

基于轻量型卷积神经网络的马铃薯种薯芽眼检测算法

黄杰, 王相友^{*}, 吴海涛, 刘书玮, 杨笑难, 刘为龙

(山东理工大学农业工程与食品科学学院, 淄博 255000)

摘要: 马铃薯种薯芽眼属于小目标物体, 识别难度大、要求高。为了在试验台(芽眼识别装置)上快速、准确地完成识别任务, 该研究提出一种基于轻量型卷积神经网络的芽眼检测模型。首先, 为了降低模型的计算量和聚焦小目标物体, 替换 YOLOv4 的主干网络 CSPDarkNet-53 为 GhostNetV2 轻量型特征提取网络; 其次, 在 YOLOv4 的颈部网络中, 使用深度可分离卷积(depthwise separable convolution, DW)模块代替普通卷积块进一步降低模型计算量; 最后, 更改边界框损失函数为具有角度代价的边界框损失函数(SIoU), 避免因预测框的位置不确定, 而影响模型收敛速度和整体检测性能。结果表明, 改进后芽眼检测模型参数量为 12.04 M, 使用笔记本电脑 CPU 检测单张图片的时间为 0.148 s, 从试验台收集的测试数据显示平均精度为 89.13%。相对于其他主干特征提取网络 CSPDarkNet-53、MobileNetV1、MobileNetV2、MobileNetV3、GhostNetV1, 其检测精度分别高出 1.85、0.75、2.67、4.17、1.89 个百分点; 与同类目标检测模型 SSD、Faster-RCNN、EfficientDet、CenterNet、YOLOv7 相比, 在检测精度上, 分别高出 23.26、27.45、10.51、18.09、2.13 个百分点, 在检测时间上, 分别降低 0.007、6.754、1.891、1.745、0.422 s, 且模型参数量具有明显优势。该研究为小目标物体检测和模型部署提供技术支撑。

关键词: 图像识别; 卷积神经网络; 马铃薯芽眼检测; 小目标; YOLOv4; GhostNetV2

doi: 10.11975/j.issn.1002-6819.202303035

中图分类号: TP391.4; S214.9; S24

文献标志码: A

文章编号: 1002-6819(2023)09-0172-11

黄杰, 王相友, 吴海涛, 等. 基于轻量型卷积神经网络的马铃薯种薯芽眼检测算法[J]. 农业工程学报, 2023, 39(9): 172-182. doi: 10.11975/j.issn.1002-6819.202303035 <http://www.tcsae.org>

HUANG Jie, WANG Xiangyou, WU Haitao, et al. Detecting potato seed bud eye using lightweight convolutional neural network (CNN)[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(9): 172-182. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202303035 <http://www.tcsae.org>

0 引言

马铃薯是全球第四大粮食农作物, 栽培范围遍布全世界^[1-3]。全球疫情期间, 欧美国家马铃薯的产量走低, 马铃薯市场需求缺口扩大。2018—2019 年中国马铃薯种植面积分别为 7.18×10^6 、 7.14×10^6 、 7.21×10^6 hm²^[4]。然而, 马铃薯种薯播种前的切块工作, 仍以人工切块为主, 存在机械化程度低、人工成本高、劳动强度大等问题。为了解决这一系列问题, 国内外许多研究机构、科研院所已经开始研究自动化切块设备^[5-9], 但大多数研究成果没有考虑芽眼的位置, 直接对其进行加工, 造成了极大的浪费。因此研究一种能够在切块装置上快速、准确、实时完成芽眼识别任务的算法极为重要。

国内外研究人员针对芽眼识别的研究, 主要使用传统图像处理方式。李玉华等^[10]提出一种基于色彩饱和度三维几何特征的马铃薯种薯芽眼识别方法, 芽眼识别准确率为 91.4%。孟令军^[7]通过提取 HSV 色彩空间中马铃薯种薯的色彩饱和度图像, 获取到马铃薯芽眼位置, 试验结果表明, 色彩饱和度提取马铃薯芽眼可行。JI 等^[11]

使用 K-means 聚类方法对马铃薯进行分割, 提取发芽马铃薯信息, 最终采用建立的模型准确率达到 84.62%。YANG 等^[12]在灰度图上使用 Canny 边缘检测器, 获得分割掩码, 完成马铃薯芽眼的检测。结果表明, 该方法的检测精度为 89.28%。以上研究方法都能较好地识别马铃薯芽眼, 但其应用场景单一、鲁棒性差, 不能在复杂环境中完成检测任务。

随着智能时代的到来, 以机器学习为代表的深度学习技术被广泛应用于农业的目标识别任务中。李就好等^[13]提出一种改进 Faster R-CNN 的田间苦瓜叶部病害检测算法, 结果表明, 检测苦瓜叶部病害的平均精度值为 78.5%, 检测时间为 0.322 s。陈柯屹等^[14]使用改进型 Faster R-CNN 识别田间棉花顶芽, 试验结果表明, 该方法平均检测精度值为 98.1%, 识别速度为 10.3 帧/s。梁喜凤等^[15]使用改进 Mask R-CNN 模型去识别番茄侧枝修剪点, 试验结果表明, 修剪点平均识别准确率为 82.99%, 检测时间为 0.319 s。LEE 等^[16]使用 Mask R-CNN 方法, 检测马铃薯, 试验结果表明平均检测精度为 90.08%。XI 等^[17]基于混沌优化的 K-means 算法实现了马铃薯种薯芽眼的快速分割, 试验结果表明, 所提算法检测单幅图像时间为 1.109 s。

以上使用 R-CNN^[18]为代表的二阶段算法, 其检测精度高, 但检测速度相对较慢。而以 SSD^[19]、YOLO^[20-23]为代表的一阶段目标检测算法, 相比二阶段检测算法在速度上存在优势。张兆国等^[24]提出了一种采用改进 YOLOv4

收稿日期: 2023-03-07 修订日期: 2023-04-08

基金项目: 山东省农业重大应用技术创新项目(SD2019NJ010)

作者简介: 黄杰, 博士生, 研究方向为计算机技术在农业工程中的应用。

Email: huangjie0306@126.com

*通信作者: 王相友, 博士, 教授, 博士生导师, 研究方向为农产品加工技术与装备。Email: wxy@sdut.edu.cn

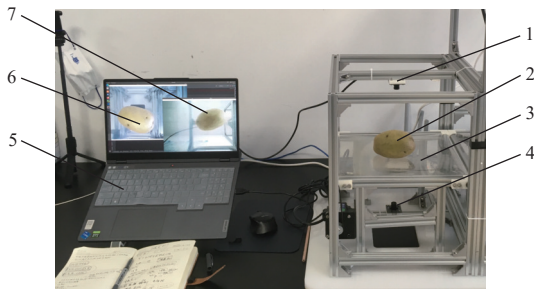
模型检测复杂环境下马铃薯的目标检测算法, 其平均识别率为 91.4%, 识别速度为 23.01 帧/s。王相等^[25]提出了一种基于改进 YOLOv4 模型的马铃薯中土块石块检测方法, 利用通道剪枝算法对模型进行剪枝处理, 其模型存储空间为 20.75 MB, 检测速度为 78.49 帧/s。孙俊等^[26]提出一种快速精准识别棚内草莓的改进 YOLOv4-Tiny 模型, 使用 GhostNetV1 作为主干网络, 并修改注意力机制, 提升识别小目标的性能, 其模型权重大小为 4.68 MB, 平均识别精度为 92.62%。WANG 等^[27]提出一种基于 YOLO 框架的新型轻量级小物体检测框架, 试验结果表明, 枯木检测精度为 89.11%, 模型权重大小为 7.6 MB。ZENG 等^[28]提出一种基于改进 YOLO 的轻量级番茄实时检测方法, 试验结构表明, 平均检测帧率为 26.5 帧/s, 检测精度达到 93%。

上述以一阶段检测算法为代表的 YOLO 目标检测模型, 在检测速度和检测精度上具有较好的效果, 同时在此框架下修改的网络结构又能降低模型权重大小, 利于部署在不同的移动设备上。鉴于此, 本研究在 YOLOv4 的基础上, 提出一种基于轻量级卷积神经网络的马铃薯种薯芽眼检测模型, 该目标检测模型替换改进前的 CSPDarkNet-53 主干为 GhostNetV2 轻量级网络, 大幅度减轻模型计算量; 其次修改颈部的标准卷积为深度可分离卷积, 进一步降低计算量, 同时丰富目标的语义信息提升小目标物体的检测能力; 最后更改边界框回归损失函数为 SIOU 回归损失函数, 提升芽眼检测模型的收敛速度和整体检测性能, 为快速、准确完成小目标物体芽眼识别任务提供基础。

1 材料与方法

1.1 数据采集

本次试验数据采集于自制试验台, 同时自制的试验台也用于芽眼检测。该试验台主要由三部分组成, 分别是便携式笔记本电脑、顶部和底部摄像头以及透明输送带, 如图 1 所示。其中便携式笔记本电脑主要用于前向推理, 顶部和底部摄像头主要负责收集马铃薯和芽眼的顶部及底部数据, 透明输送带主要帮助底部摄像头获取马铃薯的底部芽眼信息和传输马铃薯到达下一个工序位。



1. 顶部摄像头 2. 马铃薯 3. 透明输送带 4. 底部摄像头 5. 便携式笔记本 6. 顶部图像 7. 底部图像
1. Top camera 2. Potatoes 3. Transparent conveyor 4. Bottom camera 5. Portable laptops 6. Top image 7. Bottom image

图 1 数据采集试验台

Fig.1 Data acquisition test bench

该试验台中的摄像头型号为 WH-L2140.K214L, 分

辨率为 1920×1080 像素, 速率为 60 帧/s。在自然光照条件下拍摄马铃薯。拍摄时, 底部摄像头和顶部摄像头同时工作, 在同一时刻获取马铃薯底部和顶部的芽眼信息。选取 200 个带有芽眼的马铃薯, 品种为中薯 2 号。为了更加有效地利用每个马铃薯, 将每个马铃薯按照横正、竖正、横反、竖反的顺序放在试验台上进行数据采集, 如图 2 所示。共计采样 800 次, 收集 1600 张从顶部和底部拍摄的马铃薯图像。

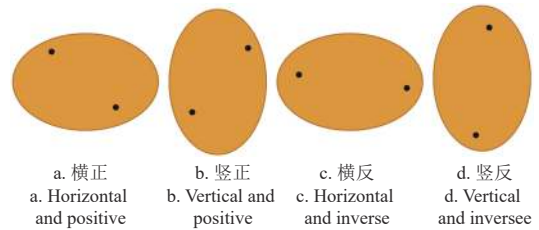


图 2 马铃薯摆放顺序

Fig.2 The order of the potatoes

1.2 数据处理

首先将采集到的 1600 张原始图片使用 LabelImg 工具进行标注, 标注内容为马铃薯和芽眼两个类别, 其中马铃薯设置为 potato, 芽眼设置为 bud。由于检测环境相对固定, 不需要增加额外的噪音提升模型的泛化能力, 由此将已经标注好的每张图片使用 python 脚本程序对图像进行简单的数据增强处理, 包括 2 次随机翻转、2 次随机旋转, 共生成 6400 张带有标签的马铃薯种薯图像。最后将数据增强后的图像按照 8:1:1 的比例随机分为训练集 (5120 幅)、验证集 (640 幅) 以及测试集 (640 幅), 用于之后的模型训练和测试任务。

1.3 芽眼检测算法

1.3.1 改进 YOLOv4 主干网络

YOLOv4^[20] 主干网络主要参考 CSPNet^[29] 网络结构, 和 YOLOv3 中的 DarkNet^[21] 网络结构, 形成现在的 CSPDarkNet-53 主干特征提取网络, 其中 53 代表有 53 次标准卷积操作。相比于其他 YOLO 系类^[22-23], YOLOv4 主干网络结构较清晰, 简单明了, 易于修改, 为不同场景的目标检测任务提供研究基础。YOLOv6、YOLOv7 主干网络前向推理时, 参数多, 计算量大, 不适宜用于边缘设备实时检测需求和储存要求。GhostNet 模型是北京华为诺亚方舟实验室提出的一种轻量级卷积神经网络^[30], 其核心思想是使用 Ghost 模块代替标准卷积达到降低模型计算量的目的, 具体工作过程如图 3 所示。

首先, 对输入特征 X , 使用 1×1 标准卷积生成少量的本征特征 Y' :

$$Y' = X \odot F' \quad (1)$$

式中 \odot 表示卷积操作; $X \in R^{H \times W \times C}$ 表示输入特征, H 、 W 、 C 分别为输入特征的长、宽和通道数; $F' \in R^{C' \times K \times K \times C'}$, 其中 $K \cdot K$ 表示标准卷积核的大小, 这里 K 取 1, $Y' \in R^{H \times W \times C'}$ 表示输出的少量本征特征, $C' < C$ 。

其次, 使用线性卷积 (深度卷积) 对本征特征进行特征信息提取, 生成 s 个 Ghost 特征 Y'' :

$$Y'' = Y' \odot F''_{dp} \quad (2)$$

式中 \odot 表示卷积操作； $F''_{dp} \in R^{1-D \cdot D \cdot SC}$ 表示线性卷积， $D = 3$ ； $Y' \in R^{H \cdot W \cdot C \cdot (S-1)}$ 表示 Ghost 特征。

最后，将本征特征与 Ghost 特征拼接，产生一个与

使用标准卷积通道数相同的特征图 Y ：

$$Y = \text{Concat}([Y', Y'']) \quad (3)$$

式中 $Y \in R^{H \cdot W \cdot C}$ 表示输出特征图（一个 Ghost 模块操作）， $C = SC'$ 。

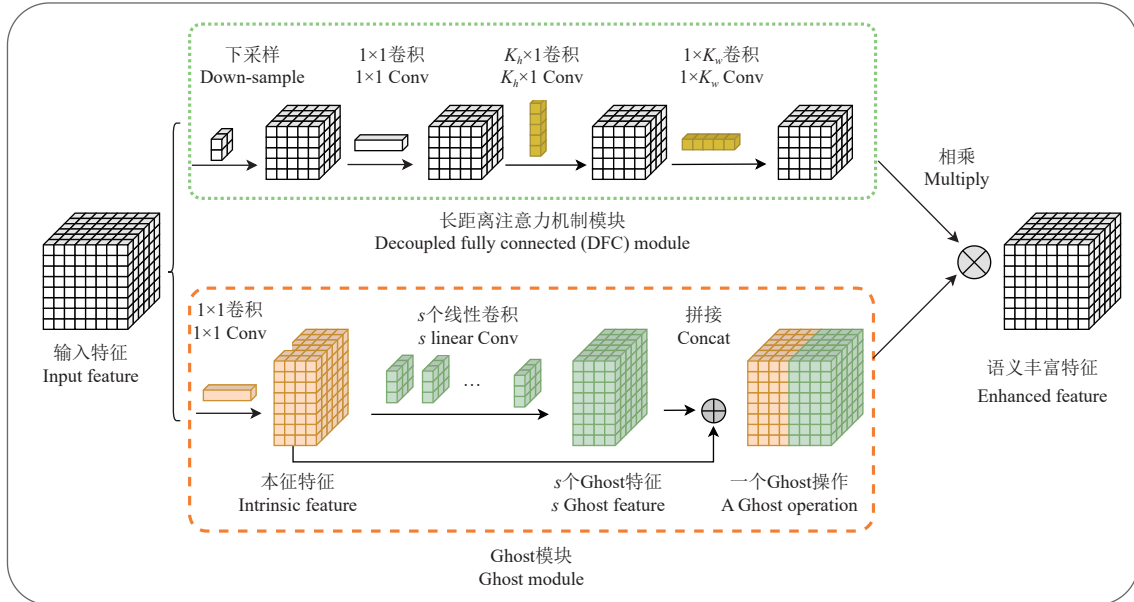


图 3 Ghost 模块和长注意力机制模块

Fig.3 Ghost module and decoupled fully connected module

经过上面 3 个步骤，Ghost 模块的计算量与标准卷积之比为

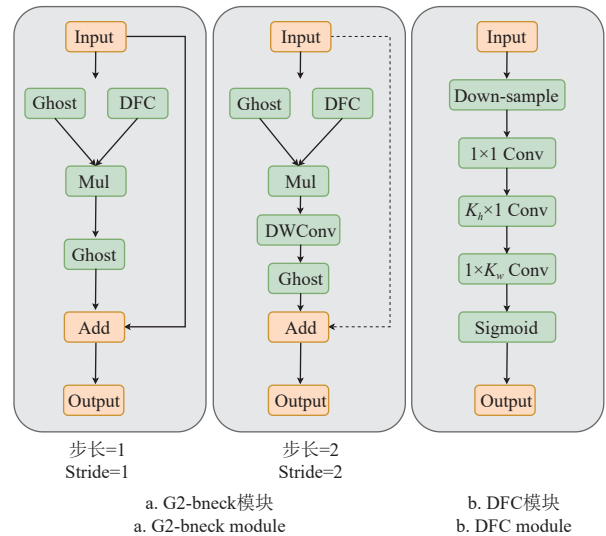
$$\frac{K \cdot K \cdot H \cdot W \cdot C' + C' \cdot (S - 1) \cdot H \cdot W \cdot D \cdot D}{C \cdot H \cdot W \cdot SC' \cdot D \cdot D} = \frac{1}{S} + \frac{S - 1}{CS} = \frac{1}{S} \quad (4)$$

由结果可知，Ghost 模块的浮点数计算量大约是标准卷积的 1/s。

在 GhostNetv2 中作者提出一种长距离注意力机制 (DFC) 模块^[31]，用于进一步捕获空间语义信息，其前向推理过程如图 3 所示。首先使用 1×1 标准卷积生成特征图，接着分别使用大小为 $K_h \times 1$ ，和 $1 \times K_w$ 的卷积核提取空间语义信息。相比普通注意力机制的复杂度 $O(H^2W^2)$ ，该方法的复杂度为 $O(K_hHW + K_wHW)$ ，最后将 Ghost 模型生成的特征与使用 DFC 机制生成的特征逐元素相乘得到语义更加丰富的特征，且不改变输出特征维度。

另外，GhostNetV2 中的 Ghostbottleneck 是由 Ghost 模块和 DFC 模块共同组成的一个逆残差瓶颈结构，如图 4 所示。图 4a 为 $\text{Stride}=1$ 时 GhostNetV2bottleneck (G2-bneck) 结构。该结构首先使用一个 Ghost 模块和 DFC 模块对输入特征进行特征提取，并扩充特征信息，然后使用逐元素相乘的方式丰富语义信息，接着将其传入下一个 Ghost 模块进行通道压缩，最后将原始输入特征与压缩特征进行拼接，完成一个 G2-bneck 操作。

当 $\text{Stride}=2$ 时，在逐元素相乘和第二个 Ghost 模块之间插入一个深度卷积模块 (DW)，对特征的宽高进行压缩，完成下采样操作，同时在右分支使用一个深度卷积和一个 1×1 的标准卷积，确保主干分支在相同尺度下完成拼接操作。



注：Ghost 代表使用一个 Ghost 模块进行前向推理；DFC 表示一个长距离注意力机制模块；Mul 表示逐元素相乘；DWConv 表示深度可分离卷积模块；Add 表示拼接操作；Down-sample 表示下采样操作；Conv 表示标准卷积操作；Sigmoid 表示归一化操作；图 b 中的虚线表示使用了深度卷积和 1×1 卷积。
Note: Ghost represents forward reasoning using a Ghost module; DFC represents a decoupled fully connected module; Mul represents element-wise multiplication; DWConv represents a depthwise separable convolution module; Add represents concatenation operations; Down-sample represents down-sampling operation; Conv means standard convolution operation; Sigmoid means normalization operation; The dotted lines in Figure b represent the use of a deep convolution kernel 1×1 convolution.

图 4 GhostNetV2bottleneck (G2-bneck) 模块和 decoupled fully connected (DFC) 模块

Fig.4 GhostNetV2bottleneck(G2-bneck) module and decoupled fully connected(DFC) module

经过试验发现^[31]，在获取空间语义信息时，直接使用DFC模块将会额外增加计算成本，因此先使用下采样，得到较小特征图，再进行一系列便宜卷积操作，最后使用Sigmoid函数将注意力值限制在(0, 1)范围内，实现加速推理，如图4c所示。

1.3.2 颈部深度可分离模块

YOLOv4的颈部网络(PANet)使用标准卷积方式获取主干中的特征信息。此种操作会进一步增加计算成本，因此，本研究使用MobileNetV1^[32]中的深度可分离卷积模块(如图5所示)代替原先的标准卷积达到减小模型整体运算量和进一步提升小目标检测能力的目的。可分离卷积计算量为：

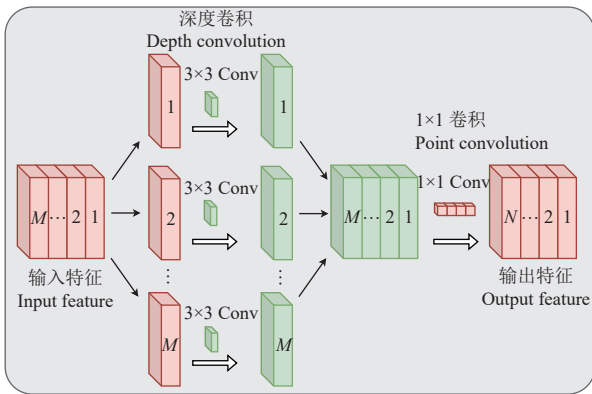


图5 深度可分离卷积计算过程

Fig.5 Depthwise separable convolution calculation procedure

$$\frac{D_H \cdot D_W \cdot M \cdot D_K \cdot D_K + D_H \cdot D_W \cdot M \cdot N}{D_H \cdot D_W \cdot M \cdot N \cdot D_K \cdot D_K} = \frac{1}{N} + \frac{1}{D_K^2} \quad (5)$$

式中 D_H 、 D_W 为特征图长宽尺寸， M 、 N 分别为特征图输入、输出通道数， D_K 为卷积核大小，一般取3。由此可知，深度可分离卷积的计算成本比标准卷积少8~9倍。

1.3.3 改进后芽眼检测模型

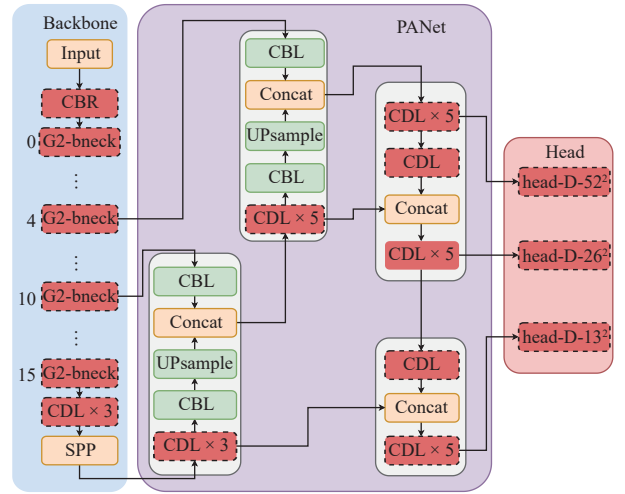
改进的芽眼检测模型主要由部分GhostNetV2网络结构和深度可分离卷积模块组成。如图6所示，带有黑色虚线外框的模块和主干网络为修改位置。模型前向推理时，首先将尺寸为 $416 \times 416 \times 3$ 的图像在通道数上进行扩张，接着使用16个G2-bneck模块(如表1所示)进行特征提取，其次提取主干G2-bneck模块中的4、10、15层特征信息与YOLOv4的颈部网络相连，经过使用深度可分离卷积模块修改的特征金字塔结构(颈部网络)，进一步加强改进YOLOv4网络的特征提取能力，且减少计算量。最后从颈部分出3个特征层，使用3种不同尺度的检测头完成目标检测任务。

1.3.4 边界框回归损失函数

边界框回归损失函数是目标检测损失函数的重要组成部分，良好的定义将为目标检测模型带来显著的提升。YOLOv4目标检测模型使用CIoU作为回归损失函数(\mathcal{L}_{CIoU})，其定义为^[33]

$$\mathcal{L}_{CIoU} = \mathcal{L}_{IoU} + \frac{(x - x_{gr})^2 + (y - y_{gr})^2}{W_g^2 + H_g^2} + \alpha v \quad (6)$$

$$\mathcal{L}_{IoU} = 1 - I_{oU} = 1 - \frac{W_i H_i}{S_u} \quad (7)$$



注：CBR由Convolution加BatchNorm加ReLU三个模块共同组成；G2-bneck表示GhostnetV2bottleneck模块；CDL×3表示先进行一次CBL操作，再进行一次CDW操作，最后再进行一次CBL操作，其中CDW表示进行一次深度可分离卷积操作，其中CBL由Convolution加BatchNorm加ReLU6三个模块共同组成；SPP表示空间金字塔池化操作；Upsample表示上采样操作；Concat表示拼接操作；CDL×5表示先进行一次CBL操作，再进行一次CDW操作，接着进行一次CBL操作，之后再进行一次CDW操作，最后进行一次CBL操作；head-D表示先进行一次CDW操作，再进行一次CBL操作。

Note: CBR consists of three modules together, Convolution plus BatchNorm plus ReLU; G2-bneck indicates GhostnetV2bottleneck module; CDL×3 indicates a CBL operation, then a CDW operation, and finally a CBL operation, where CDW indicates a depthwise separable convolution operation, where CBL consists of three modules together, Convolution plus BatchNorm plus ReLU6; SPP indicates a spatial pyramid pooling operation; Upsample indicates an upsampling operation; Concat indicates a concatenation operation; CDL×5 indicates a CBL operation, followed by a CDW operation, followed by a CBL operation, followed by another CDW operation, and finally a CBL operation; head-D indicates a CDW operation followed by a CBL operation.

图6 改进后芽眼检测模型

Fig.6 Improved bud eye detection model

表1 改进后芽眼检测模型的主干网络

Table 1 The backbone network of the improved bud eye detection model

输入尺寸 Input size	操作 Operator	扩展层 Expand layer	输出层 Output layer	注意力 Attention	步长 Stride
$416^2 \times 3$	Conv2D 3×3	—	16	—	2
$208^2 \times 3$	G2-bneck	16	16	—	1
$208^2 \times 3$	G2-bneck	48	24	—	2
$104^2 \times 3$	G2-bneck	72	24	—	1
$104^2 \times 3$	G2-bneck	72	40	1	2
$52^2 \times 3$	G2-bneck	120	40	1	1
$52^2 \times 3$	G2-bneck	240	80	—	2
$26^2 \times 3$	G2-bneck	200	80	—	1
$26^2 \times 3$	G2-bneck	184	80	—	1
$26^2 \times 3$	G2-bneck	184	80	—	1
$26^2 \times 3$	G2-bneck	480	112	1	1
$26^2 \times 3$	G2-bneck	672	112	1	1
$26^2 \times 3$	G2-bneck	672	160	1	2
$13^2 \times 3$	G2-bneck	960	160	—	1
$13^2 \times 3$	G2-bneck	960	160	1	1
$13^2 \times 3$	G2-bneck	960	160	—	1
$13^2 \times 3$	G2-bneck	960	160	1	1

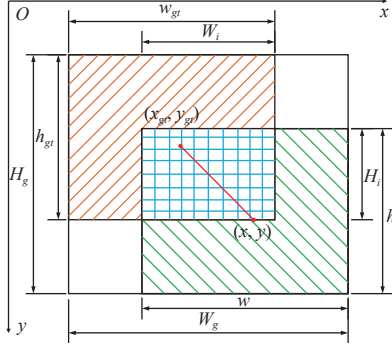
注：“—”表示当前网络层未使用该结构或者策略；1表示当前网络层使用该结构或者策略。

Note: “—” indicates that the current network layer does not use this structure or strategy; 1 indicates that the current network layer uses this structure or strategy.

$$\alpha = \frac{v}{\mathcal{L}_{IoU} + v} \quad (8)$$

$$v = \frac{4}{\pi^2} \left(\tan^{-1} \frac{w}{h} - \tan^{-1} \frac{w_{gt}}{h_{gt}} \right)^2 \quad (9)$$

式中 S_u 的值如图 7 所示。



注: $(x_{gt}, y_{gt}), w_{gt}, h_{gt}$ 为真实框的中心位置和宽高; $(x, y), w, h$ 为预测框的中心位置和宽高; $W_i \cdot H_i$ 为交集面积(网格填充区域); $W_g \cdot H_g$ 为最小边界框面积; 真实框和预测框的并集为 $S_u = wh + w_{gt}h_{gt} - W_i H_i$ 。
Note: $(x_{gt}, y_{gt}), w_{gt}, h_{gt}$ are the center position, width and height of the ground truth box; $(x, y), w, h$ are the center position, width and height of the prediction box; $W_i \cdot H_i$ is the intersect area(grid fill area); $W_g \cdot H_g$ is the area of the smallest enclosing box; $S_u = wh + w_{gt}h_{gt} - W_i H_i$ is the union area.

图 7 真实框和预测框的交集及并集面积

Fig.7 The intersection and concatenation areas of ground truth box and prediction box

除 CIoU 边界框回归损失函数外, 还有一系列基于加法的损失函数, 如 GIoU^[34]、DIoU^[33]、EIoU^[35], 都遵循以下范式:

$$\mathcal{L}_i = \mathcal{L}_{IoU} + \mathcal{R}_i \quad (10)$$

式中 \mathcal{R}_i 为惩罚项。

以上回归损失函数都依赖于边界框回归指标的聚合, 如预测框和真实框的距离、重叠区域和纵横比。而预测框在训练过程中难免存在方向不确定 (“四处游荡”) 的情况, 导致收敛速度慢和效率低下。由此使用一种考虑了角度、距离、形状的回归损失函数 (SCYLLA-IoU, SIoU^[36]) 替换 CIoU 损失函数, 提升模型的收敛速度和整体检测能力。

其中, 角度代价 Λ 描述中心点连线 (如图 7 所示) 与 x - y 轴之间的最小角度, 定义如下:

$$\Lambda = \sin \left(2 \sin^{-1} \frac{\min(|x - x_{gt}|, |y - y_{gt}|)}{\sqrt{(x - x_{gt})^2 + (y - y_{gt})^2 + \epsilon}} \right) \quad (11)$$

当中心点连线与 x 轴或 y 轴对齐时, $\Lambda = 0$; 当中心点连线与 x 轴为 45° 时, $\Lambda = 1$ 。

其中, 距离代价 Δ 描述中心点之间的距离, 其代价与角度正相关, 定义为

$$\Delta = \frac{1}{2} \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \quad (12)$$

式中 $\gamma = 2 - \Lambda$; $\rho_x = \left(\frac{x - x_{gt}}{W_g} \right)^2$; $\rho_y = \left(\frac{y - y_{gt}}{H_g} \right)^2$ 。

其中, 形状代价 Ω 描述边界框之间的大小差异, 定义为

$$\Omega = \frac{1}{2} \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (13)$$

式中 $\theta = 4$; $\omega_w = \frac{|w - w_{gt}|}{\max(w, w_{gt})}$; $\omega_h = \frac{|h - h_{gt}|}{\max(h, h_{gt})}$ 。

最后, 综合角度代价、距离代价、形状代价, 定义回归损失函数为

$$\mathcal{L}_{SIoU} = \mathcal{L}_{IoU} + \frac{\Delta + \Omega}{2} \quad (14)$$

1.3.5 试验平台及评价指标

本研究使用的算力资源来自中科视拓科技有限公司的线上服务器 (AutoDL)。处理器 (CPU) 型号为 AMD EPYC 7 642 48-Core Processor, 运行内存容量为 80 GB, 固态硬盘 (SSD) 容量为 50 GB, 内核个数为 24 个, 显卡 (GPU) 型号为 Nvidia GeForce RTX 3 090, 显存为 24 GB, 系统环境为 Ubuntu 20.04, 搭建 Python3.8 的编程语言和 Pytorch1.10.0 的深度学习框架及 CUDA11.3 的并行计算框架。

模型训练参数设置: 单机单卡, 使用 VOC07+12 数据作为预训练数据集。输入尺寸为 416×416 , 每批次样本数量为 16, 多线程进程为 8, 优化器为 Adam (adaptive moment estimation), 训练次数 (epoch) 为 200 轮, 学习率下降方式使用余弦退火 \cos (cosine annealing), 为防止过拟合设置权重衰减为 0, 初始学习率为 0.001, 权重衰减系数为 0.000 5, 动量因子为 0.937, 使用 Mosaic、Mixup 数据增强方式。

为了准确评估模型的性能, 本研究采用以下指标进行性能评估: 所有类别的平均准确率 (mean average precision, mAP), 平均精度 (average precision, AP), 准确率 (precision, P), 召回率 (recall, R), F_1 得分, 权重大小, 检测时间及参数量。其中, F_1 表示准确率 (P) 和召回率 (R) 的调和平均数, 最大值为 1, 最小值为 0, 得分越大, 模型整体性能越好。参数量以卷积块大小与输出特征尺度之间的乘积为标准, 权重大小为总 Loss 值最小时保存的参数, 检测时间为便携式笔记本电脑 CPU 上的前向推理时间, 其电脑型号为联想拯救者 Y7000P, CPU 型号为 12th Gen Intel (R) Core (TM) i5-12500H, GPU 型号为 NVIDIA GeForce RTX 3 050 Ti Laptop。

2 结果与分析

2.1 改进模型试验结果

2.1.1 不同主干网络试验结果

本试验以改进 YOLOv4 目标检测模型为基础, 更换不同的主干网络, 如 MobileNetV1、MobileNetV2、MobileNetV3、GhostNetV1。在不改变除主干以外参数的情况下, 训练目标检测模型。将得到的参数进行分析比较, 验证改进目标检测模型的可行性。由表 2 可知, 使用 GhostNetV2 模型作为 YOLOv4 的主干特征提取网络时, 其参数量为 12.04 M, 检测精度 (mAP) 为 89.13%, 检测单张图片的时间为 0.148 s, 芽眼和马铃薯的 F_1 得分分别为 0.80、0.99。相比改进前使用 CSPDarkNet-53 的主干网络, 改进后主干网络参数量约为原来的 1/3, 检测精度提升 1.85 个百分点, 检测时间减少 0.279 s, 芽眼得分略

高于改进前主干。另外与其他轻量级主干网络 MobileNetV1、MobileNetV2、MobileNetV3、GhostNetV1 相比，GhostNetV2 检测精度分别提升 0.75、2.67、4.17、1.89 个百分点，芽眼的 F_1 值分别提升 0.06、0.07、0.12、0.08。

模型参数量相比 MobileNetV1 减少 1.65 M，相比 MobileNetV2、MobileNetV3、GhostNetV1 不存在优势，但也能满足模型部署需求。检测时间与其它轻量级主干网络基本一致。

表 2 不同主干网络对比

Table 2 Comparison of different backbone networks

主干网络 Backbone network	mAP/%	参数 Params/M	时间 Time/(s·帧 ⁻¹)	芽眼 Bud				马铃薯 Potato			
				AP/%	P/%	R/%	F_1	AP/%	P/%	R/%	F_1
CSPDarkNet-53	87.28	36.10	0.427	74.25	83.77	64.52	0.73	99.89	99.03	99.35	0.99
MobileNetV1	88.38	13.69	0.136	76.78	78.90	69.35	0.74	99.98	99.67	99.69	1.00
MobileNetV2	86.46	10.81	0.143	72.93	78.70	68.55	0.73	99.99	99.35	99.67	1.00
MobileNetV3	84.42	11.72	0.134	68.87	77.16	61.29	0.68	99.98	99.03	99.35	0.99
GhostNetV1	87.24	11.42	0.125	74.85	78.95	66.53	0.72	99.64	99.67	98.70	0.99
GhostNetV2	89.13	12.04	0.148	79.03	85.57	75.11	0.80	99.24	100.0	98.11	0.99

注：mAP 表示芽眼和马铃薯的平均准确率；AP 表示平均精度；P 为准确率；R 为召回率； F_1 为调和平均数；下同。

Note: mAP indicates the average accuracy of bud eye and potato; AP indicates the average precision; P is the accuracy; R is the recall; F_1 is the harmonic mean; the same below.

整体而言，以 GhostNetV2 网络作为 YOLOv4 主干特征提取网络时，其模型在检测时间、模型参数量上明显优于改进前 YOLOv4 主干网络，在检测精度上也能高于其他轻量级主干网络。故选用 GhostNetV2 网络作为主干，提高检测精度的同时也能满足实时检测和模型部署需求。

2.1.2 不同回归损失函数试验结果

改进前 YOLOv4 目标检测模型使用 CIoU 作为边界框回归损失函数，其在训练过程中有较强的拟合能力，但在训练过程中不可避免地存在预测框方向漂浮不定的情况，导致收敛速度慢，影响模型整体检测性能。由此本试验在改进后 YOLOv4 目标检测模型基础上，更换多种边界框回归损失函数，从 mAP、AP、P、R、 F_1 得分的角度，分析比较不同损失函数对模型整体性能的影响。由表 3 可知，SIoU 回归损失函数的检测精度相比没有使用方向代价的回归损失函数，在检测芽眼时有更高的准确率。根据试验结果可知，SIoU 的检测精度相比 GIoU、CIoU、DIoU、EIoU 分别提升 2.97、4.33、2.38、3.18 个百分点。由于 SIoU 考虑了预测框的方向，在训练的过程中可以引导锚框移到目标框最近的轴上，减小损失的总自由度，提升模型的整体性能。图 8 展示了不同回归损失函数训练时的收敛情况，相比 GIoU、CIoU、DIoU、EIoU，SIoU 回归损失函数收敛速度更快，多次训练后，损失值也更低。

表 3 不同损失函数对比

Table 3 Comparison of different loss functions

损失函数 Loss function	mAP/%	AP _{bud} /%	$F_{1\text{bud}}$	AP _{potato} /%	$F_{1\text{potato}}$
GIoU	86.34	72.72	0.68	99.97	0.99
CIoU	84.80	70.81	0.71	98.85	0.98
DIoU	86.75	73.52	0.71	99.97	0.99
EIoU	85.95	71.92	0.69	99.99	1.00
SIoU	89.13	79.03	0.80	99.24	0.99

注：AP_{bud} 表示芽眼平均精度； $F_{1\text{bud}}$ 表示芽眼的 F_1 值；AP_{potato} 表示马铃薯平均精度； $F_{1\text{potato}}$ 表示马铃薯的 F_1 值。

Note: AP_{bud} means the average accuracy of bud eyes; $F_{1\text{bud}}$ means the F_1 value of bud eyes; AP_{potato} means the average precision of potatoes; $F_{1\text{potato}}$ means the F_1 value of potatoes.

2.1.3 YOLOv4 改进前后试验结果

改进前 YOLOv4 主干特征提取网络使用 CSPDarkNet-53，颈部金字塔结构使用标准卷积，回归损失函数使用

CIoU。改进后 YOLOv4 主干使用 GhostNetV2，用于减少模型参数量，同时具备较高的检测精度，如表 2 所示；颈部金字塔结构中的标准卷积替换成深度可分离卷积，进一步降低模型参数，同时也可以丰富图像语义信息，利于提取小目标信息；回归损失函数使用 SIoU，提升模型整体检测性能和加快收敛速度。图 9 展示了改进前和改进后 YOLOv4 芽眼识别结果，从图中可以发现，在检测小目标芽眼时，改进前 YOLOv4 出现较多漏检情况，而改进后 YOLOv4 能够检测大部分芽眼位置。另外从表 4 中可以发现，改进后 YOLOv4 的参数量约是改进前的 1/5，检测芽眼时的精度相较改进前提升了 0.56 个百分点，在笔记本电脑上的前向推理时间也从原来的 0.474 s 降低至 0.148 s，减少了 0.326 s。

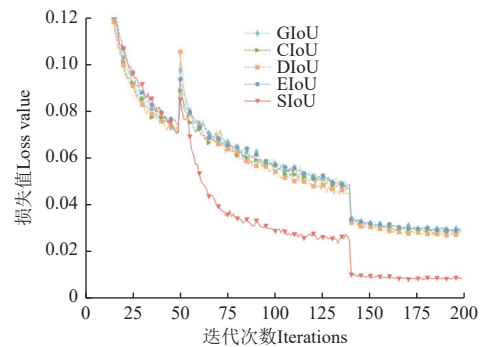


图 8 不同损失函数收敛情况

Fig.8 Different loss function convergence

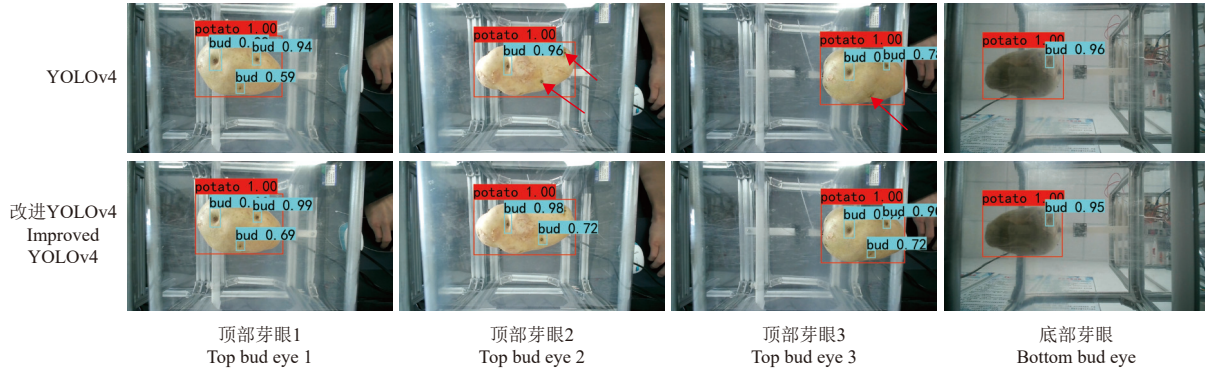
综上所述，改进后的目标检测模型在检测芽眼小目标物体时有较高的检测精度，同时也具有较少的推理参数，为模型部署提供研究基础。

2.2 消融试验

在模型主体网络结构中，由表 2 的试验结果可以发现，使用 GhostNetV2 网络结构作为改进后 YOLOv4 的主干特征提取网络有较好的检测性能。本试验在不改变 GhostNetV2 主干网络内部结构的基础上，比较改进前 YOLOv4 颈部网络结构和改进后颈部网络使用深度可分离卷积模块之间的区别。从表 5 的结果可以看出，使用深度可分离卷积模块后，模型的参数量大幅降低，约为

原先的 1/4 倍, 更加利于移动设备或者边缘设备的部署。在预测层中, 由表 3 的试验结果可以看出, 当使用具有角度代价的 SIoU 作为回归损失函数时, 其整体检测性能达到最优, 同时收敛速度也相较 GIoU、CIoU、DIoU、EIoU 快。本次试验结合主干特征提取网络 (GhostNetV2)、

颈部深度可分离卷积模块 (DW) 以及回归损失函数 (SIoU) 观测模型的权重值、平均精度值以及检测时间。从表 5 中可以看出, 当使用 SIoU 回归损失函数后, 其平均检测精度有明显提升, 相较于使用 CIoU 回归损失函数, 提高了 4.33 个百分点。



注: 方框表示预测框, 箭头表示芽眼漏检的位置。下同。

Note: The box represents the prediction frame, and the arrow represents the location of the missed detection of the bud eye. The same below

图 9 改进前后模型检测结果

Fig.9 Model detection results before and after improvement

表 4 改进前后模型试验结果

模型 Model	参数 Params/M	时间 Time/s	mAP/%	AP _{bud} /%	AP _{potato} /%
YOLOv4	64.36	0.474	89.21	78.47	99.96
改进 YOLOv4 Improved YOLOv4	12.04	0.148	89.13	79.03	99.24

另外, 本试验在改进后 YOLOv4 的基础上, 还对主干特征提取网络 GhostNetV2 中的注意力机制分布情况进行了测试, 如表 5 所示。当完全不使用注意力机制 (A1) 时, 改进后 YOLOv4 的平均精度值只有 85.96%; 当完全使用注意力机制 (A2) 时, 其平均精度值提升至 86.87%, 同时模型的权重大小上升 10.23 MB; 当部分使用注意力机制 (A3, 其分布情况如表 1 所示) 时, 其平均精度值为 89.13%, 相比不使用注意力机制, 检测精度提升 4.83 个百分点。

表 5 颈部和预测层及主干注意力机制消融试验

Table 5 Neck and prediction layer and backbone attention mechanism ablation test

方法 Methods	权重 Weights/MB	时间 Time/s	mAP/%
GhostNetV2+NODW+CIoU+A3	148.00	0.184	89.11
GhostNetV2+DW+CIoU+A3	46.40	0.148	84.80
GhostNetV2+DW+SIoU+A3	46.40	0.148	89.13
GhostNetV2+DW+SIoU+A1	40.07	0.149	85.96
GhostNetV2+DW+SIoU+A2	50.30	0.153	86.87

注: GhostNetV2 为主干特征提取网络; NODW 表示使用标准卷积模型; DW 表示使用深度可分离卷积模块; A1 表示 GhostNetV2 中主干特征提取网络全部不使用注意力机制; A2 表示 GhostNetV2 中主干特征提取网络全部使用注意力机制; A3 表示 GhostNetV2 中主干特征提取网络部分使用注意力机制, 分布情况如表 1 所示。

Note: GhostNetV2 is the backbone feature extraction network; NODW indicates that the standard convolutional model is used; DW indicates that the depth separable convolutional module is used; A1 indicates that the backbone feature extraction network in GhostNetV2 does not use the attention mechanism at all; A2 indicates that the backbone feature extraction network in GhostNetV2 uses the attention mechanism at all; A3 indicates that the backbone feature extraction network in GhostNetV2 partially uses the attention mechanism, and the distribution is shown in Table 1.

综上所述, 使用主干特征提取网络 (GhostNetV2) 加上深度可分离卷积模块 (DW) 加上回归损失函数 (SIoU) 加上在 GhostNetV2 中部分使用注意力机制组成的芽眼检测模型, 整体性能达到最优, 其平均检测精度值为 89.13%, 检测时间为 0.148 s, 模型权重大小为 46.40 MB。

2.3 不同目标检测算法识别试验

本试验使用自制试验台 (如图 1 所示) 收集 50 张马铃薯种薯图像作为测试集。在训练参数、训练数据集相同的情况下, 将测试集数据使用 SSD、Faster-RCNN、EfficientDet、CenterNet、YOLOv7、YOLOv4 及改进后 YOLOv4 目标检测模型进行前向推理。观察各个检测模型的参数量、便携式笔记本电脑 CPU 检测时间、芽眼 F_1 得分和马铃薯 F_1 得分以及 mAP 值。由表 6 可知, 本研究改进 YOLOv4 目标检测模型平均精度值 (mAP) 为 89.13%, 与 SSD、Faster-RCNN、EfficientDet、CenterNet、YOLOv7 相比, 分别提升 23.26、27.45、10.51、18.09、2.13 个百分点, 与改进前 YOLOv4 相比, 检测精度基本一致; 在模型参数量上面, 改进后检测模型相比上述 6 种目标检测模型, 占有明显优势, 仅为 12.04 M; 在检测时间上面, 改进后模型在笔记本电脑 CPU 上检测单张图像的时间为 0.148 s, 相比前 6 种检测模型, 分别减少 0.007、6.754、1.891、1.745、0.422、0.326 s。

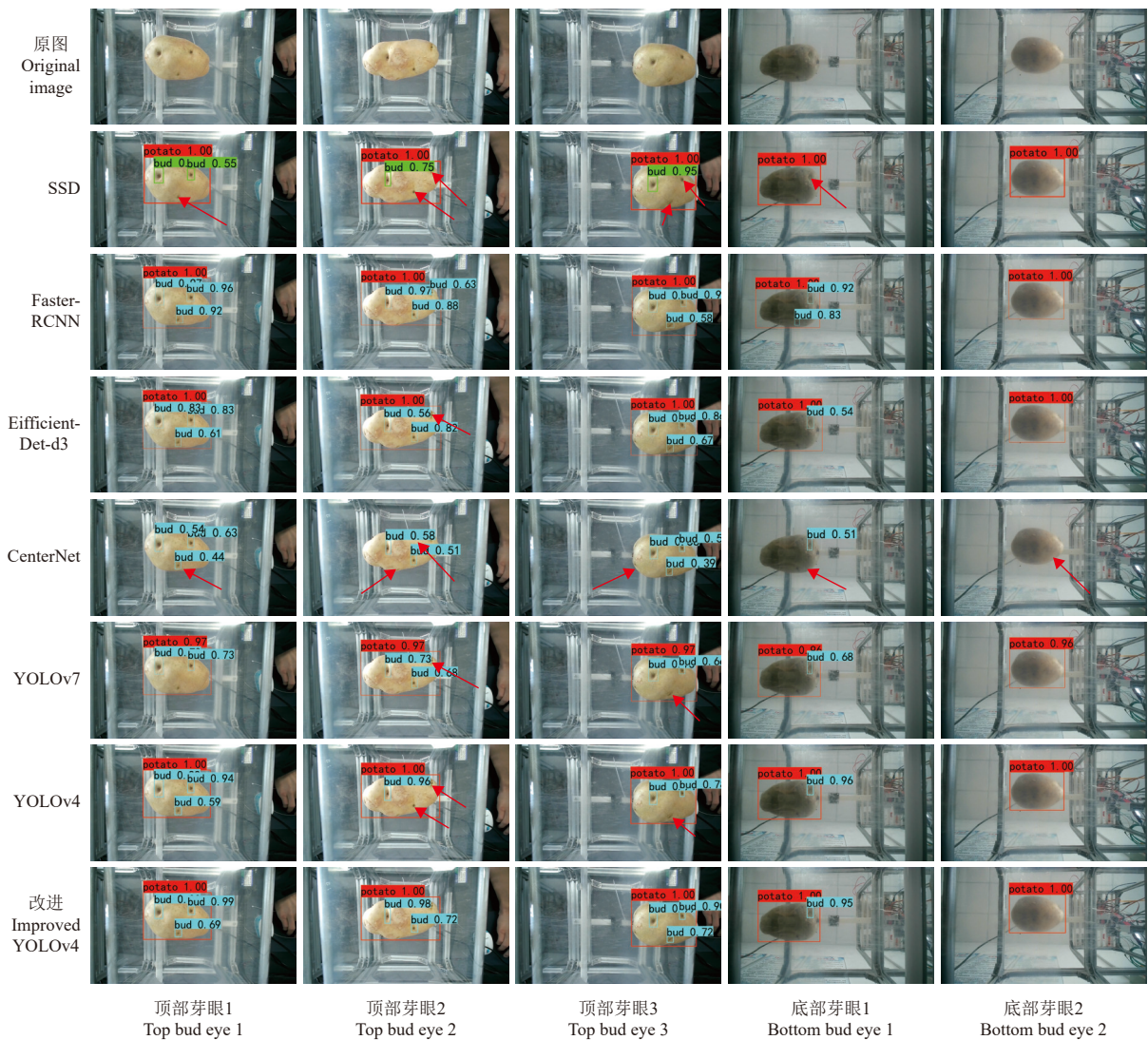
另外, 为了更加清楚地观察检测效果, 分别从上述 6 种目标检测算法中, 随机选出 5 张预测结果。如图 10 所示, 使用 SSD、YOLOv4、YOLOv7、CenterNet、EfficientDet 算法检测芽眼时, 存在漏检情况, 图中用箭头标出。值得注意的是 CenterNet 不能识别马铃薯。使用 Faster-RCNN 及改进 YOLOv4 目标检测算法时, 可以识别出大部分芽眼, 但 Faster-RCNN 使用笔记本电脑 CPU 检测单张图片的时间不能满足实时检测需求。综上所述,

基于 YOLOv4 改进的轻量型芽眼检测模型在检测时间、精度及参模型参数量和权重大小方面都具有较好的表现能力，能够满足小目标芽眼检测需求和更加利于模型部署。

表 6 不同目标检测算法芽眼识别结果对比

Table 6 Comparison of bud eye recognition results of different target detection algorithms

模型 Model	主干网络 Backbone network	参数 Params/M	检测时间 Detection time/s	F_{1bud}	$F_{1potato}$	总体平均准确率 Mean average precision/%
SSD	VGG-16	26.280	0.155	0.12	0.99	65.87
Faster-RCNN	VGG-16	137.090	6.902	0.32	0.95	61.68
EifficientDet-d3	EfficientNet	12.018	2.039	0.55	0.97	78.62
CenterNet	ResNet50	32.665	1.893	0.36	0.01	71.04
YOLOv7	MulticoncatNet	37.620	0.570	0.69	0.99	87.00
YOLOv4	CSPDarkNet-53	64.360	0.474	0.77	0.99	89.21
改进 YOLOv4 Improved YOLOv4	GhostNetv2	12.040	0.148	0.80	0.99	89.13



注：方框表示预测框，箭头表示芽眼漏检或者马铃薯漏检的位置。
Note: The box represents the prediction frame, and the arrow represents the location of missed bud eye or potato missed detection.

图 10 不同目标检测算法识别马铃薯和芽眼的结果

Fig.10 The results of different object detection algorithms to identify potato and eye

3 结 论

为了在试验台上快速、准确地完成芽眼识别任务，本文提出一种基于轻量型卷积神经网络的芽眼检测算法，

主要结论如下：

- 1) 使用轻量型主干网络 GhostNetV2 代替 YOLOv4 的主干网络 CSPDarkNet-53，明显减少网络参数量，约为改进前的 1/3，检测时间减少 0.279 s，相比 CSPDarkNet-

53、MobileNetV1、MobileNetV2、MobileNetV3、GhostNetV1 主干网络, 平均检测精度分别高出 1.85、0.75、2.67、4.17、1.89 个百分点。

2) 使用 SIOU 回归损失函数代替 CIOU 回归损失函数, 提高模型的整体检测精度, 相比 GIoU、CIOU、DIOU、EIOU 回归损失函数, 检测精度分别高出 2.97、4.33、2.38、3.18 个百分点。

3) 本文所提轻量型目标检测算法, 可以有效识别芽眼和马铃薯的位置, 平均检测精度为 89.13%, 检测时间为 0.148s, 相比 SSD、Faster-RCNN、EfficientDet、Center-Net、YOLOv7 目标检测模型, 平均精度值分别高出 23.26、27.45、10.51、18.09、2.13 个百分点, 检测时间分别减少 0.007、6.754、1.891、1.745、0.422 s, 模型参数量上占有明显优势, 仅为 12.04 M。由检测结果可知, 改进后轻量型芽眼检测模型能够满足小目标芽眼检测需求和利于模型部署, 为该类研究提供技术支撑。

[参 考 文 献]

- [1] TORRES M D, FRADINHO P, RODRIGUEZ P, et al. Biorefinery concept for discarded potatoes: Recovery of starch and bioactive compounds[J]. *Journal of Food Engineering*, 2020, 275: 109886.
- [2] SU W, XUE H. Imaging spectroscopy and machine learning for intelligent determination of potato and sweet potato quality[J]. *Foods*, 2021, 10(9): 2146.
- [3] WANG Y, SU W. Convolutional neural networks in computer vision for grain crop phenotyping: A review[J]. *Agronomy*, 2022, 12(11): 2659.
- [4] 罗其友, 伦闰琪, 高明杰, 等. 2021—2025 年我国马铃薯产业高质量发展战略路径[J]. *中国农业资源与区划*, 2022, 43(3): 37-45.
LUO Qiyu, LUN Runqi, GAO Mingjie, et al. Development of potato industry in chinaduring from 2021 to 2025[J]. *Chinese Journal of Agricultural Resources and Regional Planning*, 2022, 43(3): 37-45. (in Chinese with English abstract)
- [5] 王相友, 祝珊, 李学强, 等. 定向排列纵横切分马铃薯种薯切块机设计与试验 [J]. *农业机械学报*, 2020, 51(6): 334-345.
WANG Xiangyou, ZHU Shan, LI Xueqiang, et al. Design and experiment of directional arrangement vertical and horizontal cutting of seed potato cutter[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(6): 334-345. (in Chinese with English abstract)
- [6] 陈志伟. 基于 YOLOX 的马铃薯种薯芽眼检测与自动切块试验研究 [D]. 泰安: 山东农业大学, 2022.
CHEN Zhiwei. Research on the Detection and Automatic Cutting Test of Seed Potato Buds Based on YOLOX[D]. Taian: Shandong Agricultural University, 2022. (in Chinese with English abstract)
- [7] 孟令军. 基于图像处理的马铃薯种薯智能切块技术研究 [D]. 淄博: 山东理工大学, 2021.
- MENG Lingjun. Research on Smart Cutting Technology of Potato Seed based on Image Processing[D]. Zibo: Shandong University of Technology, 2021. (in Chinese with English abstract)
- [8] HAN. Apparatus slitting seed potatoes: KR20080019602[P]. 2009-09-08.
- [9] WILLIAM. Laseroperatedseedpotatocutter:US19990365981[P]. 2001-11-27.
- [10] 李玉华, 李天华, 牛子孺, 等. 基于色饱和度三维几何特征的马铃薯芽眼识别[J]. *农业工程学报*, 2018, 34(24): 158-164.
LI Yuhua, LI Tianhua, NIU Ziru, et al. Potato bud eyes recognition based on three-dimensional geometric features of color saturation[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2018, 34(24): 158-164. (in Chinese with English abstract)
- [11] JI Y, SUN L. Nondestructive classification of potatoes based on HSI and clustering[C]//Proceedings of the 2019 4th International Conference on Measurement, Information and Control (ICMIC). Harbin, China: IEEE, 2019: 73-77.
- [12] YANG Y, ZHAO X, HUANG M, et al. Multispectral image based germination detection of potato by using supervised multiple threshold segmentation model and Canny edge detector[J]. *Computers and Electronics in Agriculture*, 2021, 182: 106041.
- [13] 李就好, 林乐坚, 田凯, 等. 改进 Faster R-CNN 的田间苦瓜叶部病害检测[J]. *农业工程学报*, 2020, 36(12): 179-185.
LI Jiuhao, LIN Lejian, TIAN Kai, et al. Detection of leaf diseases of balsam pear in the field based on improved Faster R-CNN[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2020, 36(12): 179-185. (in Chinese with English abstract)
- [14] 陈柯屹, 朱龙付, 宋鹏, 等. 融合动态机制的改进型 Faster R-CNN 识别田间棉花顶芽[J]. *农业工程学报*, 2021, 37(16): 161-168.
CHEN Keyi, ZHU Longfu, SONG Peng, et al. Recognition of cotton terminal bud in field using improved Faster R-CNN by integrating dynamic mechanism[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE)*, 2021, 37(16): 161-168. (in Chinese with English abstract)
- [15] 梁喜凤, 章鑫宇, 王永维. 基于改进 Mask R-CNN 的番茄侧枝修剪点识别方法[J]. *农业工程学报*, 2022, 38(23): 112-121.
LIANG Xifeng, ZHANG Xinyu, WANG Yongwei, et al. Recognition method for the pruning points of tomato lateral branches using improved Mask R-CNN[J]. *Transactions of the Chinese Society of Agricultural Engineering(Transactions of*

- the CSAE), 2022, 38(23): 112-121. (in Chinese with English abstract)
- [16] LEE H, SHIN B. Potato detection and segmentation based on Mask R-CNN[J]. *Journal of Biosystems Engineering*, 2020, 45: 233-238.
- [17] 席 芮, 侯加林, 李立成. 基于混沌优化 K 均值算法的马铃薯芽眼的快速分割[J]. *农业工程学报*, 2019, 35(5): 190-196.
XI Rui, HOU Jialin, LI Licheng. Fast segmentation on potato buds with chaos optimization-based K-means algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(5): 190-196. (in Chinese with English abstract)
- [18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39: 1137-1149.
- [19] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//*Proceedings of the European Conference on Computer Vision*. Amsterdam, the Netherlands: Springer, 2016: 21-37.
- [20] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23)[2023-03-05].<https://arxiv.org/abs/2004.10934>.
- [21] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-8)[2023-03-05].<https://arxiv.org/abs/1804.02767>.
- [22] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. (2022-07-06)[2023-03-05].<https://arxiv.org/abs/2207.02696>.
- [23] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2023-03-05].<https://arxiv.org/abs/2209.02976>.
- [24] 张兆国, 张振东, 李加念, 等. 采用改进 YoloV4 模型检测复杂环境下马铃薯[J]. *农业工程学报*, 2021, 37(22): 170-178.
ZHANG Zhaoguo, ZHANG Zhendong, LI Jianian, et al. Potato detection in complex environment based on improved YoloV4 model[J]. *Transactions of the Chinese Society of Agricultural Engineering*(*Transactions of the CSAE*), 2021, 37(22): 170-178. (in Chinese with English abstract)
- [25] 王相友, 李晏兴, 杨振宇, 等. 基于改进 YOLO v4 模型的马铃薯中土块石块检测方法[J]. *农业机械学报*, 2021, 52(8): 241-247, 262.
WANG Xiangyou, LI Yanxing, YANG Zhenyu, et al. Detection method of clods and stones from impurified potatoes based on improved yolo v4 algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(8): 241-247, 262. (in Chinese with English abstract)
- [26] 孙俊, 陈义德, 周鑫, 等. 快速精准识别棚内草莓的改进 YOLOv4-Tiny 模型[J]. *农业工程学报*, 2022, 38(18): 195-203.
SUN Jun, CHEN Yide, ZHOU Xin, et al. Fast and accurate recognition of the strawberries in greenhouse based on improved YOLOv4-Tiny model[J]. *Transactions of the Chinese Society of Agricultural Engineering*(*Transactions of the CSAE*), 2022, 38(18): 195-203. (in Chinese with English abstract)
- [27] WANG X, ZHAO Q, JIANG P, et al. LDS-YOLO: A lightweight small object detection method for dead trees from shelter forest[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107035.
- [28] ZENG T, LI S, SONG Q, et al. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment[J]. *Computers and Electronics in Agriculture*, 2023, 205: 107625.
- [29] WANG C Y, LIAO H Y M, 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 (CVPRW). Seattle, WA, USA: IEEE, 2020: 1571-1580.
- [30] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1577-1586.
- [31] TANG Y H, HAN K, GUO J Y, et al. GhostNetV2: Enhance cheap operation with long-range attention[EB/OL]. (2022-11-23)[2023-03-05].<https://arxiv.org/abs/2211.12905>.
- [32] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2023-03-05].<https://arxiv.org/abs/1704.04861>.
- [33] ZHENG Z H, 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*. Palo Alto, California USA: AAAI Press, 2020: 12993-13000.
- [34] REZATOFIGHI S H, TSOI N, GWAK J, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, 2019: 658-666.
- [35] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient iou loss for accurate bounding box regression[J]. *Neurocomputing*, 2021, 506: 146-157.
- [36] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression[EB/OL]. (2022-05-25)[2023-03-05].<https://arxiv.org/abs/2205.12740>.

Detecting potato seed bud eye using lightweight convolutional neural network (CNN)

HUANG Jie, WANG Xiangyou^{*}, WU Haitao, LIU Shuwei, YANG Xiaonan, LIU Weilong

(*School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, China*)

Abstract: Manually pre-sowing of seed pieces cannot fully meet the large-scale potato production in China in recent years, due to the low-level mechanization, high labor costs, and intensity. Automated equipment can be expected to realize potato seed cutting. However, the potato seed eyes cannot be accurately positioned during processing using mechanized equipment, resulting in serious waste. Particularly, accurate and rapid target detection is highly required to identify the potato seed eyes, due to the small target objects. It is also necessary for the high recognition of small targets with fewer forward inference parameters. In this study, a target detection model was proposed to rapidly, accurately, and real-time recognize the potato seed eyes in the block-cutting equipment using a lightweight convolutional neural network (CNN). Firstly, a lightweight feature extraction network (GhostNetV2) was selected to replace the CSPDarkNet-53 in the backbone network of YOLOv4, in order to reduce the forward inference parameters of the model for the more focus on small target objects. Secondly, the depthwise separable convolution (DW) modules were used to further reduce the computational complexity in the neck network of YOLOv4. Finally, the bounding box loss function was changed to the SCYLLA-IoU (SIoU) loss function with the angle cost. The impact of the uncertain position was avoided in the prediction box on the convergence speed and the overall detection performance of the model. The experimental results indicated that the parameter size was 12.04 M, when the GhostNetV2 model was utilized as the backbone feature extraction network for the YOLOv4. The test dataset was also collected from the experimental platform. A better performance was achieved in the average precision of 89.13%, where the time required to detect a single image using a CPU on a laptop was 0.148 s. The F1 scores were 0.80 and 0.99 for the buds and potatoes, respectively. The improved backbone network presented approximately one-third of the original parameter size, with an increase in the detection accuracy of 1.85 percentage points, and a decrease in the detection time of 0.279 s, compared with the CSPDarkNet-53 backbone network before improvement. Furthermore, the GhostNetV2 backbone network improved the detection accuracy by 0.75, 2.67, 4.17, and 1.89 percentage points, compared with the lightweight backbone networks, including MobileNetV1, MobileNetV2, MobileNetV3, and GhostNetV1. The F1 values were also improved by 0.06, 0.07, 0.12, and 0.08 for the buds, respectively. The SIoU bounding box loss function showed detection accuracy improvements of 2.97, 4.33, 2.38, and 3.18 percentage points, compared with the GIoU, CIoU, DIoU, and EIoU ones, respectively. Moreover, the improved YOLOv4 object detection model shared the higher recognition accuracy, with increases of 23.26, 27.45, 10.51, 18.09, and 2.13 percentage points, respectively, compared with similar object detection models, such as SSD, Faster-RCNN, EfficientDet, CenterNet, and YOLOv7. In terms of the detection time, the improved YOLOv4 object detection model reduced the detection times by 0.007, 6.754, 1.891, 1.745, 0.422, and 0.326 s, compared with the SSD, Faster-RCNN, EfficientDet, CenterNet, YOLOv7, and YOLOv4, respectively. In model parameter size, the improved detection model was only 12.04M parameters. Overall, the finding can also provide new technical support for the recognition and model deployment of small target objects, such as the potato buds.

Keywords: image identification; convolutional neural network; potato bud eye detection; small target; YOLOv4; GhostNetV2