· 农业信息与电气技术 ·

基于旋转框定位的改进 YOLOv7 的香蕉目标检测与定位方法

伍荣达[1,2]，张世昂[5]，付根平[4]，陈天赐[3]，赖颖杰[2]，罗文轩[2]，郭晓耿[2]，朱立学[2※]

（1. 广东电网有限责任公司江门供电局，江门 529000；2. 仲恺农业工程学院机电工程学院，广州 510225；
3. 华南农业大学工程学院，广州 510642；4. 仲恺农业工程学院自动化学院，广州 510225；
5. 仲恺农业工程学院创新创业学院，广州 510225）

摘 要：针对目标检测算法无法较好适配倾斜目标，且算法参数量大难以部署到嵌入式设备等问题，提出了一种带旋转定位框的改进 YOLOv7 的香蕉目标检测和定位方法，引入 GSConv 模块降低计算复杂度和参数数量提高检测精度。香蕉目标以旋转框定位，使用五参数表示法定义旋转框，采用 Kullback-Leibler divergence 损失函数将旋转框映射到一个二维高斯分布，计算两个概率分布之间的差异，并将差异性作为损失进行优化，能更准确地衡量预测框与真实框之间的差异。此外，还采用 Criminisi 算法修复由深度相机局部空洞引起的定位错误。试验结果表明，改进后的旋转检测模型在香蕉目标检测速度和准确性上均有提升，平均精度达到 96.15%，相比 YOLOv7 模型提高了 17.04 个百分点，检测帧率提高约 40 帧/s。此外，改进模型通过旋转边界框能更准确地预测香蕉果柄位置，定位误差均值降低到 7.02 mm，平均相对误差降低到 0.65%，相比 YOLOv7 模型分别减少了 24.3 mm 和 1.96%。因此，该方法为复杂果园环境下快速、准确地识别和定位香蕉串及其果柄提供了有效解决方案。

关键词：香蕉；采摘；YOLOv7；三维定位；旋转框；深度图像修复
doi：10.11975/j.issn.1002-6819.202404143
中图分类号：S24 **文献标志码**：A **文章编号**：1002-6819(2024)-22-0115-09

伍荣达，张世昂，付根平，等. 基于旋转框定位的改进 YOLOv7 的香蕉目标检测与定位方法[J]. 农业工程学报，2024，40(22)：115-123. doi: 10.11975/j.issn.1002-6819.202404143 http://www.tcsae.org
WU Rongda, ZHANG Shi'ang, FU Genping, et al. Detecting and locating banana targets using improved YOLOv7 and rotational bounding box positioning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(22): 115-123. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202404143 http://www.tcsae.org

0 引 言

深度学习技术出现前，大多数果蔬目标检测方法基于传统机器学习，依赖手工设计的特征[1-5]进行对象定位，缺乏普适性和鲁棒性。近年来，目标检测算法逐渐应用于水果检测中，BOOGAARD 等[6]提出了一种基于深度学习的黄瓜节间长度自动测量方法，使用多视点解决遮挡问题，提升了节点聚类与编号的精度。图像处理算法在水果中应用，降低了计算复杂度，提高了水果的识别准确率[7-11]。为了实现香蕉的智能管理，研究者们探索了香蕉植株、果串和果柄的检测方法。NEUPANE 等[12]基于 Fast R-CNN 开发了香蕉植株检测方法，通过增强植被特征的图像处理技术提升检测率。FU 等[13]则开发了高效的香蕉果串和果蕾检测方法，在不同光照和遮挡条件下表现良好。尽管大规模网络提高了精度，但往往以牺牲速度为代价，尤其是在边缘设备上。近期，多种轻量级网络架构证明了 YOLO 模型中的深度可分离卷积具有优势[14-15]。例如，FU 等[16]通过修剪 YOLOv4 实现了轻量化，并在香蕉种植中的智能管理上展现出良好的实时性能。现有大多数检测方法[17-21]使用轴对齐边界框进行目标定位，但香蕉果实倾斜的特性使得传统水平边界框包含更多背景信息，导致检测精度下降，特别是在三维定位中容易放大误差。目前，基于旋转边界框的目标检测算法已经应用于遥感图像领域[22-24]。因此，研究基于旋转框的目标检测算法可以很好地解决蕉园环境中香蕉果柄的朝向问题。

农业机器人视觉检测的目标是精确定位作业对象，获取三维位置信息以支持后续智能采收。段洁利等[25]提出了基于 YOLOv5 的蕉穗识别及底部果轴定位方法，提高了香蕉采摘机器人的效率。GAO 等[26]改进了 YOLOv3 算法，加入 CBAM（convolutional block attention module）注意力机制以增强鲁棒性，并结合双目立体相机获取番茄的三维位置信息，定位误差小于 1.5%。这些研究表明，机器视觉可以有效用于采摘作业中的果实定位。深度相机虽能获取物体表面深度，但其三维定位局限性较大，王金鹏等[27]通过改进 YOLOv5s 实现油茶果的高精度检测，但是其使用双目相机对目标进行定位试验漏检率较高，对目标的表面进行定位容错率较低，为更精准定位，可以通过简化物体形状并结合改进的检测算法来获取其二维中心及相关参数，进而计算实际质心三维位置。例

收稿日期：2024-04-21 修订日期：2024-10-19
基金项目：“十四五”农业“揭榜挂帅”课题-岭南特色水果智能采收技术（2022SDZG03-5）；广东省重点领域科技研发计划项目（2019B020223003）
作者简介：伍荣达，研究方向为智能装备。
Email：m13172219939@163.com
※通信作者：朱立学，博士，教授，博士生导师，研究方向为农产品采收技术与智能农机装备。zhulixue@zhku.edu.cn

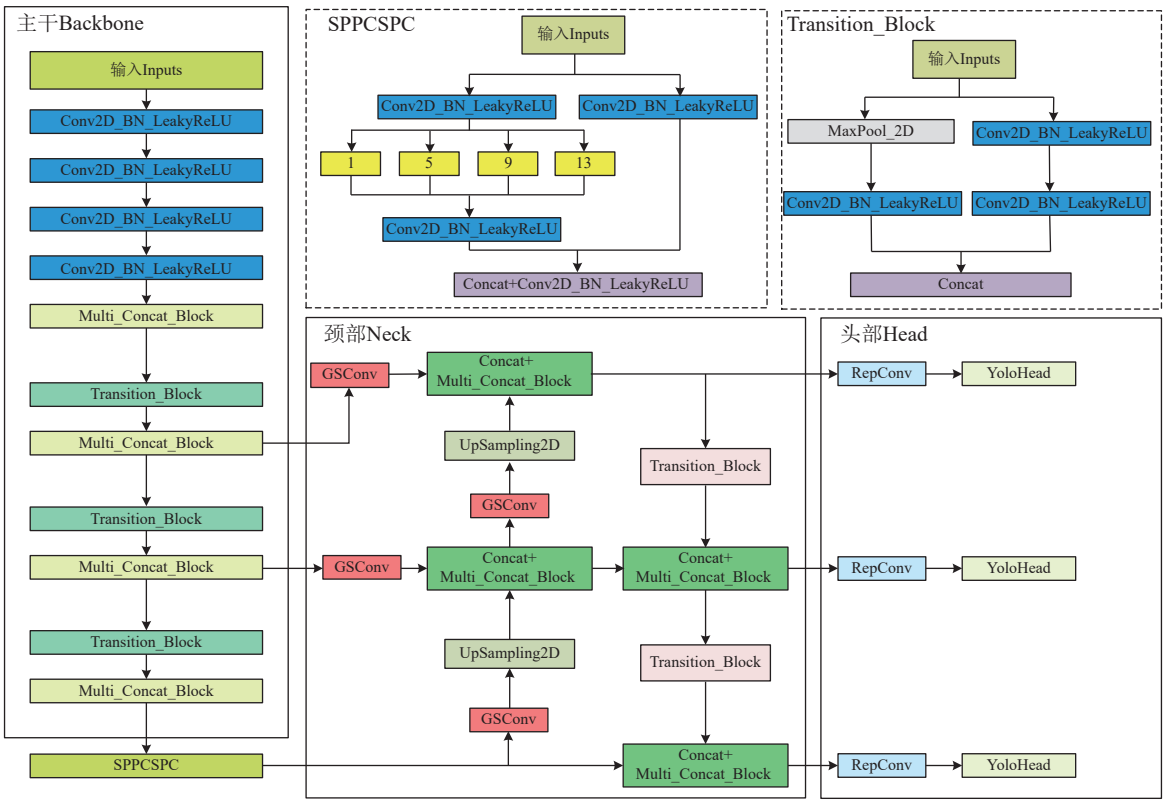
如，刘洁等[28]将橙果简化为球体，通过 YOLOv4 预测橙果质心的三维位置。研究目标质心计算方法同样适用于香蕉采摘中的三维定位，不同于类球型的柑橘，类圆柱型的香蕉果柄具有倾斜性，常规的水平检测算法不能得到其直径信息，因此需要研究带旋转角度的目标检测算法。于俊伟等[29]提出基于旋转框的 R-CNN 麦穗检测模型，该方法通过旋转框有效减少背景区域，提高了检测精度。

上述方法存在如下问题：1）基于深度学习的方法模型较大，耗费大量计算资源，难以满足实时性及模型部署等要求。2）实际蕉园中香蕉果柄常倾斜生长，水平包围框包含较多背景，导致检测精度下降，尤其在三维定位中，质心位置估计更加困难。3）深度相机获取深度图像时，受光照、遮挡等因素影响，常出现部分像素深度信息缺失的空洞现象，导致目标三维定位误差增大。基于此，本文建立复杂环境下的香蕉数据集，改进 YOLOv7 检测网络模型，针对数据样本存在的问题进行算法模型改进与优化，为香蕉机械化采摘提供研究基础。

1 材料与方法

1.1 改进 YOLOv7 旋转目标检测的网络架构

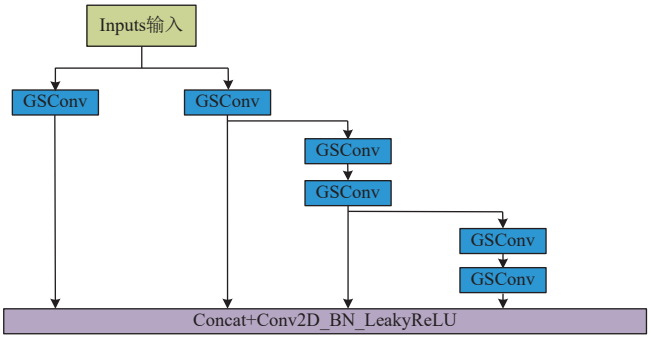
本文提出了 RL-YOLOv7 模型（rotated-lightweight-YOLOv7），通过改进 YOLOv7 模型[30]，提高了香蕉果串和果柄的识别性能。RL-YOLOv7 由 Backbone、Neck 和 Head 三部分组成：Backbone 负责从输入图像中提取特征，Neck 将不同尺度的特征融合以应对复杂场景变化，并通过加入 GSConv 模块减少网络参数和计算量，增强了多尺度特征融合，提高模型在香蕉园复杂场景下的表现力。Head 进行分类和回归，通过 KLD 损失函数对旋转边界框的 5 个参数进行回归计算，KLD 损失函数通过将旋转角度视为概率分布，解决了周期性角度问题，特别适用于长宽比大的香蕉果串检测，从而提高了定位精度。这些改进增强了模型对香蕉果串和果柄的检测能力。



注：Conv2D_BN_LeakyReLU 指卷积层、批量归一化层和激活函数，SPPCSPC 是 SPP 和 CSPN 的结合，SPPCSPC 中的 1，5，9，13 分别代表池化层的核大小，Transition_Block 是处理多尺度信息的转换模块。
Note: The Conv2d_BN_LeakyReLU refers to the convolutional layer, the batch normalization layer and the activation function, reapectively. SPPCSPC is a combination of spatial pyramid pooling (SPP) and cross stage partial Networks (CSPN). In SPPCSPC, 1, 5, 9, and 13 represent the kernel sizes of the pooling layers. The Transition_Block is a module used for processing multi-scale information transitions.

图 1 RL-YOLOv7 网络架构
Fig.1 RL-YOLOv7 network diagram

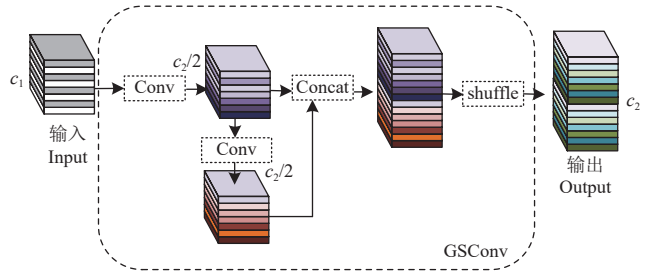
本文采用了多分支堆叠模块生成最终的输入，如图 2 所示。该模块将包含的 4 个特征层进行连接，实现特征的整合，以增强模型对多尺度目标的检测能力。这种多分支堆叠的结构实际上对应着更密集的残差结构。由于残差网络[31]具有易于优化的特点，并且可以通过增加网络的深度提高准确性，因此，在该模块内部使用了跳跃连接的残差块，以解决深度神经网络中梯度消失的问题。多分支堆叠模块通过将多个分支的特征进行整合，可以提高模型对复杂场景的表示和特征提取能力。同时，跳跃连接的运用使得信息能够直接传递到后续层，有效解决了深度增加带来的梯度消失问题，有助于提升模型的训练效果和准确性。



注：GSConv 为 GSConv 卷积，Conv2D_BN_SiLU 为 conv2d 卷积层、Batch Normalization 批量归一化层和 LeakyReLU 激活函数，Concat 表示特征拼接。
Note: GSConv represents group shuffle convolution. Conv2d_BN_SiLU represents the convolutional layer, BN refers to the batch normalization layer, and LeakyReLU activation function.Concat indicates feature concatenation.

图 2　多分支模块堆叠
Fig.2　Muliti concat block

本文引入分组洗牌卷积[32](group shuffle convolution, GSConv)，包括标准卷积、深度可分离卷积和洗牌（shuffle），如图 3 所示。该模块是一种轻量化的卷积方法，可以在减少计算量的同时提升模型的表达能力。GSConv 将输入的特征图分为两部分，一部分进行普通的卷积操作，另一部分进行深度可分离的卷积操作，然后将两部分的输出拼接起来，并进行洗牌操作，使得每个通道的信息都能充分混合。因此，GSConv 既保留了普通卷积的密集连接，又发挥了深度可分离卷积的低计算复杂度优势。



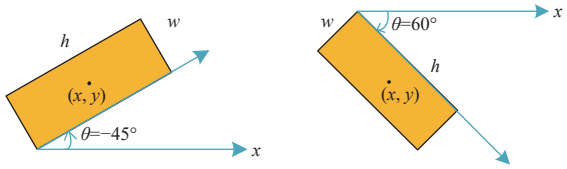
注：c_1 是输入特征图的通道数，c_2 是最终输出特征图的通道数，shuffle 指的是对特征通道进行重新排列。
Note: c_1 is the number of input feature map channels. c_2 is the number of output feature map channels. Shuffle refers to the rearrangement of feature channels.

图 3　GSConv 模块结构图
Fig.3　The structure of the group shuffle convolution (GSConv) module

1.2　旋转边界框的表示方法

传统 YOLO 方法使用水平边界框将物体包围起来，忽略了物体可能存在的旋转或倾斜特性。针对种植园中香蕉果柄具有朝向多样化的特点，本文引入旋转边界框来生成具有方向性的检测框，旋转边界框的边界不一定平行于图像的水平和垂直方向。

本研究的目标位置采用五维坐标 (x,y,w,h,θ) 表示一个旋转边界框，如图 4 所示。角度 θ 用于定义长边 h 与 x 轴之间的角度，其取值范围是 $[-90°,90°)$。其中，90°表示竖直向下，顺时针方向为正方向。由于边界值−90°和 90°实际上是重合的，但存在着 180°的差距，因此需要处理边界值的不连续性，以确保正确解释和计算旋转边界框的角度参数。



a. 负角度示意图
a. Negative Angle diagram

b. 正角度示意图
b. Positive Angle diagram

注：（x，y）为检测框的中心坐标，h，w 为检测框的长和宽，θ 为检测框的旋转角度，（°）。
Note: The (x,y) are the center coordinates of the bounding box, h and w represent the height and width of the bounding box, and θ is the rotation angle of the bounding box, (°).

图 4　旋转框在正负旋转角度下的示意图
Fig.4　Schematic diagram of the rotating frame at positive and negative rotation angles

1.3　旋转框回归损失计算

与 YOLO 系列目标检测算法相比，旋转目标检测算法面临着学习周期性角度参数的挑战，即在周期变化的边界处存在不连续性，可能导致的损失函数值突增，进而增加网络学习和训练难度。为此，本文引入 Kullback-Leibler 散度（Kullback-Leibler divergence, KLD）方法[33]。KLD 是一种用于度量两个概率分布之间差异的方法。在旋转目标检测算法中，可以将角度参数视为一个概率分布，表示目标可能的旋转角度。而周期性角度参数的不连续性则意味着在周期变化的边界处存在着概率分布的突变。为了克服这个问题，可以使用 KLD 方法度量两个相邻周期之间的概率分布差异，并将其作为损失函数的一部分。

基于 IoU 的损失是距离函数，而 KLD 是高斯分布距离的回归损失函数，可将其预测框、真实框的中心点坐标、宽、高和角度信息 B（x,y,w,h,θ）转换为二维高斯分布 $N(\mu,\boldsymbol{\Sigma})$，转换过程用下式表示：

$$\mu=(x,y)^{\mathrm{T}} \tag{1}$$

$$\begin{aligned}\Sigma^{\frac{1}{2}}&=\boldsymbol{RSR}^{\mathrm{T}}\\&=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\frac{w}{2} & 0\\ 0 & \frac{h}{2}\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\\&=\begin{pmatrix}\frac{w}{2}\cos^2\theta+\frac{h}{2}\sin^2\theta & \frac{w-h}{2}\cos\theta\sin\theta\\ \frac{w-h}{2}\cos\theta\sin\theta & \frac{w}{2}\sin^2\theta+\frac{h}{2}\cos^2\theta\end{pmatrix}\end{aligned} \tag{2}$$

式中 $\boldsymbol{\Sigma}$ 指的是真实分布的协方差矩阵，$\boldsymbol{R}$ 表示旋转矩阵，$\boldsymbol{S}$ 表示特征值的对角矩阵。

将目标物体的中心点信息表示为二维高斯分布的均值，并通过旋转矩阵将物体的宽度、高度和角度作为高斯分布的尺度信息，将旋转矩形框映射为二维高斯分布。这种转化解决了角度周期性和边界交换性问题。通过计算真实框与生成框在二维高斯分布中的距离作为回归损失，优化目标是最小化损失函数，从而使预测框更精确地匹配真实框的位置和尺度信息。KLD 损失函数的位置

回归损失可用下式表示：

$$D=\frac{1}{2}(\mu_p-\mu_t)^{\mathrm{T}}\Sigma_t^{-1}(\mu_p+\mu_t)+\frac{1}{2}\boldsymbol{T}_r\left(\Sigma_t^{-1}\Sigma_p\right)+\frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_t|}{|\boldsymbol{\Sigma}_p|}-1 \tag{3}$$

式中 D 指的是 KLD 的损失值，衡量两个概率分布之间的差异，μ_p 指的是预测分布的均值，μ_t 指的是真实分布的均值，$\boldsymbol{T}_r$ 指的是真实分布的协方差矩阵的迹（trace），$\boldsymbol{\Sigma}_t$ 指的是真实分布的协方差矩阵，$\boldsymbol{\Sigma}_p$ 指的是预测分布的协方差矩阵。

KLD 损失函数具备尺度不变性，既不同尺度下损失函数的取值都是相同的，从而减少了预测框大小不同对损失的影响，提高了检测的精度和鲁棒性。此外，当目标的长宽比增大时，KLD 损失函数的导数更加关注角度的优化，从而减小了长宽比较大的目标受角度差异较大带来的影响。因此，KLD 损失函数更适合用于检测长宽比较大的物体，可以提供较高精度的检测结果，例如香蕉果柄和香蕉串。

$$L_{kld}=1-\frac{1}{\tau+\mathrm{sqrt}(D)} \tag{4}$$

式中 L_{kld} 表示边界框的回归损失，其中包含损失值 D 的非线性函数 sqrt(D) 使损失更加平滑，τ 为调节整个损失函数的超参数。

1.4 图像数据集及处理

香蕉图像集由 640×480 分辨率的 Realsense D455 相机于 2022 年 11 月 16 日在广东省广州市白云区采集，包含 1000 张香蕉果串图片。拍摄高度为 1500～1800 mm，香蕉串长 1000～2000 mm，主要特征为绿色生长期的香蕉串和果柄，果串弯曲成簇且果柄倾斜。为了使检测框更紧密包围香蕉果柄，标注的真实框仅包括离果柄最近的竖直部分，以使检测框更紧密地包围香蕉果柄。精确的香蕉串和果柄识别有助于机械化采摘系统定位和摘取香蕉，并为碰撞检测提供支持。

本文分别使用 labelImg 和 roLabelImg 对原始图像中香蕉果柄和果串进行精确标注，如图 5 所示，标签图像中共有 2 种类别：香蕉果柄和香蕉果串。

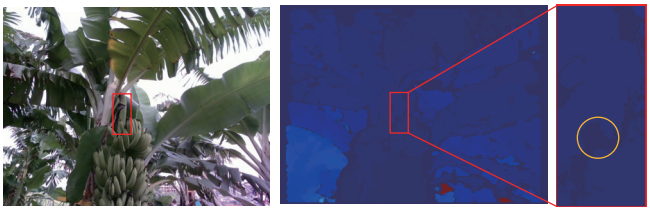
a. 水平标注 b. 旋转标注
a. Horizontal annotation b. Rotate annotation

图 5 水平框与旋转框标注方式对比图

Fig.5 Comparison chart of horizontal boundingbox and rotated boundingbox annotation methods

1 000 张人工标注的数据集按照 8∶1∶1 的比例随机划分为训练集、验证集和测试集，本数据集中香蕉串数量为 2 197，果柄数量为 1 115。此外，为了提高数据集的有效性，采取增强技术对数据进行随机增强，包括旋转、缩放和翻转等方法，以扩展数据集的实际规模。

在野外香蕉园实际应用场景中，深度相机作为一种先进传感器，为香蕉实体深度图像获取提供了强大工具。然而，实际采集中面临多重挑战，由于深度相机受光照和遮挡等环境因素影响，使得深度图中部分像素深度信息缺失，限制了香蕉实体目标三维位置（xyz 中 z 坐标）的准确获取，如图 6 所示，这将导致采摘机器人视觉系统的定位错误。



a. 彩色图 b. 深度图
a. RGB image b. Depth image

注：红框为香蕉果柄的区域，黄色圆圈为深度图上的空洞区域。
Note: The red box are the areas of the stalk of the banana, and the yellow circles are the empty areas on the depth image.

图 6 目标果柄深度图空洞现象示意图

Fig.6 Schematic diagram of hole phenomenon in the target stalk

因此本文基于 Criminisi 算法[34] 对深度图像进行修复，旨在填补深度图中的缺失区域。Criminisi 算法基于对图像结构的理解和周围信息的推断，能够智能填充空洞，提升深度图的完整性和准确性。首先，该算法通过分析深度图中的亮度和对比度变化，从而推断空洞区域可能的深度值。其次，Criminisi 算法利用邻近像素之间的关联性，通过对周围已知深度信息的利用，智能填充缺失的深度值。

1.5 香蕉果柄定位原理

在深度相机获取深度信息和二维彩色图像后，利用 RL-YOLOv7 识别模型从彩色图像中提取的香蕉果柄二维中心点位置，并将其映射到深度图中，从而获取香蕉果柄的深度信息，实现三维空间定位，如图 7 所示。香蕉果柄可近似用一个圆柱表示。

根据相似三角形原理可得：

$$\frac{f}{z+r}=\frac{w}{2r} \tag{5}$$

由式 (5) 得，

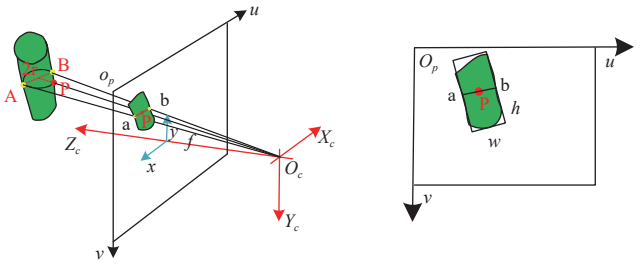
$$r=\frac{zw}{2f-w} \tag{6}$$

式中 f 为相机内参。因此，果柄几何中心的实际深度值可表示为 d=z+r。

像素坐标系与相机坐标系之间的转换关系为

$$d\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x&0&u_0\\0&f_y&v_0\\0&0&1\end{bmatrix}\begin{bmatrix}x_c\\y_c\\z_c\end{bmatrix} \tag{7}$$

式中 $\{f_x, f_y\}$ 和 $\{u_0, v_0\}$ 分别为相机的焦距和主点，$\{u, v\}$ 为像素坐标系下的坐标，$\{x_c, y_c, z_c\}$ 为相机坐标系下的坐标。



a. 相机坐标系投影原理图
a. Camera coordinate system projection principle picture

b. 图像坐标系二维成像图
b. Image coordinate system 2D imaging picture

注：P 为靠近相机侧的果柄质心外表面的空间点，AB 为果柄直径，用 2r 表示，其在相机图像坐标中的投影为 ab，用 w 表示，f 为相机的焦距，xy 为图像坐标系，o_p-uv 为像素坐标系，O_c-$X_cY_cZ_c$ 为相机坐标系。
Note: P is the space point on the outer surface of the centroid of the fruit stalk near the camera side. AB represents the diameter of the fruit stalk, denoted as 2r, and its projection in the camera image coordinates is ab, denoted by w. f is the focal length of the camera. xy represents the image coordinate system, o_p-uv represents the pixel coordinate system, and O_c-$X_cY_cZ_c$ refers to the camera coordinate system.

图 7　香蕉果柄质心三维位置计算原理图
Fig.7　The schematic diagram for calculating the three-dimensional position of the banana stalk's centroid

因此物体在空间中的三维坐标 (x_c, y_c, z_c) 表达式为

$$t=\begin{cases}x_c=\dfrac{d\cdot(u-u_0)}{f_x}\\ y_c=\dfrac{d\cdot(v-v_0)}{f_y}\\ z_c=d\end{cases}\tag{8}$$

将香蕉的朝向简化为仅绕相机坐标系的 z 轴旋转，因此可得香蕉相对于相机坐标系的姿态为

$$\boldsymbol{R}=\begin{bmatrix}\cos\theta_c & -\sin\theta_c & 0\\ \sin\theta_c & \cos\theta_c & 0\\ 0 & 0 & 1\end{bmatrix}\tag{9}$$

式中 θ_c 为相机坐标系中的角度，由式（8）和式（9）可得香蕉果柄相对于相机坐标系的位姿表达式为

$$T=\begin{bmatrix}\cos\theta_c & -\sin\theta_c & 0 & x_c\\ \sin\theta_c & \cos\theta_c & 0 & y_c\\ 0 & 0 & 1 & z_c\\ 0 & 0 & 0 & 1\end{bmatrix}\tag{10}$$

$\boldsymbol{T}$ 为姿态转换矩阵，由于图像坐标系中 θ 为长边定义法，与相机坐标系中角度 θ_c 定义不一致，因此需要进行角度转换，转换关系如下：

$$\theta_c=\begin{cases}-90+\theta,\theta\in[0,90]\\ 90+\theta,\theta\in[-90,0)\end{cases}\tag{11}$$

1.6　试验参数和评价指标

本文试验均在基于 python 版本的 Pytorch 深度学习框架下进行，将上述算法在自建的香蕉数据集中进行验证评估，所提算法在 Ubuntu18.04 操作系统运行，CPU 为 Intel(R) Xeon(R) Gold 5 218 CPU @ 2.30 GHZ，RAM 为 252 G，GPU 是 Quadro RTX 5 000，CUDA 为 10.2，具体试验训练策略：批次大小（batch size）设置为 16，模型将进行 300 个训练轮次（epochs），初始学习率设定为 0.01，使用 0.937 的动量系数以加速梯度下降并防止模型陷入局部最优解，输入图像的尺寸为 640×640 像素，确保模型处理的图像数据具有统一的大小。为验证所香蕉检测算法的有效性，本文采用召回率（recall）、平均精确率（average precision, AP）、F1、参数量、检测速度（frames per second, FPS）、FLOPs（floating point operations）作为评价指标。

在定位精度上，本文将水平检测算法和旋转检测算法作对比。在香蕉果园内进行了 5 个目标的三维定位试验，并记录试验数据，使用激光测距仪测量 z 和游标卡尺测量香蕉果柄半径 r，并记录对应点在图像的 u、v 坐标，为了降低测量误差，一次测量同时采集三组数据取平均值，并通过式（5）和式（8）计算得到目标在相机坐标系下的深度距离 z_t=z+r，作为真实值。记录试验中的 z_c 和 z_t，E_v 表示估计值与真实值之间的绝对误差，而 E_{vr} 则代表两者之间的相对误差。其计算式如下：

$$E_v=\frac{\sum_{i=1}^{m}|z_{ci}-z_{ti}|}{m}\tag{12}$$

$$E_{vr}=\frac{\sum_{i=1}^{m}\left|\dfrac{z_{ci}-z_{ti}}{z_{ci}}\right|}{m}\tag{13}$$

式中 z_c 为激光测距仪测量值，z_t 为深度相机获取的估计值，m 为同一幅图像成功识别并定位的香蕉果柄个数，i 为同一幅图像中的第 i 次个果柄。

2　结果与分析

2.1　旋转框定位的 YOLOv7 的目标检测结果

2.1.1　旋转框与水平框的效果对比图

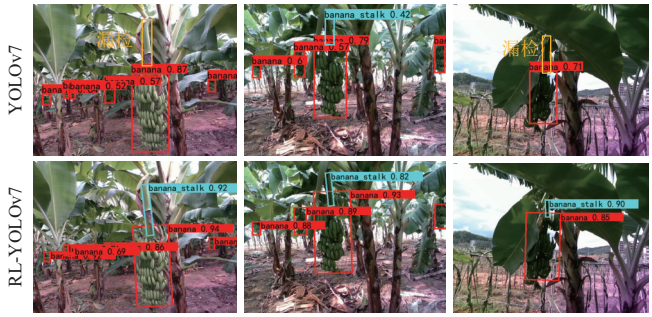
为了对比水平检测框算法 YOLOV7 和旋转检测算法 RL-YOLOV7 之间的效果，针对香蕉目标检测进行了对比，检测效果图如图 8 所示，试验结果显示 RL-YOLOv7 算法输出的预测框能够更好地贴合香蕉果柄的真实轮廓，更准确地反映香蕉果柄在蕉园中的朝向情况。当香蕉果柄以倾斜的角度密集排列时，RL-YOLOv7 算法的检测效果更好，且不易漏检。

如表 1 所示，两种检测方法对香蕉串的检测正确率非常接近，但是对香蕉果柄的检测正确率差别较大。YOLOv7 模型香蕉果柄检测正确率低于香蕉串的检测正确率，且对香蕉果柄的漏检严重，56 根香蕉果柄没有检测到，而香蕉串有 6 个没检测到。相比之下，RL-YOLOv7 有较好表现，对香蕉串的检测正确率和漏检率分别为 98.1%和 1.8%，香蕉果柄的的检测正确率和漏检率分别为 7.7% 和 2.6%，在同样检测到目标的情况下，RL-YOLOv7 的置信度比 YOLOv7 要高。

2.1.2　模型性能对比

不同模型在测试集的检测效果如表 2 所示。从表中可以看出，RL-YOLOv7 的香蕉串和香蕉果柄的 AP 值分别为94.36% 和95.29%，高于YOLOv5（AP 分别为92.61% 和 63.64%）、YOLOv7（AP 分别为 91.33% 和 63.71%）、YOLOv7-tiny（AP 分别为 94.50% 和 78.08%）。RL-

YOLOv7 具有更高的召回率和准确率，并且推理时间、参数数量和计算量与 YOLOv7-tiny 相似，平均识别准确率和 F1 比 YOLOV7 提升了 17.04 和 40.5 个百分点，同时模型减小了 32.16MB。而且可以看出，水平检测算法在香蕉果柄的准确率远远低于香蕉串的准确率，这是因为香蕉果柄较小，而且与香蕉园中的背景较为相似（如香蕉叶柄），所以更难发现香蕉果柄。而 RL-YOLOv7 引入 KLD 损失函数使得香蕉果柄的损失回归变得更加顺利，因此香蕉果柄的识别准确率提高，且稍高于香蕉串的准确率。此外 RL-YOLOv7 具有更高的召回率和准确率，更短的推理时间和较小的参数量。



a. 图像 1
a. Image 1

b. 图像 2
b. Image 2

c. 图像 3
c. Image 3

注：红色标注框为香蕉串检测对象，蓝色标注框为香蕉果柄检测对象，黄色标注框为算法漏检对象

Note: The red bounding boxes represent the detection objects of banana bunches, the blue bounding boxes represent the detection objects of banana stalks, and the yellow bounding boxes in the figure is the missing object of the algorithm.

图 8 RL-YOLOv7 与 YOLOv7 的检测结果对比

Fig.8 Comparison of detection results of RL-YOLOv7 and YOLOv7

表 1 YOLOv7 和 RL-YOLOv7 模型下的检测结果

Table 1 Detection results under YOLOv7 and RL-YOLOv7 models

类别 Category	总数 Totality	模型 Model	检测正确 Correct detection		检测错误 Error detection		漏检 Missing detection	
			个数 Number	百分比 Percent/%	个数 Number	百分比 Percent/%	个数 Number	百分比 Percent/%
香蕉串 Banana cluster	272	YOLOv7	266	97.8	0	0	6	2.2
		RL-YOLOv7	267	98.1	1	0.36	5	1.8
香蕉果柄 Banana stalk	133	YOLOv7	77	57.9	0	0	56	42.1
		RL-YOLOv7	130	97.7	0	0	3	2.6

表 2 不同检测网络模型的检测结果

Table 2 Detection results of different detection network models

模型 Model	AP/%		mAP/%	召回率 R/%	*F*1/%	推理时间 Inference time/ms	参数 Params/MB	FLOPs
	香蕉串	果柄						
YOLOv5	92.61	63.64	78.12	40.01	46.00	11.60	7.06	8.24
YOLOv7	94.50	63.71	79.11	50.11	55.00	24.75	37.20	52.56
YOLOv7-tiny	91.10	85.42	88.26	58.72	75.00	8.56	6.02	6.59
RL-YOLOv7	94.78	97.51	96.15	95.54	95.50	9.15	5.04	5.73

注：AP 为精确率，mAP 为平均精确率，*R* 为召回率，*F*1 为调和平均值，FLOPs 为浮点运算数。

Note: AP is the average precision, mAP is the mean average precision, *R* is recall, *F*1 is harmonic average, FLOPs is the floating point operations.

从表 3 中的 Baseline+KLD 可以看出，引入 KLD 损失函数后，网络的性能明显提高。

表 3 YOLOV7tiny 的消融试验

Table 3 Abltion experiment of YOLOV7tiny

模型 Models	参数 Params/MB	FLOPs/GB	mAP/%	*R*/%	*F*1/%	推理时间 Inference time/ms	FPS/(帧·s^{-1})
Baseline(CBS)	6.02	6.59	88.26	58.72	75.00	8.56	116
Baseline+ CBL	6.02	6.59	84.71	57.13	68.00	8.12	122
Baseline+Ghost	5.04	5.73	84.12	49.86	59.50	8.78	113
Baseline+ GSConv	5.04	5.73	82.74	46.00	54.87	9.19	108
Baseline+KLD, CBS	6.02	6.59	94.49	82.49	86.50	8.25	121
Baseline+KLD, CBL	6.02	6.59	94.83	94.68	92.50	8.18	122
Baseline+KLD, Ghost	5.04	5.73	93.42	82.27	86.50	8.74	114
Baseline+KLD, GSConv	5.04	5.73	96.15	95.54	95.50	9.15	109

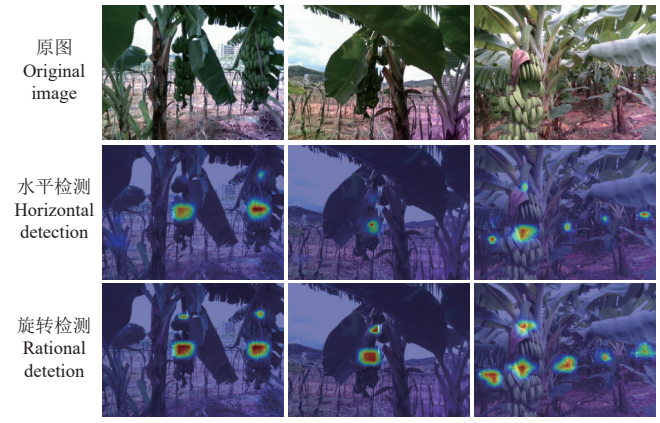
注：Baseline 为 YOLOV7tiny，CBS 为 conv2d 卷积层 、 Batch Normalization 批量归一化层 和 SiLU 激活函数的组合，CBL 为 conv2d 卷积层 、 Batch Normalization 批量归一化层 和 LeakyReLU 激活函数的组合，Ghost 表示 Ghost 卷积，GSConv 表示 GSConv 卷积，KLD 为 Kullback-Leibler 散度损失函数，FPS 为帧速率。

Note: The Baseline is YOLOV7tiny, CBS is the combination of conv2d convolution layer, Batch Normalization layer and SiLU activation function,CBL is the combination of conv2d convolution layer, Batch Normalization layer and LeakyReLU activation function number. Ghost stands for Ghost convolution, GSConv stands for group shuffle convolution, and KLD stands for kullback-leibler divergence loss function, FPS is frame per second.

对比 Baseline+KLD 在引入 CBL、Ghost 和 GSConv 模块后，GSConv 的准确率最高，说明 GSConv 比 CBL 和 Ghost 具有更强的特征提取能力。表 3 展示了香蕉数据集的试验结果。使用带有 KLD 和 GSConv 的网络具有更好的性能，与 baseline 相比 mAP 提升了 7.89 个百分点，而且参数量和计算量明显降低。

2.1.3 RL-YOLOv7 模型的特征可视化分析及效果对比

Grad-CAM 可以对模型的检测效果进行特征可视化分析为图 9，可以看出基于旋转检测的 RL-YOLOv7 的高响应区域的确集中在核心部位的部位，相比于水平检测的 YOLOv7，RL-YOLOv7 在香蕉果柄上具有更大的响应区域，而 YOLOv7 在香蕉果柄上的响应区域微乎其微，由此也可以解释表 1 中 YOLOv7 香蕉果柄漏检率高的原因，而 RL-YOLOv7 用于香蕉串检测也具有更大的响应区域，如图 9 中旋转检测的图像 3，即使检测远处的小目标也有较好的效果，可见，RL-YOLOv7 对目标对象具有更高的关注度和注意力。



a. 图像 1
a. Image 1

b. 图像 2
b. Image 2

c. 图像 3
c. Image 3

图 9 YOLOV7 与 RL-YOLOV7 的网络可视化特征对比

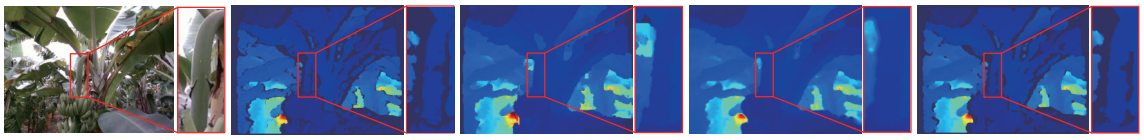
Fig.9 Comparison of network visualization features between YOLOV7 and RL-YOLOV7

2.2 深度图像修复效果和香蕉果柄定位精度对比

试验结果显示，Criminisi 算法、快速行进算法和中值

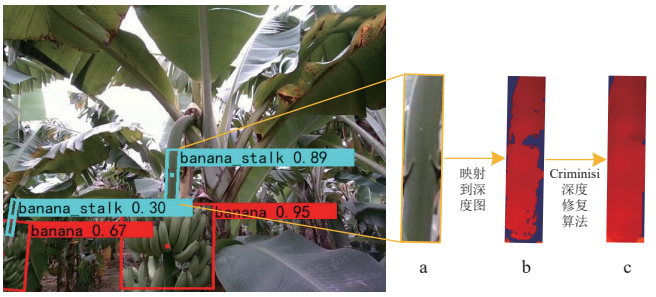
滤波算法的测量时间分别为 0.605、202.644 和 1.792 s。尽管中值滤波器对小空洞有一定修复能力，但对大空洞效果不佳，且存在噪声，如图 10 所示。快速行进算法能修复所有空洞，但在边缘产生伪影，且处理时间较长，约为 202.644 s。相比之下，Criminisi 算法能够有效修复大多数空洞，且修复后的图像边缘更清晰，保留较好的边缘信息，处理时间较短，仅为 0.605 s，具有更好的实时性。可见，本文所提方法可以弥补中值滤波和文献 [35] 中算法的不足。

考虑到实际采摘机器人在定位中只需利用香蕉果柄质心的三维信息和采摘工作范围内的香蕉，因此可以简化深度图像修复处理流程，仅对香蕉果柄的目标区域进行深度图像修复，以缩短运行时间。用网络预测得到的香蕉果柄检测区域对深度图像进行裁剪，图 11 是裁剪后的深度图像修复效果图，其深度修复处理时间为 0.006 52 s，发现 Criminisi 算法的修复效果较好，且运行时间大大减少。



a. 彩色图　b. 深度图　c. Criminisi算法　d. 快速行进算法　e. 中值滤波算法
a. RGB image　b. Depth image　c. Criminisi algorithm　d. FMM algorithm　e. Mean filteringalgorithm

图 10　深度图像修复算法效果对比
Fig.10　Comparison of the effects of the deep image inpainting algorithm



注：a 为香蕉果柄目标检测框的局部放大，b 为 a 的对应深度局部放大图，c 为 b 经过 Criminisi 深度修复后的深度局部放大图。
Note: a is a local enlargement of the target detection box for the banana stalk, b is the corresponding local enlargement of the depth image for a, and c is the local enlargement of b after depth restoration using the Criminisi method.

图 11　Criminisi 算法修复香蕉果柄目标区域图
Fig.11　Criminisi algorithm repaired target region image of banana stalk

表 4 为香蕉果柄定位结果。

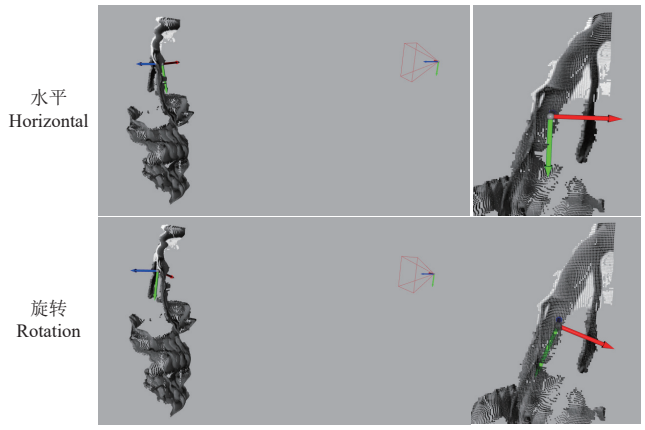
表 4　香蕉果柄定位深度值结果
Table 4　Depth value comparison of banana stalk positioning

编号 Number	实际测量 Actual measurement $z_{c/mm}$	YOLOV7			RL-YOLOV7 的		
		z_{c1} /mm	E_v/mm	E_{vr}/%	z_{c2} /mm	E_v/mm	E_{vr}/%
1	1 400.00	1 364.0	36.00	2.57	1 402.5	2.50	0.18
2	1 313.50	1 287.0	26.50	2.02	1 317.88	4.38	0.33
3	953.50	0	–	–	964.62	11.12	1.17
4	1 079.40	1 047.0	32.40	3.0	1 083.42	4.01	0.37
5	1 073.40	1 043.0	30.40	2.83	1 086.51	13.10	1.22
平均值 Mean	–	–	31.32	2.61	–	7.02	0.65

注：z_{c1} 和 z_{c2} 分别为 YOLOv7 和 RL-YOLOv7 检测的深度值；E_v、E_{vr} 分别为误差均值和平均相对误差。
Note: z_{c1} and z_{c2} are the depth values corresponding to the YOLOv7 and RL-YOLOv7, respectively. E_v and E_{vr} are the mean error and the mean error ratio respectively.

如表 4 所示，在进行深度图像修复前的 5 次试验中，本文算法所得到的误差均值 *Ev* 和误差比均值分别为 7.02 mm 和 0.65%；而水平检测的 YOLOv7 的误差均值 *Ev* 和误差比均值分别为 31.32 mm 和 2.61%，因为改进的旋转 YOLOv7 算法加上了预测的香蕉果柄半径，有效地降低香蕉果柄质心定位的误差均值和误差比均值的；除此之外，表 4 中编号 3 的深度估计值为 0，原因是深度相机受光照影响或者香蕉果柄表面存在黑色斑点等因素导致无法获取该位置的深度信息，导致出现定位错误，而 RL-YOLOv7 使用 Criminisi 算法对目标区域进行空洞修复可以有效弥补局部空洞导致的定位错误，相比于 YOLOv7，误差均值降低 24.3 mm，误差比均值下降 1.96 个百分点。

综上所述，改进的 RL-YOLOv7 模型对香蕉果柄定位能够达到试验目的，对比原始 YOLOv7 模型，其效果有所提升。图 12 是原始 YOLOv7 和改进的 RL-YOLOv7 在相机视角下的三维定位可视化图，两者之间相差了一个香蕉果柄半径的距离，且 RL-YOLOv7 可以更好拟合香蕉果柄的姿态。



a. 相机坐标系下的全局视角　b. 局部放大图
a. Global view in the camera coordinate system　b. Locally enlarged image

图 12　相机视角下的香蕉果柄定位可视化图
Fig.12　Visualization of banana stalk positioning from a camera perspective

3　结　论

为了实现果园环境下香蕉果柄的快速识别定位，本研究在 YOLOv7 模型的基础上添加了旋转角度参数和 GSConv 模块，并采用 KLD 回归损失函数替换原先的 SmoothL1 损失函数。对于采用旋转框的目标检测，本文

利用香蕉果柄定位点五维坐标和旋转角度，通过深度相机获取定位点的三维坐标。本研究得到以下结论：

1）与 YOLOv7 模型相比，改进的 RL-YOLOv7 模型对香蕉串的平均识别准确率提升了 17.04 个百分点，F1 值提升了 40.5 个百分点，且模型减小了 32.16 MB。本算法的改进模型在模型大小和平均精度值方面都得到大幅提升，且减少了模型容量大小，有利于模型的迁移应用。

2）改进的 RL-YOLOv7 模型可以有效预测香蕉果柄的三维位置，其误差均值和误差比均值分别为 7.02 mm 和 0.65%。较 YOLOv7 模型降低了 24.3 mm 和 1.96 个百分点，并且利用深度补全算法 Criminisi 能有效修复定位过程中出现的卷积空洞。改进模型的定位试验误差低，进一步减小了对香蕉果柄的定位误差。

[参 考 文 献]

[1] FU L, Tola E, Al-Mallahi A, et al. A novel image processing algorithm to separate linearly clustered kiwifruits[J]. Biosystems Engineering, 2019, 183: 184-195.

[2] REIS M J C S, MORAIS R, PERES E, et al. Automatic detection of bunches of grapes in natural environment from color images[J]. Journal of Applied Logic, 2012, 10(4): 285-290.

[3] Cubero S, Diago M P, Blasco J, et al. A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis[J]. Biosystems Engineering, 2014, 117: 62-72.

[4] WANG C, LEE W. S, ZOU X , et al. Detection and counting of immature green citrus fruit based on the local binary patterns (LBP) feature using illumination-normalized images[J]. Precision Agriculture, 2018, 19, 1062–1083.

[5] NUSKE S, WILSHUSEN K, ACHAR S, et al. Automated visual yield estimation in vineyards[J]. Journal of Field Robotics, 2014, 31, 837–860.

[6] BOOGAARD F P, RONGEN K S A H, KOOTSTRA G W. Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging[J]. Biosystems Engineering, 2020, (192): 117-132.

[7] 熊俊涛，林睿，刘振，等. 夜间自然环境下荔枝采摘机器人识别技术[J]. 农业机械学报，2017，48(11)：28-34.
XIONG Juntao, LIN Rui, LIU Zhen, et al. Visual technology of picking robot to detect litchi at nighttime under natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(11): 28–33. (in Chinese with English abstract)

[8] KUZNETSOVA A, MALEVA T, SOLOVIES V. Using YOLOv3 algorithm with pre-and post-processing for apple detection in fruit-harvesting robot[J]. Agronomy, 2020, 10(7): 1016.

[9] WEI X, JIA K, LAN J, et al. Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot[J]. Optik - International Journal for Light and Electron Optics, 2014, (125): 5684–5689.

[10] GUO Q, CHEN Y, TANG Y, et al. Lychee fruit detection based on monocular machine vision in orchard environment[J]. Sensors, 2019, 19(19): 4091.

[11] 徐胜勇，卢昆，潘礼礼，等. 基于 RGB-D 相机的油菜分枝三维重构与角果识别定位[J]. 农业机械学报，2019，50(2)：21-27.
XU Shengyong, LU Kun, PAN Lili. 3D reconstruction of rape branch and pod recognition based on RGB-D camera[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50 (2): 21–27. (in Chinese with English abstract)

[12] NEUPANE B, HORANONT T, HUNG N D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV)[J]. PloS one, 2019, 14(10): e0223906.

[13] FU L, WU F, ZOU X, et al. Fast detection of banana bunches and stalks in the natural environment based on deep learning[J]. Computers and Electronics in Agriculture, 2022, (194): 106800.

[14] LIU Y, WANG X. SAR ship detection based on improved YOLOv7-Tiny[C]//2022 IEEE 8th International Conference on Computer and Communications (ICCC). New York: IEEE, 2022: 2166-2170.

[15] ZHANG C, KANG F, WANG Y. An improved apple object detection method based on lightweight YOLOv4 in complex backgrounds[J]. Remote Sensing, 2022, 14(17): 4150.

[16] FU L, YANG Z, WU F, et al. YOLO-Banana: A lightweight neural network for rapid detection of banana bunches and stalks in the natural environment[J]. Agronomy, 2022, 12(2): 391.

[17] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPPR). Las Vegas, NV, USA: IEEE, 2016: 779-788.

[18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. 2016: 21-37.

[19] WANG Z, JIN L, WANG S, et al. Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system[J]. Postharvest Biology and Technology, 2022, 185: 111808.

[20] WANG Y, YAN G, MENG Q, et al. DSE-YOLO: Detail semantics enhancement YOLO for multi-stage strawberry detection[J]. Computers and Electronics in Agriculture, 2022, 198: 107057.

[21] FU L, FENG Y, WU J, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. Precision Agriculture, 2021, 22: 754-776.

[22] YU Y, YANG X, LI J, et al. A cascade rotated anchor-aided detector for ship detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 60: 1-14.

[23] ZHOU K, ZHANG Z, GAO C, et al. Rotated feature network for multiorientation object detection of remote-sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(1): 33-37.

[24] ZHANG Z, GUO W, ZHU S, et al. Toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(11): 1745-1749.

[25] 段洁利，王昭锐，邹湘军，等. 采用改进 YOLOv5 的蕉穗识别及其底部果轴定位[J]. 农业工程学报，2022，38(19)：122-130.
DUAN Jieli, WANG Zhaorui, ZOU Xiangjun, et al. Recognition of bananas to locate bottom fruit axis using improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(19): 122-130. (in Chinese with English abstract)

[26] GAO G, SHUAI C, WANG S. Mature tomato recognition and location algorithm based on binocular vision and deep learning[C]//International Conference on Artificial Intelligence, Virtual Reality, and Visualization (AIVRV 2022). SPIE, 2023, 12588: 209-215.

[27] 王金鹏，何萌，甄乾广，等. 基于 COF-YOLOv5s 的油茶果识别定位[J]. 农业工程学报，2024，40(13)：179-188.
WANG Jinpeng, HE Meng, ZHEN Qianguang, et al. Camellia

oleifera fruit harvesting in complex environment based on COF-YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(13): 179-188. (in Chinese with English abstract)

[28] 刘洁，李燕，肖黎明，等. 基于改进 YOLOv4 模型的橙果识别与定位方法[J]. 农业工程学报，2022，38(12)：173-182.
LIU Jie, LI Yan, XIAO Liming, et al. Recognition and location method of orange based on improved YOLOv4 model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(12): 173-182. (in Chinese with English abstract)

[29] 于俊伟，陈威威，郭园森，等. 基于改进 Oriented R-CNN 的旋转框麦穗检测与计数模型[J]. 农业工程学报，2024，40(6)：248-257.
YU Junwei, CHEN Weiwei, GUO Yuansen, et al. Improved Oriented R-CNN-based model for oriented wheat ears detection and counting[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(6): 248-257. (in Chinese with English abstract)

[30] 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[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPPR). Vancouver, BC, Canada: IEEE, 2023: 7464-7475.

[31] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.

[32] LI H , LI J, WEI H , et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicle[EB/OL]. (2022-08-17)[2023-12-01]https://arxiv.org/pdf/2206.02424.

[33] YANG X, YANG X J, YANG J R, et al. Learning highprecision bounding box for rotated object detection via kullback leibler divergence[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2021: 18381-18394.

[34] CRIMINISI A, Pérez P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[35] SETHIAN J A. Fast marching methods[J]. SIAM Review, 1999, 41(2): 199-235.

Detecting and locating banana targets using improved YOLOv7 and rotational bounding box positioning

WU Rongda[1,2] , ZHANG Shi'ang[5] , FU Genping[4] , CHEN Tianci[3] , LAI Yingjie[2] , LUO Wenxuan[2] , GUO Xiaogeng[2] , ZHU Lixue[2※]

(1. *Jiangmen Power Supply Bureau*, *Guangdong Power Grid Co.*, *Ltd.*, *Jiangmen* 529000, *China*; 2. *School of Electro Mechanical Engineering*, *Zhongkai University of Agricultre and Engineering*, *Guangzhou* 510225, *China*; 3. *College of Engineering*, *South China Agricultural University*, *Guangzhou* 510642, *China*; 4. *School of Automation*, *Zhongkai University of Agricultre and Engineering*, *Guangzhou* 510225, *China*; 5. *School of Innovation and Entrepreneurship*, *Zhongkai University of Agricultre and Engineering*, *Guangzhou* 510225, *China*)

Abstract: Current target detection cannot adapt to the tilted targets so far. Particularly, the objects are not aligned vertically or horizontally, such as the bananas in orchards that typically grow at varying angles. Furthermore, the large number of parameters are also confined to deploy on the embedded devices with limited computing resources. Moreover, the high computational requirements of advanced models have hindered their deployment on embedded systems, which are commonly used in agricultural automation. In this study, an improved YOLOv7 algorithms was proposed for the banana target detection and localization with a rotational positioning frame. GSConv module was also incorporated to reduce the computational complexity and the number of parameters in the model for the high detection accuracy. Specifically, the GSConv module was the lightweight convolutional structure to maintain the model efficiency, thus more feasible for the real-time applications on embedded platforms with constrained hardware. The banana target was firstly positioned using a rotational bounding box, and then defined by a five-parameter representation. The tilted objects were better handled, as the rotational bounding box was used to accurately represent the various angles at which bananas grow. Furthermore, the Kullback-Leibler divergence loss function was employed to map the rotational frame into a two-dimensional Gaussian distribution. As such, the difference between two probability distributions (the predicted box and the ground truth box) was calculated to optimize as the loss function. The KL divergence-based approach also provided a more precise measure on the difference between the predicted and the actual bounding box, thus improving the localization accuracy. Another, Criminisi algorithm was integrated to solve the problems on the local holes of depth camera, where depth information was missing, due to the influencing factors, such as light interference or occlusions. Among them, the image inpainting was implemented to fill the missing information, and then correct the positioning errors in the depth data, further enhancing the accuracy of three-dimensional localization in the real-world orchard environments. Experimental results show the significant improvements were achieved in the detection speed and accuracy of banana targets using the improved model. Specifically, the average detection accuracy reached 96.15%, which was an impressive 17.04% increase over the standard YOLOv7 model. Additionally, the detection frame rate of the improved model was boosted by approximately 40 frames per second, highly suitable for real-time applications in agricultural settings. Moreover, the position of banana stems was predicted to notably enhance using the rotational boundary frame. The mean positioning error was reduced to 7.02 mm, and the mean error ratio was now 0.65%, which were reduced to 24.3 mm and 1.96%, respectively, compared with the original YOLOv7. In conclusion, the improved model can offer an effective solution to the fast and accurate identification and localization of banana bunches and fruit stalks in complex orchard environments. The higher detection accuracy was also achieved to significantly reduce the computational requirements, particularly for the real-time agricultural applications on embedded devices.

Keywords: bananas; picking; YOLOv7; three-dimensional localization; rotational bounding box; depth image restoration