv.org/abs/1707.01083>.
- [21] Lu W, Chen S B, Ding C H, et al. Lwganet: a lightweight group attention backbone for remote sensing visual tasks[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2026: 7574–7582.
- [22] Chen P, He W, Qian F, et al. A synergistic CNN-transformer network with pooling attention fusion for hyperspectral image classification[J]. Digital Signal Processing, 2025, 160: 105070.
- [23] Zhu L, Wang X, Ke Z, et al. Biformer: vision transformer with bi-level routing attention[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE,2023.
- [24] Rahmzadehgervi P, Nguyen H H, Liu R, et al. TAB: transformer attention bottlenecks enable user intervention and debugging in vision-language models [J]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Honolulu: IEEE, 2025: 22551–22562.
- [25] 彭珍瑞, 朱亚斌, 王玉龙, 等. 基于改进 YOLO11s 的机车圆簧表面裂纹检测[J]. 铁道学报, 2026, 48(5):1–10.
Peng Zhenrui, Zhu Yabin, Wang Yulong, et al. Surface crack detection of locomotive coil springs based on improved YOLO11s [J]. Journal of the China Railway Society, 2026, 48(5):1–10. (in Chinese)