

基于 YOLOv3 的油茶果视觉定位系统^①



熊仕琦, 王长坤, 熊璐康

(南昌航空大学 信息工程学院, 南昌 330063)
通信作者: 王长坤, E-mail: 2076885@qq.com

摘要: 随着科技的进步, 采摘机器人各个部分的系统也日益完善. 其中, 机器人视觉定位的系统设计很大程度影响了其工作效率, 尤其是在目标检测速率、采摘果实准确率以及采摘目标环境适应度方面. 本次研究提出利用双目立体视觉系统获取油茶果目标图像, 并采集计算深度信息, 制作自己的油茶果 VOC 数据集, 采用 YOLOv3 目标检测算法来实现复杂环境下油茶果果实的识别, 并通过设计上位机界面, 直观展示对油茶果目标的定位功能. 实验发现该方法具有更高的识别率和更快的识别速度, 在复杂环境下展示了其算法的优越性.

关键词: 双目立体视觉系统; YOLOv3; VOC 数据集; 油茶果识别

引用格式: 熊仕琦, 王长坤, 熊璐康. 基于 YOLOv3 的油茶果视觉定位系统. 计算机系统应用, 2022, 31(1): 132-137. <http://www.c-s-a.org.cn/1003-3254/8274.html>

Visual Positioning System for Camellia Oleifera Fruit Based on YOLOv3

XIONG Shi-Qi, WANG Chang-Kun, XIONG Lu-Kang

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: With the advancement of technology, the systems of various parts of the picking robots have been increasingly improved. The design of the visual positioning system largely affects the work efficiency of a picking robot, especially its target detection speed, fruit picking accuracy, and target picking environment adaptation. In this study, we propose to use a binocular stereo vision system to acquire images of camellia oleifera fruit targets and then collect and calculate depth information to build our own VOC dataset of Camellia oleifera fruits. The you only look once v3 (YOLOv3) target detection algorithm is adopted to achieve Camellia oleifera fruit recognition in complex environments. The function of locating Camellia oleifera fruit targets is intuitively demonstrated by a newly designed upper computer interface. Experimental results show that compared with other methods, the proposed method has a higher recognition rate and a faster recognition speed, which demonstrates the superiority of its algorithm in complex environments.

Key words: binocular stereo vision system; YOLOv3; VOC dataset; camellia oleifera fruit recognition

油茶果是天然的油料, 是中国特有的油料之一, 多生长在中国的南方, 常见于广西、湖南、江西、浙江、河南等省区, 去年, 仅江西省的油茶林面积就达到了 1 598 万亩 (其中高产油茶林 598 万亩), 产业总产值 383.5 亿元, 油茶品牌集群建设成效初显. 江西省已将油茶产业列为推进农业结构调整的九大产业之一. 各地如何围绕“三千工程”发展目标, 全力推动油茶产

业高质量发展, 实现千万亩高产油茶林目标^[1]. 为响应《中国制造 2025》号召, 油茶果的采摘也亟待智能化的采摘设备来解决劳动力紧缺的问题, 油茶果采摘作业一旦机械智能化, 将大大提高油茶果的产量, 随着科技的进步, 采摘机器人各个部分的系统也日益完善, 视觉识别系统能够稳定的进行目标识别. 目前在日本、美国等一些发达国家, 采摘机器人的技术已经十分成

^① 基金项目: 江西省自然科学基金 (AA201920039)

收稿时间: 2021-04-02; 修改时间: 2021-04-29; 采用时间: 2021-05-07; csa 在线出版时间: 2021-12-17

熟,这对于农林业的规模化、精准化具有极大帮助.近年来很多国内外学者对采摘机器人视觉系统进行研究^[2].

Ji等在苹果采摘视觉系统采用了SVM分类器,对于套袋苹果的认识率效果不错,达到了89%,但是这种方法的识别效率并不高^[3].

Rakun等将苹果的颜色与纹理两个特征相结合来进行识别,但对于复杂的周围环境时的果实目标,无法提取纹理特征,难以识别目标物^[4].

相比于这些对果实目标识别的常规方法,深度卷积神经网络展现出了其巨大的优越性,本次研究提出利用双目立体视觉系统获取油茶果目标图像,并采集计算深度信息,采用one-stage目标检测神经网络来实现复杂环境下油茶果果实的识别并与其他目标检测算法进行实验比较,通过网络规模的压缩来加速检测的速度,利用基于YOLOv3的油茶果识别方法定位图像中的油茶果果实并返回其坐标,完成定位任务.

1 数据收集

系统的设计思路是在卷积神经网络算法的基础上,结合双目立体视觉系统实现油茶果识别.在硬件选择上,我们采用型号为CAM-AR0135-3T16的双目相机,且支持在Matlab下的GUI界面开发,并通过USB 3.0在上位机实时显示,如图1所示.其具体参数见表1.



图1 油茶果的双目视觉识别系统模型

双目立体视觉(成像原理如图2)是采用两组成像机器在不同位置摄取目标物图像,再根据不同位置摄取的目标物偏差推算出目标物的空间位置坐标的一种机器视觉形式^[5-7].在设计视觉识别系统中,可以使用双目立体视觉对摄取的图像进行分析处理得到目标精确的三维空间的坐标.

表1 双目识别系统基础参数

| 参数名称 | 参数值 |
|-------|-----------------|
| 图像传感器 | AR0135 |
| 传感器尺寸 | 1/4 |
| 镜头视角 | 水平71°,垂直55° |
| 标配镜头 | 3.6 mm焦距畸变<0.3% |
| 接口协议 | UVC协议,免驱动 |
| 分辨率 | 2560×960 |
| 整体尺寸 | 200×38 |

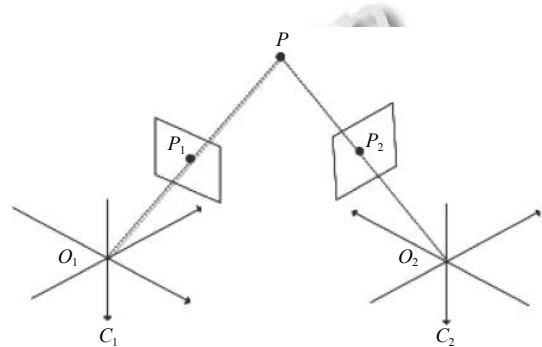


图2 双目立体视觉成像原理

二维信息可以直接从图像中获取,而深度信息可由深度计算得出,如图1所示,P点为空间某物体位置, f_1 和 f_2 为双目相机左右焦距.相机的空间转换模型如式(1)和式(2)所示.

$$s_1 \begin{bmatrix} X_1 \\ Y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} f_1 & 0 & 0 \\ 0 & f_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1)$$

$$s_2 \begin{bmatrix} X_2 \\ Y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} f_2 & 0 & 0 \\ 0 & f_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2)$$

用旋转平移矩阵来表示左右相机坐标间的关系,具体如式(3)所示:

$$\begin{bmatrix} X_r \\ Y_r \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3)$$

结合上式可以得到两相机在坐标系中投影点位置如式(4)所示:

$$\rho_2 \begin{bmatrix} X_2 \\ Y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} f_2 r_1 & f_2 r_2 & f_2 r_3 & f_2 t_x \\ f_2 r_4 & f_2 r_5 & f_2 r_6 & f_2 t_y \\ r_7 & r_8 & r_9 & t_s \end{bmatrix} \begin{bmatrix} zX_1/f_1 \\ zY_1/f_1 \\ z \\ 1 \end{bmatrix} \quad (4)$$

由此可知空间某目标点的位置如式(5)所示:

$$\begin{cases} x = zX_1/f \\ y = zY_1/f_1 \\ z = \frac{f_1(f_2t_y - Y_2t_s)}{Y_2(r_7X_1 + r_8Y_1 + f_1r_9) - f_2(r_4X_1 + r_5Y_1 + f_1r_6)} \end{cases} \quad (5)$$

由于本次研究使用的双目视觉设备左右两相机高度相等, $y_1=y_2$. 根据空间变换关系可得空间中任意一点的深度表达式, 具体见式 (6):

$$\frac{z}{f} = \frac{b}{x_1 - x_2} \Rightarrow z = \frac{fb}{x_1 - x_2} \quad (6)$$

由此我们便可以从左右两相机所采集到的油茶果图像中获取目标的三维信息.

2 基于 YOLOv3 的油茶果目标检测算法

在 YOLO 算法发表之前, 大部分表现比较好的对象检测 (object detection) 算法都是以 R-CNN 为代表两阶段算法, 这样的算法存在一个很明显的问题, 那就是速度太慢, 对实时性要求很高的场景适用性不高. YOLO 算法提出一步完成预测, 而且只在一个 CNN 网络模型中完成图片中所有位置对象的 box 和类别预测, 推理速度大大提升, 完全可以满足实时对象检测.

YOLO 是以 Darknet-53 为基础的算法, 它只需要进行一次深度卷积网络的计算过程, 就可以实现端对端的目标检测功能^[8,9].

2.1 YOLOv3 算法数学原理

YOLOv3 算法将图像分成 $S \times S$ 的网格^[10], 设 x, y 为预测框的中心坐标, w, h 为预测框宽高坐标, confidence (置信度) 指特定目标落在待特定格子的真实性程度^[11].

在对目标图像的检测中, 定义 IOU 为候选框与原标记框的交叠率. 对于 YOLOv3 网络的最终输出维度为式 (7) 所示:

$$Dimension = S \times S \times (B \times 5 + C) \quad (7)$$

我们定义置信度的置信值计算如式 (8) 所示:

$$score = Pr(Object) \times IOU_{pred}^{truth} \quad (8)$$

每个网格预测的类别信息为 $C = Pr(class_i | Object)$, 只预测一类概率, 其计算公式为式 (9):

$$\begin{aligned} Pr(class_i | Object) \times Pr(Object) \times IOU_{pred}^{truth} \\ = Pr(class_i) \times IOU_{pred}^{truth} \end{aligned} \quad (9)$$

3 系统设计

整个系统的流程图如图 3.

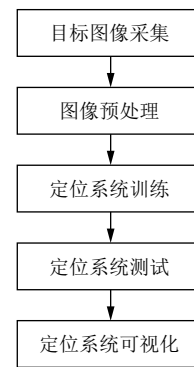


图 3 油茶果识别系统流程图

(1) 目标图像采集模块

使用可变基线 USB 3.0 双目相机采集目标图像, 对相机参数进行实验设定及调整, 使其能够实现对目标图像的获取. 图像要求无曝光情况或大量噪声干扰^[10].

通过双目视觉系统的获取的两组坐标信息, 可以通过解算实现获取目标位置信息. 通过多次的调试和对比分析, 图像摄取距离在 2 m 以内, 双目相机基线距离设定为 90.89 mm, 相机具体参数如表 2 所示.

表 2 相机标定后的主要数据

| 相机参数 | 参数值 |
|------------|----------|
| 焦距 (mm) | 4 |
| 分辨率 (mm) | 1024×768 |
| 倾斜角 (°) | 5 |
| 基线距离 (mm) | 90.89 |
| C_x (mm) | 262.29 |
| C_y (mm) | 107.47 |

(2) 定位系统训练模块

训练的一般步骤是先提取训练样本的目标特征值, 生成一组最优的网络参数设置和权重, 训练过程中不断调整网络参数设置和权重.

(3) 定位系统测试模块

定位系统测试模块是验证训练得到的神经网络的识别能力. 通过训练, 改进后的 YOLOv3 神经网络初步具备对目标的检测能力. 实际的训练的效果需要进一步进行测试, 测试过程中可以手动修正一些特殊的情况, 如识别精度不足, 个别无法识别的状况, 当网络的神经感知度得到进一步深化的时候, 识别系统对目标的识别准确率将明显提高.

(4) 可视化界面设计

信息可视化是识别系统最直接的体现, 通过人机交互界面可以直观的判断系统识别的效率. 我们可以根据需要, 对一些关键参数实施可视化处理, 如此对于

整个系统的优化以及判断会更加直观,也在一定程度上增强了系统的可靠性。

3.1 实验方案流程设计

(1) 第1步. 实验样本图像采集。

本次研究从某油茶种植园采集油茶果图片,于2019年8月3日、8月4日,进行了油茶果目标的采集工作。共采集图片质量清晰,特征明显的油茶果图像240张,共计约油茶果1200个。

(2) 第2步. 实验样本图像的预处理。

对于采集完的图像,我们还需要对其进行预处理,除了对已获取的图片进行筛选,我们还需要对实验样本图像数据增强。由于卷积神经网络对图像变换保持不变性,这就是数据增强的前提^[12-15]。

数据增强可以成倍增加数据集数量可以达到更好的训练效果,防止过拟合减少不相关特征的影响^[16]。

对于获得的油茶果图片,进行翻转、锐度变换、裁剪等操作对已有数据集进行数据增强^[17-19]。我们选择已有的240张油茶果图像,进行随机的图片变换操作,将数据集扩大到600张,共计约油茶果3000个,从中选择300张用于训练,且将测试样本分类为疏松样本、背景遮挡样本、果实遮挡样本3个部分。

(3) 第3步. 训练与测试样本目标标注。

由于监督学习可以获得更完善的目标特征信息,本文采用监督学习的方式,先把目标区域框出,加强目标特征提取效率,提高识别率。标注方法如下:

利用LabelImg图片标注工具,LabelImg是基于Python的图片标注工具,支持自主标注图片,框出目标位置和类别,并自动生成包含目标位置信息的XML文件、所有图片地址的txt文档和图片本身。于此得到了一个完整的样本VOC数据集。图4为包含目标位置信息的XML文件,图5是LabelImg图片标注工具。

(4) 第4步. 油茶果识别系统网络训练。

对已标注过的油茶果数据集,放入YOLOv3网络中进行特征提取,建立油茶果识别模型,其训练流程如图6所示。

大量油茶果图片通过YOLOv3网络进行提取特征和深度学习,保留了其目标特征的大量信息,已经实现初步的识别效果。

(5) 第5步. 油茶果识别系统检测。

训练出的油茶果模型的精密度得通过测试数据集进行试验,根据试验结果中油茶果目标的识别率来判断算法的优越性。

```
<?xml version="1.0"?>
- <annotation>
  <folder>油茶果</folder>
  <filename>fruit (2).jpg</filename>
  <path>C:/Users/Xiong/Desktop/油茶果/fruit (2).jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>400</width>
    <height>520</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>fruits</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>153</xmin>
      <ymin>86</ymin>
      <xmax>298</xmax>
      <ymin>245</ymin>
    </bndbox>
  </object>
- </annotation>
```

图4 包含目标位置信息的XML文件

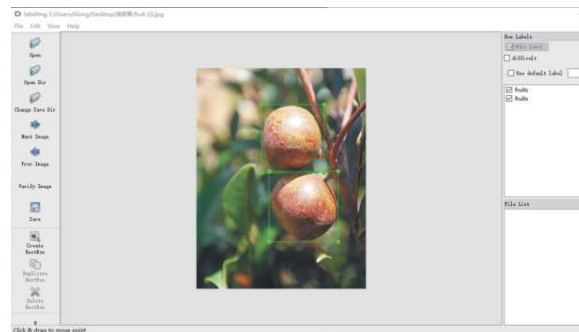


图5 LabelImg 图片标注工具

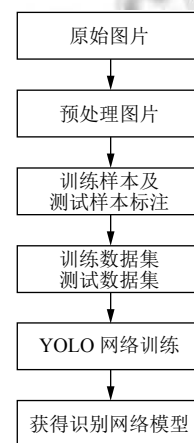


图6 油茶果识别系统网络训练过程

图7-图9为YOLOv3在不同环境下的识别效果图,图10是油茶果定位系统的可视化界面。

4 定位系统实验结果分析

YOLOv3算法在对油茶果图片的检测过程中表现

稳定,能够对测试样本中各种情况下的油茶果目标进行精准的识别并进行三维定位,油茶果定位系统运行顺畅,鲁棒性较强.此外为了对比算法的效果,此次试验还对相同条件下的其他算法进行比对,对识别的准确率和召回率进行比较.



图7 普通完整油茶果识别



图8 背景遮挡油茶果识别



图9 果实重叠遮挡油茶果识别

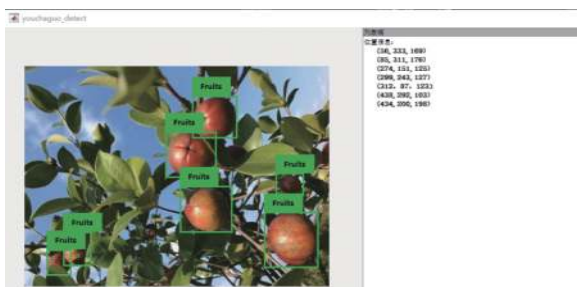


图10 油茶果定位系统可视化界面

在试验对比中我们统计了 R-CNN、FastRCNN 和 YOLO 以及 YOLOv3 目标检测常用算法的准确率、召回率和用时,如表 3.且针对不同环境情况进行分类实验,研究油茶果识别的视觉系统对不同情况下油茶果图片识别率差异.其柱状图如图 11 所示.

由表 3 和图 8 得出, YOLOv3 在 3 种场景下的准确率分别达到 98.20%、93.65% 和 90.21%. YOLOv3 算法比 R-CNN、FastRCNN、YOLO 的准确率更好.在背景遮挡与果实重叠情况时 YOLOv3 算法与其他算

法差异更加明显.证明 YOLOv3 在复杂环境下具有更精确的识别能力.

表 3 4 种算法准确率对比 (%)

| 油茶果场景 | R-CNN | FastRCNN | YOLO | YOLOv3 |
|-------|-------|----------|-------|--------|
| 疏松果实 | 90.03 | 96.93 | 92.97 | 98.21 |
| 背景遮挡 | 87.07 | 92.90 | 91.82 | 93.65 |
| 果实重叠 | 82.16 | 89.97 | 85.53 | 90.21 |

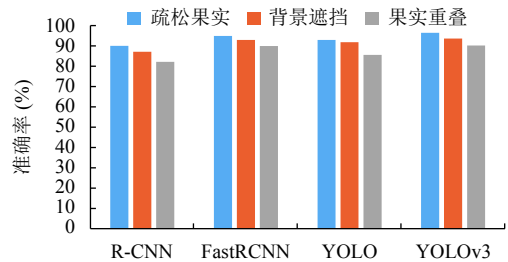


图 11 4 种算法准确率柱形图

4 种算法的召回率如表 4 所示,其数据转换成柱状图如图 12 所示.

表 4 4 种算法的召回率对比 (%)

| 油茶果场景 | R-CNN | FastRCNN | YOLO | YOLOv3 |
|-------|-------|----------|-------|--------|
| 疏松果实 | 94.17 | 95.02 | 96.23 | 96.87 |
| 背景遮挡 | 88.54 | 92.07 | 90.77 | 92.20 |
| 果实重叠 | 81.13 | 86.73 | 84.45 | 87.65 |

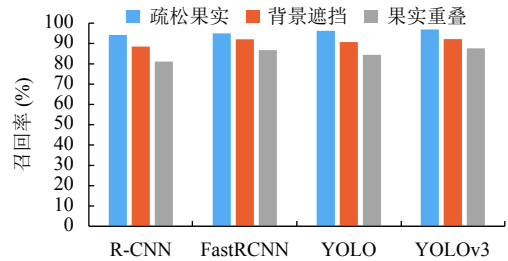


图 12 4 种算法召回率柱形图

由表 4 和图 12 得出, YOLOv3 在上述情况下的召回率分别为 96.86%、92.20%、87.65%.在复杂情况下的总召回率比 R-CNN、FastRCNN 和 YOLO 分别提高了 6.51%、0.92% 和 3.20%.分析结果可知, YOLOv3 算法总体的识别率更高,其提取的目标特征更为完全,且在复杂环境下算法具高的鲁棒性. P-R 曲线图如图 13 所示.

改进的 YOLOv3 算法与 FastRCNN 两种目标检测算法检测 300 张目标图片的用时分别为 7.87 s 和 15.43 s.比较 FastRCNN 与 YOLOv3 这两种识别率较好的算法,我们发现改进的 YOLOv3 算法速度上更快.

综上所述, YOLOv3 算法比 YOLO 和 R-CNN 算

法有更好的识别精度. 在同等条件下, YOLOv3 准确率比 YOLO 和 R-CNN 算法提高了 3.91% 和 7.6%; 而召回率则提高了 4.29% 和 1.75%, 在复杂背景条件下, 准确率提高明显, 并且在识别速度上远远高于其他算法.

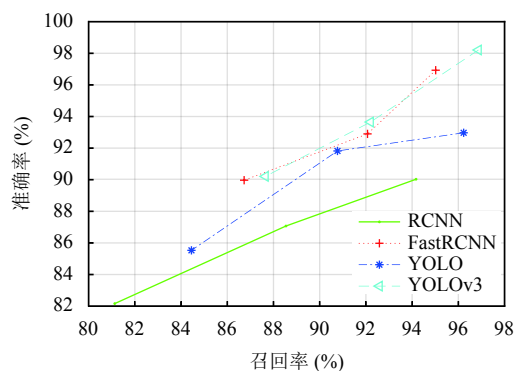


图 13 P-R 曲线图

5 结论与展望

通过实验发现了改进的 YOLOv3 算法在油茶果单种目标识别方面尤其是在一些复杂环境下具有准确度高、召回率高、识别速度快等优点. 实验还探究了双目立体视觉系统在油茶果目标定位上的可行性, 对于定位的准确性上可能还会出现偏差. 所以后续会在双目视觉系统的基础上, 油茶果的定位仅仅是个开始, 实现使用采摘机器人对成熟的油茶果采摘才是本课题研究的最终目的, 将视觉识别系统与机器臂联系起来也是研究的重点, 视觉系统可将捕捉到的图像中的目标果实进行识别和定位, 将识别到的每一处油茶果标记, 使标记的图像框位置并线性排列, 进行硬件组态和软件设计, 控制机械臂的轨迹优化运动, 可以实现对成熟油茶果的精准采摘^[20].

参考文献

- 黄光文, 张祖姣, 沈玉平, 等. 油茶自然落籽特性的研究. 湖南科技学院学报, 2019, 40(10): 32-34.
- 杨碧玉. 油茶产业“三千工程”破题开局. 江西日报, 2021-04-01(007).
- 黄健, 张钢. 深度卷积神经网络的目标检测算法综述. 计算机工程与应用, 2020, 56(17): 12-23.
- 许必青, 宫婧, 孙知信. 基于卷积神经网络的目标检测模型综述. 计算机技术与发展, 2019, 29(12): 87-92. [doi: 10.3969/j.issn.1673-629X.2019.12.016]
- 吕朦. 基于多尺度卷积神经网络的图像分类算法研究 [硕士学位论文]. 武汉: 湖北工业大学, 2019.
- ZahediNasab R, Mohseni H. Neuroevolutionary based convolutional neural network with adaptive activation

- functions. Neurocomputing, 2020, 381: 306-313. [doi: 10.1016/j.neucom.2019.11.090]
- Sharpe SM, Schumann AW, Boyd NS. Detection of Carolina geranium (*Geranium carolinianum*) growing in competition with strawberry using convolutional neural networks. Weed Science, 2019, 67(2): 239-245. [doi: 10.1017/wsc.2018.66]
- Altabay W, Noori M, Alarjani A, et al. Nano-delamination monitoring of BFRP nano-pipes of electrical potential change with ANNs. Advances in Nano Research, 2020, 9(1): 1-13.
- Yucel O, Aydin ES, Sadikoglu H. Comparison of the different artificial neural networks in prediction of biomass gasification products. International Journal of Energy Research, 2019, 43(11): 5992-6003. [doi: 10.1002/er.4682]
- Chau VTN, Phung NH. Enhanced CNN models for binary and multiclass student classification on temporal educational data at the program level. Vietnam Journal of Computer Science, 2021, 8(2): 311-335. [doi: 10.1142/S2196888821500135]
- 刘红铮. 基于深度学习的花朵图像识别算法研究与实现 [硕士学位论文]. 南昌: 南昌航空大学, 2019.
- Song ZN, Sui HG, Hua L. A hierarchical object detection method in large-scale optical remote sensing satellite imagery using saliency detection and CNN. International Journal of Remote Sensing, 2021, 42(8): 2827-2847. [doi: 10.1080/01431161.2020.1826059]
- 李兵兵. 基于卷积神经网络的芒果识别视觉系统研究 [硕士学位论文]. 成都: 西华大学, 2019.
- 刘振宇, 丁宇祺. 自然环境中被遮挡果实的识别方法研究. 计算机应用研究, 2020, 37(S2): 333-335, 339.
- Gopi N, Rao PRK. Automated mango fruit recognition by multi-task convolutional neural networks for harvest robot. International Journal of Engineering and Advanced Technology (IJEAT), 2020, 9(6): 214-219. [doi: 10.35940/ijeat.F1383.089620]
- Liu DJ, Shen J, Yang HS, et al. Recognition and localization of actinidia arguta based on image recognition. EURASIP Journal on Image and Video Processing, 2019, 2019(1): 21. [doi: 10.1186/s13640-019-0419-6]
- Huang YQ, Zheng JC, Sun SD, et al. Optimized YOLOv3 algorithm and its application in traffic flow detections. Applied Sciences, 2020, 10(9): 3079. [doi: 10.3390/app10093079]
- Katarzyna R, Paweł M. A vision-based method utilizing deep convolutional neural networks for fruit variety classification in uncertainty conditions of retail sales. Applied Sciences, 2019, 9(19): 3971. [doi: 10.3390/app9193971]
- 许凯. 基于图像识别的苹果果实检测技术. 实验室研究与探索, 2016, 35(10): 36-39. [doi: 10.3969/j.issn.1006-7167.2016.10.010]
- 高峰. 自然环境下的柑橘目标识别方法研究 [硕士学位论文]. 北京: 北方工业大学, 2019.