

温室远程监控系统人机交互与番茄识别研究

张美娜^{1,2} 王 潇³ 梁万杰² 曹 静² 张文字^{2,4}

(1. 南京农业大学工学院, 南京 210031; 2. 江苏省农业科学院农业信息研究所, 南京 210014;
3. 南京农业大学人工智能学院, 南京 210031; 4. 江苏大学农业工程学院, 镇江 212013)

摘要: 为提升设施农业远程监控系统的数据可视化与信息化程度,设计了一种温室远程监控系统,该系统主要由巡检机器人、移动通信网络、云服务器与远程监控中心组成,实现了温室端与远程监控中心端之间的文本、图像、视频3类数据传输。综合应用机器学习、深度学习算法实现人机交互与温室端番茄识别任务。基于 Haar 级联算法与 LBPH 算法实现了管理员人脸识别,识别成功率达 90%;基于 YOLO v3 与 ResNet-50 算法分别识别手部与手部关键点,单手、双手的识别置信度分别为 0.98 与 0.96;基于提取的食指指尖坐标与左右手部候选框中心点坐标实现了手指交互与图像尺寸缩放的功能。应用 Swin Small + Cascade Mask RCNN 网络模型,针对农业数据集有限的问题,对比分析了应用迁移学习方法前后的番茄检测效果。试验结果表明,应用迁移学习方法后,模型收敛速度有所提升且收敛后的损失值均有所下降;同时,IoU 为 0、0.5、0.75 时的平均精度(mask AP)分别提升了 7.8、6.4、7.2 个百分点,模型性能更优。

关键词: 温室; 远程监控; 人机交互; 番茄识别; 深度学习; Swin Transformer

中图分类号: S62; TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2022)10-0363-08

OSID:



Human-computer Interaction and Tomato Recognition in Greenhouse Remote Monitoring System

ZHANG Meina^{1,2} WANG Xiao³ LIANG Wanjie² CAO Jing² ZHANG Wenyu^{2,4}

(1. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China
2. Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China
3. College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China
4. School of Agriculture Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: In order to improve the data visualization and information level, a kind of greenhouse remote monitoring system was designed and developed, which included inspection robot, mobile communication network, cloud server and remote monitoring center. Three kinds of data transmitted between the greenhouse and the remote monitoring center, including text, image and video. Machine learning and deep learning algorithms were used for human-computer interaction and tomato recognition tasks. On the one hand, administrator face recognition was achieved based on Haar cascade and LBPH algorithms, and the recognition success rate was 90%. Then YOLO v3 and ResNet-50 algorithms were used to recognize the hand and the key points of hand respectively, and the recognition confidence of single-hand and two-hand was 0.98 and 0.96, respectively. Based on the extracted coordinates of the forefinger and the center points of the left and right hand candidate frame, finger interaction and image size scaling were realized. On the other hand, the model framework of Swin Small + Cascade Mask RCNN was used for tomato recognition. Aiming at the problem of limited agricultural data set, the effect of tomato detection before and after applying transfer learning method was compared and analyzed. By using transfer learning method, the experimental results showed that the convergence rate of the model was increased and the loss value was decreased. In terms of semantic segmentation, AP value was used to evaluate the model performance, when IoU was set to be 0, 0.5 and 0.75, test results showed that the mask average

收稿日期: 2022-06-16 修回日期: 2022-07-11

基金项目: 江苏省农业科技自主创新项目(CX(20)1005、CX(20)3073)

作者简介: 张美娜(1986—),女,副研究员,博士,主要从事智能农机装备传感监测与控制系统研究,E-mail: zmnll22@163.com

通信作者: 张文字(1983—),男,副研究员,博士,主要从事作物系统模拟等智慧农业关键技术研究,E-mail: research@wwery.cn

precisions were improved by 7.8 percentage points, 6.4 percentage points and 7.2 percentage points, respectively after using transfer learning method.

Key words: greenhouse; remote monitoring; human-computer interaction; tomato recognition; deep learning; Swin Transformer

0 引言

我国是设施农业大国,设施种植面积占世界总面积的 85% 以上,但与发达国家相比,我国设施农业的投入产出比、生产效率、机械化率与现代化管理水平仍然较低^[1-4]。温室远程监控是一种现代化管理手段,依靠多源传感器、智能装备、物联网以及人机交互等多项技术支撑^[5-8],如荷兰 Priva Topcrop Monitor 作物监测系统、美国 Sensaphone 系列环境与网络监控报警系统以及以色列 BF-Agritech 植物生态监测系统等。近年来,我国规模化温室面积稳步发展,基本配备有环境管控系统,但多依赖国外进口,受限于封装的软件系统,温室内监测的数据开放和利用不便。

传统的远程监控端人机交互方式主要依靠键盘、鼠标等外设,互动性不强;同时,数据展示方式以静态文本与监控视频为主,原始数据较多而综合性、分析性、决策性的数据较少,缺少数据信息与真实世界的交融。国外研究学者开展了相关研究,如 KING 等^[9]开发了 ARVino 可视化系统,实现了在实景视频中叠加地理信息系统 (Geographic information system, GIS) 数据与葡萄种植数据,为葡萄栽培者提供了三维可视化的数据与分析。HUUSKONEN 等^[10]设计了一种增强现实 (Augmented reality, AR) 系统,远程监控端的操作员不仅可以获得实时可视化信息,还可以通过配套的可穿戴设备控制农机行驶。国内以计算机视觉技术为基础的人机交互系统在农业场景中的应用较少。

作物生理、形态等参数的获取首先建立在作物目标识别的基础上,国内外在该领域的研究主要集中在应用机器学习与深度学习算法从作物图像中识别目标。KOIRALA 等^[11]在果实检测方法综述中指出深度学习方法优于传统机器学习、神经网络、浅层卷积神经网络算法,果实识别精度普遍达到 85% 以上。用于目标检测的深度学习方法主要分为两阶段 (2-stage) 和单阶段 (1-stage) 检测算法。两阶段检测算法包括 R-CNN (Region-convolutional neural network)、SPP- NET (Spatial pyramid pooling networks)、Fast R-CNN、Faster R-CNN、R-FCN (Region-fully convolutional network)、Mask R-CNN 等。单阶段检测模型算法包括 SSD (Single shot

multibox detector)、YOLO (You only look once) v1 ~ v5 等^[12]。相关研究主要以离线采集目标图像进行识别检测为主^[13-16],极少涉及远程实时监控的场景。同时,新型深度神经网络模型算法仍不断涌现,使其在农业领域的探索应用不断拓展。

综上所述,本文设计一种温室远程监控系统,在远程监控端集成图像融合算法、人脸识别与手势识别算法,增强远程监控系统的多类型环境数据可视化与人机交互效果。同时,应用深度学习网络模型算法,实时、在线识别温室端监控视频中的番茄目标,以期管理者提供作物生长信息奠定基础。

1 监控系统构建与识别算法设计

1.1 温室远程监控系统结构

温室远程监控系统设计需求主要包括:①温室内通过“机器换人”实现温室环境与作物生长监测的无人化。②通过远程监控中心能够可视化温室内各类型的监测数据(文本、图像、视频等),具有预警提示功能。③安全、互动性强的人机交互系统。针对上述需求设计一种温室远程监控系统,如图 1 所示,系统主要由温室巡检平台、通信网络和远程监控中心 3 部分组成。

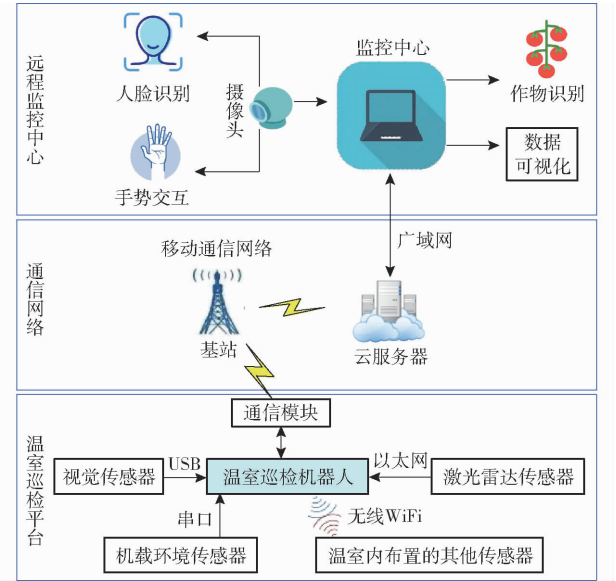


图 1 系统结构示意图

Fig. 1 Architecture of system

温室巡检平台以巡检机器人(松灵 SCOUT2.0 型)为运动载体,搭载控制器、多源传感器与通信模块实现温室内环境与作物的动态监测。控制器为

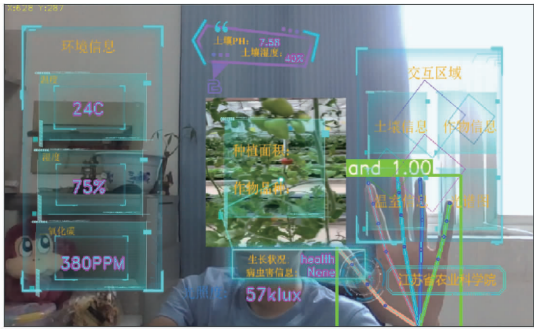


图3 人机交互界面

Fig. 3 Human-computer interaction interface

Fisherface 以及局部二值模式 (Local binary patterns, LBP), 其中 Eigenface 算法的局限性为: 待识别图像中人脸尺寸接近特征脸中人脸的尺寸, 且待识别人脸图像必须为正面人脸图像; Fisherface 算法增加了线性判别分析, 对人脸差异的识别效果有所提升, 但对光照、人脸姿态的变化不敏感; LBP 算法将人脸图像分为若干子区域, 并在子区域内根据 LBP 值统计其直方图, 以直方图作为其判别特征, 对光照有明显的鲁棒性^[17]。综上所述, 系统以 OpenCV 库为支持, 采用 Haar 级联算法结合局部二进制模式直方图 (Local binary pattern histogram, LBPH) 算法分别实现人脸检测与有权限管理员人脸识别。

首先采用 Haar 级联算法实现人脸检测, Haar 级联算法是一种用于在图像上定位对象的检测方法, 具有识别速度快的优点^[18]。算法原理概括为: ①使用 Haar-like 特征做检测。②使用积分图 (Integral image) 对 Haar 特征求值进行加速。③使用 AdaBoost 算法选择关键特征, 进行人脸和非人脸分类。④使用 Cascade 级联把强分类器级联到一起, 提高准确率。

然后, 采用 LBPH 算法^[19]实现管理员人脸识别, 算法实现概括为: ①对待识别整幅图像进行 LBP 编码。②将图像分区并获取各区域 LBP 编码直方图, 转换为直方图矩阵 HIST。③计算待识别人脸图像直方图矩阵与已训练好的模型直方图矩阵之间的距离, 若距离差小于设定阈值则判断为管理员, 否则判断为非管理员。LBPH 算法具有不受光照、缩放、旋转和平移影响的优点。

1.3.2 手势识别算法

识别手部并定位手部关键点位置用于实现人机交互操作, 即界面切换与图像尺寸缩放。随着目标检测技术的发展, 很多学者将传统基于人工建模方式的手势识别转换成目标检测问题。考虑实时性需求, 选择 YOLO 系列目标检测算法, 其中, YOLO v3 在手势识别中应用较为广泛^[20-21], 为此, 本系统采用 YOLO v3 算法^[22]先进行手部识别, 通过 PyTorch

框架实现。识别后的手部图像将被截取输入到后续网络, 进一步提取手部关键点信息。

手部关键点检测与面部关键点检测类似, 应用面部检测表现突出的 ResNet-50 网络^[23]检测手部 21 个关键点并输出坐标矩阵。ResNet 网络有效解决了深度神经网络的退化问题, 在神经网络发展史上具有里程碑意义。模型训练过程中, 为解决传统 L2 损失函数对离群值敏感的问题, 采用 FENG 等^[23]提出的 Wing loss 作为损失函数计算式, 为

$$\text{wing}(x) = \begin{cases} w \ln(1 + |x|/\theta) & (|x| < w) \\ |x| - D & (|x| \geq w) \end{cases} \quad (1)$$

其中 $D = w(1 - \ln(1 + w/\theta))$ (2)

式中 $\text{wing}(x)$ ——Wing loss 函数

w ——增大损失函数的值

$|x|$ ——误差 D ——常数

θ ——约束非线性区域曲率

最后, 以输出的食指指尖坐标为圆心绘制 (eclipse() 函数) 动态圆形加载图标, 同时, 用直线连接左右手部候选框中心点坐标 (Line() 函数), 根据直线距离的长短控制图像尺寸。

1.4 作物识别算法

1.4.1 算法模型

远程监控中心端的温室作物视频监控界面具备作物识别功能, 可为作物尺寸、成熟度等生长参数的进一步提取奠定基础。农业环境复杂, 传统 CNN 算法通过加深网络层数提高图像理解能力容易出现过拟合, 性能提升有限。另一方面, Transformer 作为一种基于注意力的编码器-解码器架构, 从自然语言处理 (NLP) 领域拓展至计算机视觉 (CV) 领域。视觉 Transformer 架构 Swin Transformer 在 ImageNet、COCO 和 ADE20k 等多个数据集上取得了非常优异的性能。为此, 应用以 Swin Transformer^[24]为主干网络提取特征图, 再进入 Cascade Mask RCNN^[25]后续网络进行果实语义分割的网络模型, 从而实现番茄作物识别, 模型整体结构示意图如图 4 所示。

Swin Transformer 网络模型主要考虑了视觉信号的特点, 解决了传统 Transformer 模型多尺度建模与计算量巨大的问题, 能够灵活处理不同尺度的图像, 采用滑动窗口机制, 将注意力的计算限制在每个窗口内, 因此大大降低了计算复杂度。Swin Transformer 网络模型输出与典型卷积模型的特征图分辨率一致, 已成为一种基本骨干网络。这里选用 Swin Small 为主干网, 其中 Stage 1~4 的 Block 数量分别为 2、2、18、2, 维度参数 C 为 96。

Cascade Mask RCNN 主要参考了实例分割主流算法 Mask RCNN^[26]模型结构, 在每个级联阶段添加

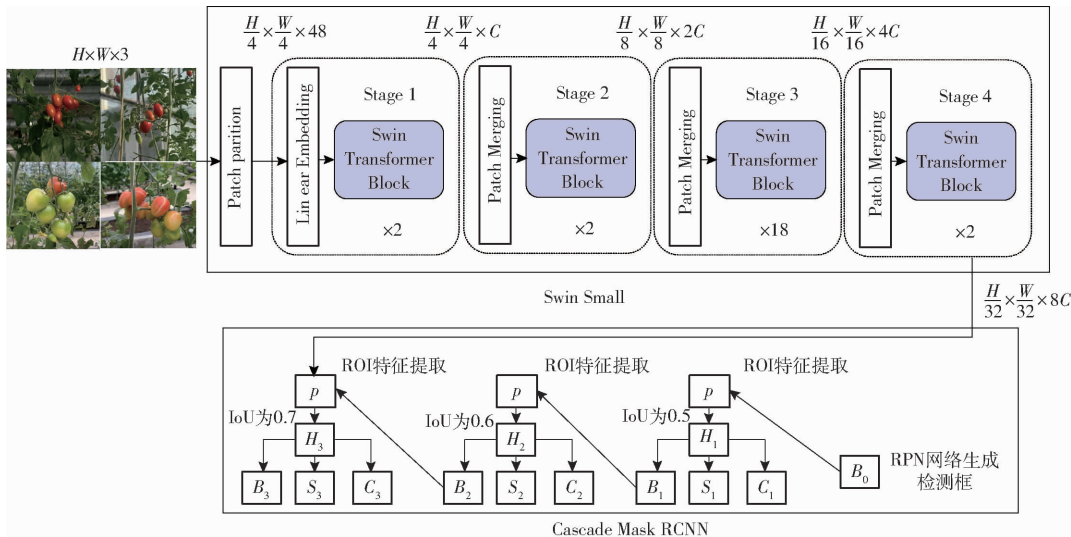


图 4 番茄识别网络模型示意图

Fig. 4 Diagram of network model for tomato recognition

一个预测目标掩码的分支,从而实现更精确的实例分割。主干网获取的特征图进入 Cascade Mask RCNN 后续的区域生成网络 (Region proposal network, RPN) 和级联的检测网络分别实现目标识别和定位两阶段任务,通过使用不同的交并比 (Intersection over union, IoU) 阈值训练多个级联的检测器能够有效提高目标检测精度。模型中, IoU 梯度分别设置为 0.5、0.6、0.7。

1.4.2 损失函数

在模型训练阶段采用多种损失函数结合的方式评估模型。其中,图 4 中 B_0 处 Bounding Box 采用 Smooth L1 作为损失函数。分类性能采用交叉熵损失函数。

Cascade Mask RCNN 模型不同 IoU 梯度下的分类性能和语义分割性能仍然采用交叉熵损失函数评估; Bounding Box 采用 GIoU^[27] 作为损失函数。

1.4.3 迁移学习应用

农业场景中,精细领域的数据集仍处于稀缺状态,制作大批量的高质量数据集所需的时间与人力成本较高,迁移学习是解决样本不足的有效方法之一。利用官方发布的 coco 预训练模型作为源模型,为后续番茄数据集模型训练的学习起点,以充分利用源模型的知识基础。

2 试验

2.1 系统通信试验

温室巡检机器人端控制器将采集到的环境传感器数据转换为字符串并连接为字符串报文进行发送,主要包括温室温度 (℃)、温室相对湿度 (%)、光照强度 (lx)、二氧化碳浓度、土壤湿度 (%)、土壤 pH 值、病虫害染病程度 (%)、作物健康状况 (健康、

不健康) 共 8 类数据,设置文本数据发送周期为 1 h。

视频数据应用 OpenCV 库的图像压缩函数 imencode() 对图像帧进行处理,转换为字节流后传输,解决了基于像素访问进行传输数据量大的问题,接收端收到数据后利用 imdecode() 函数进行解压缩操作。图像数据传输与视频数据传输方法相同。系统通信试验结果表明,温室巡检机器人发送端到远程监控中心端之间的通信延迟均在 200 ms 以内。

2.2 人机交互试验

2.2.1 人脸识别试验

人脸检测训练数据来源于 OpenCV 官方训练的 XML 文件 haarcascade_frontalface_default.xml。同时,在不同环境拍摄不同表情、不同姿态管理员面部图像 80 幅形成训练数据集,并进行验证,用 20 幅管理员面部图像作为验证集,识别成功率为 90%。

2.2.2 手势识别试验

手势识别数据集来自 TV-Hand 中 2 万幅手部图像^[28]; 手部特征点检测数据集来自 Large-scale Multiview 3D Hand Pose Dataset 中的 5 万幅图像^[29]。YOLO v3 与 ResNet-50 网络超参设置如表 2 所示。模型在 AMD EPYC 7302 16-Core Processor, GeForce RTX 3090, 23 GB 显存, 搭载 Linux 系统的环境下训练。Wing loss 损失函数参数中 w 和 θ 分别设置为 5 和 2。

手部识别试验结果表明,20 次不同角度下的单手识别置信度均大于 0.98; 同时,20 次双手识别置信度均大于 0.96。手部关键点检测的 ResNet-50 网络模型,训练 146 次的损失函数变化曲线如图 5 所示,损失函数最终收敛在 0.1 以下。

经手势识别后,人机交互与图像尺寸缩放的功能实现如图 6 所示。

表 2 网络超参数设置

Tab.2 Network hyperparameter settings		
超参数	YOLO v3	ResNet - 50
单批次图像数	9	16
模型输入图像尺寸/(像素×像素)	416×416	
优化器动量		0.9
epochs	320	150
训练时学习率更新参数	10, 20, 30	
学习率衰减指数		0.95
初始学习率	2×10^{-5}	1×10^{-3}

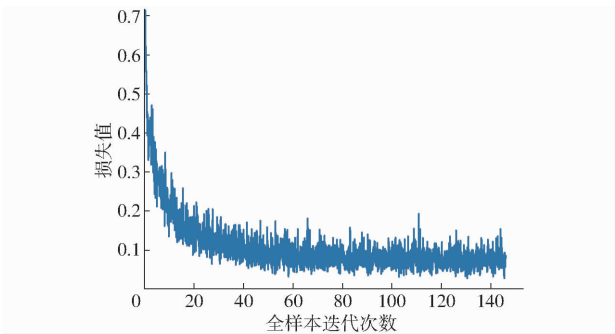


图 5 手势识别网络损失函数变化曲线
Fig. 5 Loss function curve of gesture recognition network

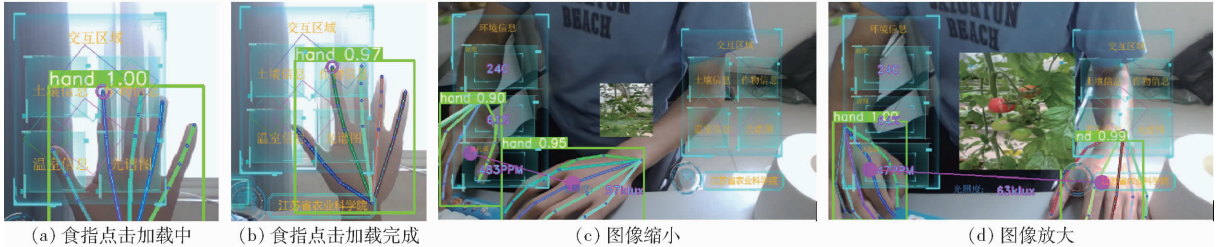


图 6 手势识别交互实现
Fig. 6 Human-computer interaction by gesture

2.3 作物识别试验

2.3.1 数据集与超参数设置

基于温室巡检机器人在温室采集不同角度、不同遮挡情况以及不同成熟度共计 1 800 幅番茄生长图像制作数据集。首先,采用基于 PaddlePaddle 的高效图像分割工具 PaddleSeg 人工标注图像;然后将标注后得到 VOC 格式的 json 文件经算法批量转换为 Lableme 格式的 coco 数据集;最后,将标注的番茄果实数据集划分为训练集、测试集、验证集,按比例 8:1:1 进行随机分配。模型同样在 AMD EPYC 7302 16-Core Processor, GeForce RTX 3090, 23 GB 显存, 搭载 Linux 系统的环境下训练。Cascade Mask RCNN 网络中 IoU 梯度分别设置为 0.5、0.6、0.7;网络超参数设置为:单批次图像数使用迁移学习算法前为 2,使用迁移学习算法后为 1,选用 Adamw 优化器,学习率为 1×10^{-5} ,权重衰减为 0.05,预热学习率为 500。使用迁移学习之后将单批次图像数的值由原来的 2 改为 1,该超参数为一次训练所选取的样本数,影响模型的优化程度和速度,在 GPU 内存受限的情况下取值较小。

2.3.2 迁移学习应用效果分析

使用迁移学习方法前后损失函数变化曲线如图 7 所示,损失值统计如表 3 所示。

在训练速度方面,由图 7 可知,使用迁移学习前后的模型都在迭代 15 000 次后收敛,但使用迁移学习后的单批次图像数调整为 1,即花费了更少的训练时间。由表 3 可知,使用迁移学习方法后不同 IoU 阈值下损失值均小于未使用迁移学习方法的损

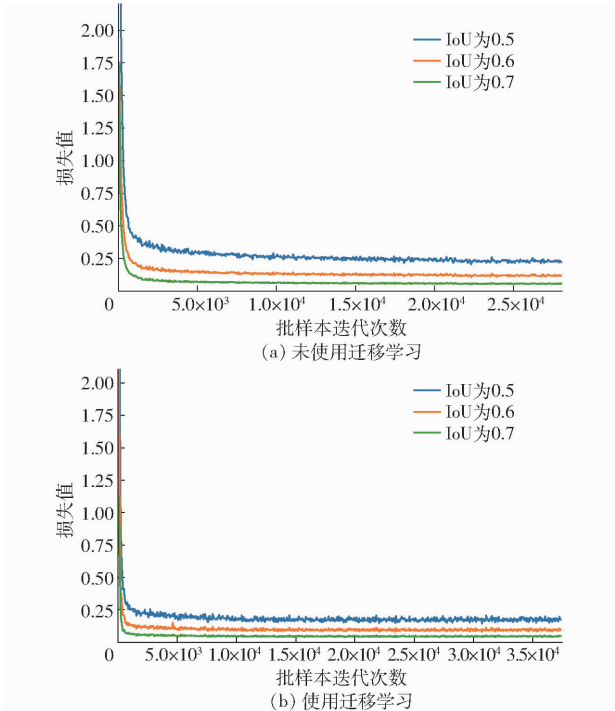


图 7 使用迁移学习前后损失函数变化曲线
Fig. 7 Loss functions curves before and after using transfer learning method

表 3 使用迁移学习前后损失值对比
Tab.3 Comparison of loss value before and after using transfer learning method

类型	IoU		
	0.5	0.6	0.7
未使用迁移学习	0.24	0.13	0.07
使用迁移学习	0.18	0.10	0.05

失值。因此,应用迁移学习方法后,模型在收敛速度

和最终收敛的损失值上更优。

同时,采用平均精度 mask AP (Average precision)评估使用迁移学习前后的模型语义分割性能,mask AP 越大,模型效果越好。mask AP 计算结果如表 4 所示。

表 4 使用迁移学习前后语义分割的 mask AP 对比
Tab.4 Comparison of mask AP value before and after using transfer learning method %

类型	IoU		
	0	0.5	0.75
未使用迁移学习	39.0	67.1	40.8
使用迁移学习	46.8	73.5	48.0

由表 4 可知,使用迁移方法后,IoU 取 0、0.5、0.75 时的 mask AP 提升了 7.8、6.4、7.2 个百分点,因此,该模型精度优于未使用迁移学习方法的模型。番茄实物最终的检测效果如图 8 所示。



图 8 检测效果
Fig.8 Detection results

3 结论

(1)设计了一种温室远程监控系统。该系统以巡检机器人作为传感器数据采集与通信载体将温室 内作物及环境数据发送至通信模块;通信系统采用 移动通信网络加云服务器的网络架构实现温室端与 远程监控中心端的文本、图像及视频 3 种类型数据 的传输。

(2)集成了人脸识别与手部关键点识别算法实 现远程监控端的人机交互。其中,管理员人脸识别 基于 Haar 级联算法与 LBPH 算法,识别成功率达 90%。手部关键点识别基于 YOLO v3 与 ResNet – 50 算法模型,单手部识别置信度均大于 0.98,双手 识别置信度均大于 0.96;同时,基于提取的食指指 尖坐标与左右手部候选框中心点坐标实现了人机交 互与图像尺寸缩放的功能。

(3)应用最新的 Swin Small + Cascade Mask RCNN 深度学习网络模型进行远程视频中番茄的 检测与语义分割。同时,采用迁移学习方法解决 农业场景数据集不足的问题。试验结果表明,使用 迁移学习方法后,模型的收敛速度更快、损失值 更低,语义分割精度有所提升,IoU 取 0、0.5、0.75 时的 mask AP 分别提升了 7.8、6.4、7.2 个百分 点。

参 考 文 献

[1] 孙锦,高洪波,田婧,等. 我国设施园艺发展现状与趋势[J]. 南京农业大学学报,2019,42(4):594 – 604.
SUN Jin,GAO Hongbo,TIAN Jing, et al. Development status and trends of protected horticulture in China[J]. Journal of Nanjing Agricultural University,2019,42(4):594 – 604. (in Chinese)

[2] 刘霓红,蒋先平,程俊峰,等. 国外有机设施园艺现状及对中国设施农业可持续发展的启示[J]. 农业工程学报,2018,34(15):1 – 9.
LIU Nihong,JIANG Xianping,CHENG Junfeng, et al. Current situation of foreign organic greenhouse horticulture and its inspiration for sustainable development of Chinese protected agriculture[J]. Transactions of the CSAE,2018,34(15):1 – 9. (in Chinese)

[3] 辜松. 我国设施园艺生产作业装备发展浅析[J]. 现代农业装备,2019,40(1):4 – 11.
GU Song. An analysis on general situation of operating machines for facility horticulture production in China[J]. Modern Agricultural Equipment,2019,40(1):4 – 11. (in Chinese)

[4] 齐飞,李恺,李邵,等. 世界设施园艺智能化装备发展对中国的启示研究[J]. 农业工程学报,2019,35(2):183 – 195.
QI Fei,LI Kai,LI Shao, et al. Development of intelligent equipment for protected horticulture in world and enlightenment to China[J]. Transactions of the CSAE,2019,35(2):183 – 195. (in Chinese)

[5] VILLA-HENRIKSEN A, EDWARDS G T C, PESONEN L A, et al. Internet of things in arable farming: implementation, applications, challenges and potential[J]. Biosystems Engineering,2020,191:60 – 84.

[6] TANG Y, DANANJAYAN S, HOU C, et al. A survey on the 5G network and its impact on agriculture: challenges and opportunities[J]. Computers and Electronics in Agriculture,2021,180:105895.

[7] KUMAR A, SINGH V, KUMAR S, et al. IoT enabled system to monitor and control greenhouse[J]. Materials Today: Proceedings,2022,49:3137 – 3141.

[8] VASCONEZ J,KANTOR G,AUAT CHEEIN F A. Human-robot interaction in agriculture:a survey and current challenges[J]. Biosystems Engineering,2019, 179:35 – 48.

[9] KING G R,PIEKARSKI W,THOMAS B H. ARVino-outdoor augmented reality visualisation of viticulture GIS data[C]//Fourth

- IEEE and ACM International Symposium on Mixed and Augmented Reality,2005:52 – 55.
- [10] HUUSKONEN J,OKSANEN T. Augmented reality for supervising multirobot system in agricultural field operation[J]. IFAC-PapersOnLine,2019,52(30):367 – 372.
- [11] KOIRALA A,WALSH K,WANG Z, et al. Deep learning-method overview and review of use for fruit detection and yield estimation[J]. Computers and Electronics in Agriculture,2019, 162:219 – 234.
- [12] ZOU Z,SHI Z,GUO Y, et al. Object detection in 20 years:a survey[J]. arXiv:1905.05055v2,2019.
- [13] WANG D, HE D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning[J]. Biosystems Engineering,2021,210:271 – 281.
- [14] GENE-MOLA J,SANZ-CORTIELLA R,ROSELL-POLO J R, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry[J]. Computers and Electronics in Agriculture,2020,169:105165.
- [15] 李天华,孙萌,丁小明,等. 基于 YOLOv4 + HSV 的成熟期番茄识别方法[J]. 农业工程学报,2021,37(21):183 – 190.
LI Tianhua,SUN Meng,DING Xiaoming, et al. Tomato recognition method at the ripening stage based on YOLOv4 and HSV [J]. Transactions of the CSAE,2021, 37(21):183 – 190. (in Chinese)
- [16] 何斌,张亦博,龚健林,等. 基于改进 YOLOv5 的夜间温室番茄果实快速识别[J]. 农业机械学报,2022,53(5):201 – 208.
HE Bin,ZHANG Yibo,GONG Jianlin, et al. Fast recognition of tomato fruit in greenhouse at night based on improved YOLOv5 [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):201 – 208. (in Chinese)
- [17] 王浩. 当今人脸识别技术的现状与未来研究方向[J]. 计算机与数字工程,2021,49(9):1905 – 1911.
WANG Hao. Current situation and future research direction of face recognition technology [J]. Computer & Digital Engineering,2021,49(9):1905 – 1911. (in Chinese)
- [18] VIOLA P,JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision,2004, 57(2):137 – 154.
- [19] AHONEN T,HADID A,PIETIKAINEN M. Face description with local binary patterns:application to face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2006,28(12):2037 – 2041.
- [20] 凌利,陶俊,吴瑰. 基于 YOLOv3 的手势识别技术[J]. 江汉大学学报(自然科学版),2021,49(5):79 – 87.
LING Li,TAO Jun,WU Gui. Gesture recognition technology based on YOLOv3[J]. Journal of Jiangnan University(Natural Science Edition),2021,49(5):79 – 87. (in Chinese)
- [21] 王粉花,黄超,赵波,等. 基于 YOLO 算法的手势识别[J]. 北京理工大学学报,2020,40(8):873 – 879.
WANG Fenhua,HUANG Chao,ZHAO Bo, et al. Gesture recognition based on YOLO algorithm[J]. Transactions of Beijing Institute of Technology,2020,40(8):873 – 879. (in Chinese)
- [22] REDMON J,FARHADI A. YOLOv3:an incremental improvement[J]. arXiv:1804.02767, 2018.
- [23] FENG Z H,KITTLER J, AWAIS M, et al. Wing loss for robust facial landmark localisation with convolutional neural networks [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2018.
- [24] LIU Z,LIN Y,CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[J]. arXiv:2103.14030, 2021.
- [25] CAI Z,VASCONCELOS N. Cascade R – CNN:high quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2021,43(5):1483 – 1498.
- [26] HE K,GKIOXARI G,DOLLAR P, et al. Mask R – CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(2):386 – 397.
- [27] REZATOFIGHI H, TSOI N,GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2019.
- [28] NARASIMHASWAMY S,WEI Z,WANG Y, et al. Contextual attention for hand detection in the wild[C]//IEEE International Conference on Computer Vision (ICCV),Seoul, 2019.
- [29] GOMEZ-DONOSO F,ORTS-ESCOLANO S,CAZORLA M. Large-scale multiview 3D hand pose dataset[J]. Image and Vision Computing,2019,81(1):25 – 33.