

doi:10.6041/j.issn.1000-1298.2020.S1.001

# 基于改进 YOLOv3 - tiny 的田间行人与农机障碍物检测

李文涛 张岩 莫锦秋 李彦明 刘成良

(上海交通大学机械与动力工程学院, 上海 200240)

**摘要:** 为实现农机自主作业中的避障需求,本文针对室外田间自然场景中因植被遮挡、背景干扰而导致障碍物难以检测的问题,基于嵌入式平台应用设备,提出了农机田间作业时行人和农机障碍物检测的改进模型,更好地平衡了模型的检测速度与检测精度。该改进模型以 You only look once version 3 tiny (YOLOv3 - tiny) 为基础框架,融合其浅层特征与第 2 YOLO 预测层特征作为第 3 预测层,通过更小的预选框增加小目标表征能力;在网络关键位置的特征图中混合使用注意力机制中的挤压激励注意模块 (Squeeze and excitation attention module, SEAM) 与卷积块注意模块 (Convolutional block attention module, CBAM),通过强化检测目标关注以提高抗背景干扰能力。建立了室外环境下含农机与行人的共 9 405 幅图像的原始数据集。其中训练集 7 054 幅,测试集 2 351 幅。测试表明本文模型的内存约为 YOLOv3 与单次多重检测器 (Single shot multibox detector, SSD) 模型内存的 1/3 和 2/3;与 YOLOv3 - tiny 相比,本文模型平均准确率 (Mean average precision, mAP) 提高 11 个百分点,小目标召回率 (Recall) 提高 14 百分点。在 Jetson TX2 嵌入式平台上本文模型的平均检测帧耗时 122 ms,满足实时检测要求。

**关键词:** 田间障碍物检测; 注意力机制; YOLOv3 - tiny; 嵌入式

**中图分类号:** TP391.4      **文献标识码:** A      **文章编号:** 1000-1298 (2020) S1-0001-08

## Detection of Pedestrian and Agricultural Vehicles in Field Based on Improved YOLOv3 - tiny

LI Wentao ZHANG Yan MO Jinqiu LI Yanming LIU Chengliang

(School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** The real-time detection of pedestrian and agricultural vehicles is very important for the navigation and path planning of autonomous agricultural vehicles. In the field, obstacles are difficult to be detected due to crops occlusion and background interference. A real-time pedestrian and agricultural vehicles detection model in natural field scene was proposed, which effectively improved the feasibility of pedestrian and agricultural vehicles visual detection to embedded platform in the independent operation of agricultural machinery. This detection model was improved based on You only look once version 3 tiny (YOLOv3 - tiny). A third prediction layer was got by merging the features of YOLOv3 - tiny's shallow layer and the features of second YOLO prediction layer, thus more smaller anchors resulted in the detection ability improvement of small targets. Both the squeeze and excitation attention module (SEAM) and the convolutional block attention module (CBAM) were applied in the key feature maps of the network, thus the model's anti-background disturbance capability was increased. A data set included 9 405 images of pedestrian and agricultural vehicles with different shooting angles and natural field scenes was set, and 7 054 images were used for training while the remained 2 351 images were used for testing. Tests showed that the memory size of the improved model was reduced to 1/3 and 2/3 of that of the YOLOv3 and single shot multibox detector (SSD) models, the improved model's mean average precision (mAP) was increased by 11 percentage points, and the small target recall (R) rate was increased by 14 percentage points while compared with that of YOLOv3 - tiny. On the Jetson TX2 embedded hardware platform, the single frame detection time of the improved model was 122 ms, which can meet the requirements of real-time detection.

**Key words:** field object detection; attention mechanism; YOLOv3 - tiny; embedded hardware

收稿日期: 2020-08-12 修回日期: 2020-09-14

基金项目: 国家重点研发计划项目 (2019YFB1312301)

作者简介: 李文涛 (1994-), 男, 博士生, 主要从事机电系统集成、机器视觉研究, E-mail: wentao.li@sjtu.edu.cn

通信作者: 李彦明 (1971-), 男, 副教授, 主要从事非结构化环境下智能装备自动驾驶与自主作业研究, E-mail: ymli@sjtu.edu.cn

## 0 引言

农机自主作业中障碍物检测是作业环境感知中不可忽视的重点,准确并及时地得到障碍物的类别和位置信息将有力促进农机自主作业中路径规划、避障的精准性与合理性。按照所使用的传感器种类可以将障碍物检测分为基于视觉、毫米波雷达、激光雷达、激光测距仪和基于多传感器融合等多种方法。宋怀波等<sup>[1]</sup>在单目相机的基础上结合最小错误率贝叶斯决策与 Hough 变换相结合的算法,实现了农田等非结构化场景中的障碍物识别。杨福增等<sup>[2]</sup>利用两台相同的相机经过立体矫正、区域匹配、三角测距等步骤实现了农田中多种障碍物的检测。孙柯等<sup>[3]</sup>利用双毫米波雷达协同工作得到了障碍物的宽度和距离信息。季宇寒等<sup>[4]</sup>利用激光雷达设计了农田环境点云采集系统,为农机导航避障提供了数据。袁文涛等<sup>[5]</sup>通过对激光测距仪的数据进行解析、坐标转换和边缘拟合,测量出了障碍物的宽度和距离等信息。宋正根等<sup>[6]</sup>将毫米波雷达数据与相机数据融合,以毫米波雷达得到的数据作为种子点,对相机采集到的图像进行分割,从而得到障碍物的详细信息。在上述障碍物检测方法中,基于视觉的检测方法具有价格低廉、信息丰富和检测范围广等优点,被广泛地应用于理论研究和实机应用探索。

非结构农田环境下目标障碍物的周围存在植被强遮挡、自然背景干扰等问题<sup>[7]</sup>,这些干扰影响视觉算法对田间障碍物的检测。在目标检测领域,基于卷积神经网络的算法因其高准确率与鲁棒性在近年来不断涌现并取得广泛应用。这些算法主要分为一阶检测算法、二阶检测算法。二阶检测算法如快速区域卷积神经网络(Faster region-convolutional neural networks, Faster R-CNN)<sup>[8]</sup>等将目标检测过程分成定位、分类两个过程。一阶检测算法如 YOLO(You only look once)、SSD、MobileNets 等<sup>[9-10]</sup>则一步完成定位与分类。毛亮等<sup>[11]</sup>将深度稠密(Depthwise dense block, DDB)模块结合 MobileNets 得到一种自然场景下的挖掘机实时检测网络模型,并能够应用到嵌入式设备上。农业生产活动中,对于农机而言田间动态障碍物主要为行人与协同作业的其他农机。行人与农机因为其障碍物属性不同,障碍物的安全距离、多并发优先级等避障的策略也有所不同。如能准确检测出行人与农机的位置与类别属性,将为农机在线路径规划与避障策略提供更为完整的信息。据此,薛金林等<sup>[12]</sup>通过构建 5 层网络对小样本数据集的农机和行人进行了分类并取得较高的分类准确率;刘慧等<sup>[13]</sup>以 MobileNets 为基础

骨架改进 SSD 网络完成果园中行人的检测任务。相对于其它一阶检测算法,YOLO 系列具有简洁高效、准确度高、泛化能力强等特点<sup>[14]</sup>。YOLOv3-tiny 是 YOLO 系列的简化版本之一,对硬件要求更低、检测速度更快,但是与 YOLOv3 相比存在检测精度下降、漏检率升高的问题<sup>[15]</sup>。刘军等<sup>[16]</sup>通过对 YOLOv3-tiny 增加浅层特征预测层并减少卷积层,在一定程度上降低了 YOLOv3-tiny 对城市交通中单类车辆障碍物检测的漏检率。

本文以农田中的非结构、非固定的行人和其它农机(相对于正在自主作业的农机)两类目标为检测任务,并基于农机感知任务中采用的嵌入式主控器具有较差算力的背景,在保证田间障碍物检测实时性的基础上提高检测准确率。本文以一阶检测模型 YOLOv3-tiny 为基础网络,改进构建适用田间行人与农机检测的模型。其中针对 YOLOv3-tiny 在局部遮挡、自然背景干扰时的漏检加剧、准确率下降的现象,融合其浅层特征与第 2 YOLO 预测层特征作为第 3 预测层得到更小网格的小目标 YOLO 检测层;采用混合注意力机制的模块增加抗背景干扰能力。本文在具体数据集上对比本文模型和其它一阶检测模型在台式机、Jetson TX2 嵌入式平台的运行准确性和实时性,并在抽取形成的小目标数据子集中验证本文模型对 YOLOv3-tiny 漏检的改善能力。

## 1 田间行人与农机障碍物的检测方法

### 1.1 YOLOv3-tiny 模型的漏检现象

对于农机算力较弱的嵌入式控制器,YOLOv3-tiny 相较 YOLOv3 具备完全的轻量效应优势,如在 Jetson TX2 或 Jetson Nano 等其他相当级别的嵌入式设备中检测速度能够达到 10 f/s<sup>[17]</sup>,满足农机作业的实时检测要求。

YOLOv3-tiny 检测速度的大幅提升导致检测精度不高。YOLOv3-tiny 是由深度为 107 层的 YOLOv3 模型裁剪而来<sup>[18]</sup>。为了追求检测速度同时考虑到网络层数比较浅,YOLOv3-tiny 完全删除了原模型的残差模块,减少了大量卷积层和多尺度特征融合层。在农田环境下整个网络在前向传播过程中,被遮挡与受背景干扰的小目标本身特征就比较少,而 YOLOv3-tiny 缺少了残差结构,较深层的特征图中对图像细节信息表达能力又比较弱,从而导致这些目标的细节特征在整个模型的深层网络中消失。图 1a 是输入的原始图像,图 1b~1d 为对应原始图像在不同深度的卷积层中的特征图。由图 1 可知,即使经数据增强训练和光照矫正预处理,基于 YOLOv3-tiny 的目标检测也产生了漏检。YOLOv3-

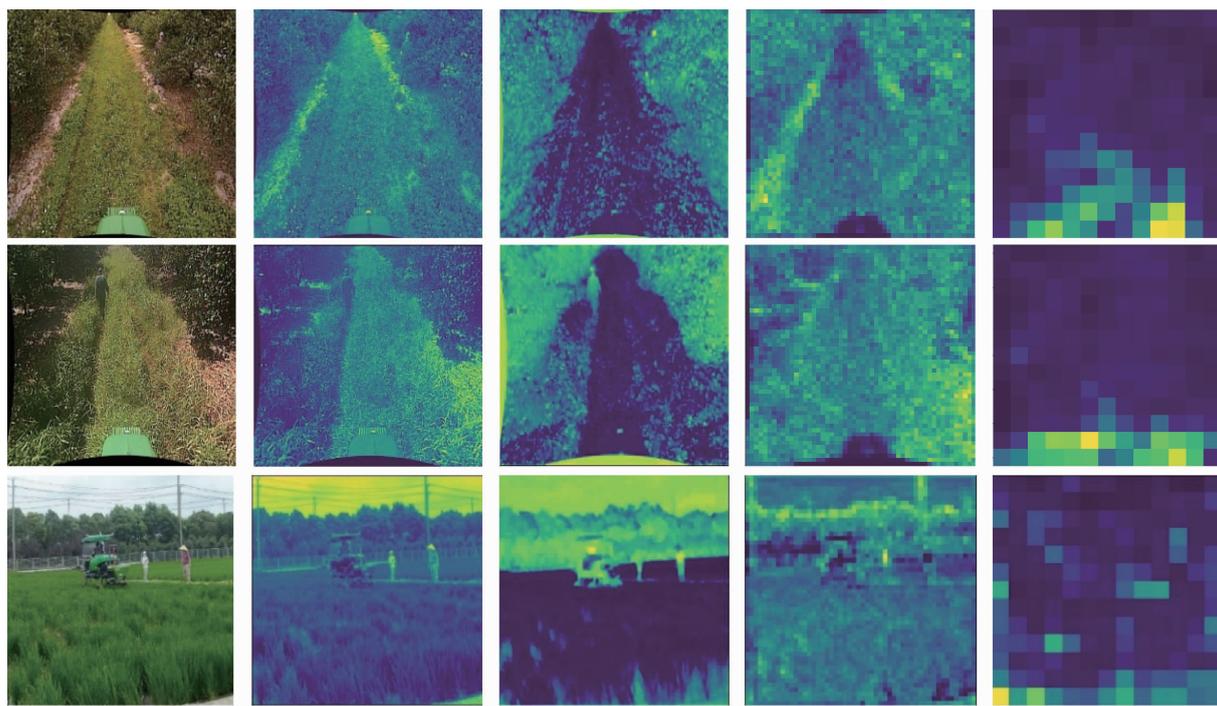


图 1 YOLOv3 - tiny 的漏检现象

Fig. 1 Undetection under YOLOv3 - tiny

tiny 在田间检测中常因为植被遮挡、自然背景干扰而导致漏检比较远的农机和行人。

1.2 融合浅层特征的改进 YOLOv3 - tiny 模型

图 2 中虚线框内除 SC 模块以外的网络结构为 YOLOv3 - tiny, 由 11 个 CBL 模块、2 个卷积层 (Conv)、6 个最大池化层 (Max pooling)、1 个上采样层组成, 最终得到 2 个 YOLO 预测层 y1 与 y2, 一共 22 层网络。网络中每层的输出特征图尺寸表示为“横向网格数 × 纵向网格数 × 通道数”。CBL 模块由卷积层 (Conv)、BN 归一化模块与 LeakReLU 激活层按顺序组合而成。激活函数 LeakReLU 可以有效减少训练过程中出现梯度消失的情况, BN 归一化模块使得激活函数 LeakReLU 的每个批次输入都正态地分布在 LeakReLU 函数的有效区域。输入图像的尺寸首先被调整为 416 像素 × 416 像素, 经过 CBL 模块与最大池化层处理得到尺寸为 13 × 13 的 y1 预

测层。y1 预测路径中第 7 个 CBL 模块引出卷积池化后并与 y1 预测路径中第 5 个 CBL 模块引出融合 (即同维拼接), 随后再卷积得到尺寸为 26 × 26 的 y2 预测层。YOLO 预测层是依据金字塔原理<sup>[19]</sup>在不同尺度下得到不同尺寸的目标特征信息图。

不同于常规 YOLOv3 应用中需用比较深的网络来拟合几十甚至上百个目标, 本文检测任务只包含行人和农机两类障碍物, YOLOv3 - tiny 中残差模块的缺位与大量卷积层的减少并不会对本文检测任务的检测准确率有显著影响。但为了克服 YOLOv3 - tiny 中浅层低维特征在网络深层丢失并因此产生的漏检, 本文构建融合浅层信息的 YOLO 预测层 y3。如图 2 所示, 将 YOLOv3 - tiny 的浅层网络与 y2 预测层之前的拼接层进行融合构建一个新的尺寸为 52 × 52 的 y3 预测层。y3 预测层的特征信息来自浅层信息与 y2 结构路线拼接前的特征图。融合了浅

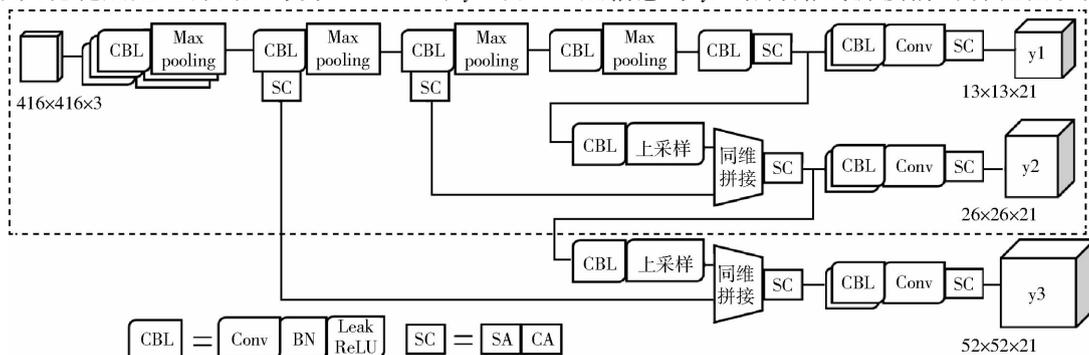


图 2 YOLOv3 - tiny 及其改进模型

Fig. 2 YOLOv3 - tiny and it's improved model

层特征图的  $y_3$  包含了更多的图像低维特征信息。

如图3所示,  $y_1$  小尺度YOLO层  $13 \times 13$  的每个网格生成的预测框比较大,  $y_2$  其次, 而远距离行人、农机, 需要更小的预测框在  $y_3$  预测层预测。在本文构建的预测层  $y_3$  中, 输入原图像被划分为  $52 \times 52$  个网格, 每个网格与  $y_3$  预测层的每个通道一一对应。每个网格根据对应通道信息生成3个尺寸预先设置好的预选框, 在训练时不断调整大小并优选出合适的预选框作为输出结果<sup>[12]</sup>。各个预测层通道包含了每个网格最终的预测参数。如图3中以  $y_1$  预测层为例, 每个通道分别由预测框中心坐标  $(t_x, t_y)$ 、预测框长宽尺寸  $(t_w, t_h)$ 、预选框的预测置信度  $p_0$ 、农机的预测得分  $s_1$  和行人的预测得分  $s_2$  组成。每个网格生成3个预选框, 每个预选框包含以上7个参数, 故  $y_1$ 、 $y_2$ 、 $y_3$  预测层预测输出通道维数均为21。

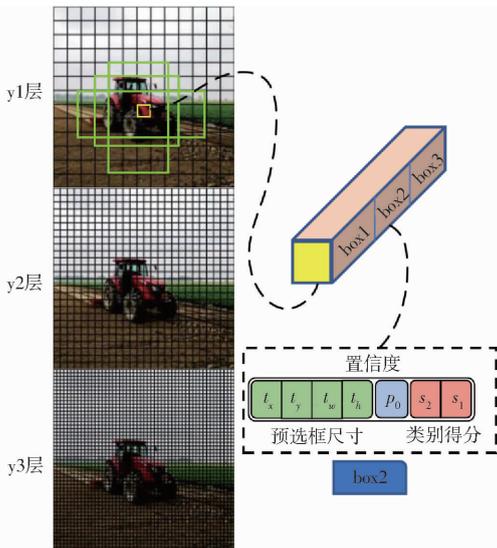


图3 预测层空间结构

Fig.3 Spatial structure of prediction layer

### 1.3 先空间后通道的混合注意力机制

注意力机制不特定指某种具体算法, 而是一种图像处理的思想。注意力机制源自人的视觉神经。在处理视野中的景象时, 人类的视觉系统会将关注点集中在目标区域然后在该区域投入更多注意力以

获取细节信息, 而不会像计算机处理图像时逐行像素扫描处理<sup>[20]</sup>。传统的局部图像特征提取、滑动窗口等方法<sup>[21-22]</sup>都可以看作注意力机制。农业视觉检测应用中, 尤其是随着深度学习的发展, 也引入了注意力机制类型的算法<sup>[23-24]</sup>。

本文混合使用通道注意力机制与空间注意力机制, 对YOLOv3-tiny卷积特征图中容易出现的农田障碍物的特征分配高权重, 对自然背景分配低权重, 从而使模型输出更加关注既定农田障碍物, 检测率得到有效提升, 解决局部遮挡、自然背景混淆的小目标农机与行人容易漏检的问题。

采用SENet的挤压激励(Squeeze and excitation, SE)模块<sup>[25]</sup>作为通道注意力机制模块。如图4a所示, 通道注意力机制分为压缩、激发及特征重标定3个阶段。压缩阶段通过全局平均将输入特征图在空间上压缩为  $1 \times 1$  的空间, 保留通道维度  $C$ 。激发阶段由2个全连接层(Fully connected, FC)、1个ReLU激活层和1个Sigmoid激活层组成。其中第1个全连接层中的神经元个数是  $C/r$  ( $r$ 为预设的降维倍数), 将其将通道维数从  $C$  维降至  $C/r$  维。第2个全连接层神经元个数为  $C$ , 将通道维数从  $C/r$  升至  $C$  维。相比直接用一个全连接层, 这样构建具有更多的非线性, 可更好地拟合通道间复杂的相关性, 同时极大地减少了参数量和计算量<sup>[25]</sup>。特征重标定阶段则是将激发阶段的结果与原输入相乘。

图4b为所采用的与通道注意力机制相似的CBAM空间注意力机制<sup>[26]</sup>, 首先对通道分别进行全局平均池化与最大池化操作, 然后将两者结果在通道维度上拼接得到  $H \times W \times 2$  的特征图, 再通过一个  $3 \times 3$  的卷积核得到  $H \times W \times 1$  的特征图, 之后通过Sigmoid激活函数层得空间注意力特征图  $H \times W \times 1$ , 将其与原输入的每个通道的特征图  $H \times W \times 1$  相乘并拼接, 得输出特征图  $H \times W \times C$ 。

经过训练的模型会在每个使用注意力机制的地方对关注的通道、空间区域分配高分权重, 对不关心的区域起到抑制作用。本文将两种注意力机制采用

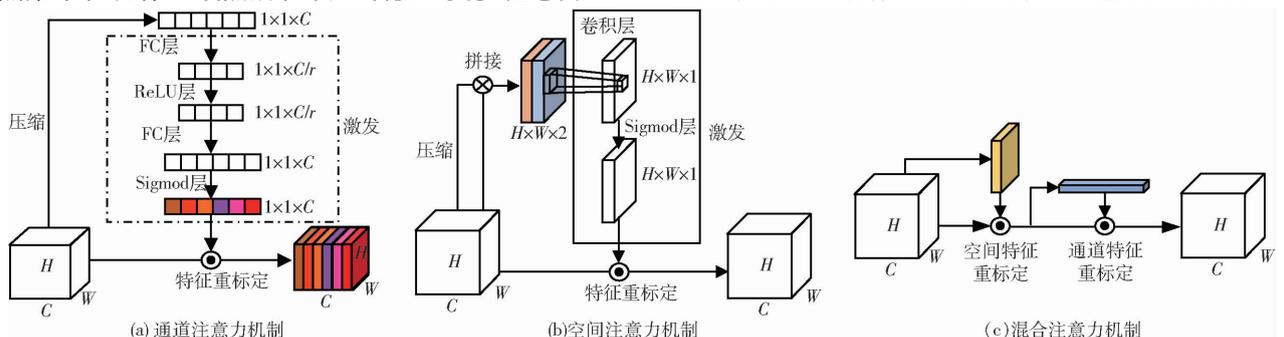


图4 注意力机制

Fig.4 Attention mechanism

如图 4c 所示的先空间后通道的组织顺序混合构建为一个混合注意力机制模块。

如图 2 所示,将混合注意力机制 SC 模块加入已经改进的 YOLOv3-tiny 模型中。SC 模块由空间注意力 (Spatial attention, SA) 模块与通道注意力 (Channel attention, CA) 模块组成。距离 YOLO 预测层近的特征图对预测影响比较高,同时考虑到如果对所有 CBL 层使用混合注意力模块将增加更多且无效的计算量。故对模型中的 3 个 YOLO 预测层前的 CBL 层、拼接层以及浅层融合层和中层融合层作为混合注意力机制的输入。最终输出的  $y_1$ 、 $y_2$ 、 $y_3$  预测层中将更多的权重分配给了关注的图像信息,每个通道中的 21 维信息将包含更准确和更多的农机与行人障碍物信息。

## 2 网络训练与检测实验

### 2.1 数据集的制作

为了保证有足够的验证模型检测能力,同时增加实际使用中模型的泛化能力。从多渠道制作和收集了 9 405 幅图像形成原始数据集,其中仅含农机图像 4 151 幅,仅含行人图像 4 838 幅,既含行人又含农机的图像 416 幅。所有图像中总共包含农机类目标 4 567 个,行人类目标 5 254 个。行人数据集大部分来自卡耐基梅隆大学开源的农区果园行人数据集,该数据集以农作物植被与田园土壤为背景,采集时间覆盖白天全时段,行人姿态多样、标注规范,符合本文的检测场景。农机图像源自本课题合作单位第一拖拉机股份有限集团、网络收集和自行采集。自行采集设备为 XW200 型相机,RGB 图像尺寸为 1 920 像素  $\times$  1 080 像素。采用 LabelImg 软件对图像中的农机与行人进行标注,标注文件格式为 txt 文件,主要包含障碍物目标框的类别、中心坐标值、高、宽 4 个参数。按 3:1 的比例将原数据集分为 7 054 幅图像的训练集与 2 351 幅图像的测试集,其中测试集包含 1 305 幅小目标图像。各数据集中农机与行人目标数的比例为 1:1。训练过程中为了增强模型抗干扰能力与泛化能力,对原数据集的每幅图像按 50% 的概率做水平翻转、旋转剪切、透视变换和模糊滤波数据的随机组合<sup>[20]</sup>,实现数据增广,最终每个周期将训练约 18 800 幅图像。

### 2.2 损失函数

损失函数是在 YOLOv3-tiny 的原损失函数<sup>[18]</sup>基础上,通过将其定位误差重新定义为 GIoU 定位误差得到的新损失函数,即

$$L = \sum_{n=1}^3 L_{y_n} \quad (1)$$

其中

$$L_{y_n} = \lambda_{GloU} \sum_{i=0}^{C_{h_n} C_{w_n} - 3} \sum_{j=0}^3 \mathbb{1}_{ij}^{obj} F_{GloU(box_{ij}, box_{0_{ij}})} + \lambda_o \sum_{i=0}^{C_{h_n} C_{w_n} - 3} \sum_{j=0}^3 \mathbb{1}_{ij}^{obj} F_o + \lambda_c \sum_{i=0}^{C_{h_n} C_{w_n} - 3} \mathbb{1}_i^{obj} \sum_{k=0}^2 F_c \quad (2)$$

式中  $L$ ——总体损失函数

$n$ ——第  $n$  个预测层,  $n = 1, 2, 3$

$L_{y_n}$ ——第  $n$  个预测层的损失

$C_{h_n}$ ——第  $n$  个预测层的长度

$C_{w_n}$ ——第  $n$  个预测层的宽度

$i$ ——当前预测层的网格序号

$j$ ——当前预测层的第  $i$  个网格的预选框序号

$\mathbb{1}_{ij}^{obj}$ ——当前预测层的第  $i$  个网格的第  $j$  个预选框,有目标时为 1,没有目标时为 0

$\mathbb{1}_i^{obj}$ ——当前预测层的第  $i$  个网格任意一个预选框,有目标时为 1,没有目标时为 0

$\lambda_{GloU}$ ——定位误差损失权重

$\lambda_o$ ——置信度损失权重

$\lambda_c$ ——分类损失权重

$F_{GloU(box_{ij}, box_{0_{ij}})}$ ——检测框  $box_{ij}$  与真实框  $box_{0_{ij}}$  的定位误差损失函数

$F_o$ ——置信度损失函数

$F_c$ ——分类损失函数

$k$ ——检测目标类别,本文中农机  $k = 0$ ,行人  $k = 1$

YOLOv3-tiny 中损失函数包含检测目标的定位损失、置信度损失以及分类损失。YOLOv3-tiny 定位误差函数采用 MSE 回归策略。MSE 因对边框尺度相当敏感而不利于模型收敛。在 YOLOv3-tiny 的一些应用中因为检测框与真实框的交叠率 (Intersection over union, IoU) 与定位误差相关性更高,转而采用 IoU 损失函数评判定位误差。但当检测框与真实框没有重合时, IoU 损失函数为零,这将导致定位误差在损失函数的梯度为零,无法优化<sup>[27]</sup>。基于以上两种原因,本文采用广义交叠率 (Generalized intersection over union, GIoU) 定位误差。GIoU 保留了 IoU 的优点且在检测框与真实框不重合的情况不会使定位误差损失为零。GIoU 一定程度上能够缓解田间目标检测时因植被遮挡与背景干扰导致的定位精度不高。

$$I_{oU} = \frac{b_{A \cap B}}{b_{A \cup B}} \quad (3)$$

$$G_{IoU} = I_{oU} - \frac{\left| \frac{b_D}{b_{A \cup B}} \right|}{|b_D|}$$

$$(0 \leq I_{oU} \leq 1; -1 \leq G_{IoU} \leq 1, G_{IoU} \neq 0) \quad (4)$$

式中  $b_{A \cap B}$ ——网络预测的预选框与真实框的重叠面积

$b_{A \cup B}$ ——真实框与预选框的总包围面积

$b_D$ ——真实框与预选框的最小包络矩形框面积

$G_{IoU}$ ——广义交叠率

$I_{oU}$ ——交叠率

### 2.3 实验平台

实验平台为台式计算机,32 GB 内存、11 GB 显存、GTX 2080ti 显卡、Intel i7 - 9700F 芯片,在 Ubuntu 18.04 环境的 Pytorch 3.7 框架下搭建了所有实验涉及的模型框架。

实验的测试平台有 2 套。其中台式机测试平台与训练平台为同一台,在台式机训练好模型后,OpenCV 环境中调用检测障碍物。另一个测试平台为 Jetson TX2 嵌入式测试平台,主控制器 Jetson TX2 搭载 NVIDIA Pascal,算力达到每秒浮点运算 1 万亿次 (Tera floating-point operations per second, TFLOPS),256 个 CUDA 核心,内存 8 GB,存储空间 32 GB。在其 Ubuntu 18.04 系统中安装 CUDA 8.0 并行计算框架、cudnn 5.1 深度神经网络加速库以及 OpenCV 开发框架。在训练好模型后,将模型导入嵌入式测试平台,在 OpenCV 环境中调用检测障碍物。

### 2.4 模型的训练

模型训练开始时,通过 HE 等<sup>[28]</sup>的方法取得良好的模型初始值,以减少训练时间成本并加快模型收敛。在训练过程中,为了使模型更加贴合农机与行人的形态,对前文中提到的 YOLO 网格预选框进行高、宽 K-means 方法聚类<sup>[29]</sup>,得到 9 个预选框尺寸,其中(12,19)、(12,42)、(17,60)对应 y1 预测层,(33,48)、(25,89)、(39,132)对应 y2 预测层,(73,182)、(220,265)、(324,347)对应 y3 预测层,使得预选框的高宽比以及实际尺寸都更加接近数据集中农机与行人的尺寸。损失函数中的定位损失也将更容易下降收敛。设置 batch-size(每批次输入网络的图像数)为 32,同时采用多尺度训练,每次输入网络的图像顺序依次为 384 像素 × 384 像素、416 像素 × 416 像素、448 像素 × 448 像素、480 像素 × 480 像素、512 像素 × 512 像素。设置训练周期上限为 450 ~ 500,损失下降至 1 左右结束训练。

## 3 结果与分析

### 3.1 评价标准

使用平均准确率 mAP 与召回率  $R$  作为模型检

测精度的评判标准。使用模型内存、平均检测帧耗时作为检测速度的评价标准。 $R$  能够准确表达测试集中漏检的数量, $R$  越大,漏检越少。mAP 是综合评价漏检和误检的百分系数,mAP 越大,说明检测漏检越少,误检也越少<sup>[30]</sup>。模型内存大小表达了检测时的参数计算量。平均检测帧耗时则直接表现了模型在某设备上的检测速度。

### 3.2 结果分析

针对测试数据集分别使用一阶检测领域比较常用的 YOLOv3、SSD、YOLOv3 - tiny 模型与本文模型进行对比,结果如表 1 所示。采用 YOLOv3 - tiny 与本文模型对数据集中相对像素面积小于 32 像素 × 32 像素的小目标行人与农机共 1 303 幅图像形成的小目标数据子集进行比较,比较结果如表 2 所示。由表 1 可知,相比于 YOLOv3 与 SSD 模型,本文模型平均准确率分别下降了 5.6 个百分点和 3.6 个百分点;本文模型的内存约为 YOLOv3 的 1/3、SSD 的 2/3;本文模型在 Jetson TX2 平台检测的平均帧耗时为 YOLOv3 和 SSD 的 1/3 和 1/5。

表 1 整个数据集中的不同模型检测能力比较

Tab.1 Detection capabilities of different models under whole data set

模型	模型内存/MB	平均准确率/%	台式计算机检测平均帧耗时/ms	Jetson TX2 检测平均帧耗时/ms
YOLOv3	248.1	92.1	182	418
YOLOv3 - tiny	34.8	75.5	50	83
SSD	105.2	90.1	201	514
本文模型	69.9	86.5	93	122

表 2 小目标数据子集中不同模型检测能力比较

Tab.2 Detection capabilities of different models under small target data subset

模型	平均准确率/%	召回率/%
YOLOv3 - tiny	57.1	52.8
本文模型	66.5	67.1

由表 1 可知,相比 YOLOv3 - tiny 模型,本文模型在整个数据集中平均准确率提高了 11 个百分点。在 Jetson TX2 检测平台上本文模型的平均图像检测耗时为 122 ms,是 YOLOv3 - tiny 的 1.5 倍,但能够满足农机在 1.5 ~ 2.0 m/s 行进速度实时的检测需求。虽然本文模型内存是 YOLOv3 - tiny 的 2 倍,但当检测工作从台式计算机移植到嵌入式 Jetson TX2 设备时,本文模型在 Jetson TX2 检测的平均每帧耗时增大至在台式计算机检测耗时的 1.3 倍,而 YOLOv3 - tiny 增大至 1.7 倍,在移植过程中本模型增加的检测耗时是 4 个模型中最小的。故本文模型

在满足田间作业需求的前提下实现了速度与精度的平衡,并适用于嵌入式平台。

由表 2 可知,相比 YOLOv3-tiny 模型,本文模型在小目标数据集中平均准确率上升 9.4 个百分点,召回率上升 14.3 个百分点。本文模型在增设 y3 预测层和混合注意力机制后,内存只比原模型增加 35.1 MB,却缓解了植被遮挡、背景干扰等带来的小目标漏检问题。图 5 是采用本文模型与 YOLOv3-tiny 的对比检测实例,实验图像尺寸为 1 980 像素 ×

1 080 像素,来源于上海松江与河南许昌实验基地实地拍摄以及卡耐基梅隆大学田间行人公开数据集。红色框内为本文模型检测得出的行人或农机,蓝色框为 YOLOv3-tiny 模型得到的行人和农机。其中图 5a、5c、5d、5i 中因背景干扰和植被遮挡而导致小目标行人与农机漏检的情况得到了改善。图 5g、5h、5f 中农机定位的准确性也得到了提升。对于行人与农机同时出现的场景如图 5b、5g、5h,通过本文模型漏检现象也得到了有效改善。

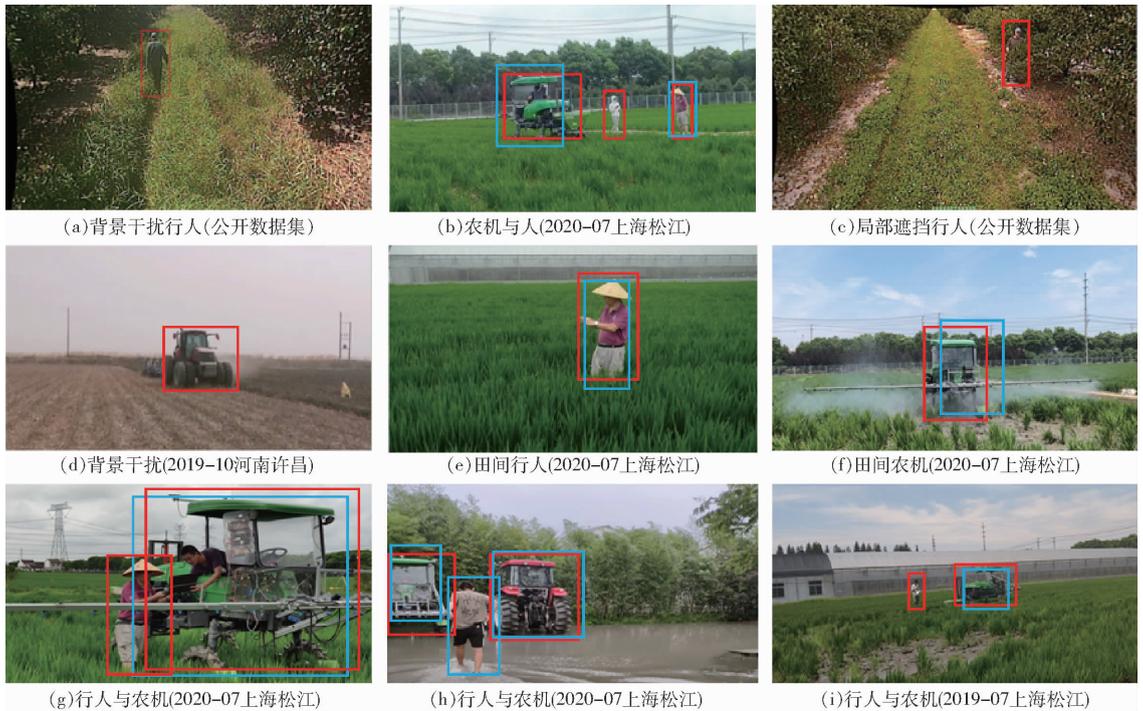


图 5 分别采用本文模型和 YOLOv3-tiny 的检测实例

Fig. 5 Detection examples using proposed model and YOLOv3-tiny

## 4 结论

(1) 本文模型的内存约为 YOLOv3 模型的 1/3、SSD 模型的 2/3、YOLOv3-tiny 的 2 倍。

(2) 本文模型和 YOLOv3-tiny 在 Jetson TX2 嵌入式平台上检测的平均帧耗时远小于 YOLOv3 和 SSD。本文模型的平均每帧检测耗时 122 ms 与 YOLOv3-tiny 的 83 ms 均能够满足农机作业正常行进速度的实时检测需求。当检测工作从台式机移植到嵌入式平台时,本文模型在 Jetson TX2 检测

的平均每帧耗时增大至台式计算机的 1.3 倍,小于 YOLOv3-tiny 的 1.7 倍,同时检测耗时波动也是 4 个模型中最小的,故本文模型具有较高的鲁棒性。

(3) 本文模型在整个数据集的平均准确率达到 86.5%,分别比 YOLOv3、SSD 低 5.6、3.6 个百分点,但比 YOLOv3-tiny 高 11 个百分点。在小目标数据子集中本文模型平均准确率达到 66.5%,比 YOLOv3-tiny 高 9.4 个百分点,召回率高 14.3 个百分点。

## 参 考 文 献

- [1] 宋怀波,何东健,辛湘俊. 基于机器视觉的非结构化道路检测与障碍物识别方法[J]. 农业工程学报, 2011, 27(6): 225-230. SONG Huaibo, HE Dongjian, XIN Xiangjun. Unstructured road detection and obstacle recognition algorithm based on machine vision[J]. Transactions of the CSAE, 2011, 27(6): 225-230. (in Chinese)
- [2] 杨福增,刘珊,陈丽萍,等. 基于立体视觉技术的多种农田障碍物检测方法[J/OL]. 农业机械学报, 2012, 43(5): 168-172, 202. YANG Fuzeng, LIU Shan, CHEN Liping, et al. Detection method of various obstacles in farmland based on stereovision technology[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(5): 168-172, 202. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20120529&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20120529&journal_id=jcsam). DOI: 10.6041/j.

issn. 1000-1298. 2012. 05. 029. (in Chinese)

- [3] 孙柯, 吴开华, 王亚涛, 等. 基于毫米波雷达的植保无人机避障系统研究[J]. 传感器与微系统, 2020, 39(6): 73-76.  
SUN Ke, WU Kaihua, WANG Yatao, et al. Research on obstacle avoidance system of plant protection UAV based on millimeter wave radar[J]. Transducer and Microsystem Technologies, 2020, 39(6): 73-76. (in Chinese)
- [4] 季宇寒, 徐弘祯, 张漫, 等. 基于激光雷达的农田环境点云采集系统设计[J/OL]. 农业机械学报, 2019, 50(增刊): 1-7.  
JI Yuhan, XU Hongzhen, ZHANG Man, et al. Design of point cloud acquisition system for farmland environment based on LiDAR[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 1-7. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=2019s001&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=2019s001&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2019.S0.001. (in Chinese)
- [5] 袁文涛, 刘卉, 胡书鹏. 面向自动驾驶拖拉机的农田障碍物识别研究[J]. 农机化研究, 2018, 40(10): 247-251.  
YUAN Wentao, LIU Hui, HU Shupeng. Farmland obstacle identification for automatic navigation tractor[J]. Journal of Agricultural Mechanization Research, 2018, 40(10): 247-251. (in Chinese)
- [6] 宋正根, 彭竟德, 肖臻. 基于毫米波雷达和摄像头的农用车辆障碍物检测[J]. 现代信息科技, 2019, 3(14): 46-48.  
SONG Zhenggen, PENG Jingde, XIAO Can. Obstacle detection of agricultural vehicles based on millimeter wave radar and camera[J]. Modern Information Technology, 2019, 3(14): 46-48. (in Chinese)
- [7] 蔡道清, 李彦明, 覃程锦, 等. 水田田埂边界支持向量机检测方法[J/OL]. 农业机械学报, 2019, 50(6): 22-27, 109.  
CAI Daoqing, LI Yanming, QIN Chengjin, et al. Detection method of boundary of paddy fields using support vector machine[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(6): 22-27, 109. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20190603&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190603&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2019.06.003. (in Chinese)
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with rregion proposal networks[C]// Advances in Neural Information Processing Systems, 2015: 91-99.
- [9] LIU W, ANQUELOY D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [10] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv, 2017: 1704.04861.
- [11] 毛亮, 薛月菊, 朱婷婷, 等. 自然场景下的挖掘机实时监测方法[J]. 农业工程学报, 2020, 36(9): 214-220.  
MAO Liang, XUE Yueju, ZHU Tingting, et al. Method for the real-time monitoring of the excavator in natural scene[J]. Transactions of the CSAE, 2020, 36(9): 214-220. (in Chinese)
- [12] 薛金林, 闫嘉, 范博文. 多类农田障碍物卷积神经网络分类识别方法[J/OL]. 农业机械学报, 2018, 49(增刊): 35-41.  
XUE Jinlin, YAN Jia, FAN Bowen. Classification and identification method of multiple kinds of farm obstacles based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(Supp.): 35-41. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=2018s006&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=2018s006&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2018.S0.006. (in Chinese)
- [13] 刘慧, 张礼帅, 沈跃, 等. 基于改进 SSD 的果园行人实时检测方法[J/OL]. 农业机械学报, 2019, 50(4): 29-35, 101.  
LIU Hui, ZHANG Lishuai, SHEN Yue, et al. Real-time pedestrian detection in orchard based on improved SSD[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 29-35, 101. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20190403&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190403&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2019.04.003. (in Chinese)
- [14] 刘芳, 刘玉坤, 林森, 等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J/OL]. 农业机械学报, 2020, 51(6): 229-237.  
LIU Fang, LIU Yukun, LIN Sen, et al. Fast recognition method for tomatoes under complex environments based on improved YOLO[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 229-237. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20200624&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200624&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2020.06.024. (in Chinese)
- [15] 马立, 巩笑天, 欧阳航空. Tiny YOLOv3 目标检测改进[J]. 光学精密工程, 2020, 28(4): 988-995.  
MA Li, GONG Xiaotian, OUYANG Hangkong. Improvement of Tiny YOLOv3 target detection[J]. Optics and Precision Engineering, 2020, 28(4): 988-995. (in Chinese)
- [16] 刘军, 后士浩, 张凯, 等. 基于增强 Tiny YOLOv3 算法的车辆实时检测与跟踪[J]. 农业工程学报, 2019, 35(8): 118-125.  
LIU Jun, HOU Shihao, ZHANG Kai, et al. Real-time vehicle detection and tacking based on enhanced Tiny YOLOv3 algorithm[J]. Transactions of the CSAE, 2019, 35(8): 118-125. (in Chinese)
- [17] HOSSAIN S, LEE D. Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices[J]. Sensors, 2019, 19(15): 3371.
- [18] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv, 2018: 1804.02767.
- [19] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [20] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 60.
- [21] RUBLEE E, RABAUD V, BRADSK G. ORB: an efficient alternative to SIFT or SURF[J]. Proceedings of the IEEE International Conference on Computer Vision, 2011: 2564-2571.

- winter wheat yield and quality[J]. Transactions of the CSAE, 2019, 35(18): 78–85. (in Chinese)
- [25] WEST D W, HOFFMAN G J, FISHER M J. Photosynthesis, leaf conductance, and water relations of cowpea under saline conditions[J]. Irrigation Science, 1986, 7(3): 183–193.
- [26] PANG H C, LI Y Y, YANG J S, et al. Effect of brackish water irrigation and straw mulching on soil salinity and crop yields under monsoonal climatic conditions[J]. Agricultural Water Management, 2010, 97(12): 1971–1977.
- [27] QU L Y, QURORESHI A M, KOIKE T. Root growth characteristics, biomass and nutrient dynamics of seedlings of two larch species raised under different fertilization regimes[J]. Plant and Soil, 2003(255): 293–302.
- [28] ASHRAF M. Salt tolerance of cotton: some new advances[J]. Critical Reviews in Plant Sciences, 2002, 21(1): 1–30.
- [29] 唐胜强, 余冬立. 灌溉水质对土壤饱和导水率和入渗特性的影响[J/OL]. 农业机械学报, 2016, 47(10): 108–114. TANG Shengqiang, SHE Dongli. Influence of water quality on soil saturated hydraulic conductivity and infiltration properties [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(10): 108–114. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20161015&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20161015&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2016.10.015. (in Chinese)
- [30] 刘梅先, 杨劲松, 李晓明, 等. 滴灌模式对棉花根系分布和水分利用效率的影响[J]. 农业工程学报, 2012, 28(增刊1): 98–105. LIU Meixian, YANG Jinsong, LI Xiaoming, et al. Effects of drip irrigation strategy on cotton root distribution and water use efficiency[J]. Transactions of the CSAE, 2012, 28(Supp.1): 98–105. (in Chinese)
- [31] MI G H, CHEN F J, WU Q P, et al. Ideotype root architecture for efficient nitrogen acquisition by maize in intensive cropping systems[J]. Scientia China Life Science, 2010, 53: 1369–1373.
- [32] 卫新东, 汪星, 汪有科, 等. 黄土丘陵区红枣经济林根系分布与土壤水分关系研究[J/OL]. 农业机械学报, 2015, 46(4): 88–97. WEI Xindong, WANG Xing, WANG Youke, et al. Root distribution and water dynamics of jujube plantations in loess hilly regions[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(4): 88–97. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20150414&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20150414&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2015.04.014. (in Chinese)
- [33] NARAYAN D. Root growth and productivity of wheat cultivars under different soil moisture condition[J]. International Journal of Ecology and Environmental Sciences, 1991, 17(1): 19–26.
- [34] BOWEN G D. Tree roots and the use of soil nutrients[M]//BOWEN G D, NAMBIAR E K S. Nutrition of plantation forests. London: Academic Press, 1984: 147–177.
- [35] 杨志谦, 王维敏. 秸秆还田后碳、氮在土壤中的积累与释放[J]. 土壤肥料, 1991(5): 43–46. YANG Zhiqian, WANG Weimin. Accumulation and release of carbon and nitrogen in soil after straw returning[J]. Soil and Fertilizer Sciences, 1991(5): 43–46. (in Chinese)

(上接第 8 页)

- [22] BORII A, CHENG Mingming, QIBIN H, et al. Salient object detection: a survey[J]. arXiv preprint arXiv, 2019: 1411.5878v6.
- [23] ZENG Weihui, LI Miao. Crop leaf disease recognition based on self-attention convolutional neural network[J]. Computers and Electronics in Agriculture. 2020, 172: 105341.
- [24] XU Lele, LI Ye, XU Jinzhong, et al. Two-level attention and score consistency network for plant segmentation[J]. Computers and Electronics in Agriculture, 2020, 170: 105281.
- [25] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.
- [26] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision, 2018: 3–19.
- [27] REZATOFIGHI R, TSOI N, SAVARESE S, et al. Generalized intersection over union: a metric and a loss for bounding box regression[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 658–666.
- [28] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026–1034.
- [29] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263–7271.
- [30] ZHANG P, SU W. Statistical inference on recall, precision and average precision under random selection [C] // 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, 2012: 1348–1352.