

DOI: 10.13733/j.jcam.issn.2095-5553.2024.06.035

周康乔, 刘向阳, 郑特驹. 改进 YOLOv5 的小目标多类别农田害虫检测算法研究[J]. 中国农机化学报, 2024, 45(6): 235-241

Zhou Kangqiao, Liu Xiangyang, Zheng Teju. Research on improved YOLOv5 small target multi-class farmland pest detection algorithm [J]. Journal of Chinese Agricultural Mechanization, 2024, 45(6): 235-241

# 改进 YOLOv5 的小目标多类别农田害虫检测算法研究\*

周康乔, 刘向阳, 郑特驹

(河海大学理学院, 南京市, 211100)

**摘要:** 针对农田害虫图像中感兴趣目标特征不明显、小目标居多导致的目标检测精度较低的问题, 提出一种基于 YOLOv5 改进的小目标多类别农田害虫目标检测算法。首先, 在主干网络最后两个 C3 卷积块特征融合部分引入 Swin Transformer 窗口注意力网络结构, 增强小目标的语义信息和全局感知能力; 其次, 在颈部网络的 C3 卷积块后添加通道注意力机制和空间注意力机制的可学习自适应权重, 使网络能够关注到图像中关于小目标的特征信息; 最后, 由于 YOLOv5 自身的交并比函数存在收敛速度较慢且精确率较低的问题, 引入 SIOU 函数作为新的边界框回归损失函数, 提高检测的收敛速度和精确度。将所提出的算法在包含 28 类农田害虫公开数据集上进行试验, 结果表明, 改进后的算法在农田害虫图像数据集上的准确率、召回率和平均准确率分别达到 85.9%、76.4%、79.4%, 相比于 YOLOv5 分别提升 2.5%、11.3%、4.7%。

**关键词:** 农田害虫检测; 小目标; YOLOv5; 注意力机制; 损失函数

中图分类号: S763.3; TP391 文献标识码: A 文章编号: 2095-5553 (2024) 06-0235-07

## Research on improved YOLOv5 small target multi-class farmland pest detection algorithm

Zhou Kangqiao, Liu Xiangyang, Zheng Teju

(College of Science, Hohai University, Nanjing, 211100, China)

**Abstract:** Aiming at the problem of low target detection accuracy caused by the lack of obvious features of the interested target and the majority of small targets in the farmland pest images, a small target multi-category farmland pest target detection algorithm based on YOLOv5 was proposed. Firstly, the Swin Transformer window attention network structure was introduced into the feature fusion part of the last two C3 convolution blocks of the trunk network to enhance the semantic information and global awareness of small targets. Secondly, the learnable adaptive weights of the channel attention mechanism and the spatial attention mechanism were added to the C3 convolution block of the neck network, so that the network could pay attention to the feature information of small targets in the image. Finally, since the intersection ratio function of YOLOv5 itself had the problem of slow convergence speed and low accuracy rate, SIOU function was introduced as a new boundary box regression loss function to improve the convergence speed and accuracy of detection. The proposed algorithm was tested on the open data set of 28 farmland pests. The results showed that the accuracy rate, recall rate and average accuracy of the improved algorithm in the farmland pest image data set reached 85.9%, 76.4% and 79.4%, respectively, which were 2.5%, 11.3% and 4.7% higher than that of YOLOv5.

**Keywords:** farmland pest detection; small goal; YOLOv5; attention mechanism; loss function

## 0 引言

中国是世界上最大的农业生产国之一, 也是农业有害生物问题的高发区。农业害虫每年都会对农作物造

成很大的危害<sup>[1]</sup>。如果没有专业知识, 人们很难识别害虫, 而错误地使用农药往往会对受虫害影响的地区造成二次损害<sup>[2]</sup>。近年来, 计算机技术发展迅速, 农业害虫识别作为研究的热点, 其中包括一些基于计算机视觉的

收稿日期: 2022 年 12 月 10 日 修回日期: 2023 年 1 月 13 日

\* 基金项目: 云南省重大科技专项计划项目资助(202002AE090010)

第一作者: 周康乔, 男, 1998 年生, 江苏徐州人, 硕士研究生; 研究方向为深度学习与目标检测。E-mail: 1207581563@qq.com

通讯作者: 刘向阳, 男, 1976 年生, 山东青岛人, 副教授, 硕导; 研究方向为智能计算、机器学习。E-mail: liuxy@hhu.edu.cn

方法。Larios 等<sup>[3]</sup>开发了一种采用级联特征直方图方法对石蝇幼虫进行分类的系统。Zhu 等<sup>[4]</sup>通过分析翅膀图像的颜色直方图和灰度共生矩阵,提出一种昆虫分类方法。在 100 种鳞翅目昆虫数据库中对该方法进行测试,识别率高达 71.1%。Wang 等<sup>[5]</sup>设计了昆虫自动识别系统。使用人工神经网络和支持向量机作为模式识别方法对昆虫进行分类。利用人工神经网络对具有不同特征的八阶和九阶昆虫图像进行测试,系统稳定性良好,准确率为 93%。Faithpraise 等<sup>[6]</sup>提出了一种基于 k-means 聚类 and 对应滤波器相结合的害虫检测系统。Xia 等<sup>[7]</sup>使用分水岭算法将害虫图像从背景中分离出来,然后通过马氏距离提取害虫的颜色特征,对移动和嵌入式设备捕获的害虫图像进行分类。以常见温室害虫粉虱、蚜虫和蓟马为对象,在低分辨率图像下,粉虱、蓟马、蚜虫与人工鉴定的相关性较高,分别为 Q 934、Q 925 和 Q 945。Wang 等<sup>[8]</sup>提出一个局部 Chan-Vese 模型来完成图像分割任务。Xie 等<sup>[9]</sup>开发了具有先进多任务稀疏表示和多核学习的昆虫识别系统。对 24 种常见作物害虫的试验结果表明,该方法在昆虫种类分类方面表现良好。

然而,上述害虫识别方法在很大程度上依赖于人工选择的害虫特征,这些特征对模型性能有很大影响。深度学习技术在图像识别工作中已经取得了良好的效果,利用深度卷积神经网络可以自动学习不同害虫的特征,不需要人工选择的害虫特征。目前计算机视觉方面的目标检测算法可以分为 2 类:一类是一阶段检测算法(one-stage),直接在一个阶段里完成寻找目标出现的位置和目标类别的预测;另一类是二阶段检测算法(two-stage)<sup>[10, 11]</sup>,二阶段算法通常在第一阶段专注于感兴趣区域的生成,得到建议框,然后在第二阶段专注于对建议框进行类别的分类与位置的回归,寻找检测目标更确切的位置。二阶段检测以 Girshick 等<sup>[12]</sup>提出的 R-CNN 为代表,包括 Fast R-CNN<sup>[13]</sup>、Faster R-CNN<sup>[14]</sup>、MASK R-CNN<sup>[15]</sup>和 R-FCN<sup>[16]</sup>。一阶段检测以 Redmon 等<sup>[17]</sup>提出的 YOLO 为代表,除此之外还有 Swin Transformer<sup>[18]</sup>和 SSD<sup>[19]</sup>。一阶段与二阶段的主要不同之处在于没有候选框生成阶段。

由于农田害虫图像中害虫尺寸通常较小,携带的信息量较少,且图像背景信息较为复杂,一般的深度学习目标检测算法直接应用于农田害虫图像会出现目标检测精度较低的问题。针对上述出现的问题,本文提出一种基于注意力机制和新的边界框损失函数的 YOLOv5 的农田害虫目标检测算法。首先,本文为了避免主干网络中随着网络结构的加深,小目标的语义信息在深层特征图中丢失问题,在主干网络的 C3 卷积块引入 Swin Transformer 网络结构;其次,为了提高 Neck

网络对目标的特征提取能力,在颈部的 C3 卷积块后引入 CBAM<sup>[20]</sup>模板;最后,引入 SIOU 函数作为新的边界框回归损失函数,提高模型的收敛速度和精确率。

## 1 材料与方法

### 1.1 材料

本文采用 2022 年中国高校大数据教育创新联盟和泰迪杯数据挖掘委员会提供的公开数据集作为试验的检测基准。该数据集中图像尺寸较大,数据集包含对农作物生长有危害的 28 种农田害虫目标,害虫的尺寸较小且分布不均衡,某些类别害虫目标数量少且分辨率较为模糊,不易检测,如图 1 所示。对数据集进行数据清洗和增强<sup>[21, 22]</sup>操作后获得 2 200 张标注准确的图像,可以用于研究。其中最多的 156 类(八点灰灯蛾)共有 288 次标记,占总标记的 24.26%;而最少的 430 类(豆野螟)和 673 类(干纹冬夜蛾)均小于 10 次标记;各类害虫标记数量并不平衡,且每张图片的类别数不相同,此时模型对于数目少的类别的学习效果有可能不如数目多的类别的学习效果。因此人工划分训练集和验证集,以确保验证集中每种害虫的图片至少存在一张,进一步验证算法对每一类别害虫的预测精度。具体划分方法:若包含某种害虫的图片少于或等于 10 张,则随机选择 1 张图片划分入验证集,其余图片划分入训练集<sup>[23]</sup>;若包含某种害虫的图片多于 10 张,则以 1:9 的比例划分入验证集和训练集。

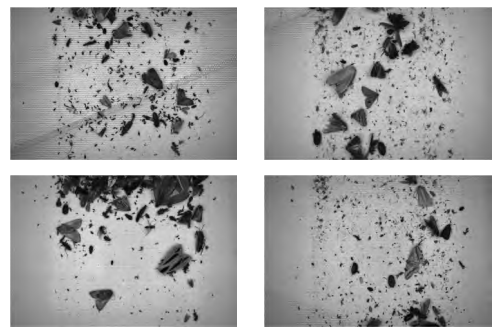


图 1 试验数据集害虫图像

Fig. 1 Experimental data set images of pests

按照上述方法对 2 200 张图片进行划分,最后训练集中包含 1 760 张图片,测试集中包含 440 张图片。

### 1.2 试验环境及参数设置

本试验使用的操作系统为 Ubuntu 18.04 LTS, GPU 为 2 块 NVIDIA A10 24G, CUDA 版本为 11.6,深度学习框架为 pytorch 1.10。预训练权重采用 Yolov5s.pt,输入图片大小为 640 像素×640 像素, batch-size 为 16。采用随机梯度下降算法训练 300epoch<sup>[24]</sup>。其中初始学习率为 0.01;采用余弦退火衰减策略调整学习率,余弦函数动态降低学习率为

0.1;权重衰减度为 0.937;预热训练时轮数的学习率为 0.0005;预热训练时的梯度动量为 3;预热训练时偏置 b 的学习率为 0.8。Loss 函数的定位框损失函数的增益比为 3;分类损失的增益比为 0.5;正样本的权重为 1;目标距离的权重为 1;负样本的权重为 1。

### 1.3 YOLOv5 原理

YOLOv5<sup>[25]</sup>网络模型主要分为如下 4 个模块:输入端、主干网络(backbone)、颈部网络(neck)和预测端(prediction)。其中 YOLOv5 在输入端使用了 Mosaic 方法,随机将训练集中的任意 4 张图像拼接在一起生成一张新的图像,增强图像的输入,同时在输入端自适应计算锚框的尺寸,自适应缩放图像缩的尺寸。在主干网络中引入一种新的 SPPF 结构对处理后的特征图在通道方向进行拼接,同时引入一种新的 C3 网络结构,有效减少了信息丢失。在 Neck 部分使用 FPN+PAN 结构,首先 FPN 通过将上采样后的特征图和 backbone 中的低层特征图进行 concat,增强网络学习图像特征的能力。然后 PAN 对 FPN 获得的图像通过自下向上传递强的定位信息,同时使用两者达到互补效果,增强模型的特征提取能力。

YOLOv5 网络进行通用目标检测时,虽然其精度较好,但是在小目标农田害虫检测中仍存在不理想的地方:一是由于主干网络对于小目标的特征提取能力较弱;二是由于 Neck 网络在连续多次采样过程中会出现目标信息丢失严重的情况,特别是在融合不同层级特征时,浅层特征会受到深层特征的干扰,而浅层特征通常包含小目标丰富的位置信息,这样会造成定位失准。对此,本文提出了更适用于小目标农田害虫检测的改进 YOLOv5 网络结构。

### 1.4 主干网络的改进

在主干网络中随着网络结构的加深,经过多次卷积操作,农田害虫图像中小目标应该具有的大部分目标特征信息在深层特征图中可能会丢失。所以,在主干网络最后两个 C3 卷积块特征融合部分借鉴 Swin Transformer 的思想,将其替换掉原有的 Bottleneck 模块。替换后的 C3STR 结构作为一个辅助模块,借助窗口自注意力模块增强小目标的语义信息和特征表示,改进后的卷积结构图 2 所示。

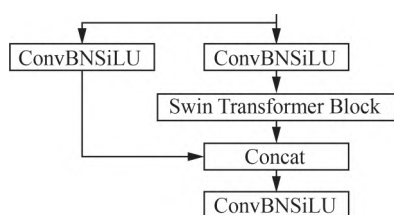


图 2 C3STR 结构

Fig. 2 C3STR structure

Swin Transformer Block 将 Transformer Block 中多头自注意力模块(MSA)改进成窗口多头自注意力模块(W-MSA)和滑动窗口多头自注意力模块(SW-MSA),两个模块成对出现,再经过多层感知机 MLP,其中每个模块内部均采用残差连接。具体结构如图 3 所示。

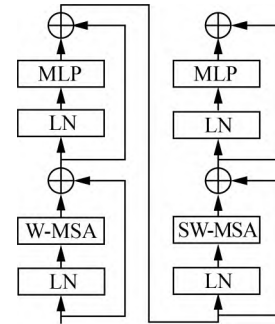


图 3 Swin Transformer Block 结构

Fig. 3 Swin Transformer Block structure

其中局部窗口大小为 7,多层感知机隐藏层的嵌入维度为 4。多头自注意力机制<sup>[26]</sup>的计算过程引入相对位置编码,具体公式如式(1)所示。

$$Attention(Q, K, V) = \text{softmax}(QK^T / \sqrt{d} + B)V \quad (1)$$

式中:  $Q, K, V$ ——对应 Query、Key 和 Value 矩阵;

$d$ ——输入特征图的 channels;

$B$ ——相对位置偏差,由网络学习得到。

与传统的 Transformer<sup>[27]</sup>中的多头自注意力模块相比,W-MSA 和 SW-MSA 模块和通过划分局部窗口控制每一个窗口中计算区域的方法降低计算复杂度和网络计算量,同时利用 Shifted Windows 实现跨窗口的信息交互。

### 1.5 Neck 网络的改进

为提高 Neck 网络对目标的特征提取能力,更好地抑制连续多次采样过程中出现的目标信息丢失严重的情况,得到具有方向感知和位置感知信息的特征图,对融合后的特征图进行更新,在 Neck 的 C3 卷积块后引入卷积注意力机制模板(CBAM),CBMA 模块通过卷积核池化操作计算出特征图在空间维度和通道维度的不同权重,使网络能够更加关注到图像中关于检测目标的特征信息,如图 4 所示。假设特征图有  $C$  个通道,通道注意力机制对特征图进行全局最大池化和全局平均池化两种池化操作,分别得到  $C$  个  $1 \times 1$  大小的特征图,再将得到的两个特征图分别进行 MLP 操作后加在一起,最后经过 Sigmoid 函数获得特征图在每一个通道的权重。通道注意力具体如式(2)所示。

$$M_c(F) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (2)$$

式中:  $\sigma$ ——使用 Sigmoid 函数近激活;

- $F_{\max}^c$  —— 对特征图进行全局最大池化；
- $F_{\text{avg}}^c$  —— 对特征图进行全局平均池化；
- $W_0, W_1$  —— 对应第 1 个和第 2 个全连接层的权重,为了降低计算参数,通道注意力模块在 MLP 的第一个全连接层中采用了一个降维系数  $r$ ；
- $M_c(F)$  —— 通道注意力得到每一个通道权重。

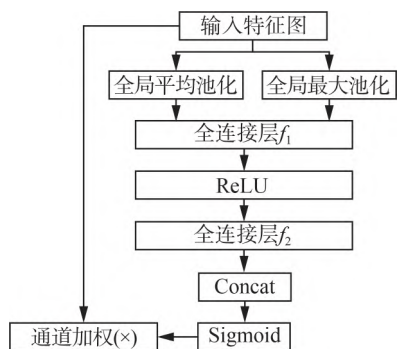


图 4 通道注意力机制模块

Fig. 4 Channel attention mechanism module

如图 5 和式(3)所示,假设特征图在每个通道上的尺寸为  $H \times W$ ,空间注意力机制模块对学习到的特征图在通道方向上进行平均和最大池化操作,分别得到 1 个  $H \times W$  大小的特征图,再将两个特征图在通道方向上进行拼接,得到一个  $2 \times H \times W$  大特征图,最后经过卷积核为  $7 \times 7$  的卷积操作和 Sigmoid 激活函数获得学习到的特征图在每一空间位置上的权重。

$$M_s(F) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\max}^s])) \quad (3)$$

式中:  $f^{7 \times 7}$  —— 空间注意力模块的 MLP 操作采用的是  $7 \times 7$  卷积操作。

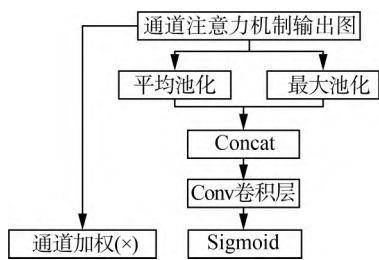


图 5 空间注意力机制模块

Fig. 5 Spatial attention mechanism module

### 1.6 损失函数的改进

YOLOv5 自身的交并比函数为 CIoU<sup>[28]</sup> 边界框回归损失函数,只考虑到 IoU loss、中心点损失和长宽比例损失,没有考虑到真实框与预测框之间的方向,导致收敛速度较慢,对此本文考虑到期望的真实框和预测框之间的向量夹角,引入 SIoU 函数重新定义相关损失函数,SIoU<sup>[29]</sup> 函数具体包含四个部分。

角度损失(Angle cost),定义如式(4)和图 6 所示。

$$\Lambda = 1 - 2 \times \sin^2 \left[ \arcsin \left( \frac{c_k}{\xi} \right) - \frac{\pi}{4} \right] \quad (4)$$

式中:  $c_k$  —— 真实框和预测框中心点的高度差；  
 $\xi$  —— 真实框和预测中心点的距离。

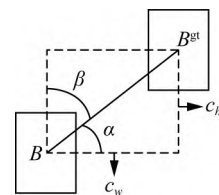


图 6 角度损失计算

Fig. 6 Calculation of angle loss

距离损失(Distance cost),定义如式(5)所示。其中  $\rho_x = \left( \frac{b_{c_x}^{\text{gt}} - b_{c_x}}{c_w} \right)^2, \rho_y = \left( \frac{b_{c_y}^{\text{gt}} - b_{c_y}}{c_h} \right)^2, \gamma = 2 - \Lambda, (b_{c_x}^{\text{gt}}, b_{c_y}^{\text{gt}})$  和  $(b_{c_x}, b_{c_y})$  分别表示真实框和预测框中心点坐标,  $(c_w, c_h)$  为真实框和预测框最小外接矩形的宽和高。

$$\Delta = \sum_{t=x,y} (1 - e^{-\rho^t}) = 2 - e^{-\rho_x} - e^{-\rho_y} \quad (5)$$

形状损失(Shape cost),定义如式(6)所示。其中  $W_w = \frac{|\omega - \omega^{\text{gt}}|}{\max(\omega, \omega^{\text{gt}})}, W_h = \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})}$ ,  $(\omega, h)$  和  $(\omega^{\text{gt}}, h^{\text{gt}})$  分别为预测框和真实框的宽和高,  $\theta$  控制对形状损失的关注程度,一般取 4。

$$\Omega = \sum_{t=\omega,h} (1 - e^{-w^t})^\theta = (1 - e^{-w_\omega})^\theta + (1 - e^{-w_h})^\theta \quad (6)$$

IoU 损失(IoU cost),定义如式(7)所示。

$$IoU = \frac{\text{交集 } A}{\text{并集 } B} \quad (7)$$

最终 SIoU 损失函数定义如式(8)所示。改进后的 YOLOv5 网络结构如图 7 所示。

$$Loss_{\text{SIoU}} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (8)$$

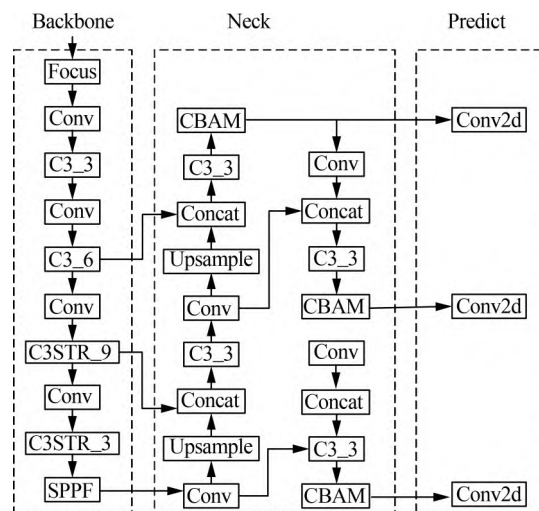


图 7 改进后 YOLOv5 网络结构

Fig. 7 Network structure of the improved YOLOv5

### 1.7 评价指标

本文对算法的性能评估,以准确率(Precision,  $P$ )、

召回率(Recall,  $R$ )和平均准确率( $mAP$ )为主要评价指标。当真实框与预测框的  $IoU$  值大于 0.5 时即认为检测正确。评价指标计算公式如式(9)~式(11)所示。

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$mAP = \int_0^1 P(R) dR \quad (11)$$

其中,正确预测框  $TP$  表示真实框与预测框正确匹配,两者间的  $IoU$  大于 0.5;误检框  $FP$  表示将背景预测为目标;漏检框  $FN$  表示本应被模型检测出的目标被预测为背景。

## 2 结果与分析

### 2.1 不同检测试验结果及分析

为了验证本文提出的改进 YOLOv5 算法模型的性能,在公开数据集上将其与其他主流目标检测算法进行了试验对比,试验结果如表 1 所示。

表 1 不同算法的检测结果对比  
Tab. 1 Comparison of results of different algorithms

算法	$P$	$R$	$mAP$
SSD	0.712	0.570	0.632
YOLOv3	0.785	0.622	0.683
YOLOv5	0.834	0.651	0.747
TPH-YOLOv5	0.831	0.673	0.773
本文算法	0.859	0.764	0.794

由表 1 可知,本文改进的算法模型 YOLOv5 + C3STR+CBMA+SIOU 相比于其他主流目标检测算法,准确率、召回率和平均准确率均有所提升。与 SSD、YOLOv3 相比, $mAP$  值分别提升 16.7%、9.4%。

TPH-YOLOv5 是基于 Transformer 改进的 YOLOv5 模型,在 YOLOv5 的基础上又添加了一个用来检测小目标的预测头,为其设置初始锚框,并同时对其对 YOLOv5 的 head 部分进行改进,将 Transformer Prediction Heads 集成到 YOLOv5 中,代替原来的 prediction heads,能有效提高小目标的平均检测精确度。本文模型与 TPH-YOLOv5 相比,准确率提升了 2.8%,召回率提升了 9.1%,平均准确率  $mAP$  提升了 2.1%。

由表 1 可知,本文算法相比于原始的 YOLOv5 模型,改进后的 YOLOv5 算法模型对农田害虫检测的准确率、召回率、平均正确率的指标参数都有所提升。将改进前后的模型在数据集上的准确率、召回率、 $mAP$  等参数变化进行了可视化展示。一般来说,准确率和召回率为负的相关的互斥关系,但本文改进后的算法

在数据集上准确率相较于原始值提升了 2.5%,召回率提升了近 11.3%,平均准确率  $mAP$  提升了 4.7%,在准确率和召回率中达到了最优的平衡状态,因此本文提出的 YOLOv5 改进模型是适合用于小目标农田害虫的检测实际应用中,如图 8 和图 9 所示。

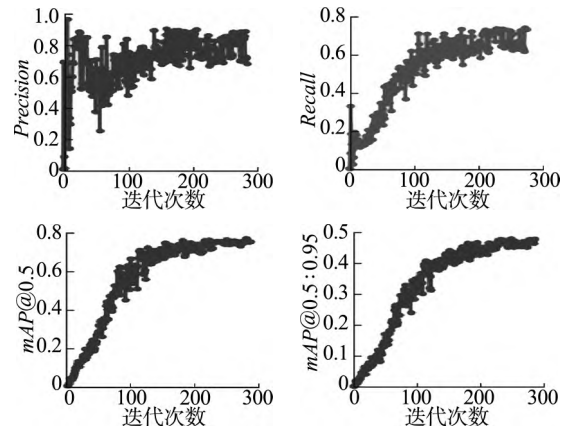


图 8 YOLOv5s 改进前参数

Fig. 8 Parameters before YOLOv5s improvement

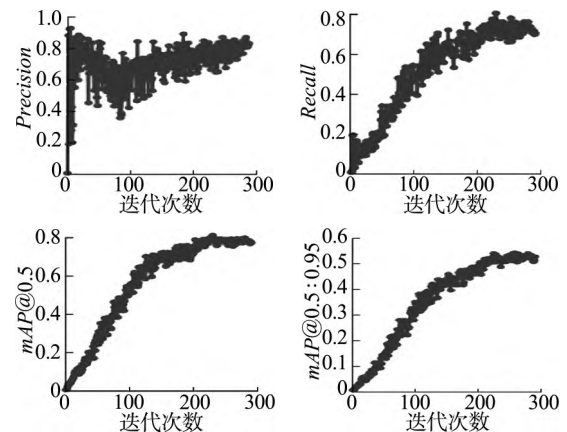


图 9 YOLOv5s 改进后参数

Fig. 9 Parameter after YOLOv5s improvement

对数据集进行训练得到的结果,如图 10 所示。

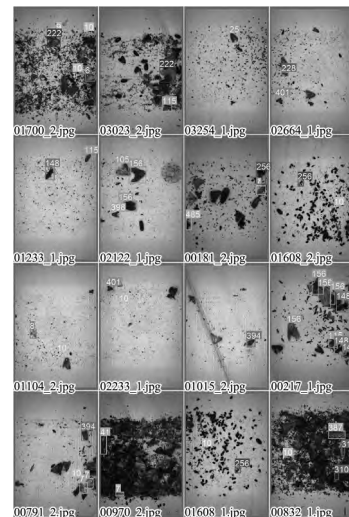


图 10 改进后农田害虫检测

Fig. 10 Improved pest detection in farmland

从图 10 可以得知,改进后的算法对农田害虫的检测结果精度更高,检测结果更全面;同时改进后的算法对小目标检测具有良好的识别性,比原始算法具有更高的置信度。对于训练样本中某些标记次数较多的害虫类别,如 148 类(黄足猎蝽)、156 类(八点灰灯蛾)和 256 类(蟋蟀)等,不存在误检和漏检情况。但由于训练样本中有些类别害虫标记次数过少,如 430 类(豆野螟)和 673 类(干纹冬夜蛾)等,网络对于这些类别学习能力不足,仍然会出现部分害虫误检和漏检情况。

## 2.2 消融试验结果及分析

通过本文对原始 YOLOv5s 模型的三个改进,在公开数据集上进行了消融试验,证明所提出的改进模型的有效性,试验结果如表 2 所示。

表 2 消融试验结果对比  
Tab. 2 Comparison of ablation results

算法	<i>P</i>	<i>R</i>	<i>mAP</i>
YOLOv5	0.834	0.651	0.747
YOLOv5+C3STR	0.818	0.676	0.763
YOLOv5+SIOU	0.828	0.626	0.735
YOLOv5+CBAM	0.813	0.621	0.695
YOLOv5+C3STR+SIOU	0.847	0.754	0.773
YOLOv5+C3STR+CBAM	0.842	0.743	0.782
YOLOv5+CBAM+SIOU	0.850	0.692	0.748
本文算法	0.859	0.764	0.794

其中,C3STR 是指在主干网络的 C3 卷积块特征融合部分引入 Swin Transformer 模块;CBAM 是指在 Neck 的 C3 卷积块后引入卷积注意力机制模板 CBAM;SIOU 是指在 YOLOv5 中引入 SIOU 边界框回归损失函数。

由表 2 可知,本文所提出改进方法中,在 YOLOv5s 主干网络中加入 C3STR,*mAP* 可以提升 1.6%;加入 C3STR 并进入 SIOU 损失函数,*mAP* 可以提升 2.6%;加入 C3STR 并在 Neck 中加入 CBAM,*mAP* 可以提升 3.5%;当所有改进方法同时加入原始 YOLOv5s 模型后,模型整体的 *mAP* 可以提升 4.7%,达到 79.4%。

## 3 讨论

本文针对小目标多类别农田害虫图像识别问题提出的改进后的 YOLOv5 模型,对图像数据集进行预处理并进行数据增强处理,克服部分类别的训练样本数据量不足的问题。模型能够较好地识别定位出小目标和大目标的害虫图像,检测效率高。模型中采用自适应锚框,能够根据图像数据选择合适的锚框并进行检测,很好地解决农业害虫图像多尺度检测的问题,模型适应性

强,对害虫图像数据的鲁棒性较好。模型的模块化和可迁移性较好,下一批害虫图像数据可以添加进入模型的训练中,不需要重新训练全部数据,在之前训练权重基础上进行训练。在处理更大的数据量,模型的迁移性好。

本文是针对农业病虫害图像的识别,模型可以推广到其他的包含小目标检测的问题。模型可以进一步考虑剔除一些不是病虫的昆虫,减小其噪声的干扰。

## 4 结论

针对农田害虫图像中目标检测存在的问题,本文提出了改进后的 YOLOv5 算法。

1) 首先,主干网络中随着网络结构的加深,经过多次卷积操作,小目标应该具有的大部分目标特征信息可能会丢失。故在主干网络最后两个 C3 卷积块引入 Swin Transformer 网络结构,借助窗口自注意力模块增强小目标的语义信息和特征表示。

2) 其次,在颈部的 C3 卷积块后引入 CBAM 模板,CBAM 模块通过卷积核池化操作计算出特征图在空间维度和通道维度的不同权重,使网络能够更加关注到图像中关于检测目标的特征信息。

3) 最后,由于 YOLOv5 自身的交并比函数为 CIoU 边界框回归损失函数,没有考虑到真实框与预测框之间的方向,导致收敛速度较慢,且对小目标的检测精度不准确。对此本文考虑到期望的真实框和预测框之间的向量夹角,引入 SIOU 函数重新定义相关损失函数。

4) 经过试验对比,本文算法相比于原始的 YOLOv5s,平均检测准确率在公开数据集上提升 4.7%,可以表明本文算法在农田害虫图像目标检测领域的有效性。

本文算法也存在一定的局限性:改进后的模型会使得网络结构变复杂,使得网络的训练时间增加,检测速度下降,实时性变差。未来的研究方向是采用网络剪枝、权重量化等手段使模型更加轻量化,继续优化硬件资源占用,提高检测速度,实现农田害虫的实时监测在工业中的实际应用。

## 参 考 文 献

- [1] 肖忠毅. 农业生产数字化转型的实践机制研究[D]. 无锡: 江南大学, 2022.
- [2] 牛霆葳. 基于机器视觉的农田害虫自动识别方法研究[D]. 天津: 天津科技大学, 2015.
- [3] Larios N, Deng H, Zhang W, et al. Automated insect identification through concatenated histograms of local appearance features; feature vector generation and region detection for deformable objects [J]. Machine Vision and Applications, 2008, 19(2): 105-123.

- [4] Zhu L Q, Zhang Z. Auto-classification of insect images based on color histogram and GLCM [C]. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, 2010, 6: 2589–2593.
- [5] Wang J, Lin C, Ji L, et al. A new automatic identification system of insect images at the order level [J]. Knowledge-Based Systems, 2012, 33: 102–110.
- [6] Faithpraise F, Birch P, Young R, et al. Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters [J]. Int. J. Adv. Biotechnol. Res, 2013, 4(2): 189–199.
- [7] Xia C, Chon T S, Ren Z, et al. Automatic identification and counting of small size pests in greenhouse conditions with low computational cost [J]. Ecological Informatics, 2015, 29: 139–146.
- [8] Wang X F, Huang D S, Xu H. An efficient local Char-Vese model for image segmentation [J]. Pattern Recognition, 2010, 43(3): 603–618.
- [9] Xie C, Zhang J, Li R, et al. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning [J]. Computers and Electronics in Agriculture, 2015, 119: 123–132.
- [10] 文斌, 曹仁轩, 杨启良, 等. 改进YOLOv3算法检测三七叶片病害[J]. 农业工程学报, 2022, 38(3): 164–172.  
Wen Bin, Cao Renxuan, Yang Qiliang, et al. Detecting leaf disease for Panax notoginseng using an improved YOLOv3 algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(3): 164–172.
- [11] 周逸博, 马毓涛, 赵艳茹. 基于YOLOv5s和Android的苹果树皮病害识别系统设计[J]. 广东农业科学, 2022, 49(10): 155–163.  
Zhou Yibo, Ma Yutao, Zhao Yanru. Design of mobile app recognition system for apple bark disease based on YOLOv5s and Android [J]. Guangdong Agricultural Sciences, 2022, 49(10): 155–163.
- [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, 1: 580–587.
- [13] Girshick R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015, 1: 1440–1448.
- [14] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28.
- [15] He K, Gkioxari G, Dollár P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961–2969.
- [16] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks [J]. Advances in Neural Information Processing Systems, 2016, 29.
- [17] 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.
- [18] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012–10022.
- [19] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21–37.
- [20] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018, 1: 3–19.
- [21] 赵文博, 周德强, 邓干然, 等. 基于改进YOLOv5的甘蔗茎节识别方法[J]. 华中农业大学学报, 2023, 42(1): 268–276.
- [22] Wan J, Chen B, Yu Y. Polyp detection from colorectum images by using attentive YOLOv5 [J]. Diagnostics, 2021, 11(12): 2264.
- [23] 汪斌斌, 杨贵军, 杨浩, 等. 基于YOLO\_X和迁移学习的无人机影像玉米雄穗检测[J]. 农业工程学报, 2022, 38(15): 53–62.  
Wang Binbin, Yang Guijun, Yang Hao, et al. UAV images for detecting maize tassel based on YOLO\_X and transfer learning [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(15): 53–62.
- [24] Xue Z, Lin H, Wang F. A small target forest fire detection model based on YOLOv5 improvement [J]. Forests, 2022, 13(8): 1332.
- [25] Redmon J, Farhadi A. Yolov3: An incremental improvement [J]. arXiv Preprint arXiv: 1804.02767, 2018.
- [26] Parmar N, Vaswani A, Uszkoreit J, et al. Image transformer [C]. International Conference on Machine Learning, 2018: 4055–4064.
- [27] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. arXiv Preprint arXiv: 2004.10934, 2020.
- [28] Yan B, Fan P, Lei X, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5 [J]. Remote Sensing, 2021, 13(9): 1619.
- [29] Gevorgyan Z. SIoU Loss: More powerful learning for bounding box regression [J]. arXiv Preprint arXiv: 2205.12740, 2022.

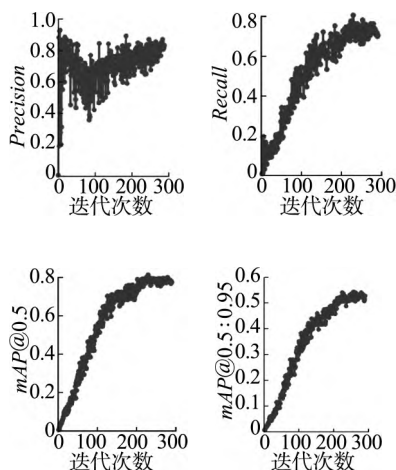
# 论文精选:改进 YOLOv5 的小目标多类别 农田害虫检测算法研究

## 研究概况

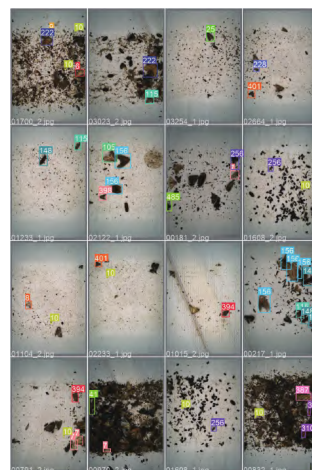
针对农田害虫图像中感兴趣目标特征不明显、小目标居多导致的目标检测精度较低的问题,提出一种基于 YOLOv5 改进的小目标多类别农田害虫目标检测算法。首先,在主干网络最后两个 C3 卷积块特征融合部分引入 Swin Transformer 窗口注意力网络结构,增强小目标的语义信息和全局感知能力;其次,在颈部网络的 C3 卷积块后添加通道注意力机制和空间注意力机制的可学习自适应权重,使网络能够关注到图像中关于小目标的特征信息;最后,由于 YOLOv5 自身的交并比函数存在收敛速度较慢且精确率较低的问题,引入 SIOU 函数作为新的边界框回归损失函数,提高检测的收敛速度和精确度。将所提出的算法在包含 28 类农田害虫公开数据集上进行试验,结果表明,改进后的算法在农田害虫图像数据集上的准确率、召回率和平均准确率分别达到 85.9%、76.4%、79.4%,相比于 YOLOv5 分别提升 2.5%、11.3%、4.7%。该研究由云南省重大科技专项计划项目(202002AE090010)等资助。该研究通讯作者为河海大学的刘向阳副教授。

## 主要技术指标

试验类别(神经网络模型)	检测准确率	检测召回率	检测平均准确率
SSD	0.712	0.570	0.632
YOLOv3	0.785	0.622	0.683
YOLOv5	0.834	0.651	0.747
TPH-YOLOv5	0.831	0.673	0.773
YOLOv5+C3STR	0.818	0.676	0.763
YOLOv5+SIOU	0.828	0.626	0.735
YOLOv5+CBAM	0.813	0.621	0.695
YOLOv5+C3STR+SIOU	0.847	0.754	0.773
YOLOv5+C3STR+CBAM	0.842	0.743	0.782
YOLOv5+CBAM+SIOU	0.850	0.692	0.748
Our approach	0.859	0.764	0.794



YOLOv5s 改进后参数



试验害虫检测结果图像