

# 基于改进 YOLOX 的猕猴桃分类识别与空间定位<sup>\*</sup>

刘博文 徐卫平 徐 钦 陈华伟  
(贵州师范大学机械与电气工程学院 贵阳 550025)

**摘 要:**复杂自然环境下果实的快速、准确识别和定位是猕猴桃机械臂采摘的关键和难点。面向猕猴桃实时采摘应用,开展了基于深度学习的猕猴桃分类识别算法、空间定位模型构建和机械臂采摘实验研究。优选 Anchor\_Free 的一阶段 YOLOX 算法模型,并通过引入注意力机制、调整通道数、增加特征融合、堆叠特征融合对该模型特征融合环节加以改进,提高了非结构化环境下猕猴桃果实的分类识别精度。实验数据表明,改进 YOLOX 算法对独立果实、树叶遮挡、树干遮挡、果实重叠的平均精度(AP)分别提升至 99.38%、98.43%、93.82% 和 92.98%,平均精度均值(mAP)达到 96.15%,其平均  $F_1$  值提升至 94.24%,且在 640×640 分辨率下每幅图像的平均检测时间为 0.019 s。进一步地,利用图像深度信息对识别目标进行三维空间位置求解,并经物理实验验证,实验中果实定位坐标误差 $\leq 3.3\%$ 。算法在识别率、检测时间和定位误差 3 方面取得了较优的综合性能,满足猕猴桃机械臂采摘中果实的实时识别及定位需求。

**关键词:**机械臂采摘;目标分类识别;深度学习;YOLOX;空间定位

**中图分类号:** S126; TP249      **文献标识码:** A      **国家标准学科分类代码:** 510.40; 520.20

## Classification identification and spatial localization of kiwifruit based on the improved YOLOX

Liu Bowen Xu Weiping Xu Qin Chen Huawei

(School of Mechanical and Electrical Engineering, Guizhou Normal University, Guiyang 550025, China)

**Abstract:** Fast and accurate identification and localization of fruits in complex natural environments is the key and difficult point of kiwifruit robotic arm picking. For kiwifruit real-time picking applications, deep learning-based kiwifruit classification and recognition algorithms, spatial localization model construction and robotic arm picking experimental research were carried out. The one-stage YOLOX algorithm model of Anchor\_Free is preferred, and the feature fusion link of this model is improved by introducing an attention mechanism, adjusting the number of channels, increasing feature fusion, and stacking feature fusion, which improves the accuracy of classifying and recognizing kiwifruit fruits in unstructured environments. The experimental data show that the AP of the improved YOLOX algorithm for no occluded, leaf occluded, branch occluded and fruit occluded is improved to 99.38%, 98.43%, 93.82% and 92.98%, respectively, the mAP reaches 96.15%, its average  $F_1$  value is improved to 94.24%, and its average detection time is 0.019 s/space at 640×640 resolution. Further, the three-dimensional spatial position of the recognized target is solved using the image depth information and verified by physical experiments, in which the fruit localization coordinate error is  $\leq 3.3\%$ . The algorithm achieves superior overall performance in terms of recognition rate, detection time and localization error to meet the real-time fruit recognition and localization requirements in kiwifruit robotic arm picking.

**Keywords:** robotic arm picking; target categorization and identification; deep learning; YOLOX; spatial localization

## 0 引 言

我国是猕猴桃种植大国,猕猴桃产量和种植面积居

世界首位<sup>[1]</sup>。在西南山地,贵州省猕猴桃种植面积位居全国第3,其猕猴桃是脱贫攻坚、精准扶贫的重要开发作物<sup>[2]</sup>。目前,猕猴桃主要通过人工采摘,在人口老龄化、农

收稿日期:2023-10-05

<sup>\*</sup> 基金项目:国家自然科学基金(72061006)、贵州省林业局林业科学技术研究项目(黔林科合[2022]26号)资助

村剩余劳动力缺口大等背景下,猕猴桃采摘成本急剧增加。机械臂自动化采摘猕猴桃对于缓解劳动力需求、降低采摘成本、提高采摘效率有极大的需求和现实意义。

快速识别与精准定位是猕猴桃采摘机械臂实现高效、准确采摘的关键技术,而受果园的自然光线、枝干枝叶遮挡、果实颜色与背景相似等因素影响,猕猴桃在复杂自然环境下的检测面临巨大困难和挑战。近年来,国内外学者对非结构自然环境下猕猴桃果实识别进行了深入研究。傅隆生等<sup>[3]</sup>采用竖直向上拍摄方式获取猕猴桃果实图像,以猕猴桃果萼为参考,对夜间猕猴桃果实进行识别,但仅能识别果萼处在图像中心区域的猕猴桃;陈礼鹏等<sup>[4]</sup>以猕猴桃果萼为识别对象对猕猴桃图像进行多果识别,为后续机械臂路径规划提供信息支持,但准确率较低,识别速度较慢;Fu等<sup>[5]</sup>针对猕猴桃线性排列生长情况,利用果萼和猕猴桃边缘信息提出一种全新猕猴桃识别算法,该算法识别率高,但仅限于猕猴桃果实呈线性排列情况。上述传统图像处理方法受图像、人为主观性影响,鲁棒性和泛化性差,并且算法流程长,细节处理繁琐,识别速度相对较慢,很难满足猕猴桃实地采摘需求。

随着卷积神经网络(convolutional neural networks, CNN)<sup>[6]</sup>迅速发展,基于锚点(Anchor\_Base)的Faster R-CNN<sup>[7]</sup>、SSD<sup>[8]</sup>以及无锚点(Anchor\_Free)的YOLOX<sup>[9]</sup>等目标检测算法已在果实识别、定位、分类等视觉应用领域获得了较好的效果<sup>[10-12]</sup>。Wen等<sup>[13]</sup>针对水果检测场景、果实成熟度检测、品质检测,基于Mask R-CNN<sup>[14]</sup>提出多任务多场景的柑橘检测算法,获得0.35 s的检测速度,模型准确率达到91.56%。Zhong等<sup>[15]</sup>采用改进Faster R-CNN目标检测算法进行辣椒识别,模型准确率达到87.3%,并根据检测结果对辣椒进行高度测量。An等<sup>[16]</sup>采用基于YOLOX改进的SDNet目标检测算法对草莓成熟度进行实时检测,其平均识别率达到94.26%,识别速度仅为0.030 5 s。Liu等<sup>[17]</sup>采用NIR图像与RGB图像构建了融合数据集,并使用Faster R-CNN识别算法进行猕猴桃识别,获得0.135 s的识别速度,模型准确率为90.7%。与传统图像处理方法相比,深度学习卷积神经网络识别方法具有鲁棒性高,泛化能力强,识别速度快等优点。

猕猴桃采摘机械臂的平均采摘时间决定采摘效率,其平均采摘时间包括猕猴桃识别、定位、采摘、摆放以及误检导致采摘失败所浪费的时间<sup>[18]</sup>。精度高且识别快的目标识别算法是机器人采摘成功应用的前提,也是当前智能采摘研究领域的热点<sup>[19]</sup>。傅隆生等<sup>[20]</sup>优化LeNet模型对田间多簇猕猴桃进行分类识别,模型精度为89.29%,但识别单个果实需0.27 s,速度较低。穆龙涛等<sup>[21]</sup>采用以改进AlexNet为骨干提取网络的Faster R-CNN目标检测算法对不同光照猕猴桃识别,模型识别精度达到96%,但未考虑树枝遮挡等果园实景进行分类识别,且识别一张图像需

1 s。Fu等<sup>[22]</sup>为提高果实识别速度,基于YOLOv3\_tiny提出DY3TNet模型对猕猴桃进行识别,其识别速度达到0.034 s,但识别率仅为90.05%。

目前基于卷积神经网络的猕猴桃识别算法精度与识别速度相较其他果实识别<sup>[16,23-24]</sup>偏低,而且对树枝遮挡、铁丝遮挡和果实堆叠等实际情况展开分类识别研究较少,不能满足猕猴桃机械臂采摘需求。为此,本文通过模型对比,选择YOLOX作为基础模型,对其特征融合部分进行通道数调整、引入注意力机制以及迭代等改进,并对猕猴桃进行多标签识别和定位分析。相比于原始YOLOX算法,改进YOLOX算法在识别精度与召回率上均具有提升,且整体性能优于同类算法,能更好地满足自然生长条件下猕猴桃的自动采摘需求。

## 1 图像获取与预处理

猕猴桃图像数据的构建是应用深度学习进行猕猴桃检测的前提条件。猕猴桃采摘期短(贵州省集中于9~10月),因未能到果园实地采集图像,因此,自主构建数据集,综合了3处数据来源。1)从文献[17]中获取未进行数据增强的猕猴桃原始图像90张;2)通过爬虫程序从网络下载识别度较好(分辨率>300×300)的猕猴桃图片20张;3)模拟猕猴桃生长环境,在校园内挂接采购的猕猴桃水果,并使用D435i相机自主拍摄,采集640×640像素图像30张。共获取猕猴桃图像140张共4 533个猕猴桃目标,其中包含竖直向上和水平等多角度拍摄,白天、夜晚、顺光、逆光以及夜间补光等复杂场景下的猕猴桃图像,如图1所示。

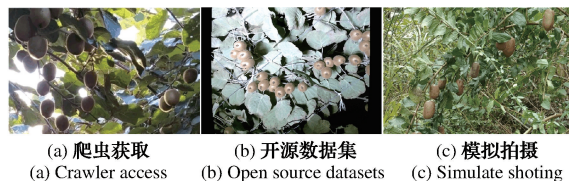


图1 猕猴桃数据集

Fig. 1 Kiwifruit dataset

为体现自然环境的复杂性并提高模型的鲁棒性,对原始数据集通过噪点、亮度调节与旋转方式扩增至840张,其中猕猴桃目标个数达到27 198个。使用LabelImg将其标注为独立果实(no occluded, NO)、树枝和铁丝遮挡(branch occluded, BO)、果实遮挡(fruit occluded, FO)、树叶遮挡(leaf occluded, LO)四类。并按照训练集:测试集=9:1的比例进行划分。

## 2 方法

### 2.1 基准模型YOLOX的选择

2016年,Redmon等<sup>[25]</sup>为减少模型检测时间提出一阶段目标检测算法YOLO并引起关注。学者们在YOLO上改进优化出v2~v4<sup>[26-28]</sup>以及YOLOX等目标检测算

法,应用于各种目标检测场景。

与常见目标检测算法不同,YOLOX 采用 Anchor\_Free 机制,直接回归目标中心点及其宽高,减少了超参数调节、预测框筛选成本。在检测头部分,YOLOX 采用解耦检测头分别回归类别和位置信息,避免了耦合头检测时类别和位置信息同时回归带来的信息冲突。

由于解耦头单独回归类别、置信度以及位置信息,YOLOX 损失由  $L_{cla}$ 、 $L_{conf}$  以及  $L_{reg}$  三部分组成如式(1)所示。其中  $L_{cla}$  和  $L_{conf}$  采用二值交叉熵损失计算如式(2)、(3)所示,而  $L_{reg}$  采用 IOULoss 损失计算如式(4)所示。

$$Loss = \frac{L_{cla} + \lambda L_{reg} + L_{conf}}{N_{pos}} \quad (1)$$

式中: $L_{cla}$  为分类损失; $L_{reg}$  为定位损失; $L_{conf}$  为置信度损失; $\lambda$  为定位损失的平衡系数; $N_{pos}$  为正样品个数。

$$L_{cla}(O, C) = - \frac{\sum_{i \in pos} \sum_{j \in cla} (O_{ij} \ln(\hat{C}_{ij}) + (1 - O_{ij}) \ln(1 - \hat{C}_{ij}))}{N_{pos}} \quad (2)$$

其中, $O_{ij} \in \{0, 1\}$ ,表示预测框  $i$  中是否真实存在第  $j$  类目标; $C_{ij}$  为模型预测值; $\hat{C}_{ij}$  为  $C_{ij}$  通过 Sigmoid 函数得到的目标概率; $N_{pos}$  为正样本个数。

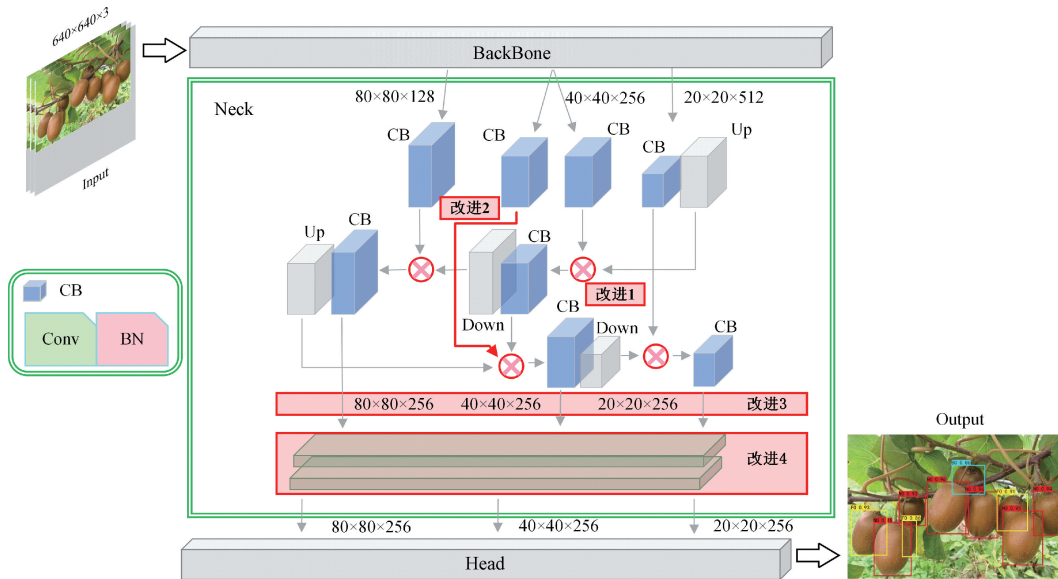


图2 改进 YOLOX 网络结构

Fig. 2 Improvement of YOLOX network structure

### 1) 引入注意力机制

当融合不同尺度特征图时,往往需要将其调整至相同尺度再进行融合。YOLOX 将所有尺度的特征图视为同等贡献,但不同尺度的特征图对识别结果有着不同的贡献。为使模型在特征融合时,更关注贡献大的特征层,提高模型对猕猴桃果实特征的识别能力。在特征图融合时,引入可训练权重  $w$  实现加权融合注意力机制如式(5)所示。

$$L_{conf}(o, c) = - \frac{\sum_i (o_i \ln(\hat{c}_i) + (1 - o_i) \ln(1 - \hat{c}_i))}{N} \quad (3)$$

其中, $o_i \in \{0, 1\}$ ,表示预测框  $i$  是否真实包含目标; $c_i$  为模型预测值; $\hat{c}_i$  为  $c_i$  通过 Sigmoid 函数得到的目标概率; $N$  为正负样本个数。

$$L_{reg} = 1 - \frac{A \cap B}{A \cup B} \quad (4)$$

式中: $A$  为预测边界框; $B$  为真实边界框。

### 2.2 YOLOX 模型的改进

深度学习目标检测网络中,骨干特征提取网络(Backbone)和检测头(Head)之间的特征融合部分(Neck)为承上启下的关键环节。Neck 需要对 Backbone 提取到的特征进行融合并合理利用,帮助 Head 进行下一步的工作。Xu 等<sup>[29]</sup>通过改进 YOLOv3 特征融合等部分对绿芒果进行识别,其调和平均值  $F_1$  达到 97.7%;Zheng 等<sup>[30]</sup>通过改进 YOLOv4 特征融合部分等部分对绿色柑橘进行检测,其识别精度达到 91.55%;为进一步提升模型识别精度,避免误检造成时间浪费,提高机械臂采摘效率,本文对 YOLOX 做 4 方面的改进,如图 2 所示。

$$Output = \sum_i \frac{w_i \cdot Input_i}{\delta + \sum_j w_j} \quad (5)$$

其中可训练权重  $w$  通过 ReLU 激活函数确保  $w \geq 0$ ,  $\delta = 0.0001$  为避免数值不稳定的极小数值, $i, j$  为特征图个数。

### 2) 增加特征融合

Neck 环节的关键是特征融合,而 YOLOX 原始 Neck 结构中存在单边输入的节点,并没有融合特征,导致该节

点在后续特征融合时提供较小的贡献。为增加特征融合的多样性,提升对猕猴桃不同分类图像的识别能力。去除单边输入的节点,并在相同大小的特征图融合时,从原始输入节点添加一条额外的边,以便在不增加太多成本的情况下融合更多的特性。

### 3) 调整通道数

在卷积神经网络中,每个通道代表某个方面的特征<sup>[31]</sup>。通道数越多提取的特征越多,从而有助于提升网络的性能。但通道数过多也会导致参数量的增加,从而增加计算量、内存消耗以及出现过拟合。对于猕猴桃识别,图像中的某种特征可能增加识别率,而另外一些特征可能并不重要。因此,本文对特征融合部分通道数从 $2^5 \sim 2^{10}$ 依次设置进行1 000次训练,并将不同通道数设置下的平均精度均值(mAP)进行对比,其对比结果如表1所示。从对比结果可知,在通道数为 $2^5 \sim 2^7$ 时,模型猕猴桃平均识别率逐步提升,当通道数为 $2^8$ 时,模型识别率达到最大值94.79%,当通道数增加至为 $2^9 \sim 2^{10}$ 时,通道数过多,出现过拟合,平均识别率下降。

表1 通道数及对应识别精度

Table 1 Channel number and corresponding recognition accuracy

| 序号 | 通道数      | mAP/% |
|----|----------|-------|
| 1  | $2^5$    | 93.12 |
| 2  | $2^6$    | 93.72 |
| 3  | $2^7$    | 93.80 |
| 4  | $2^8$    | 94.79 |
| 5  | $2^9$    | 94.54 |
| 6  | $2^{10}$ | 94.07 |

### 4) 特征融合迭代

为充分融合特征以及获取深层次的猕猴桃果实特征信息,将改进后的 Neck 特征融合模块视为整体卷积模块进行堆叠迭代。改进后特征融合部分如图3所示。

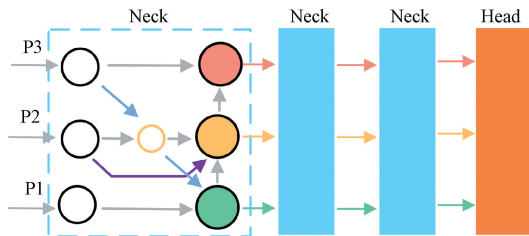


图3 特征融合部分

Fig. 3 Feature fusion section

## 2.3 模型训练

实验均在PC上配置完成,实验环境为Intel Core i5-12600K CPU;NVIDIA RTX 3060 12 G显卡;32 G运行内存;Cuda版本11.6。学习率为0.01,SGD优化器,动量为0.9,Epoch为1 500,Batch\_size为4。为提高模型的鲁棒性以及精度,采用多尺度训练策略以及Mosica<sup>[28]</sup>和

Mixup<sup>[32]</sup>数据增强策略。每10次迭代后,对权重进行保存,并实时监测模型训练状况,记录模型各项指标。

为验证模型效果,使用调和平均值 $F_1$ 、平均精度(average precision, AP)、mAP、召回率(recall, R)以及漏检率(miss rate, MR)作为性能评价指标对模型进行性能评估,如式(6)~(11)所示。

$$R = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

$$MR = \frac{FN}{TP + FN} \times 100\% \quad (8)$$

$$AP = \int_0^1 P(R) dR \times 100\% \quad (9)$$

$$F_1 = \frac{2PR}{P + R} \times 100\% \quad (10)$$

$$mAP = \frac{1}{k} \sum_{i=0}^k AP \times 100\% \quad (11)$$

式中:预测框正确检测目标数量为 $TP$ ;预测框错误检测目标数量为 $FP$ ;漏检标签框数量为 $FN$ ;准确率 $P$ 为所有预测框中检测正确的比例; $R$ 为所有标签框中被正确检测所占比例; $MR$ 为所有标签框中未被正确检测所占比例; $F_1$ 为准确率和召回率的调和平均值; $AP$ 为猕猴桃每个类别的平均精度; $mAP$ 为 $k(k=4)$ 类猕猴桃 $AP$ 的均值。

## 2.4 空间坐标转换

猕猴桃图像数据通过模型识别后,会生成含有猕猴桃果实的目标预测框,以预测框的中心点表示果实在2D RGB图像中位置。然而机械臂采摘需要三维世界坐标,深度相机获取的Depth图像中包含对应RGB图像中每个像素点的深度信息,基于图像深度信息,再按式(12)可实现2D像素坐标 $\{u, v\}$ 和3D空间坐标 $\{X, Y, Z\}$ 之间的坐标转换。

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R, t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (12)$$

式中; $K$ 为相机内部参数矩阵; $[R, t]$ 为相机外部参数矩阵; $Z_c$ 为深度信息。

## 3 实验与验证

### 3.1 对比实验

#### 1) 基础模型对比

准确快速的深度学习目标检测模型是猕猴桃采摘机械臂的核心技术。本文对主流深度学习目标检测算法Faster R-CNN、SSD以及YOLOX进行了纵向对比实验。Faster R-CNN(ResNet50)、SSD(ResNet50)与YOLOX模型的mAP、Loss曲线如图4所示。

从图4可知,Faster R-CNN、SSD在100与400次迭代左右模型收敛,YOLOX在1 200次迭代后才开始收敛,

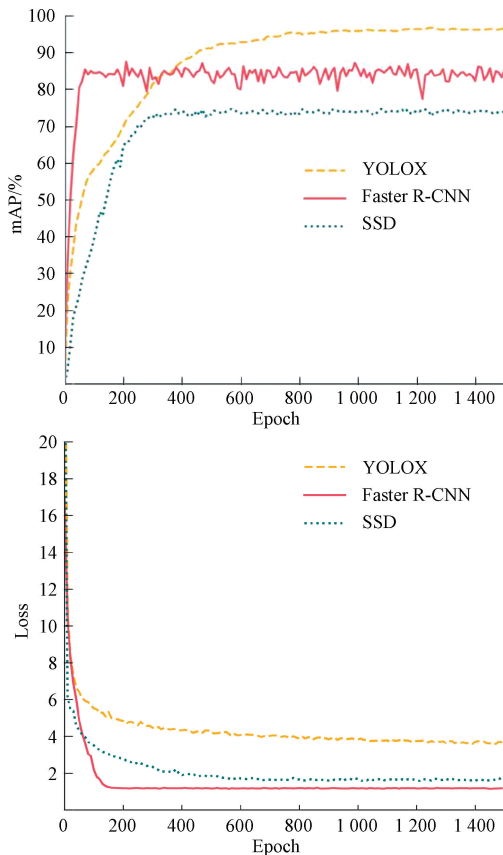


图4 模型 mAP、Loss 曲线

Fig. 4 mAP、Loss curves of different models

1 500 次迭代后 3 种模型全部收敛。3 个模型中, YOLOX 的 mAP 最高(95.58%), Faster R-CNN 次之(86.11%), 而 SSD 最低(73.47%)。在检测速度上, 二阶段目标检测 Faster R-CNN 检测 1 张图像需要 0.644 s, 一阶段目标检测 SSD 需要 0.013 s, 而 YOLOX 仅需要 0.012 s, 如表 2 所示。综上所述, 对比主流的基于 Anchor\_Base 一阶段目标检测算法(SSD)与二阶段目标检测算法(Faster R-CNN), 一阶段 Anchor\_Free 目标检测算法 YOLOX 在 mAP 以及速度上均有优势。为准确且快速识别非结构化自然环境下的猕猴桃选用 YOLOX 作为猕猴桃识别基础模型。

表 2 纵向对比实验结果

Table 2 Results of longitudinal comparative experiments

|             | 阶段 | 方法           | mAP/% | 检测 1 张图像用时/s |
|-------------|----|--------------|-------|--------------|
| Anchor_Base | 二  | Faster R-CNN | 86.11 | 0.644        |
|             | 一  | SSD          | 73.47 | 0.013        |
| Anchor_Free |    | YOLOX        | 95.58 | 0.012        |

## 2) YOLOX 模型改进对比

为验证改进后 YOLOX 模型对猕猴桃识别效果, 对改进前后 YOLOX 网络进行测试验证, 实验结果如表 3、图 5 所示。

表 3 改进前后 YOLOX 检测结果对比

Table 3 Comparison of YOLOX detected results before and after improvement

| 分类   | AP/%  |       | 检测 1 张图像用时/s |       |
|------|-------|-------|--------------|-------|
|      | 原模型   | 改进后   | 原模型          | 改进后   |
| NO   | 99.35 | 99.38 |              |       |
| BO   | 94.20 | 93.82 |              |       |
| FO   | 93.55 | 92.98 | 0.012        | 0.019 |
| LO   | 95.21 | 98.43 |              |       |
| Mean | 95.58 | 96.15 |              |       |

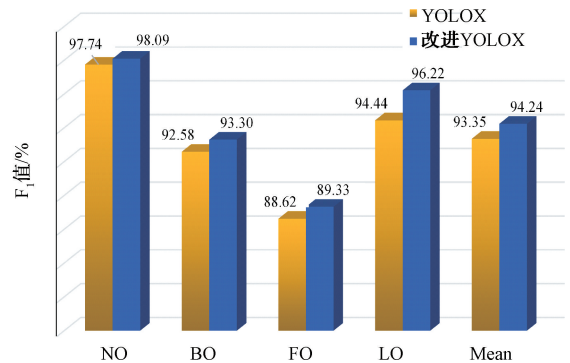


图 5 改进前后 YOLOX  $F_1$  值

Fig. 5 Comparison of YOLOX  $F_1$  values before and after improvement

表 3 为模型平均精度表现, 虽然对 BO、FO 两类目标, 改进模型比原模型分别下降 0.38% 和 0.57%, 但是对 LO 类的 AP 以及 mAP 分别提高了 3.2% 和 0.57% 分别达到 98.43% 和 96.15%。调和平均值  $F_1$  上, 由图 5 可知, 改进后的 YOLOX 模型对 4 类猕猴桃均有提高, NO、BO、FO 和 LO 分别提高 0.35%、0.72%、0.71% 和 1.78%,  $F_1$  值平均提升 0.89% 达到 94.24%。经过实验对比表明改进特征融合部分后的 YOLOX 模型获取到了猕猴桃更高级特征融合使模型对猕猴桃树 LO 识别更准确, 并且  $F_1$  值的全线提高表明改进后的模型对于 4 类猕猴桃识别不但有更高识别率而且具有更高的召回率。在识别速度上, 检测一张  $640 \times 640$  的图像, 改进模型只需 0.019 s 达到 52 fps, 比原模型(0.012 s)仅降低了 0.007 s, 仍满足实时检测要求。

## 3) 同类文献对比

近年来, 国内外学者对非结构化下猕猴桃果实识别进行了深入研究, 而猕猴桃检测的主流算法为 Anchor\_Base 机制的 Faster R-CNN 和 YOLOv4 等, 本文算法与文献[17, 33-34]检测结果对比如表 4 所示。由表 4 中可以看出, 本文提出的改进 YOLOX 模型对 NO、FO、BO 以及 LO 4 类猕猴桃识别精度分别为 99.38%、92.98%、93.82%、98.43%, 平均识别精度为 96.15%, 相较于同类研究, 4 类猕猴桃识别精度均有显著提升。其中, LO 类识别精度提升最大, FO 类提升较少, 相比 YOLOv4 模型分别提升

表 4 同类文献检测结果对比  
Table 4 Comparison of results with similar literature detected

| 方法     |                    |     |             | 不同类别的识别结果/% |       |       |       | mAP/% | 检测 1 张图<br>像用时/s |
|--------|--------------------|-----|-------------|-------------|-------|-------|-------|-------|------------------|
|        |                    |     |             | NO          | FO    | BO    | LO    |       |                  |
| 文献[34] | Faster R-CNN+ZFNet | 二阶段 | Anchor_base | 96.70       | 86.50 | —     | 82.50 | 92.30 | 0.274            |
| 文献[17] | Faster R-CNN+VGG16 | 二阶段 | Anchor_base | 96.70       | 87.40 | —     | 83.60 | 91.70 | 0.134            |
| 文献[33] | YOLOv4             | 一阶段 | Anchor_base | 95.70       | 92.40 | 90.50 | 88.20 | 91.90 | 0.025            |
| 本文     | 改进 YOLOX           | 一阶段 | Anchor_Free | 99.38       | 92.98 | 93.82 | 98.43 | 96.15 | 0.019            |

7.96%、0.58%。表 4 也显示,单幅图像检测速度(0.019 s)在同类文献中显著占优,改进 YOLOX 模型更快更准。

3.2 消融实验

通过加权注意力机制、堆叠特征融合以及增加特征融合对 YOLOX 的 Neck 部分进行改进。为验证每个改进点对模型的作用,以 YOLOX 为 Benchmark 进行消融实验。实验结果如表 5 所示,其中“√”表示使用该模块;T 表示增加特征融合;D 表示迭代特征融合部分;Z 表示增加注意力机制。

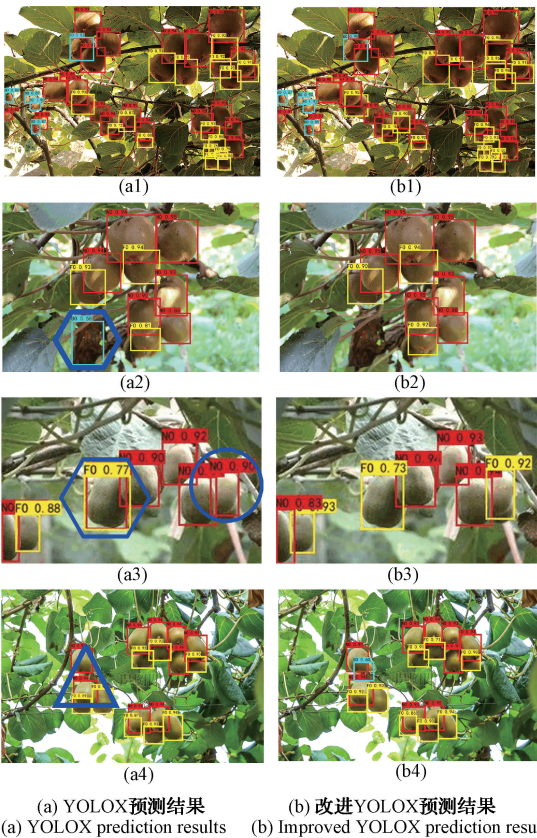
表 5 消融实验  
Table 5 Ablation experiment

| 模型    |   |   |   | mAP/% | R/%   | 检测 1 张图<br>像用时/s |
|-------|---|---|---|-------|-------|------------------|
| YOLOX | T | D | Z |       |       |                  |
| √     |   |   |   | 95.58 | 92.48 | 0.012            |
| √     | √ |   |   | 95.53 | 92.55 | 0.013            |
| √     | √ | √ |   | 95.57 | 93.96 | 0.019            |
| √     | √ | √ | √ | 96.15 | 94.19 | 0.019            |

从表 5 可知,增加特征融合后,mAP 略有下降(−0.05%),但召回率 R 略有提升(+0.07%);对特征融合部分进行迭代后,mAP 持平,但 R 提升至 93.96%(+1.48%),对应 MR 为 6.04%;加入注意力机制后,模型的 mAP 与 R 均有较大提升,分别为 96.15%(+0.57%)与 94.19%(+1.71%),MR 为 5.81%。实验数据表明,单独采用特征融合对模型性能提升效果不大;采用特征融合迭代能显著提高召回率,降低漏检率;增加注意力机制既能提高识别率又能降低漏检率。由此可知,只有在多重改进组合的情况下,才能提高模型的综合性能。此外,改进模型的单帧检测速度略有下降,但仍能满足实时检测(0.019 s)。

3.3 模型验证

为更直观地体现 YOLOX 改进前后的差异,使用 4 张未训练图像进行验证检测,结果如图 6 所示。由图 6(a1)与(b1)可以看出,两种模型检测结果一致均能识别受自然光线、枝干枝叶遮挡、果实颜色与背景相似等影响的猕猴桃果实。图 6(a2)~(a4)中,YOLOX 模型出现误检、漏检以及错误分类情况,如六边形框内出现将枯叶误检成猕猴桃及预测框重叠情况,圆形框内出现将 FO 类错误分成 NO 类的情况以及三角形框内未识别被树枝与果实严重遮挡的猕猴桃出现漏检情况。图 6(b2)~(b4)中,由于对



(a) YOLOX 预测结果 (b) 改进 YOLOX 预测结果  
(a) YOLOX prediction results (b) Improved YOLOX prediction results

图 6 模型预测结果

Fig. 6 Models predict results

YOLOX 模型特征融合部分引入注意力机制、增加特征融合、迭代等改进使模型获得更多猕猴桃特征信息的同时对猕猴桃特征信息更加敏感,从而避免了猕猴桃误检、漏检以及误分类情况,两种图像中的目标果实均成功检出。实验结果表明,改进 YOLOX 模型可以同时分类并识别复杂自然环境下的猕猴桃果实,分类检测准确性高,对光线、树叶遮挡、果实重叠等难识别情况均具有较强的鲁棒性。与原始 YOLOX 模型相比,改进模型识别精度更高,召回率更高、漏检率与误检率更低,能有效避免误检造成时间浪费,提高机械臂采摘效率,满足猕猴桃采摘机械臂在实际应用中的采摘要求。

3.4 猕猴桃空间定位与采摘

猕猴桃空间坐标计算流程如图 7 所示,Inter

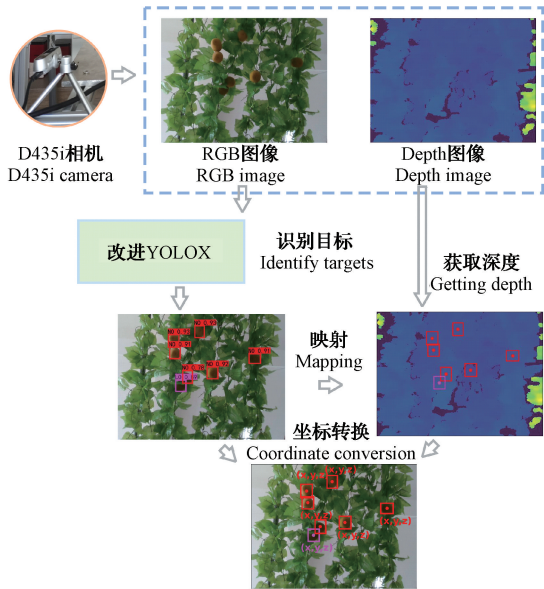


图7 猕猴桃空间坐标计算流程

Fig. 7 Kiwifruit space coordinates acquisition process

RealSense D435i 相机同时拍摄获取 RGB 和对应的 Depth 图像,RGB 图像送入改进 YOLOX 模型,识别获取每个猕猴桃果实(识别框的中心像素点)的 2D 图像坐标,读取

Depth 图像中的像素深度信息,最后通过式(12)计算每个猕猴桃果实的 3D 空间坐标。

为验证猕猴桃识别和定位的准确性,定制了一款 6 自由度最大工作距离为 1 200 mm 的机械臂,并模拟猕猴桃生长环境,搭建了猕猴桃机械臂采摘平台。采摘平台如图 8 所示,为更好地识别定位猕猴桃目标并使结构光相机 D435i 获取较高精度(1 m 内短距离工作范围),将相机固定至距离猕猴桃果实 650~950 mm 的机架上,自下向上获取猕猴桃图像。按前述算法计算单个猕猴桃果实的 3D 坐标,以此为机械臂关键数据输入控制机械臂/末端执行器移动到该坐标(图 8(b)),最后控制末端执行器对果实进行抓取(图 8(c))。

在模拟采摘实验中,分别摘取场景中不同的果实,记录末端执行器的实际采摘点坐标。由于存在图像识别精度误差、式(12)中相机内参误差、机械臂以及 D435i 相机硬件自带误差等原因,导致计算坐标与实际抓取坐标之间存在误差,表 6 为 9 组实验结果进行误差分析。 $Z$  向误差极值为 +6 mm,误差限为 0.8%, $X$  与  $Y$  坐标误差极值分别为 +1.8 mm(1.9%)和 +2.3 mm(3.3%),其综合误差  $\leq 3.3\%$ ;9 组实验结果表明  $Z$  轴误差范围在  $[-5\text{ mm}, +6\text{ mm}]$ , $X$  与  $Y$  轴误差范围在  $[-1.6\text{ mm}, +1.8\text{ mm}]$  与  $[-1.9\text{ mm}, +2.3\text{ mm}]$ ,基本满足了猕猴桃采摘的定位需求,反之也验证了改进算法的准确性。



图8 机械臂采摘平台

Fig. 8 Robotic arm picking platform

表6 空间坐标误差分析

Table 6 Spatial coordinate error analysis

| 序号 | 计算坐标/mm<br>( $X, Y, Z$ ) | 实际坐标/mm<br>( $X, Y, Z$ ) | 误差/mm<br>( $X, Y, Z$ ) |
|----|--------------------------|--------------------------|------------------------|
| 1  | (-110.44, 54.53, 699)    | (-111.54, 52.43, 698)    | (+1.1, +2.1, +1)       |
| 2  | (-94.91, 228.28, 792)    | (-96.71, 230.18, 786)    | (+1.8, -1.9, +6)       |
| 3  | (-182.33, 71.76, 831)    | (-181.33, 69.46, 827)    | (-1, +2.3, +4)         |
| 4  | (-123.96, -157.27, 833)  | (-123.16, -159.37, 836)  | (-0.8, +2.1, -3)       |

续表

| 序号 | 计算坐标/mm<br>(X,Y,Z)   | 实际坐标/mm<br>(X,Y,Z)   | 误差/mm<br>(X,Y,Z) |
|----|----------------------|----------------------|------------------|
| 5  | (-112.17,83.45,908)  | (-112.97,81.85,911)  | (+0.8,+1.6,-3)   |
| 6  | (-51.00,55.46,706)   | (-49.50,57.26,710)   | (-1.5,-1.8,-4)   |
| 7  | (-116.48,129.14,854) | (-114.88,127.44,859) | (-1.6,+1.7,-5)   |
| 8  | (-210.33,-13.46,848) | (-208.93,-14.36,842) | (-1.4,+0.9,+6)   |
| 9  | (-218.43,67.94,904)  | (-217.73,66.36,899)  | (-0.7,+1.6,+5)   |

4 结 论

猕猴桃果实的识别、分类以及空间定位是其机械臂自动化采摘的关键技术。对主流目标检测模型进行对比分析,在猕猴桃分类检测任务中,YOLOX 模型 mAP 为 95.58%,1 张 640×640 分辨率图像的平均检测时间为 0.012 s,具有综合最优的猕猴桃果实目标检测速度和分类准确率。为进一步提高猕猴桃识别精度以及识别速度,减少误检带来的时间浪费,提升机械臂采摘效率,通过引入加权注意力机制、调整通道数、堆叠特征融合等,改进 YOLOX 模型的特征融合环节。改进模型的 mAP 和  $F_1$  值分别提高至 96.15%和 94.24%,1 张图像识别用时为 0.019 s,相比原始模型,改进模型拥有更高的识别精度、更低的误检率,有效减少误检带来的时间浪费,有效提升了机械臂采摘效率。

对比和验证实验表明,改进 YOLOX 模型综合性能更优于目前主流猕猴桃识别算法,在非结构自然环境下的猕猴桃果实识别中,具有较高的识别率、检测速度以及鲁棒性。空间定位验证与采摘实验中,Z 坐标误差在 0.8%内,X、Y 坐标误差分别在 1.9%、3.3%内,并准确抓取猕猴桃果实,满足猕猴桃采摘机械臂在实际应用中的采摘需求。为进一步提高模型应用场景,后续将通过补充和细化数据集,优化网络机构框架等策略,提高深度学习模型的识别和分类能力,并将进一步完善实验场景和实验样机,早日实现从实验室采摘向田间实验的过渡。

参 考 文 献

[1] 韩茹梦,李瑞鹏,涂美艳,等. 中国和新西兰猕猴桃生产现状的比较分析[J]. 热带农业科学,2023,43(9): 122-129.  
HAN R M, LI R P, TU M Y, et al. Comparative analysis of kiwi fruit production in China and new zealand[J]. Chinese Journal of Tropical Agriculture, 2023,43(9):122-129.

[2] 韩世明,高成密,张佳乐,等. 贵州省猕猴桃产业发展问题及对策[J]. 农业技术与装备,2022(1):61-64.  
HAN SH M, GAO CH M, ZHANG J L, et al. Problems and countermeasures in the development of kiwifruit industry in Guizhou province [J].

Agricultural Technology & Equipment, 2022 (1): 61-64.

[3] 傅隆生,孙世鹏, VÁZQUEZ-ARELLANO M,等. 基于果萼图像的猕猴桃果实夜间识别方法[J]. 农业工程学报,2017,33(2):199-204.  
FU L SH, SUN SH P, VÁZQUEZ-ARELLANO M, et al. Kiwifruit recognition method at night based on fruit calyx image[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33 (2): 199-204.

[4] 陈礼鹏,穆龙涛,刘浩洲,等. 基于猕猴桃果萼图像的多目标果实识别方法[J]. 计算机工程与设计,2018, 39(6):1738-1744.  
CHEN L P, MU L T, LIU H ZH, et al. Multi-target fruit recognition method based on kiwifruit calyx image[J]. Computer Engineering and Design, 2018,39(6):1738-1744.

[5] FU L, TOLA E, AL-MALLAHI A, et al. A novel image processing algorithm to separate linearlyclustered kiwifruits[J]. Biosystems Engineering, 2019, 183:184-195.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[7] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.

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

[9] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021[J]. Computer Science, 2021, ArXiv. 2107. 08430.

[10] 张耀,吴一全,陈慧娟. 基于深度学习的视觉同时定位与建图研究进展[J]. 仪器仪表学报, 2023, 44 (7): 214-241.

- ZHANG Y, WU Y, CHEN H. Research progress of visual simultaneous localization and mapping based on deep learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(7): 214-241.
- [11] 张立豪, 李鹏, 刘轩宇. 基于深度迁移学习的变工况气体泄漏检测[J]. 仪器仪表学报, 2023, 44(6): 177-187.
- ZHANG L H, LI P, LIU X Y. Gas leak detection for variable conditions based on deep transfer learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 177-187.
- [12] 孙备, 党昭洋, 吴鹏, 等. 多尺度互交叉注意力改进的单无人机对地伪装目标检测定位方法[J]. 仪器仪表学报, 2023, 44(6): 54-65.
- SUN B, DANG ZH Y, WU P, et al. Multi scale cross attention improved method of single unmanned aerial vehicle for ground camouflage target detection and localization [J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 54-65.
- [13] WEN C X, ZHANG H, LI H H, et al. Multi-scene citrus detection based on multi-task deep learning network[C]. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020: 912-919.
- [14] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 386-397.
- [15] ZHONG S, XU W, ZHANG T, et al. Identification and depth localization of clustered pod pepper based on improved Faster R-CNN[J]. IEEE Access, 2022, 10: 93615-93625.
- [16] AN Q, WANG K, LI Z, et al. Real-time monitoring method of strawberry fruit growth state based on YOLO improved model[J]. IEEE Access, 2022, 10: 124363-124372.
- [17] LIU Z, WU J, FU L, et al. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion [J]. IEEE Access, 2020, 8(1): 2327-2336.
- [18] WILLIAMS H A M, JONES M H, NEJATI M, et al. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms [J]. Biosystems Engineering, 2019, 181: 140-156.
- [19] 郑太雄, 江明哲, 冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. 仪器仪表学报, 2021, 42(9): 28-51.
- ZHENG T X, JIANG M ZH, FENG M CH. Vision based target recognition and location for picking robot: A review [J]. Chinese Journal of Scientific Instrument, 2021, 42(9): 28-51.
- [20] 傅隆生, 冯亚利, ELKAMIL T, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205-211.
- FU L SH, FENG Y L, ELKAMIL T, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks [J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(2): 205-211.
- [21] 穆龙涛, 高宗斌, 崔永杰, 等. 基于改进 AlexNet 的广域复杂环境下遮挡猕猴桃目标识别[J]. 农业机械学报, 2019, 50(10): 24-34.
- MU L T, GAO Z B, CUI Y J, et al. Kiwifruit detection of far-view and occluded fruit based on improved AlexNet [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(10): 24-34.
- [22] FU L, FENG Y, WU J, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. Precision Agriculture, 2020, DOI:10.1007/s11119-020-09754-y.
- [23] 王芳, 崔丹丹, 李林. 基于深度学习的采摘机器人目标识别定位算法 [J]. 电子测量技术, 2021, 44(20): 162-167.
- WANG F, CUI D D, LI L. Target recognition and positioning algorithm of picking robot based on deep learning [J]. Electronic Measurement Technology, 2021, 44(20): 162-167.
- [24] ZHANG Y, ZHANG W, YU J, et al. Complete and accurate holly fruits counting using YOLOX object detection[J]. Computers and Electronics in Agriculture, 2022, 198: 107062.
- [25] 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, 2016: 779-788.
- [26] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [27] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. ArXiv E-prints ArXiv, 1804. 02767, 2018.
- [28] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint ArXiv, 2004. 10934, 2020.
- [29] XU Z F, JIA R S, SUN H M, et al. Light-

- YOLOv3: Fast method for detecting green mangoes in complex scenes using picking robots[J]. Applied Intelligence, 2020, 50(12): 4670-4687.
- [30] ZHENG Z, XIONG J, LIN H, et al. A method of green citrus detection in natural environments using a deep convolutional neural network[J]. Frontiers in Plant Science, 2021, 12: 705737.
- [31] 张绍荣, 赵紫宁, 莫云, 等. 特征提取对通道选择方法的影响研究[J]. 国外电子测量技术, 2020, 39(9): 1-6.  
ZHANG SH R, ZHAO Z N, MO Y, et al. Research on the influence of feature extraction on channel selection method [J]. Foreign Electronic Measurement Technology, 2020, 39(9): 1-6.
- [32] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: beyond empirical risk minimization [J]. ArXiv Preprint ArXiv:1710.09412, 2017.
- [33] SUO R, GAO F, ZHOU Z, et al. Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking[J]. Computers and Electronics in Agriculture, 2021, 182: 106052.
- [34] FU L, FENG Y, MAJEED Y, et al. Kiwifruit detection in field images using Faster R-CNN with ZFNet[J]. IFAC-Papers on Line, 2018, 51(17): 45-50.

## 作者简介

刘博文, 硕士研究生, 主要研究方向为机器视觉、数字图像处理。

E-mail: lbowen529@126.com

徐卫平, 硕士, 教授, 硕士生导师, 主要研究方向为农业机械。

E-mail: 2609550082@qq.com

徐钦, 硕士研究生, 主要研究方向为机器视觉、数字图像处理。

E-mail: rainofblue@163.com

陈华伟(通信作者), 博士, 教授, 硕士生导师, 主要研究方向为数字图像处理、数字化设计与制造。

E-mail: chwei0130@126.com