

基于改进 YOLOv4 的棉花检测算法^①



刘正波, 鲍义东, 孟庆伟

(贵州航天智慧农业有限公司, 贵阳 550081)

通讯作者: 刘正波, E-mail: 1290260@qq.com

摘要: 为提高自动化采棉机械的采摘效率和智能化水平, 避免误采摘、漏采摘, 采用以复杂背景下实现单个棉花检测为目标, 提出一种改进的 YOLOv4 目标检测算法. 使用 K-means 算法进行聚类锚框尺寸的筛选, 得到适合棉花数据集的精细化锚框尺寸. 同时在 YOLOv4 算法中引入注意力机制, 在其网络结构中添加 SENet (Squeeze-and-Excitation Networks) 模块. 在模型训练时, 首先在公开数据集上训练取得预训练权重, 在预训练模型上使用棉花数据集微调参数, 并使用数据增强方式扩充原始数据集, 在预训练模型上再次训练. 实验结果表明, 本文提出的 YOLOv4 改进算法, 能够很好的实现田间环境下的棉花检测.

关键词: 棉花采摘; 检测; YOLOv4; 聚类; SENet; 模型; 数据增强

引用格式: 刘正波, 鲍义东, 孟庆伟. 基于改进 YOLOv4 的棉花检测算法. 计算机系统应用, 2021, 30(8): 164-170. <http://www.c-s-a.org.cn/1003-3254/8155.html>

Cotton Detection Algorithm Based on Improved YOLOv4

LIU Zheng-Bo, BAO Yi-Dong, MENG Qing-Wei

(Guizhou Aerospace Intelligent Agriculture Co. Ltd., Guiyang 550081, China)

Abstract: To improve the efficiency and intelligence of automatic cotton-picking machines and avoid false and missed picking, we propose an improved YOLOv4 target detection algorithm to detect single cotton in complex backgrounds. The K-means algorithm is used to screen the size of the clustering anchor frame and obtain the refined size suitable for the cotton data set. The attention mechanism is also introduced to the YOLOv4 algorithm, and the Squeeze-and-Excitation Networks (SENet) module is located in the network structure. During model training, the weights of pre-training are obtained by training on an open data set, and fine-tuning parameters of the cotton data set are applied to the pre-training model. Furthermore, the original data set is expanded through data enhancement and the pre-training model has been trained again. Experimental results show that the improved YOLOv4 algorithm proposed in this study can effectively realize cotton detection in the field environment.

Key words: cotton picking; detection; YOLOv4; clustering; SENet; model; data enhancement

我国是世界第一大产棉国, 同时也是世界第一大棉花进口国, 这也造就了我国成为世界纺织大国的地位^[1]. 棉花作为种植作物中产业链最长的大田经济作物, 集大宗农产品和纺织工业原料于一身, 在国民经济中占有重要地位. 棉花产业属于劳动密集型产业, 传统

的棉花种植采摘技术落后, 需要大量的人力物力, 随着劳动力成本的增长, 棉花采摘机械化成为必然趋势. 我国采棉机的发展起步较晚, 虽然在不断进步, 但是对棉花采摘的采净率和除杂率上还有待提高. 传统采棉机相比人工采摘虽然在采摘效率上有较大提升, 但采摘

① 收稿时间: 2020-11-18; 修改时间: 2020-12-21, 2021-02-02, 2021-02-23; 采用时间: 2021-03-03; csa 在线出版时间: 2021-07-31

的同时降低了棉花的品质,并且采棉机在采摘结束后仍有较多棉花在棉杆上,导致单位面积的棉花减产^[2]。因此,智能采棉机成为了降低劳动成本、提高棉花产量和棉花质量的必要途径。智能采棉机是根据棉花自然成熟周期进行多批次精准采摘,采摘关键在于对单个棉花实现精准识别与定位。传统机器视觉在采用图像预处理、阈值分割、特征提取等方法上对作物识别和检测上取得了较好效果,但是对于多个目标重叠、障碍物遮掩等情况还是会出现漏采和误采。

近年来,计算机视觉技术发展迅速,视觉技术广泛应用于农业领域,解决农作物或杂草检测、作物疾病检测等难题。在棉花检测领域,目前研究集中对采棉机器人配备的摄像机采集的棉花图片进行识别与分割,少有研究考虑到田间实际采摘中复杂环境。Wang等^[3]提出一种基于RGB颜色空间分布的棉花检测算法用于采棉机器人,并采用Freeman链码提高识别准确率。韦皆顶等^[4]提出基于HSV彩色模型的棉花饱和度图像分割算法,能够很好地将棉花和背景分割开来,并且受光照影响较小,但无法处理田间棉花叶片遮挡等情况。刘坤等^[5]提出一种基于RGB彩色空间模型和Hough变换的棉桃识别技术,能准确识别自然环境中被遮挡的棉花。陈钦政等^[6]结合最大类间方差法提出了一种在OHTA颜色空间下基于支持向量机的棉花图像分割算法,提高了分割的准确度和实用性,但对复杂环境采棉适用度不高。尽管文献[3-7]已提出一系列高精度棉花检测和分割算法,但是这些方法都采用手工提取特征方式进行识别,算法鲁棒性不强,并且没有综合考虑田间棉花采摘复杂背景、叶片遮挡和棉花重叠等情况,难以用于田间棉花采摘。因此本文提出一种基于改进YOLOv4的棉花检测算法,应用于采棉机器人进行田间棉花采摘。

1 算法理论基础

1.1 YOLOv4 算法

YOLOv4^[7]算法在YOLOv3^[8]算法基础上,从数据处理、训练方法、激活函数、损失函数、网络结构等方面做出不同程度优化,使算法模型在目标检测速度和精度上达到当前最优平衡。

YOLOv4在YOLOv3损失函数基础上使用CIoU(Complete Intersection over Union)损失替代MSE(Mean Square Error)损失作为回归框预测损失。损失函

数可分为回归框预测损失 L_{CIoU} 、置信度损失 L_{Conf} 、分类损失 L_{cls} 三部分。

$$LOSS = L_{CIoU} + L_{Conf} + L_{cls} \quad (1)$$

回归框预测损失 L_{CIoU} 表示为:

$$L_{CIoU} = 1 - IoU - \frac{\rho^2(b, b^{gt})}{c^2} + av \quad (2)$$

其中, IoU 为预测框与真实框的交并比, b 为预测框中心点, b^{gt} 为真实框中心点, ρ 为两个框中心点间的欧式距离, c 是同时包含预测框和真实框的最小封闭区域的对角线距离, av 是对长宽比的惩罚项。 v 表示度量长宽比的相似性系数, 定义为:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

其中, w^{gt} 为真实框宽度, h^{gt} 为真实框高度, w 为预测框宽度, h 为预测框高度。 a 为长宽比的惩罚项系数, 具体定义为:

$$a = \frac{v}{(1 - IoU) + v} \quad (4)$$

置信度损失 L_{Conf} 为:

$$L_{Conf} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[\bar{C}_i^j \log(C_i^j) + (1 - \bar{C}_i^j) \log(1 - \bar{C}_i^j) \right] + \lambda_{nobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{nobj} \left[\bar{C}_i^j \log(C_i^j) + ((1 - \bar{C}_i^j) \log(1 - \bar{C}_i^j)) \right] \quad (5)$$

式(5)中, S^2 表示网格划分网格数, B 是每个网格中先验框个数, I_{ij}^{obj} 表示预测边界框包含目标, I_{ij}^{nobj} 表示预测的边界框不包含目标, C_i^j 为预测置信度, \bar{C}_i^j 为实际置信度, λ_{nobj} 为预先设定的参数值。分类损失 L_{cls} 表示为:

$$L_{cls} = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{C \in classes} \left(\left[\bar{P}_{i(c)}^j \log P_{i(c)}^j \right] + (1 - \bar{P}_{i(c)}^j) \log(1 - P_{i(c)}^j) \right) \quad (6)$$

式中, c 表示检测目标所属分类, $P_{i(c)}^j$ 和 $\bar{P}_{i(c)}^j$ 分别表示单网格目标属于类别 c 的实际概率与预测概率。

YOLOv4目标检测算法的网络结构如图1所示,主要包括特征提取网络CSPDarkNet53、空间金字塔池化网络SPP和特征聚合网络PANet三部分。CSPDarkNet53

作为 YOLOv4 算法的主干特征提取网络,是在 DarkNet53 网络的基础上改进而来. CSPDarkNet53 包含 5 个 CSPNet 模块, 首先将输入特征划分为两部分, 对输入特征进行卷积和批量归一化处理, 使用 Mish 激活函数激活, 最后跨阶段融合两部分特征, 在复杂度增加较小的情况下提高了模型整体的稳定性和准确性, 输出更加平滑并且很好的解决了梯度饱和的问题. CSPDarkNet53

将 DarkNet53 的 5 个残差模块替换为 CSPNet 模块, CSPNet 模块中主干提取部分和残差边分离的结构能够更好的提取图像特征, 在保持轻量化的同时保持较高准确率、降低计算量. 主干特征提取网络 CSPDarkNet53 充分提取图像特征后, 将后 3 个 CSPNet 模块提取的不同尺度图像特征传入 SPP 网络和 PANet 网络进行特征融合.

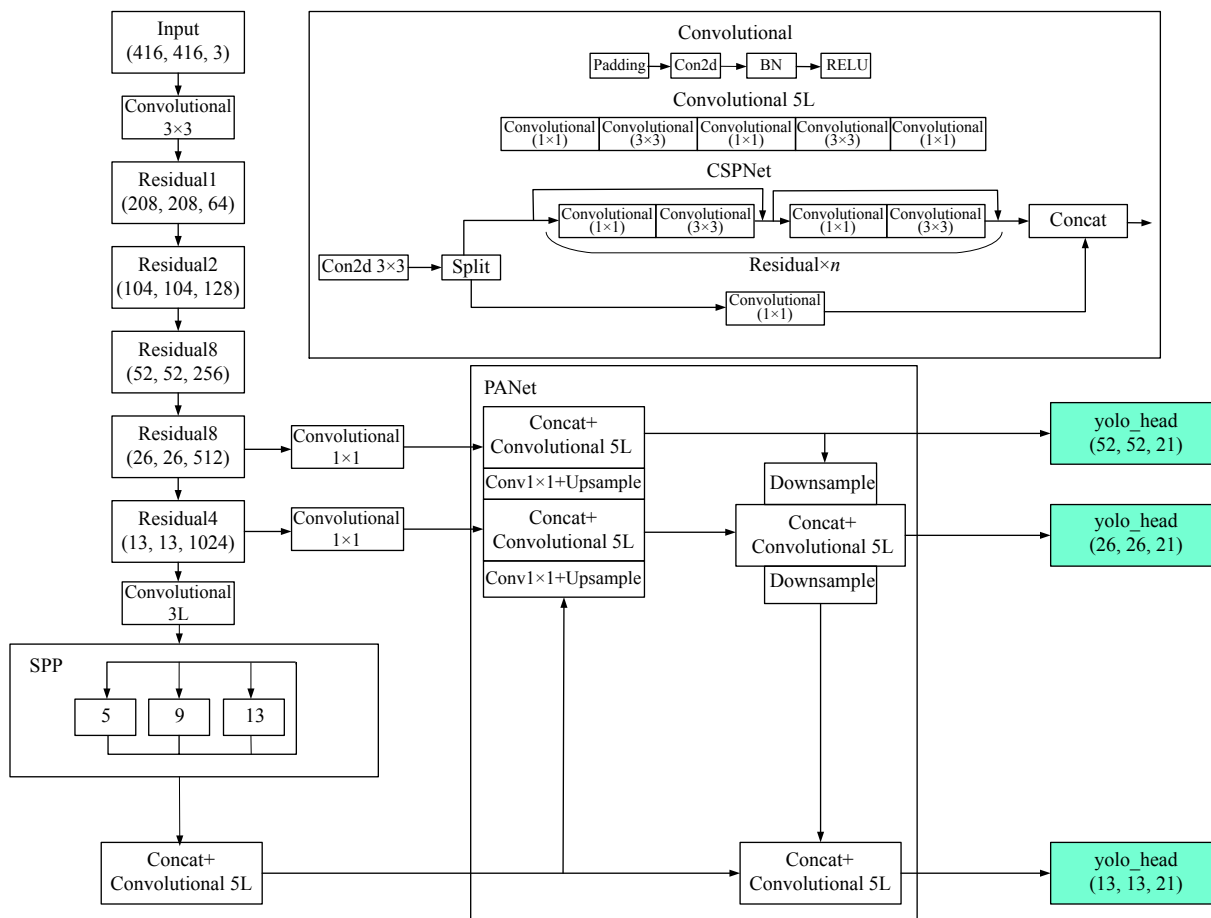


图1 YOLOv4 算法网络结构

对 CSPDarkNet53 网络最后一个特征层进行一次特征卷积后, 特征传入 SSP 网络, 利用 SSP 网络中 4 种不同尺度最大池化核充分融合图像特征, 极大增加感受野, 分离得到显著的上下文特征来. PANet 结构是由卷积操作、上采样操作、特征层融合、下采样操作构成的循环金字塔结构. 对传入的图像特征进行反复的特征提取, 自底而上融合从主干网络不同特征层提取的特征, 形成强定位特征, 进一步提高特征提取和定位能力. 从 PANet 提取的特征输出为 3 个不同尺度

YOLOHead 预测层, 预测层信息包括目标得分、预测锚框尺寸坐标、分类置信度等. 利用 YOLOHead 预测层的信息对锚框的位置、尺寸做出适当调整, 利用目标得分筛除掉部分锚框, 最后通过非极大值抑制将 3 个预测层的结果合并, 得到最终预测结果.

1.2 锚框计算

YOLOv4 目标检测算法的默认锚框是在 Coco 数据集上使用聚类算法计算得出的通用尺寸, Coco 数据集中包含各种不同的检测目标, 目标的尺寸比例差别较

大,因此锚框的尺寸也不尽相同,在实际应用场景中需要根据数据集,计算相应的锚框尺寸.由于棉花图片数据库网络资源有限,本文采用自建的棉花数据集,棉花在不同图像中的形状基本一致,因此可以针对棉花的实际形状大小计算相应锚框尺寸.

采用 K-means 聚类算法计算锚框尺寸,具体计算步骤如图 2 所示.首先读取数据集中归一化的锚框尺寸大小,锚框的宽、高组成坐标点作为待分类数据,并随机初始化 9 个聚类中心.计算各个锚框坐标点与聚类中心锚框之间的交并比 (Intersection over Union, IoU),以 $(1-IoU)$ 作为坐标之间的相对距离,距离最小的聚类中心为对应锚框的分类,把各个类别的中心坐标作为下一次迭代的聚类中心.不断重复上述过程,直至聚类中心不再变化,把此时的 9 个聚类中心作为最终的得到的锚框尺寸.使用 K-means 算法计算得到的锚框尺寸更适合当前数据集,实现单株棉花的精准检测.

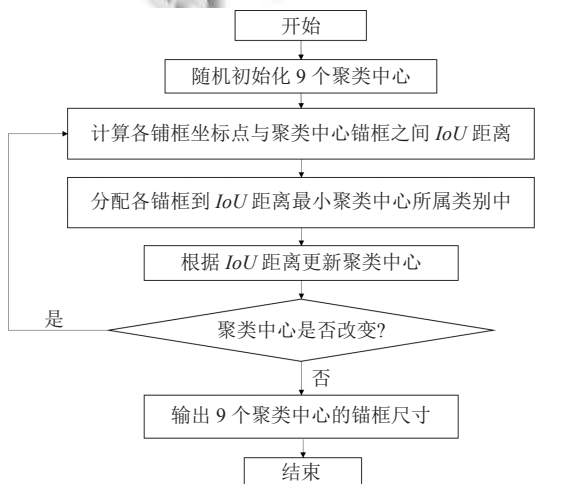


图 2 K-means 算法计算锚框尺寸

1.3 数据增强

自然环境下棉花检测任务面向复杂田间环境,检测精度不仅受到叶片遮挡、多簇棉花重叠等因素的影响,还需要考虑天气、棉花尺寸形态、自然光照等自然因素.因此除了要在数据采集的过程中考虑以上因素,尽可能多的采集各种不同场景和自然条件下的棉花图片作为数据集,也要通过数据增强方法对数据集图片做出处理,提高模型的鲁棒性.

数据增强分为离线增强和在线增强两种方式.离线数据增强主要用于数据量较小的数据集,直接处理原始数据集图片,新数据图片数量扩增为原数据集图

片的 n 倍 (n 为数据增强因子),主要用来防止过拟合.在线增强是在读取一批次数据集图片后,对图片数据进行增强操作,提高模型在不同场景中泛化能力.主要处理流程是先对数据集图片作离线增强,对图片进行翻转、裁剪、旋转、缩放、平移和添加噪声等操作,得到的数据集大小为原数据集的 3 倍.在训练过程中,YOLOv4 自带的 Mosaic 数据增强操作对片进行处理. Mosaic 数据增强的效果如图 3 所示,首先读入 4 张图片,对图片作翻转、缩放、色域变换等操作,然后以矩阵的形式截取图片的固定区域并将其拼接为一张图片,极大的丰富检测物体的背景,提升检测算法的鲁棒性.



图 3 Mosaic 数据增强

1.4 网络模型改进

在田间实际场景中对棉花进行检测时,由于复杂背景的影响,以及叶片和枝丫对棉花的遮挡,常规的目标检测算法受到置信度的影响,往往会漏检部分棉花从而造成损失.为了让检测网络学习到细致的局部信息,并融合全局信息做出更加准确的判断,避免误检测和漏检测,可采取增大全局感受野的方式,捕获更多的图像特征描述图片.

对 YOLOv4 的特征提取网络做出改进,改进后网络结构如图 4 所示,将 SENet (Squeeze-and-Excitation Networks) 结构嵌入到 YOLOv4 主干特征提取网络 CSPDarkNet53 的 Residual 模块中,得到 SE-Residual 残差块.用 SE-Residual 替换原有的 Residual 模块,加大特征提取网络的全局感受野,增强特征提取能力,从而能有效减少环境背景和障碍物遮挡对棉花检测的影响,避免误检测和漏检测. SENet 是 Hu 等^[9]在 2017 年提出,网络结构如图 4 所示,主要包括 Squeeze、Excitation 和 Scale 三部分. SENet 中的 Squeeze 部分用全局池化层顺着空间维度进行特征压缩,在各个通道上得到具有全局感受野的实数. Excitation 模块接着使用两个全连接层结构对各个通道的重要性进行预测,得到不同通道重要性.最后 Scale 部分将提取得到

的通道重要性作用到全局特征上. SENet 在减少参数的同时,对各个通道的特征进行加权操作,各个通道的

信息交互,强化有效信息,抑制无效信息,提高网络准确率.

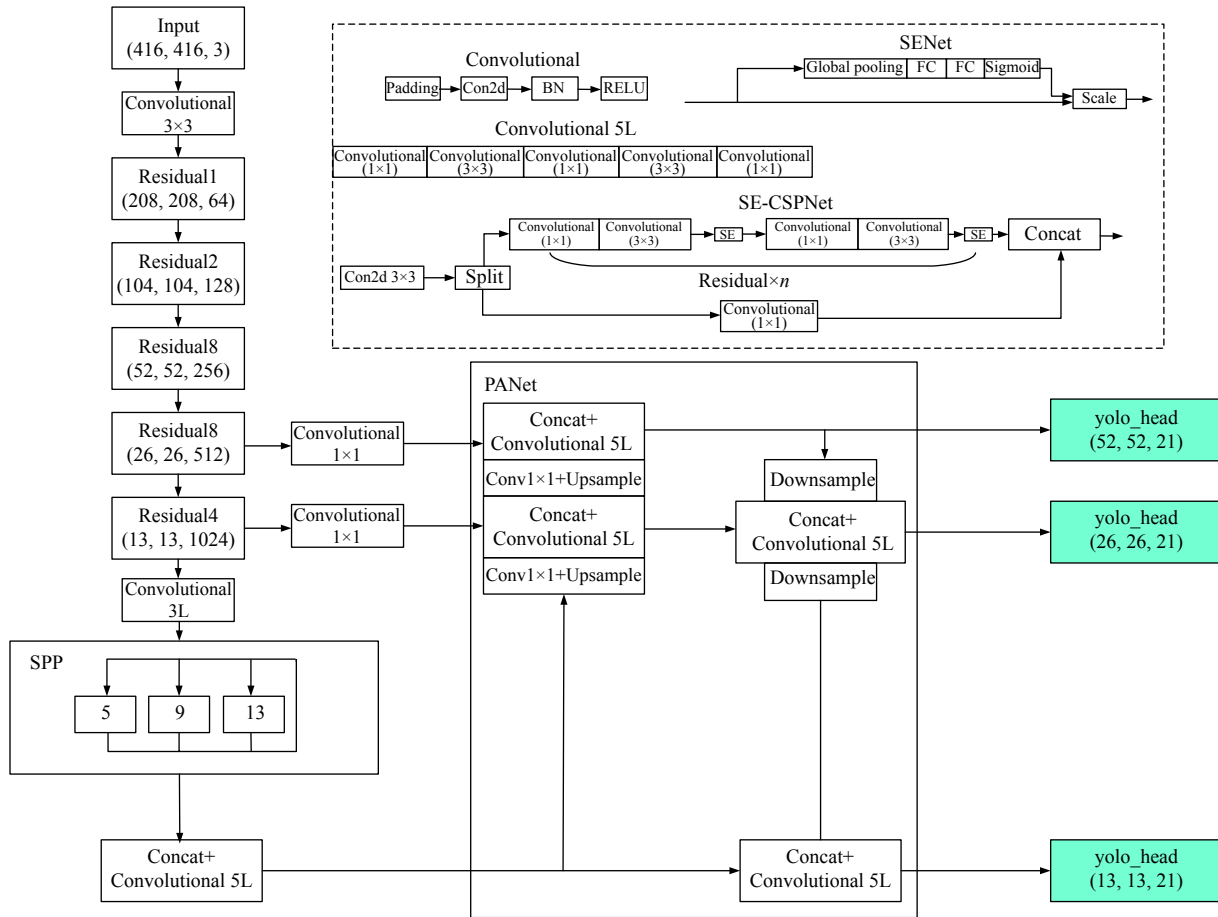


图4 YOLOv4改进算法网络结构

2 实验结果与分析

2.1 实验数据集

实验数据集主要在棉田内实际拍摄取得,为充分反映棉田内真实的棉花采摘情况,在不同天气情况下从不同的角度对棉花植株的棉花进行拍摄.并通过互联网获取部分棉花图片对数据集进行扩充,经过筛选后选取 6000 张相关图片作为初始数据集,然后使用离线数据增强方法对数据集图片翻转、缩放、平移和加噪声操作,进一步扩充数据集,最终得到的实验数据集包含 18000 张相关图片.如图 5 所示,数据集图片主要包括单个棉花、多个棉花重叠、遮掩棉花和非棉花 4 种情况.

数据集获取后需要对其进行标注,所使用标注软件为 labelImg,该软件免费且操作简单,只需对图片中

单个棉花的区域使用锚框标注为 Cotton. 选用标注格式为 PASAL VOC2007, 然后利用随机函数将数据集随机分成训练集和测试集,其中训练集和测试集图片数量比例为 5:1,在训练集中再细分为训练集和验证集比例为 9:1.

2.2 实验分析

试验在高性能计算平台上进行,平台操作系统为 Windows10,实验开发平台的配置为:CPU 选用 Intel(R) Core (TM) i7-8700, GPU 选用 NVIDIA GeForce RTX 2070,内存为 16 GB. 加载软件环境有 Pycharm、Python3.7、Cuda10.2. 模型通过 PyTorch 深度学习框架搭建.

在深度学习模型的训练上,采用迁移学习的方式通过预训练模型进行训练. 迁移学习的方式减少了构

建深度学习模型所需的训练数据和计算力,降低了模型的训练时间,能很好解决小数据集容易在复杂网络结构上的过拟合问题.首先对公开数据集 Coco 数据集来获得预训练模型,然后用棉花数据集基于预训练模型进行迁移学习.迁移学习先冻结特征提取网络前端训练 50 个 Epoch,利用已经训练好具有较强特征提取能力的预训练模型先进行特征提取,然后根据训练集/测试集对特征融合网络和分类检测网络的参数进行微调,最后对整个网络的参数进行训练和调整并输出最后的分类检测结果.



图5 棉花数据集

采用查准率 P 和查全率 R 作为评价检测算法性能的指标,查准率 P 和查全率 R 的计算公式为:

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

$$R = \frac{TP}{TP + FN} \quad (8)$$

其中,真正例 TP 表示在棉花存在的网格单元中,同时满足置信度大于 0.8、预测框与真实框的 IoU 大于 0.5、预测类别正确 3 个条件.假正例 FP 中包含两种情况,一种是在棉花存在的网格单元中,置信度大于 0.8 但不是真正例,另一种是在棉花不存在的网格单元中,置信度大于 0.8.假反例 FN 代表在棉花存在的网格单元中,置信度小于 0.8.

在网络的训练上对每种算法模型重复 3 次,每次训练 650 个 Epoch,具体训练参数设置为:批处理数为 16,梯度优化选用 Adam 算法,初始学习率为 0.001,每 50 次迭代后学习率衰减为原来的 1/10,动量为 0.9,衰减系数为 0.0005,分类置信度阈值为 0.8, IoU 阈值为 0.5.

改进算法的训练的损失值、查准率和查全率曲线如图 6 所示,在使用迁移学习方法后,损失值在 200 次迭代后趋近 0,查准率和查全率曲线在迭代 300 次后逐渐趋于平稳,并且稳定后的查准率和查全率都在 90% 以上,结果如图 6 所示.

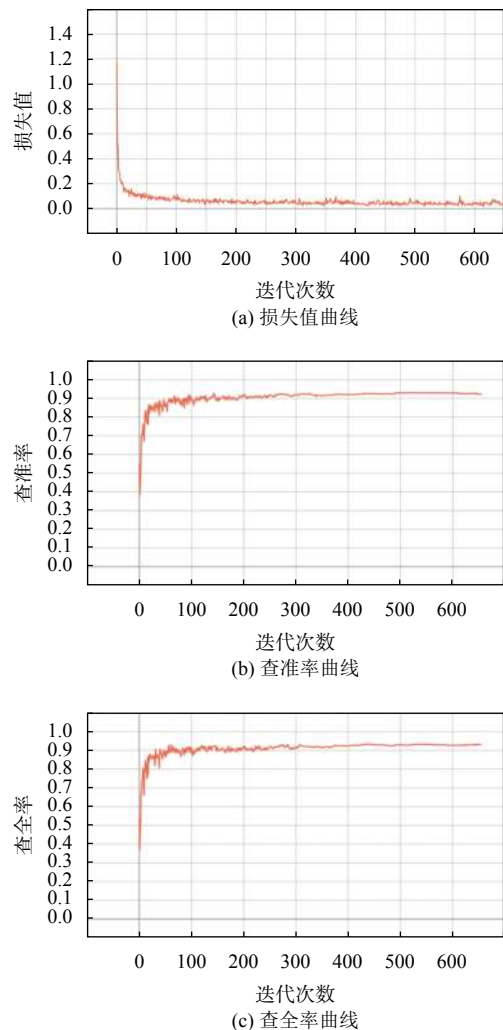


图6 YOLOv4 改进算法

本文提出的 YOLOv4 改进算法与 YOLOv3 算法、YOLOv4 算法及其改进结构的性能对比如表 1 所示,对比结果表明对 YOLOv4 目标检测算法在网络结构和锚框上的改进,相较于 YOLOv3 算法和原 YOLOv4 算法在性能上有较明显的提升.本文在 YOLOv4 算法中加入的新锚框计算方法,在查准率上的提升较小,在查全率上有 3% 左右的性能提升.在网络结构中添加 SENet 模块的方法,在查准率上有明显的性能提升,在查全率上也有较明显的性能提升.综合结果表明,本文

提出的 YOLOv4 改进算法, 在复杂环境下也有很好的性能表现, 相较于传统方法能够更好的适应棉花采摘的实际田间环境.

表 1 算法检测性能对比 (%)

算法	新锚框	SENet	查准率	查全率
YOLOv3	×	×	72.15	87.62
YOLOv4	×	×	86.87	88.65
YOLOv4(1)	√	×	87.55	90.26
SE-YOLOv4	×	√	90.39	89.97
本文算法	√	√	93.84	92.26

3 结论

采用 K-means 算法进行锚框计算, 通过在 YOLOv4 中嵌入 SENet 结构得到 YOLOv4 改进算法, 融合 YOLOv4 与 SENet 的优势, 增加网络的感受野, 提升网络的特征提取能力. 实验结果表明, 在复杂田间环境下, 本文算法仍能够准确识别目标, 减少误采摘和漏采摘, 得到较高的查准率.

参考文献

- 1 卢秀茹, 贾肖月, 牛佳慧. 中国棉花产业发展现状及展望. 中国农业科学, 2018, 51(1): 26–36. [doi: 10.3864/j.issn.0578-1752.2018.01.003]
- 2 赵春江. 智慧农业发展现状及战略目标研究. 农业工程技术, 2019, 39(6): 14–17. [doi: 10.3969/j.issn.1673-5404.2019.06.003]
- 3 Wang Y, Zhu XR, Ji CY. Machine vision based cotton recognition for cotton harvesting robot. First IFIP TC 12 International Conference on Computer and Computing Technologies in Agriculture. Wuyishan, China. 2008. 1421–1425.
- 4 韦皆顶, 费树岷, 汪木兰, 等. 基于 HSV 彩色模型的自然场景下棉花图像分割策略研究. 棉花学报, 2008, 20(1): 34–38. [doi: 10.3969/j.issn.1002-7807.2008.01.007]
- 5 刘坤, 费树岷, 汪木兰, 等. 基于改进随机 Hough 变换的蜜桃识别技术. 农业机械学报, 2010, 41(8): 160–165. [doi: 10.3969/j.issn.1000-1298.2010.08.033]
- 6 陈钦政, 赖惠成, 王星, 等. 一种基于支持向量机的棉花图像分割算法. 计算机工程, 2013, 39(5): 266–269. [doi: 10.3969/j.issn.1000-3428.2013.05.058]
- 7 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv: 2004.10934, 2020.
- 8 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv preprint arXiv: 1804.02767, 2018.
- 9 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7132–7141.