

基于改进 YOLO v5s 的复杂环境下蔗梢分叉点识别与定位

李尚平 卞俊析 李凯华 任泓宇

(广西民族大学电子信息学院, 南宁 530006)

摘要: 甘蔗蔗梢分叉点的精确识别与高度定位是实现甘蔗收获机切梢器实时控制的关键技术之一,也是提高甘蔗收获机械化水平和降低甘蔗含杂率的重要途径。针对甘蔗地环境复杂、光照变化大、蔗梢分叉点相互遮挡等问题,首先通过田间调查,并现场测试、分析甘蔗生长点、甘蔗分叉点及相互关系的特征规律,采集图像的甘蔗分叉点的统计分析,并结合现场对甘蔗分叉点高度的测量统计分析,发现其均具有明显的正态统计特征。接着,提出了一种基于改进 YOLO v5s 的蔗梢分叉点识别方法。该方法采用单目和双目相机在广西大学扶绥农科基地采集甘蔗图像数据,并进行数据预处理和标注,构建了甘蔗蔗梢分叉点数据集。然后,在 YOLO v5s 中引入 BiFPN 特征融合结构和 CA 注意力机制,以增强不同层次特征的交互和表达能力,并使用 GSConv 卷积和 Slim-Neck 范式设计,在原始模型主干网络中引入 Ghost 模块替换原始普通卷积,来降低模型的计算量和参数量,提高模型的运行效率。最后,通过在现场采集的数据集上进行训练和测试,验证了该方法的有效性和优越性。实验结果表明,该方法在甘蔗蔗梢分叉点数据集上平均精确率达到 92.3%、召回率 89.3% 和检测时间 19.3 ms,相比原始 YOLO v5s 网络,平均精确率提高 5 个百分点,召回率提高 4 个百分点,参数量降低 43%,模型内存占用量减少 5.5 MB,检测时间减少 0.7 ms。最后,根据甘蔗分叉点具有明显的正态统计特征的规律,利用该特征结合双目视觉的定位算法,可为开展甘蔗收获机切梢的特征识别、切梢器高度定位及实时控制研究奠定理论及技术基础。

关键词: 蔗梢分叉点; YOLO v5s; 生长高度统计规律; 目标识别

中图分类号: S24 文献标识码: A 文章编号: 1000-1298(2023)11-0247-12 OSID: 

Identification and Height Localization of Sugarcane Tip Bifurcation Points in Complex Environments Based on Improved YOLO v5s

LI Shangping BIAN Junxi LI Kaihua REN Hongyu

(School of Electronic Information, Guangxi Minzu University, Nanning 530006, China)

Abstract: The precise identification and height positioning of the bifurcation points of sugarcane tips is one of the key technologies for achieving real-time control of sugarcane harvester cutters, and is also an important way to improve the mechanization level of sugarcane harvesting and reduce sugarcane impurity content. In response to the complex environment of sugarcane fields, significant changes in lighting, and mutual obstruction of sugarcane bifurcation points, the field investigations, on-site testing and analysis of the characteristics of sugarcane growth points, sugarcane bifurcation points, and their interrelationships were firstly conducted, statistical analysis of sugarcane bifurcation points in images was collected, and combined with on-site measurement and statistical analysis of the height of sugarcane bifurcation points, it was found that they all had obvious normal statistical characteristics. Secondly, a sugarcane tip bifurcation point recognition method was proposed based on improved YOLO v5s. In this method, monocular and binocular cameras were used to collect sugarcane image data in Fusui Agricultural Science Base of Guangxi University, and data preprocessing and labeling were carried out to build a data set of sugarcane tip bifurcation points. Then BiFPN feature fusion structure and CA attention mechanism were introduced into the backbone network of YOLO v5s to enhance the interaction and expression ability of different levels of features, and using GSConv convolution, Slim-Neck normal form design, and the

收稿日期: 2023-07-20 修回日期: 2023-09-04

基金项目: 广西民族大学科研项目(302210506)和广西创新发展重大项目(桂科 AA22117006)

作者简介: 李尚平(1956—),男,教授,博士生导师,主要从事农业机械化工程研究,E-mail: spli501@vip.sina.com

通信作者: 李凯华(1985—),男,实验师,主要从事农业机械自动化研究,E-mail: 78281062@qq.com

Ghost module was introduced into the original model backbone network to replace the original ordinary convolution in Neck, in order to reduce the computational and parameter complexity of the model and improve its operational efficiency. Finally, the effectiveness and superiority of this method were verified through training and testing on on-site collected datasets. The experimental results showed that this method achieved an average accuracy of 92.3%, a recall rate of 89.3%, and a detection time of 19.3 ms on the sugarcane tip bifurcation point dataset. Compared with the original YOLO v5s network, the average accuracy was improved by 5 percentage points, the recall rate was improved by 4 percentage points, the parameter quantity was reduced by 43%, the model size was reduced by 5.5 MB, and the detection time was reduced by 0.7 ms. Finally, based on the obvious normal statistical characteristics of sugarcane bifurcation points, this feature can be combined with binocular vision positioning algorithms to lay a theoretical and technical foundation for conducting research on feature recognition of sugarcane harvester cuttings, height positioning of cuttings, and real-time control.

Key words: sugarcane tip bifurcation points; YOLO v5s; statistical law of sugarcane branching point growth; image feature recognition

0 引言

甘蔗是一种重要的经济作物,广泛应用于食品、饮料、医药等领域。广西是我国最主要的甘蔗种植区,2020 年广西甘蔗种植占我国甘蔗总种植面积的 64.7%,然而广西甘蔗机械化收获的普及情况严重滞后,虽然广西综合机械化水平达到了 60%,但是甘蔗机器收割率仍然停留在很低的水平,甘蔗收获机械化已经成为制约广西甘蔗全程机械化推广的最大短板^[1]。机械收割的甘蔗含杂率高、糖分损失过快等因素导致糖厂限制机收蔗进厂数量。因此降低甘蔗收获的含杂率和损失率是推进甘蔗收获机械化的关键问题。根据统计表明,甘蔗收获机机收的含杂率中,蔗梢的含杂率为主要因素,约为 50%,根据农艺要求,甘蔗切梢的最佳位置为甘蔗生长点。

然而,甘蔗生长点通常被蔗叶包裹,人眼无法直接识别,只有通过拨开外层的蔗叶才能确定生长点位置。经过大量田间调研测量,发现甘蔗生长点位置在蔗梢分叉点下方 35~40 cm 处的位置。因此本文拟通过识别蔗梢分叉点再结合双目测距算法,通过蔗梢分叉点的高度,间接确定生长点高度。目前,深度学习技术已经在目标检测领域取得了显著的成果,尤其是基于卷积神经网络(CNN)的检测算法^[2],如 Faster R-CNN^[3-4]、YOLO^[5]、SSD^[6]等,在各种数据集上都表现出了优异的性能。国内外在甘蔗智能方面做了大量研究。文献[7]使用基于深度学习的卷积神经网络对田间甘蔗叶片病害进行识别;文献[8]基于物联网和机器学习实现了甘蔗螟虫虫害的预测,李尚平等^[9]首次将基于深度学习的目标检测技术运用在甘蔗茎节识别上,其通过改进 YOLO v3 网络,建立了一个智能蔗节识别模型,并应用到切种机内部的摄像头。文献[10]通过 YOLO v4 和支持向量机实现了甘蔗幼苗的识别与定位。

然而在复杂环境下农作物目标检测仍然面临着诸多挑战,如目标尺寸变化大、形态多样、相互遮挡、光照不均等问题。针对这些问题,国内外学者进行了一些相关研究。文献[11]针对复杂田间作业环境下自动提取图像特征,快速准确检测不同生育期玉米苗期;文献[12]将密集的架构整合、空间金字塔池化和 MIou 函数应用到修改后 YOLO v3 模型,改进后的模型在枝叶遮挡、光照变化、番茄果实相互重叠等干扰因素的情况下,模型对番茄平均预测率达到了 98.5%。

本文提出一种基于改进 YOLO v5s 的目标检测方法,针对复杂田间环境,对蔗梢分叉点的检测进行优化。在广西大学扶绥农科基地采集甘蔗图像数据,并进行数据预处理和标注,构建甘蔗蔗梢分叉点数据集。通过在该数据集上进行训练和测试,验证该方法的有效性和优越性。

1 甘蔗图像获取与预处理

1.1 图像采集方法

蔗梢图像数据在广西大学扶绥农科基地拍摄采样获得,甘蔗品种为中蔗 44 号。拍摄时间分别为早上、中午和傍晚。本文模拟收获机切刀位置固定摄像机高度,分别在甘蔗地拍摄 12 段视频,通过对视频进行分割,共采集 1 500 幅包含蔗梢分叉点的图像。图 1 和图 2 为不同形态和光照下蔗梢分叉点图像。

1.2 图像数据预处理

为了提高训练集的图像质量,更好地提取蔗梢分叉点在不同天气和时间条件下的特征,本文对训练集图像进行了数据增强。考虑到光照和天气等因素的不确定性^[13],以及田间环境的复杂性,为了增强蔗梢分叉点模型的鲁棒性和泛化能力,本文对训练集图像分别采用亮度调整、色度调整和模糊处理

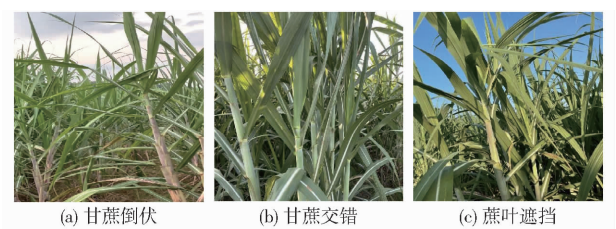


图 1 不同形态蔗梢分叉点图像

Fig. 1 Photos of branching points in different forms of sugarcane tips



图 2 不同光照下蔗梢分叉点图像

Fig. 2 Photos of sugarcane tip bifurcation points under different lighting conditions

的数据增强方法。亮度和色度调整可以模拟不同环境下的光照情况,模糊处理可以模拟设备晃动或机器抖动造成的图像模糊。然后,通过 Labelimg 可视化标注工具对图像进行手工标记,按 7:1:2 比例划分为训练集、验证集和测试集。数据集划分结果如表 1 所示。

表 1 数据集划分
Tab.1 Dataset division

处理方式	数据集		
	训练集	验证集	测试集
原始数据集	675	150	300
亮度与色度调整	1 350	300	
高斯模糊	675	150	

使用原始 YOLO v5s,在不同的数据集训练 150 轮得到最优的数据模型后,在测试集上检测模型性能,其结果如表 2 所示。

表 2 不同训练数据所得模型测试集性能评估结果
Tab.2 Performance evaluation results of model test sets obtained from different training data %

数据集类型	准确率	召回率	平均精确率
原始数据集	83.9	85.1	86.7
原始数据集 + 亮度与色度调整	83.2	88.1	89.8
原始数据集 + 模糊处理	79.3	82.1	83.5
原始数据集 + 亮度与色度调整 + 模糊处理	83.0	85.3	87.3

由表 2 可知,在原始数据集上增加亮度和色度调整后,准确率下降 0.7 个百分点,召回率上升 3 个百分点,平均精确率上升 3.1 个百分点。而在原始数据集上增加模糊处理后,准确率、召回率、平均精

确率三者都有降低,准确率降低 4.6 个百分点,召回率降低 3 个百分点,平均精确率降低 3.2 个百分点。在原始数据集上,同时增加亮度、色度调整和模糊处理后,召回率和平均精确率都有所上升,前者升高 0.2 个百分点,后者升高 0.6 个百分点,但准确率降低 0.9 个百分点。

为了分析蔗梢分叉点的高度分布特征,本文对标注的边界框数据进行了统计。由于采用单目摄像机模拟收获机切梢器的切割高度拍摄图像,将摄像机固定在一定高度进行拍摄,因此可以通过边界框坐标中心 y 的值反映蔗梢分叉点的相对高度。本文对边界框坐标中心 y 的值进行归一化处理,并绘制了直方图(图 3)。该直方图呈现一个正态分布的钟形曲线,该图两侧的尾部会逐渐变低,而中心部分则比较平缓,这说明数据呈现出明显的正态分布特征。

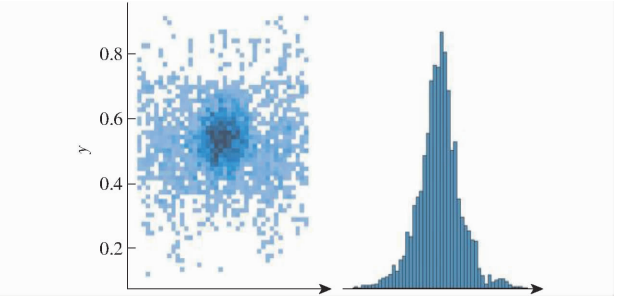


图 3 蔗梢分叉点边界框坐标中心 y 的统计直方图
Fig. 3 Statistical histogram of coordinate center y of boundary box at sugarcane tip bifurcation points

进一步验证蔗梢分叉点高度的分布规律,本文使用随机抽样的方法,在广西大学扶绥农科基地开展了现场测量统计实验,甘蔗品种为中蔗 44 号,随机选取 150 组,测量甘蔗分叉点与地面的高度,并生成如图 4 所示的蔗梢分叉点高度正态分布图。

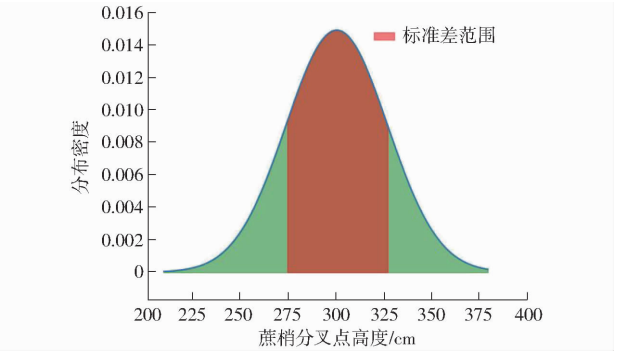


图 4 蔗梢分叉点高度正态分布图
Fig. 4 Normal distribution diagram of height of sugarcane tip bifurcation points

从图 4 中可以看出蔗梢分叉点高度存在明显的正态分布特征。蔗梢分叉点峰值约为 300 cm。150 组蔗梢分叉点高度平均值约为 300.9 cm。标准差约为 26.65 cm。图中橙色部分显示的区域为 1 个标

准差范围内的数据,约有 74% 的数据分布在左右 1 个标准差范围内,与正态分布 68% 的理论值较为符合。

图 5(红色方框为切梢器)所示是常用的整杆式甘蔗收获机,收获机高度调节需要人工操作液压系统来控制切梢器的升降,依靠人工分辨判断,其实时响应性较差,如果蔗梢分叉点高度变化幅度较大,收获机常常无法有足够的响应时间。因此在后续的蔗梢分叉点动态识别定位模型中,本文拟利用图像识别定位算法,并结合甘蔗分叉点高度的正态分布规律进行高度定位。通过设定合理的切刀初始高度和变化范围,结合蔗梢分叉点的生长规律的数据统计分析,来确定切梢器的高度定位和调整范围,实现对收获机切梢器的实时控制。考虑到甘蔗生长密集且混乱,切梢器无法以每个蔗梢生长点位置变化做实时调整。因此本文提出了一种高度定位优化算法,该算法流程如图 6 所示。



图 5 常用的整杆式甘蔗收获机

Fig. 5 Commonly used whole rod sugarcane harvesters

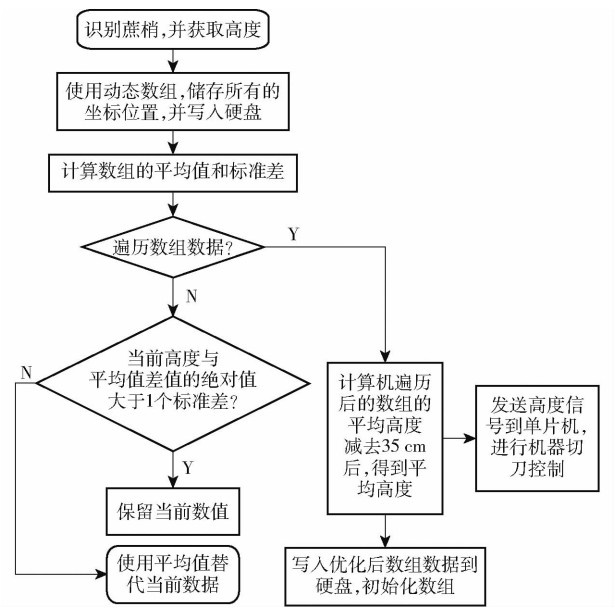


图 6 蔗梢分叉点高度定位优化算法

Fig. 6 Optimization algorithm for height positioning of sugarcane tip bifurcation points

该算法首先使用动态数组来存储每帧检测到的蔗梢分叉点的高度。接下来,求出数组的平均值和

标准差,并遍历数组进行判断。如果检测到的蔗梢分叉点高度与平均值的差的绝对值大于 1 个标准差,将用平均值替代该高度;否则,保留原始高度。最后,计算数组的平均值作为蔗梢分叉点平均高度,并从中减去蔗梢分叉点与生长点之间的平均差值 35 cm。最终,将得到的高度发送到单片机。如图 7 所示,可以看到检测到的 3 个蔗梢分叉点高度。原始数据标准差为 38 cm,平均值约为 258.7 cm。在这组数据中,高度 297 cm 与平均值差的绝对值大于标准差,因此将其替换为平均值。而高度 221 cm 和 258 cm 小于标准差,可以保留不变。经过本文高度算法优化,标准差缩小为 21.56 cm。

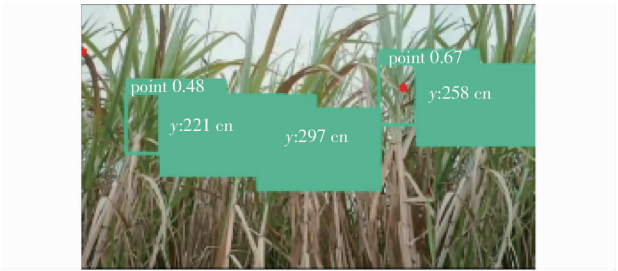


图 7 双目识别和高度定位效果图

Fig. 7 Effect diagram of binocular recognition and height positioning

经过后续的实验测试,该算法可以有效降低极端数据的影响,大幅减少数据的方差,本文进行 150 组数据对比后,采用该算法后数据的平均方差从 45 cm² 降低到 15 cm²。同时获取的蔗梢分叉点高度,也可用作数据集对甘蔗生长数据进行统计和分析。

1.3 验证蔗梢分叉点与生长点的规律

为了进一步验证蔗梢分叉点与生长点之间存在的位置规律关系,本文随机选取了 100 组甘蔗样本。选取的甘蔗品种为中蔗 44 号。针对每一组样本,测量了甘蔗分叉点与生长点的位置关系,并得出了对应的位置关系图。

如图 8 所示,图中蓝色圆点为测量的蔗梢分叉点与生长点的距离差值,红色线段为这些蓝色圆点的连线。经测量这组数据平均值为 34.39 cm。可知蔗梢分叉点到甘蔗生长点平均值为 35 cm 的距离

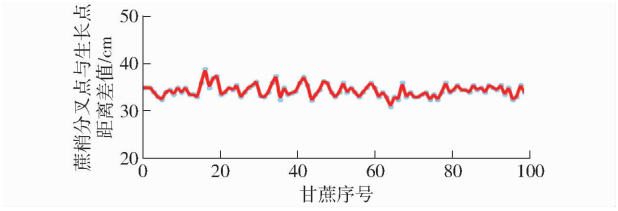


图 8 蔗梢分叉点与生长点距离差值的散点图

Fig. 8 Scatter plot of distance difference between bifurcation point and growth point of sugarcane tips

较为符合。

2 改进 YOLO v5s 网络模型

2.1 YOLO v5 网络架构

YOLO v5 是一种单阶段目标检测算法^[14]，与前代的 YOLO v4 模型相比，YOLO v5 的整体结构没有太大的差异，主要是新增了 Focus 和 SPP 结构。Focus 结构可以将输入图像分成 4 部分，并将它们拼接在一起，从而减少计算量和内存消耗。SPP 结构可以通过多尺度池化来增加感受野和特

征多样性。考虑到模型轻量化需求，选择最轻量化的 YOLO v5s 模型，其模型主要分为：输入层（Input）、主干网络（Backbone）、多尺度特征融合网络（Neck）、输出端（Prediction）。在 YOLO v5s 中引入 BiFPN 特征融合结构和 CA 注意力机制，以增强不同层次特征的交互和表达能力，使用 GSConv 卷积和 Slim-Neck 范式设计，并在主干网络中引入 Ghost 模块替换原始普通卷积，来降低模型计算量和参数量，提高模型运行效率。本文改进的整体结构图如图 9 所示。

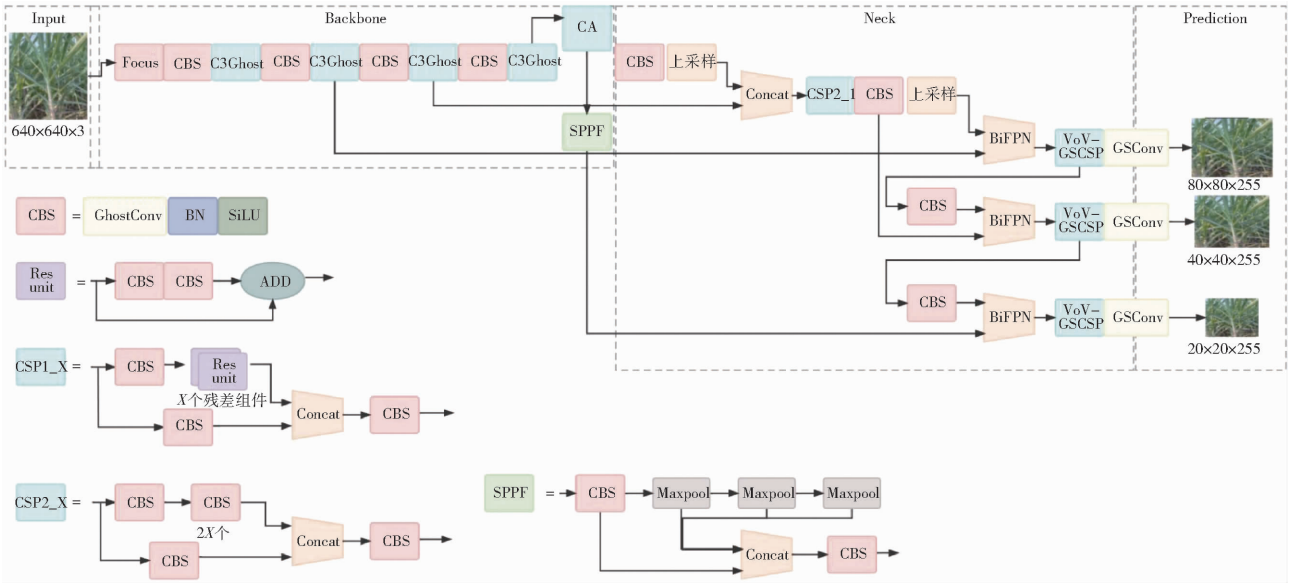


图 9 本文改进 YOLO v5s 网络结构图
Fig. 9 Improved overall network structure diagram of YOLO v5s

2.2 结合 BiFPN

在设计蔗梢分叉点实时检测模型时，要充分考虑到甘蔗地里复杂多变的情况和广西多丘陵的自然环境。因此，在 YOLO v5s 架构的基础上，对其主干网络进行优化，有效提取图像的深层特征，提高其精度以提高识别效率。

本文将 YOLO v5s 中原来的 PaNet 层^[14]改进为 BiFPN 层^[15]，从而实现自上而下与自下而上的深浅层特征双向融合，增强不同网络层之间特征信息的传递^[16]。BiFPN 的内部结构包括上采样、下采样和特征融合 3 个关键组件。上采样和下采样通过使用双线性插值和卷积操作来调整特征的尺寸，使得不同层级的特征具有相同的尺寸。特征融合则通过使用卷积操作和注意力机制来将特征进行融合，从而生成最终的特征金字塔。注意力机制根据特征的重要性动态地调整特征的权重，使得高质量的特征能够更好地传播到其他层级。整个 BiFPN 模块由多个这样的组件组成，形成一个多层级的特征融合网络。

针对本文蔗梢分叉点的识别，采用 BiFPN 结构

具有以下 2 点优势：BiFPN 通过引入跨层连接和特征加权的方式，使得特征金字塔的各个层级能够更好地平衡，提高对不同尺度目标的检测能力，从而提高对蔗梢分叉点小目标的检测效果；BiFPN 网络结构通过自上而下和自下而上的方式，将不同层级的特征图进行融合，可以提供更全局和更具语义的特征表示，有助于在复杂环境中准确地检测蔗梢分叉点。为更有效地体现 BiFPN 结构对网络的贡献，本文使用 Grad-CAM 对图像进行可视化对比^[17]。Grad-CAM 可通过热力图的方式对目标的标识进行可视化显示。将两种网络通过 Grad-CAM 测试后，其对目标识别的热力图如图 10 所示。可以观察到 BiFPN 结构相比原生 YOLO v5s 的 PaNet 结构，对蔗梢分叉点的热力值更高，位置也更为精确，效果更可靠。

2.3 CA 注意力机制

常用的通道注意力可以提高模型性能，但往往忽略位置信息，而位置信息对于生成具有空间选择性的注意力映射非常重要。针对蔗梢分叉点检测任务中存在的环境背景复杂、甘蔗交错严重等问题，以

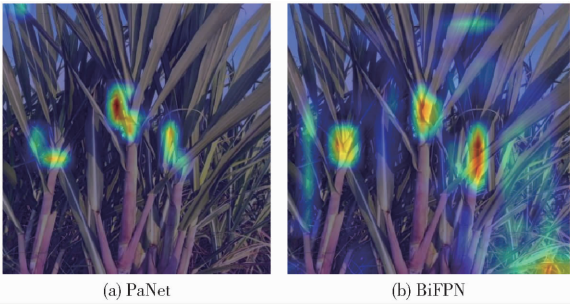


图 10 PaNet 和 BiFPN 热力图对比

Fig. 10 Comparison of PaNet and BiFPN thermal maps

及对甘蔗分叉点位置识别的精确性要求,本文在原有的模型中,在普通卷积之后嵌入了 CA 注意力机制,以增强网络的特征提取能力。该注意力机制通过准确的位置信息对通道关系和长期依赖性进行编

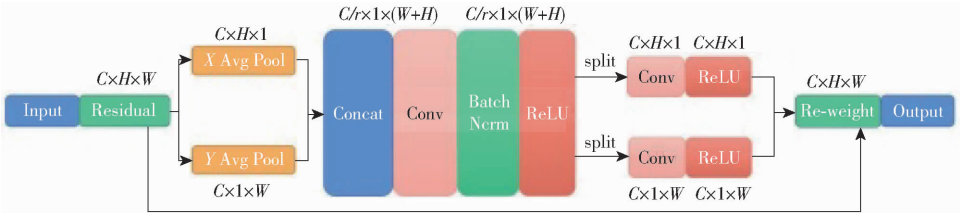


图 11 CA 注意力机制结构图

Fig. 11 CA attention mechanism structure diagram

Coordinate Attention 生成是通过对编码后的特征图进行卷积操作,生成一个注意力图。具体而言,首先使用一个卷积层对编码后的特征图进行降维,以减少计算量。然后,再通过另一个卷积层将降维后的特征图映射为注意力图。这个映射过程实质上是对特征图中的每个位置进行加权处理,其中的权重是根据该位置的坐标信息来计算的。通过这样的注意力图生成过程,CA 注意力机制能够根据图像的空间坐标信息,对不同位置的特征进行加权处理,提高对重要区域的关注度,从而增强模型的性能和鲁棒性。

Coordinate 信息嵌入将空间坐标信息与原始特征图融合,生成一个新的特征图;Coordinate Attention 生成通过卷积操作,根据坐标信息生成一个注意力图,用于加权处理特征图中的不同位置。这两个步骤共同构成了 CA 注意力机制。

2.4 融入 GSConv 和 Slim - Neck

GSConv 和范式设计 Slim - Neck 在无人驾驶领域有非常好的表现^[19]。无人驾驶领域中对模型的识别速度和准确率都有很高的要求。考虑本文对蔗梢分叉点的检测,后期需要应用在实时收割中,同样需要对检测速度和准确率有着较高的要求。所以本文将这种 GSConv 和范式设计 Slim - Neck 融入到 YOLO v5s 的网络模型中。该方法提供了一种实现检测器更高计算成本效益的方法,既减轻了模型复

码,使网络在检测过程中更加关注蔗梢分叉点的特征目标,同时抑制背景信息的干扰。

CA (Coordinate Attention) 通过精确的位置信息对通道关系和长期依赖性进行编码,如图 11 所示,具体操作分为 Coordinate 信息嵌入和 Coordinate Attention 生成 2 个步骤^[18]。Coordinate 信息嵌入是通过将输入特征图与空间坐标信息进行融合,为每个像素位置分配一个唯一的坐标编码。具体而言,对于一个输入特征图大小为 $H \times W \times C$,其中 H 和 W 分别表示高度和宽度, C 表示通道数,可以将每个像素位置的坐标值 (x, y) 编码为一个二维向量 $(x/W, y/H)$ 。然后,将该编码向量与原始特征图进行拼接,得到一个新的特征图,其大小为 $H \times W \times (C + 2)$ 。

杂度,又保持了精度。同时 GSConv 网络可以更好地平衡模型准确性和速度。GSConv 的网络结构图如图 12 所示。

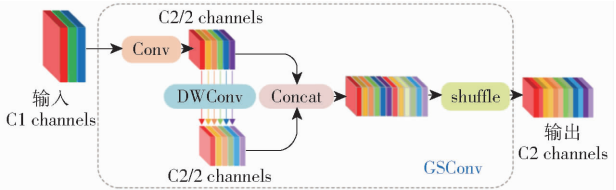


图 12 GSConv 网络结构图

Fig. 12 GSConv network structure diagram

GSConv 模块是一种用于目标检测任务的卷积模块,通过引入全局上下文信息来提高模型的感受野大小和语义理解能力。它的工作原理是将全局上下文特征图与输入特征图相加,然后进行一系列卷积操作,提取更丰富的语义信息。

Slim - Neck 模块是一种用于目标检测任务的特征融合模块,通过减少网络的参数量和计算量来提高模型的速度和效率。它的工作原理是将低维特征图与输入特征图相加,然后进行一系列卷积操作,提取更丰富的语义信息。

这两个模块分别嵌入到 YOLO v5 的 Neck 网络的不同部分。GSConv 模块嵌入到 C3 和 C4 阶段,用于提取更丰富的语义信息。而 Slim - Neck 模块嵌入到 C5 阶段,用于进行特征融合和降维,提高模型的速度和效率^[20]。

范式设计 Slim-Neck 如图 13 所示。该设计首先使用轻量级卷积方法 GSConv 来代替标注卷积网络^[21]。其计算成本为标准卷积的 60%~70%,但其对模型学习能力的贡献与后者不相上下。然后,在 GSConv 的基础上继续引入 GSbottleneck,图 13a 展示了 GSbottleneck 模块的结构。

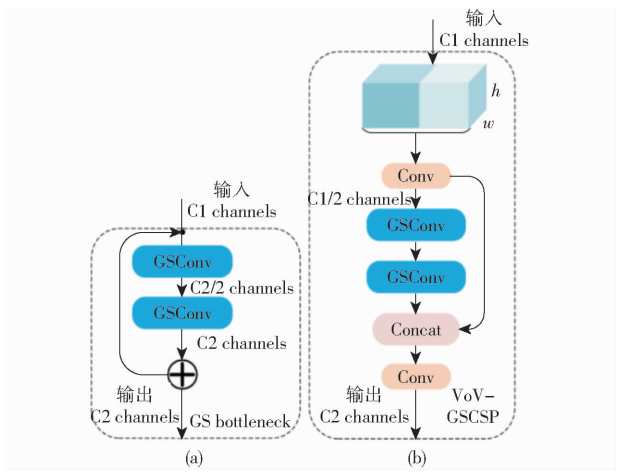


图 13 范式设计 Slim-Neck
Fig. 13 Paradigm design Slim-Neck

同样,使用一次性聚合方法来设计跨级部分网络(GSCSP)模块 VoV-GSCSP。VoV-GSCSP 模块降低了计算和网络结构的复杂性,但保持了足够的精度。图 13b 为 VoV-GSCSP 的结构。此外,如果使用 VoV-GSCSP 代替 Neck 的 CSP,其中 CSP 层由标准卷积组成,FLOPS(浮点运算次数)将比后者平均值减少 15.72%。

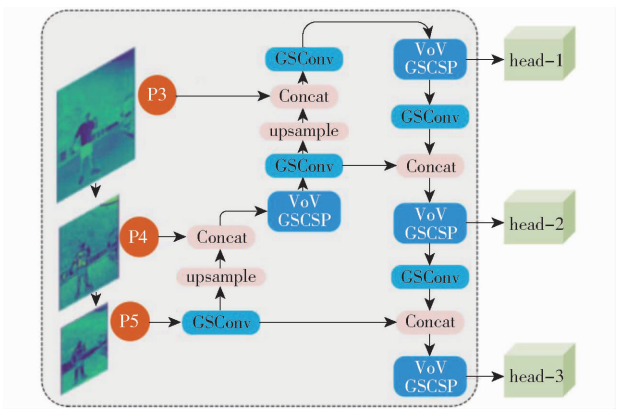


图 14 Slim-Neck 嵌入 YOLO v5 结构图
Fig. 14 Structure diagram of Slim-Neck embedded YOLO v5

YOLO 系列检测器由于检测效率高,在行业中应用更为广泛^[21-22]。本文使用 Slim-Neck 模块来嵌入到 YOLO v5 的 Neck 网络中(图 14)。

2.5 融入 Ghost 模块

为了进一步压缩 YOLO v5s 模型的大小,方便后期模型能够部署在计算资源有限的硬件设备上。本文在原始模型的主干网络中引入 Ghost 模块替换

原始普通卷积。

Ghost 模块是一种用于减少模型计算量和参数数量的模块。如图 15 所示,它的工作原理是通过将输入特征图分成多个子特征图,并使用不同的权重来处理它们,然后将它们合并成输出特征图。这种方式可以减少计算量和参数数量,同时保持模型的性能。在 YOLO v5s 主干网络中应用 Ghost 模块有 3 个优点:①减少计算量:Ghost 模块可以将输入特征图分成多个子特征图,并使用不同的权重处理它们,从而减少计算量,提高模型的推理速度。②减少参数数量:Ghost 模块通过共享部分权重来处理不同的子特征图,从而减少了参数数量,减少模型的存储空间和计算复杂度。③保持模型性能:Ghost 模块通过使用不同的权重处理特征图并将它们合并,Ghost 模块可以保持模型的表达能力和感受视野,从而保持模型的性能。

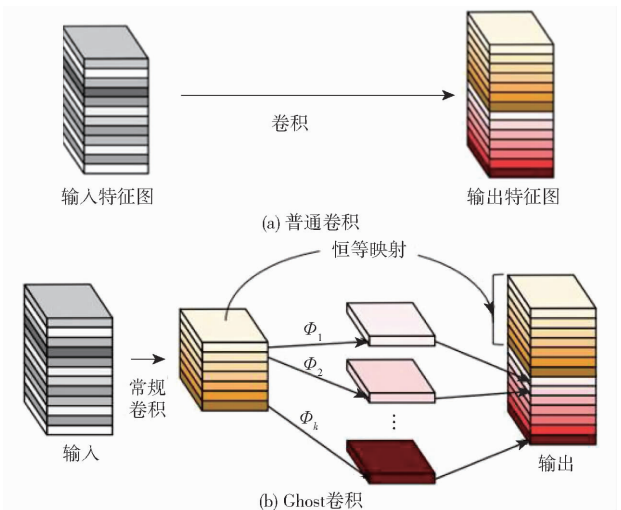


图 15 普通卷积和 Ghost 卷积
Fig. 15 Ordinary convolutions and Ghost convolutions

3 结果与分析

3.1 训练平台

本文训练基于联想 R9000X 型计算机搭载 R7-5800H 处理器,16 GB 内存,3060 移动端显卡,在 Windows 11 操作系统下搭建 Pytorch 深度学习框架,使用 Python 编写程序并调用 CUDA 和 OpenCV 等所需的库,实现本文模型的训练和测试。模型训练时 8 个样本作为一个批处理单元(Batch size),训练周期设置为 150,学习率初始为 0.001。

3.2 训练过程与结果

网络训练损失值(Loss)见图 16,由图 16 可知,网络在前 50 轮训练时,损失值快速下降,训练 120 轮之后,损失值基本趋于稳定。因此本研究将训练 150 轮结束后输出模型,确定为本文识别模型。

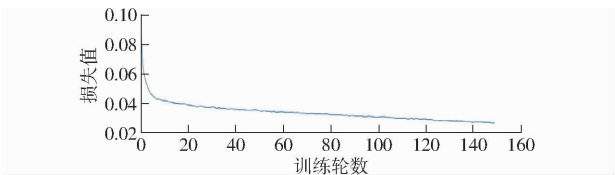


图 16 网络训练损失曲线

Fig. 16 Network training loss curve

3.3 评价指标

本文重点从运算精确率和效率两方面对模型进

行评估。使用准确率、召回率、平均精确率及检测时间、模型内存占用量等指标作为评价指标。考虑到后期模型在硬件上的部署,模型体积(占用空间内存)和参数量也作为本文评价指标之一。

3.4 模型性能消融实验

为了验证模型改进后的效果,将改进的 YOLO v5s 模型和原模型进行各项指标对比,对比结果如表 3 所示。

表 3 不同模型网络性能对比

Tab. 3 Comparison of network performance of different models

算法	准确率/ %	召回率/ %	平均精确 率/%	参数量	浮点运算 次数	检测时 间/ms	模型内存占 用量/MB
YOLO v5s	83.0	85.3	87.3	7 012 822	1.58×10^{10}	20.0	14.0
YOLO v5s_BiFPN	88.4	86.2	92.0	7 078 367	1.60×10^{10}	22.6	14.5
YOLO v5s_CA	88.2	82.1	89.4	7 038 470	1.58×10^{10}	18.6	14.1
YOLO v5s_GSConv + Slim - Neck	88.2	88.1	91.0	5 835 702	1.27×10^{10}	17.3	11.8
YOLO v5s_BiFPN + CA	89.1	88.4	92.3	6 663 055	1.55×10^{10}	20.6	14.3
YOLO v5s_BiFPN + CA + GSConv + Slim - Neck	85.1	89.6	92.1	5 975 007	1.29×10^{10}	18.8	12.1
YOLO v5s_BiFPN + CA + GSConv + Slim - Neck + Ghost	85.6	89.3	92.3	4 041 159	7.70×10^9	19.3	8.5

为统一标准,同时尽可能简化模型,实验均使用相同卷积通道数与 CSP 模块层数的 YOLO v5s 版本模型为原始模型。表中 YOLO v5s 表示使用原始 YOLO v5s 模型,YOLO v5s_BiFPN 表示使用 BiFPN 特征融合结构的 YOLO v5s 模型。YOLO v5s_CA 表示嵌入 CA 注意力机制的 YOLO v5s 模型。YOLO v5s_GSConv + Slim - Neck 表示使用融入 GSConv 卷积和 Slim - Neck 范式设计的 YOLO v5s 模型。YOLO v5s_BiFPN + CA 表示同时使用 BiFPN 特征融合结构和 CA 注意力机制的 YOLO v5s 模型,YOLO v5s_BiFPN + CA + GSConv + Slim - Neck 表示融入了 BiFPN 特征融合机构和 CA 注意力机制的基础上进一步融入了 GSConv 卷积和 Slim - Neck 范式设计的 YOLO v5s 模型。YOLO v5s_BiFPN + CA + GSConv + Slim - Neck + Ghost 表示在 YOLO v5s_BiFPN + CA + GSConv + Slim - Neck 的基础上,在主干网络融合 Ghost 模块的 YOLO v5s 模型。

由表 3 可知,单独使用 CA 注意力机制和 GSConv + Slim - Neck 范式设计的结构模型的平均精确率相对于原始 YOLO v5s 模型都有上升,前者上升 2.1 个百分点,后者上升 3.7 个百分点,但是嵌入 CA 注意力机制的召回率却比原始网络降低 3.2 个百分点。同时融入 BiFPN 特征融合结构和 CA 注意力机制的平均精确率达到 92.3%,比原始 YOLO v5s 模型平均精确率提高 5 个百分点,准确率提升 6.1 个百分点,召回率提高 3.1 个百分点,但检测时间多 0.6 ms,模型内存占用量也增加 0.3 MB。

但在 YOLO v5s_BiFPN + CA 的基础上再加上轻量化 GSConv 卷积网络和范式设计 Slim - Neck 模型后,模型的检测时间减少 1.2 ms,内存占用量减少 1.2 MB,参数量也降低 10.3%,召回率更是比原始 YOLO v5 模型提高 4.3 个百分点,而平均精确率只下降 0.2 个百分点。在此基础上,在主干网络中融入 Ghost 模块,模型平均精确率会提高到 92.3%,同时模型内存占用量减少到 8.5 MB。参数量也大幅减少,虽然模型检测时间多 0.5 ms。满足收获机对模型检测速度的要求。

由表 3 可知,在 YOLO v5s 中使用 BiFPN 结构可以有效提升模型的检测性能。使用 BiFPN 的模型参数量相比于原始模型虽然略有提升但其检测时间几乎相同且检测精度的提升收益较大。原因在于 BiFPN 混合了多尺寸的特征图,增加了针对小目标的特征图尺寸并进行加权融合,同时在训练中 BiFPN 对目标的权重数值是以一种动态可学习的方式进行调整,因此随着训练的深入其得到了比 PaNet 结构更多的特征。从实验结果上 YOLO v5s 模型上融入 BiFPN 特征融合结构、CA 注意力机制、GSConv 卷积、Slim - Neck 范式设计和 Ghost 模块的模型。大幅降低参数和模型内存占用量的同时,提高了对小目标检测的精度,更适合本文对模型轻量化和高精确率的需求。

3.5 与不同主流模型的性能对比

将本文改进的 YOLO v5s 模型与目前常用的目标检测模型(SSD、Faster R - CNN、YOLO v7、YOLO

v3) 进行对比,结果如表 4 所示。

表 4 不同主流模型检测结果对比

模型	mainstream models				
	准确率/ %	召回 率/%	平均精 确率/%	检测时 间/ms	模型内存 占用量/MB
SSD	89.6	80.6	88.3	50.2	90.6
Faster R - CNN	85.2	79.5	84.1	100.3	324.0
YOLO v7	86.9	84.3	89.1	30.1	74.8
YOLO v3	88.0	84.3	88.3	38.9	48.1
改进 YOLO v5s	85.6	89.3	92.3	19.3	8.5

由表 4 可知,双阶段检测网络模型 Faster R - CNN 的识别精度低且网络参数多,推理时间最长,单阶段目标检测模型 SSD、YOLO v3 和 YOLO v7 在推理时间上相比 Faster R - CNN 略有提高。本文改进的模型平均精确率与表中前 4 个模型相比,分别提高 4、8.2、3.2、4 个百分点。检测时间和模型内存占用量,本文改进的模型也有很大的提升,可以更好地满足本文对模型轻量化和检测速度的要求。

3.6 模型优化后检测实例

本文对不同光线条件、不同环境背景、不同遮挡情况下的两个场景进行了蔗梢分叉点实例检测的对比分析^[23]。从图 17 中可以看出,YOLO v5s 模型和 YOLO v5s_BiFPN、YOLO v5s_CA、YOLO v5s_GSConv + Slim - Neck、Faster R - CNN、SSD、YOLO v7、YOLO v3 模型都存在漏检的问题,其中 YOLO v5s_CA 模型在逆光情况下的漏检最为严重。SSD 模型在左图

中还存在错检的情况。YOLO v5s_BiFPN + CA + GSConv + Slim - Neck 和 YOLO v5s_BiFPN + CA + GSConv + Slim - Neck + Ghost 网络模型具有最好的检测效果,能够识别出最多的蔗梢分叉点。特别左图中,这两个模型能够识别出右上方遮挡和暗光下的蔗梢分叉点的模型。但 YOLO v5s_BiFPN + CA + GSConv + Slim - Neck + Ghost 网络模型具有更小的参数量和模型内存占用量。综合考虑该模型在蔗梢分叉点实例检测中表现最佳,既保证了识别准确度,又有效地避免了漏检情况,还满足了本文对模型轻量化的要求。

3.7 蔗梢分叉点定位原理与流程

双目摄像头是一种模拟人类双眼的视觉传感器,可以测量物体的深度距离,现在已经广泛应用在各个领域中(图 18)。相对于 TOF 和激光雷达双目摄像头不需要主动发射光源,其硬件成本低,具有较高的经济性和可行性。另一方面考虑到田间光线变化大,甘蔗生长错乱等因素,相对于单目摄像头,双目摄像头可以利用立体视觉的特征,区分不同的物体和背景。单目摄像头则容易受到光照、遮挡、颜色等因素的影响,导致识别错误或者漏检,误差较大。双目测距的误差率也可满足甘蔗收获机对定位误差的要求。

因此,本文使用双目镜头配合 SGBM (Semi-global block matching) 算法实现蔗梢分叉点定位功能,该方法主要流程如下:首先,本文对双目摄像头标定确定摄像头的内部参数(如焦距、主点等)和外

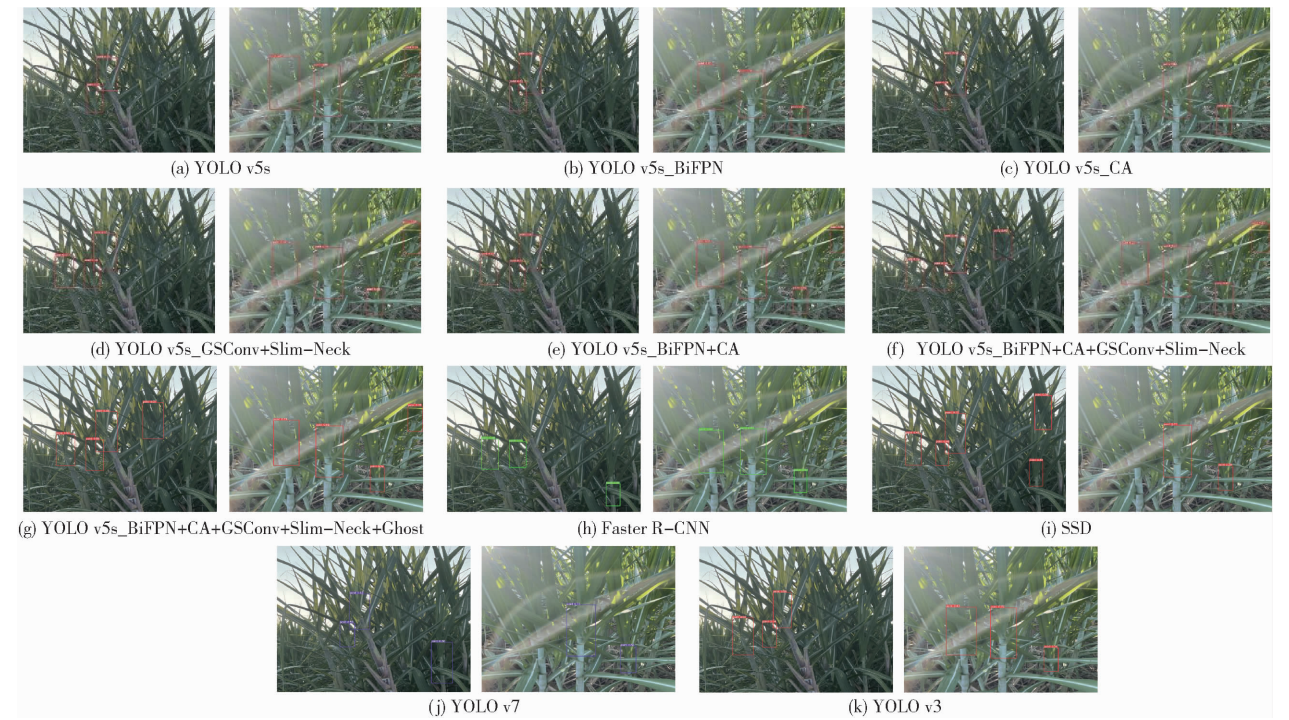


图 17 不同网络检测实例对比

Fig. 17 Comparison of detection examples for different networks



图 18 双目相机拍摄图像

Fig. 18 Taking pictures with binocular camera

部参数(如摄像头之间的距离、旋转矩阵等),之后通过双目视觉系统,同时采集左右两个摄像头的图像。对左右图像进行预处理,包括去噪、灰度化、图像矫正等操作,以便后续的立体匹配算法能够准确地找到对应的像素点。然后使用 SGBM 算法进行立体匹配。该算法通过比较左右图像中的像素点,找到最佳匹配点,从而计算出视差图。视差图表示了左右图像中对应像素点之间的距离差异,推测物体离相机的距离。然后根据视差图,可以计算出每个像素点的三维坐标。利用相机的内参、视差值和三角测量原理,将像素点映射到三维空间中,得到物体在空间中的坐标。最后,根据物体在三维空间中的坐标,获取物体真实高度信息。

3.8 蔗梢分叉点高度定位效果分析

为验证双目相机在田间的实际定位效果,本文使用双目摄像头模拟甘蔗收获机高度,拍摄了图像,拍摄的甘蔗品种为中蔗 44 号,并对拍摄到的蔗梢分叉点高度进行了实际测量。根据收获机的作业情况,本文选取成像范围在 50 ~ 150 cm 之间。另外考虑到光照情况,本文分别进行了顺光和逆光的拍摄。表 5、6 分别为顺光下和逆光下双目测距的定位效果。

从表 5、6 中可以分析出,在相机与物体的距离在 1 m 以下时,无论是顺光还是逆光,测量的高度与实际高度的误差都比较小,有着很高的精度,但随着相机与物体的距离增大,测量的误差也在逐步增大。造成这个现象的原理是当目标距离相机比较近的时候,目标在成像平面内的有效像素点比较多,这些像素点会更简单和更准确地找到最佳匹配点,所以测距精度较高;但当目标距离相机越来越远时,目标在成像平面内的有效像素点变少,匹配难度增大,匹配误差变大,所以测距精度也相应变低。另一方面,在同样距离下,双目相机在顺光的测量效果明显好于逆光的测量效果。这是因为逆光会导致被测物体的轮廓和特征变得模糊不清。当光线从背后照射到物体上时,物体的轮廓可能会被光线的干扰模糊化,使得双目系统难以准确地提取特征点。但总体而言,本文的双目测量差距在 1 m 内的距离时,平均相对误差小

表 5 顺光下双目测距定位效果					
Tab.5 Effect of binocular ranging and positioning under smooth light condition					
相机与蔗梢分叉点距离/cm	蔗梢分叉点实际高度/cm	测量值/cm	绝对误差/cm	相对误差/%	平均相对误差/%
50	310	309	1	0.32	0.429
	280	278	2	0.71	
	320	319	1	0.31	
	290	290	0	0	
	315	313	2	0.63	
	316	315	1	0.31	
	298	297	1	0.33	
	296	296	0	0	
	298	296	2	0.67	
	305	303	2	1.01	
100	310	304	6	1.93	1.547
	280	276	4	1.42	
	320	313	7	2.18	
	290	285	5	1.72	
	315	310	5	1.58	
	316	312	4	1.26	
	298	293	5	1.68	
	296	294	4	1.35	
	298	294	4	1.34	
	305	303	2	1.01	
150	310	295	15	4.84	4.976
	280	269	11	3.93	
	320	302	18	5.63	
	290	279	11	3.79	
	315	301	14	4.44	
	316	300	16	5.06	
	298	280	18	6.04	
	296	279	17	5.74	
	298	281	17	5.70	
	305	291	14	4.59	

于 5%,在 1.5 m 时平均相对误差也在 8% 左右,符合甘蔗收获机对蔗梢高度测量误差的要求。

3.9 蔗梢分叉点识别与定位流程图

实现蔗梢分叉点识别与定位的流程图如图 19 所示。该流程图可以分为蔗梢分叉点的识别系统和定位系统。蔗梢分叉点识别系统就是使用本文优化后的模型来得到蔗梢分叉点的像素坐标;定位模型则利用左右图像进行立体匹配,使用 SGBM 算法计算左右图像的视差图^[24],根据视差图和相机参数,计算出蔗梢分叉点的三维位置信息^[25],从而得到蔗梢分叉点的高度,最后结合图 6 所述的蔗梢分叉点高度优化算法,输出甘蔗生长点位置高度信息。

4 结论

(1)提出了一种基于改进 YOLO v5s 的蔗梢分

表 6 逆光下双目测距定位效果

Tab.6 Effect of binocular ranging and positioning under backlight

相机与蔗梢分叉点距离/cm	蔗梢分叉点实际高度/cm	测量值/cm	绝对误差/cm	相对误差/%	平均相对误差/%
50	310	308	2	0.64	0.659
	280	277	3	1.07	
	320	318	2	0.63	
	290	289	1	0.34	
	315	313	2	0.63	
	316	314	2	0.63	
	298	297	1	0.33	
	296	295	1	0.34	
	298	296	2	0.67	
	305	301	4	1.31	
100	310	295	15	4.83	4.275
	280	270	10	3.57	
	320	303	17	5.31	
	290	278	12	4.13	
	315	300	15	4.76	
	316	302	14	4.43	
	298	286	12	4.03	
	296	284	12	4.05	
	298	286	12	4.03	
	305	294	11	3.61	
150	310	280	30	9.68	8.192
	280	261	19	6.79	
	320	295	25	7.81	
	290	268	22	7.59	
	315	292	23	7.30	
	316	296	20	6.33	
	298	270	28	9.40	
	296	269	27	9.12	
	298	273	25	8.39	
	305	276	29	9.51	

叉点识别模型,该模型在 YOLO v5s 中引入 BiFPN 特征融合结构和 CA 注意力机制,以增强不同层次特征的交互和表达能力,并使用 GSConv 卷积和

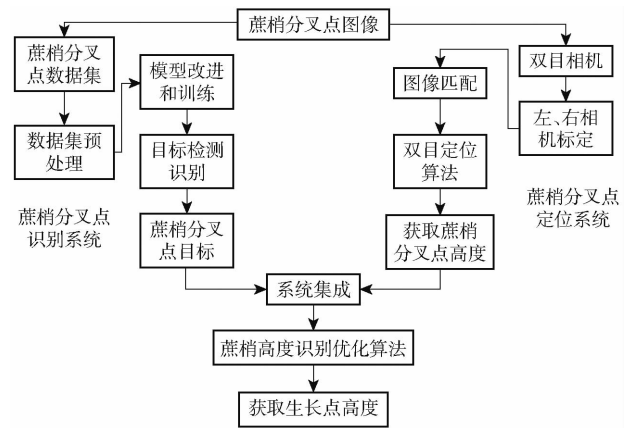


图 19 蔗梢分叉点识别与定位流程图

Fig. 19 Flowchart for identification and location of sugarcane tip bifurcation points

Slim-Neck 范式设计卷积,并在主干网络中引入 Ghost 模块替换 Neck 中原始普通卷积,提高模型的运行效率。采用本文改进后的模型在蔗梢分叉点验证集上进行实验验证,实验结果表明,本文改进后的模型相对于原始 YOLO v5s 网络,平均精确率提高 5 个百分点,召回率提高 4 个百分点,参数量降低 43%,模型内存占用量减少 5.5 MB,检测时间达到 19.3 ms。结果表明本文提出的模型具有更高的检测精度和效率,为后续进一步开展甘蔗收获机切梢器的实时定位调整提供了应用研究基础。

(2)采用单目和双目相机在广西大学扶绥农科基地采集了甘蔗图像数据,并进行了数据预处理和标注,构建了甘蔗蔗梢分叉点数据集,为开展甘蔗收获机切梢的智能识别及实时控制研究提供了数据支持。

(3)通过对采集图像的甘蔗分叉点的统计分析,并结合现场对甘蔗分叉点、生长点高度的测量统计分析,发现其均具有明显的正态统计特征,可通过图像识别和定位的方法结合甘蔗分叉点生长数据的统计和分析,来确定切梢器的位置及调整范围。

参 考 文 献

[1] 陆盼. 广西甘蔗机械化收获发展现状调查及影响因素研究[D]. 南宁:广西大学,2022.

LU Pan. Investigation on the development status of mechanized sugarcane harvesting in Guangxi and research on influencing factors[D]. Nanning: Guangxi University, 2022. (in Chinese)

[2] ZAIDI S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.

[3] FU L, MAJEED Y, ZHANG X, et al. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting[J]. Biosystems Engineering, 2020, 197: 245-256.

[4] ZHANG J, KARKEE M, ZHANG Q, et al. Multi-class object detection using Faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting[J]. Computers and Electronics in Agriculture, 2020, 173: 105384.

[5] LU S, CHEN W, ZHANG X, et al. Canopy-attention-YOLO v4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation[J]. Computers and Electronics in Agriculture, 2022, 193: 106696.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// Computer Vision-ECCV 2016: 14th

- European Conference, Amsterdam, Proceedings, Part I 14. Springer International Publishing, 2016: 21 – 37.
- [7] SAMMY V M, BOBBY D G, RUJI P M. Sugarcane disease recognition using deep learning[C]//IEEE Eurasia Conference on IOT, Communication and Engineering, 2019: 575 – 578.
- [8] RANA M N, ARFAN J, RANA M D S. IoT and machine learning based stem borer pest prediction[J]. Intelligent Automation & Soft Computing, 2022, 31(3): 1377 – 1392.
- [9] 李尚平, 李向辉, 张可, 等. 改进 YOLO v3 网络提高甘蔗茎节实时动态识别效率[J]. 农业工程学报, 2019, 35(23): 185 – 191.
- LI Shangping, LI Xianghui, ZHANG Ke, et al. Improving YOLO v3 network to improve real-time dynamic identification efficiency of sugarcane stem nodes[J]. Transactions of the CSAE, 2019, 35(23): 185 – 191. (in Chinese)
- [10] 李威, 李尚平, 潘家枫, 等. 基于机器学习的中耕期甘蔗幼苗识别定位[J]. 中国农机化学报, 2021, 42(11): 130 – 137.
- LI Wei, LI Shangping, PAN Jiafeng, et al. Machine learning based identification and localization of sugarcane seedlings during the middle tillage period[J]. Chinese Journal of Agricultural Machinery Chemistry, 2021, 42(11): 130 – 137. (in Chinese)
- [11] QUAN L, FENG H, LV Y, et al. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R – CNN[J]. Biosystems Engineering, 2019, 184: 1 – 23.
- [12] LAWAL M O. Tomato detection based on modified YOLO v3 framework[J]. Scientific Reports, 2021, 11(1): 1447 – 1458.
- [13] 闫彬, 樊攀, 王美茸, 等. 基于改进 YOLO v5m 的采摘机器人苹果采摘方式实时识别[J]. 农业机械学报, 2022, 53(9): 28 – 38, 59.
- YAN Bin, FAN Pan, WANG Meirong, et al. Real time recognition of apple picking method using improved YOLO v5m harvesting robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 28 – 38, 59. (in Chinese)
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759 – 8768.
- [15] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//The IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 10781 – 10790.
- [16] 高腾, 张先武, 李柏. 深度学习在安全帽佩戴检测中的应用研究综述[J]. 计算机工程与应用, 2023, 59(6): 13 – 29.
- GAO Teng, ZHANG Xianwu, LI Bo. Review on the application of deep learning in helmet wearing detection[J]. Computer Engineering and Application, 2023, 59(6): 13 – 29. (in Chinese)
- [17] 夏烨, 雷晓晖, 祁雁楠, 等. 基于改进 Ghost – YOLO v5s – BiFPN 算法检测梨树花序[J]. 智慧农业(中英文), 2022, 4(3): 108 – 119.
- XIA Ye, LEI Xiaohui, QI Yannan, et al. Detection of pear tree inflorescence based on improved Ghost – YOLO v5s – BiFPN algorithm[J]. Smart Agriculture (Chinese and English), 2022, 4(3): 108 – 119. (in Chinese)
- [18] 胡新荣, 龚闯, 张自力, 等. 基于改进 Deeplab v3 + 的服装图像分割网络[J]. 计算机工程, 2022, 48(7): 284 – 291.
- HU Xinrong, GONG Chuang, ZHANG Zili, et al. Clothing image segmentation network based on improved Deeplab v3 + [J]. Computer Engineering, 2022, 48(7): 284 – 291. (in Chinese)
- [19] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [J]. arXiv Preprint, arXiv:2206.02424, 2022.
- [20] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993 – 13000.
- [21] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNET: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, 2020: 1571 – 1580.
- [22] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//The IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779 – 788.
- [23] 张志远, 罗铭毅, 郭树欣, 等. 基于改进 YOLO v5 的自然环境下樱桃果实识别方法[J]. 农业机械学报, 2022, 53(增刊 1): 232 – 240.
- ZHANG Zhiyuan, LUO Mingyi, GUO Shuxin et al. Cherry fruit recognition method based on improved YOLO v5 in natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(Supp. 1): 232 – 240. (in Chinese)
- [24] ZHONG F, QUAN C. Stereo-rectification and homography-transform-based stereo matching methods for stereo digital image correlation[J]. Measurement, 2021, 173: 108635.
- [25] YIN Z, XIONG J. Stereovision measurement of layer geometry in wire and arc additive manufacturing with various stereo matching algorithms[J]. Journal of Manufacturing Processes, 2020, 56: 428 – 438.