

基于改进 YOLOv5 网络的复杂背景图像中茶尺蠖检测

胡根生, 吴继甜, 鲍文霞*, 曾伟辉

(安徽大学农业生态大数据分析与应用技术国家地方联合工程研究中心, 合肥 230601)

摘要: 茶叶的产量和品质深受病虫害的影响。茶尺蠖是一种常见的茶叶害虫, 精确检测茶尺蠖对茶叶病虫害防治有重要意义。由于茶尺蠖和茶树枝、枯死茶叶的颜色、纹理相近, 茶尺蠖的体积小、形态多变、被遮挡等问题, 现有方法检测茶尺蠖的精度不高。该研究提出一种基于深度学习的复杂背景图像中茶尺蠖检测方法, 该方法使用 YOLOv5 为基线网络, 利用卷积核组增强对茶尺蠖的特征提取, 在不增加计算量的条件下减少复杂背景对茶尺蠖检测结果的干扰; 使用注意力模块关注茶尺蠖整体, 根据茶尺蠖的大小和形状自适应调节感受野, 降低因目标大小形状不一导致的漏检; 使用 Focal loss 损失函数减少前景和背景的类不平衡对检测结果的影响。试验结果表明, 所提方法用于复杂背景图像中茶尺蠖的检测, 可以达到 0.94 的召回率, 0.96 的精确度和 92.89% 的平均精度均值。与基线网络相比, 该方法的平均精度均值提高了 6.44 个百分点。使用相同的数据集和预处理的对比分析表明, 该方法优于 SSD、Faster RCNN 和 YOLOv4 等其他经典深度学习方法, 平均精度均值比 SSD、Faster RCNN、YOLOv4 分别高 17.18 个百分点、6.52 个百分点和 4.78 个百分点。该方法可实现对茶尺蠖的智能检测, 减少人力成本, 有助于实现精准施药, 提高茶叶的产量和品质。

关键词: 农业; 算法; 目标检测; 深度学习; 卷积核组; 注意力模块; 茶尺蠖

doi: 10.11975/j.issn.1002-6819.2021.21.022

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2021)-21-0191-08

胡根生, 吴继甜, 鲍文霞, 等. 基于改进 YOLOv5 网络的复杂背景图像中茶尺蠖检测[J]. 农业工程学报, 2021, 37(21): 191-198. doi: 10.11975/j.issn.1002-6819.2021.21.022 <http://www.tcsae.org>

Hu Gensheng, Wu Jitian, Bao Wenxia, et al. Detection of Ectropis oblique in complex background images using improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(21): 191-198. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.21.022 <http://www.tcsae.org>

0 引言

中国是茶叶生产和消费大国。近年来中国茶叶生产快速发展, 茶叶种植面积扩大, 茶叶产量不断增长, 在国民经济中占有相当大的比重^[1]。茶叶生长易受病虫害影响, 导致产量和质量大幅度下降, 给茶农造成较大损失^[2]。尽早发现病虫害并采取相应的防治措施, 对减少茶叶产量损失、提高茶叶品质、减少农药使用和防止环境污染等有着重要意义。

传统上对于茶尺蠖的检测, 主要是请植保专家到茶园进行目视判断。由于大部分茶园地处偏僻, 地势陡峭, 人工检测会耗费大量的物力和财力^[3-4]。随着数码相机和互联网的普及, 使得研究者开发出自然场景图像中茶尺蠖虫害的高精确度的自动诊断技术, 方便茶农快捷地诊断茶尺蠖虫害成为可能。随着计算机和人工智能技术的快速发展, 机器学习已被广泛应用于植物图像中的病虫害检测。在机器学习方法中, 害虫检测技术大多基于形

状、颜色和纹理等特征, 通过 K 最近邻 (K-Nearest Neighbor, KNN)^[5]、支持向量机 (Support Vector Machine, SVM)^[6-8] 和 Adaboost 等^[9] 对害虫进行检测, 具有计算简单、参数量少、抗噪性能好等优点。上述传统机器学习方法需要手工抽取图像中病虫害的特征, 具有主观性, 易受背景信息影响。野外拍摄的茶尺蠖图像背景复杂, 图像中的茶树枝、枯叶等背景目标与茶尺蠖的形状颜色相近, 手工方法难以抽取茶尺蠖的本质特征, 因而利用传统机器学习方法检测复杂背景图像中的茶尺蠖目标难以获得好的检测结果。

深度学习方法能够自动地从训练数据集中提取目标的特征。在深度学习方法中, 由于卷积神经网络含有多层隐含层, 能够学习到更加高级的数据特征表示, 无需手动设计特征, 在解决目标检测等问题上有很大的优势^[10]。常用的目标检测网络有 SSD (Single Shot multiBox Detector)^[11]、RetinaNet^[12]、Faster RCNN (Faster Region with CNN) 系列^[13-15] 和 YOLO (You Only Look Once) 系列^[16-19] 等。研究者使用卷积神经网络对害虫进行检测^[20-24], 以 Fast RCNN 和 SSD 在害虫检测技术中的应用为例^[25-26], 根据卷积层提取到的特征向量, 分别通过窗口分类和窗口回归, 得到每个候选区域中害虫的类别以及原始坐标。此类方法有效地提升了害虫检测的准确率, 能够对害虫目标进行定位。由于野外拍摄的茶尺蠖图像的前景与背景及其相似, 前景茶尺蠖的形态、尺度多变,

收稿日期: 2021-06-18 修订日期: 2021-10-08

基金项目: 安徽省高等学校自然科学研究重大项目 (KJ2020ZD03); 农业生态大数据分析与应用技术国家地方联合工程研究中心开放课题项目 (AE201902)

作者简介: 胡根生, 博士, 教授, 研究方向为农业信息处理与应用。

Email: hugs2906@sina.com

※通信作者: 鲍文霞, 博士, 副教授, 研究方向为农业信息处理与模式识别。

Email: bwxia@ahu.edu.cn

直接使用上述网络检测复杂图像中的茶尺蠖的准确率有待进一步提高。

YOLO 系列目标检测网络是目标检测网络的一个重要分支,其在农业目标检测中的应用也越来越广泛^[27-32]。例如张博等^[27]将空间金字塔池化与改进的 YOLOv3 深度卷积神经网络相结合,提出了一种基于空间金字塔池化的深度卷积神经网络农作物害虫种类识别算法,首先对测试图像上的害虫进行检测定位,然后对检测定位出的害虫进行种类识别。Wu 等^[30]开发了一种基于通道修剪的 YOLO v4 算法用于自然环境下的苹果花检测。结果表明,该算法对变化的果树品种和光照方向具有较好的鲁棒性。Suo 等^[31]使用 YOLOv3 和 YOLOv4 对多类猕猴桃检测,可以避免被树枝或线遮挡的果实作为可采摘目标。YOLOv5 不仅拥有比 YOLOv3 和 YOLOv4 更快的检测速度,并且能够有效检测被遮挡的目标,通过调节网络感受野大小和增强网络特征提取能力,可以更准确的检测到复杂背景中和被遮挡的目标。

由于野外自然场景中拍摄的茶尺蠖图像背景复杂,茶尺蠖的颜色、纹理与茶树杆和枯死茶叶相近,茶尺蠖的体积小、形态多变、被遮挡,因而准确检测复杂背景图像中的茶尺蠖是一大难题。为提高茶尺蠖检测的准确率,本研究提出一种基于改进 YOLOv5 网络的复杂背景

图像中茶尺蠖检测方法。

1 材料与方法

1.1 图像获取和预处理

本研究所用的图像采集于天井山茶园。该茶园位于安徽省省会合肥市的南部,巢湖之滨。其地理坐标为北纬 31°14'37",东经 117°36'16",海拔 40 m,采集时间分别为 2019 年 10 月和 2020 年 10 月,天气晴朗。图像采集设备为佳能 EOS 80D 手持数码单反相机,拍摄时相机距离茶树冠层约 0.4 m,图像大小为 6 000×4 000 像素,共拍摄了自然场景中茶尺蠖的原始图像 109 张。从原始图像上裁剪出 112 张大小不一的茶尺蠖图像,按照 6:1:3 的比例构造训练集、验证集和测试集,其中训练集有 67 张图像,验证集有 11 张图像,测试集有 34 张图像。

为了提高小样本条件下网络的泛化性能,防止训练图像数量不足导致网络过拟合,本研究对训练图像和验证图像进行增广。增广方式包括:翻转、改变对比度和加入高斯噪声。经过增广后,训练集共有 268 张图像,验证集共有 44 张图像,表 1 给出了增广前后训练集和测试集的图像数以及图像中的茶尺蠖数量。图 1 给出了增广后的茶尺蠖图像示例。使用 Labelimg 工具对图像进行标注,生成 xml 文件,文件中包含了图像名称、大小和茶尺蠖在图像中的位置信息。

表 1 增广前后的数据集

Table 1 Before and after augmentation of data sets

项目 Item	训练集 Training set		验证集 Validation set		测试集 Test set	
	图像数 Number of images	茶尺蠖数 Number of <i>Ectropis oblique</i>	图像数 Number of images	茶尺蠖数 Number of <i>Ectropis oblique</i>	图像数 Number of images	茶尺蠖数 Number of <i>Ectropis oblique</i>
增广前 Before augmentation	67	280	11	46	34	151
增广后 After augmentation	268	1 120	44	184	34	151

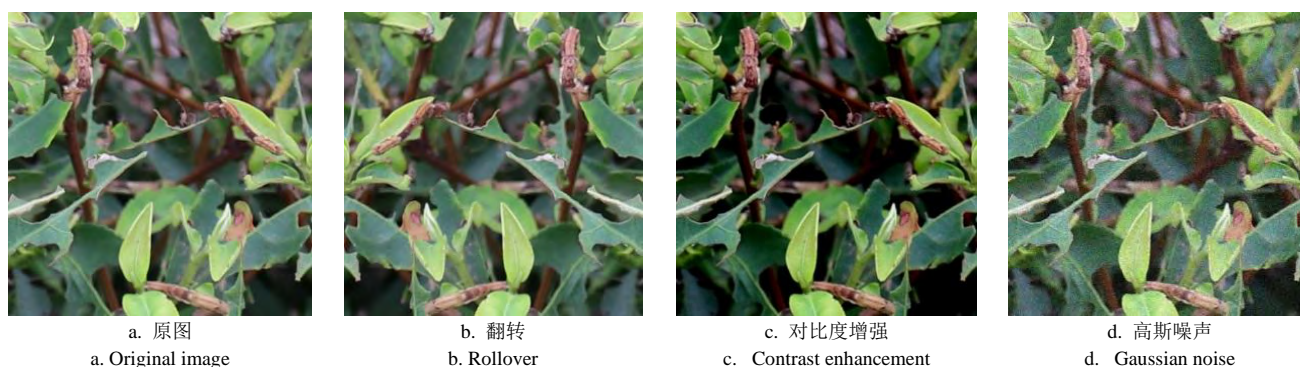


图 1 茶尺蠖图像增广结果

Fig.1 Image augmentation results of *Ectropis oblique*

由于自然场景图像中的茶尺蠖呈现灰褐色,茶杆颜色为深褐色。为减小茶杆等复杂背景对茶尺蠖检测的影响,本研究对测试图像的对比度进行调节,将图像的 R、G 和 B 三通道的强度值[0.3,1]映射到新图像的[0,1]中。经过对比度增强处理后,图像中的茶杆、茶叶缝隙等背景的颜色变为黑色,茶尺蠖的颜色和纹理特征与茶杆等背景的区别更大。

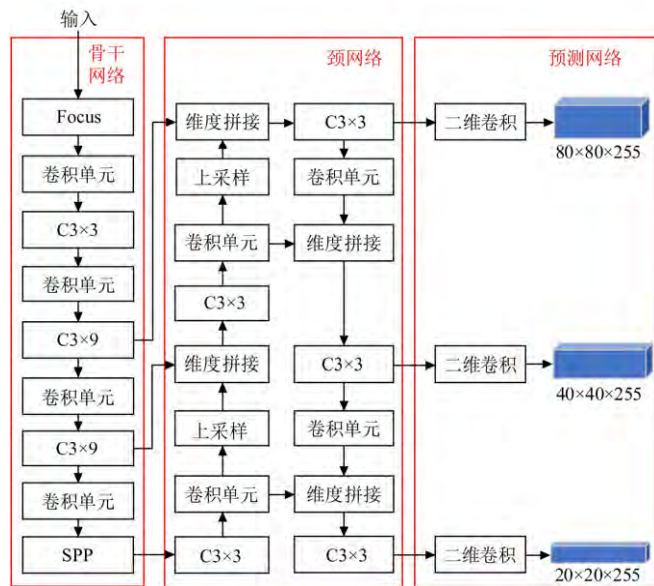
1.2 茶尺蠖检测网络

1.2.1 YOLOv5 网络

YOLOv5 的网络结构可分为三部分:骨干网络、颈网络和预测网络。在骨干网络中,Focus 模块对输入图像进行切片操作,切片后的特征图通道数扩充了 4 倍,特征图经过卷积操作得到二倍下采样特征图。卷积单元包括二维卷积、批标准化和 SiLu (激活函数)。C3 模块实

现从特征图上提取特征,通过调整 C3 模块的宽度和深度,可以获得参数量不同的 4 个网络,分别为 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x,对应参数量分别为 7.30×10^6 、 2.14×10^7 、 4.71×10^7 、 8.78×10^7 。由于参数量越少,实时性越好,考虑到农业应用的时效性,本研究使用参数量最少速度最快的 YOLOv5s,图 2 给出 YOLOv5s 的网络结构示意图。SPP 模块通过不同核大小的池化抽取不同尺度特征,再进行叠加进行特征融合。

YOLOv5 在颈网络中使用 FPN (Feature Pyramid Networks) 结构,将顶层特征和底层特征融合,以提高网络检测不同尺度目标的能力。在预测网络输出分类结果和目标的坐标。



注: Focus 为切片操作; C3 为 YOLOv5 网络中的瓶颈层,其后数字为 C3 的个数; SPP 为空间金字塔池化结构; 20×20×255、40×40×255 和 80×80×255 为预测网络输出特征图的长、宽和维度。

Note: Focus is the slice operation; C3 is the bottleneck layer in YOLOv5 network, followed by the number of C3; SPP is a spatial pyramid pool structure; 20×20×255, 40×40×255 and 80×80×255 are the length, width and dimensions of the output feature graph of the predicted network.

图 2 YOLOv5s 网络结构

Fig.2 Network structure of YOLOv5s

1.2.2 卷积核组

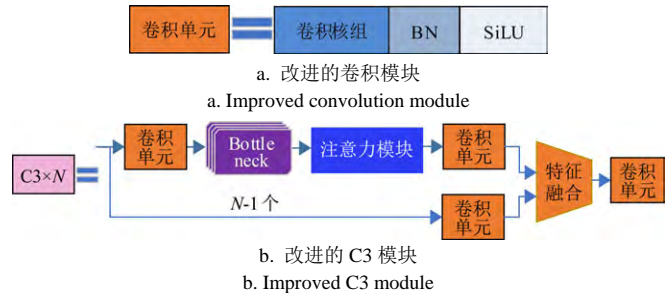
由于茶尺蠖图像是在野外自然场景中拍摄,图像中背景复杂,茶尺蠖和枯死茶叶、茶杆的颜色纹理相近,存在部分茶尺蠖被背景遮挡,YOLOv5 骨干网络不能提取清晰的特征。为了减少复杂背景的影响,本研究通过增广训练本来提高网络的泛化性,同时将骨干网络中所有的 3×3 卷积核替换成卷积核组,以增强网络对茶尺蠖的特征提取能力。

卷积核组 (convolution kernel group) 由 3 个并行的 3×3、1×3 和 3×1 的卷积核组成,这 3 个并行的卷积核对输入图像以相同的步长进行卷积,分别产生大小和通道数相同的特征图,对应的特征图求和后获得输出特征图。3 个并行的卷积核增强了网络对茶尺蠖特征的提取能力。设 I 是输入图像, $K^{(1)}$ 、 $K^{(2)}$ 和 $K^{(3)}$ 分别是 3×3、1×3 和 3×1 的卷积核,则

$$I * K^{(1)} + I * K^{(2)} + I * K^{(3)} = I * (K^{(1)} + K^{(2)} + K^{(3)}) = I * K \quad (1)$$

式中 K 为等效的卷积核; * 为卷积运算。

因而 3 个并行的卷积核等效于一个新的权重不同的 3×3 的卷积核,如图 3 所示。测试时使用训练好的等效卷积核,对比替换前的 3×3 卷积核,在不增加额外计算量的条件下增强了网络对茶尺蠖特征的提取能力。



注 BN 为批标准化; SiLU 为激活函数; Bottleneck 为一个 1×1 的卷积后接一个 3×3 的卷积,其中 1×1 的卷积将通道数减半,3×3 的卷积将通道数加倍; N 为 C3 的个数。

Note: BN is batch normalization; SiLU is the activation function; Bottleneck is 1×1 convolution followed by a 3×3 convolution, where 1×1 convolution halves the number of channels, and 3×3 convolution doubles the number of channels; N is the number of C3.

图 3 茶尺蠖检测网络

Fig.3 Detection network of *Ectropis oblique*

1.2.3 注意力模块

由于茶尺蠖的体积小且形态多变等问题对 YOLOv5 检测算法产生干扰,本研究在 YOLOv5s 网络中添加注意力模块 (attention block) [33],通过对输入特征进行加权,强化茶尺蠖目标信息,弱化背景信息,以达到对茶尺蠖的关注,减少茶尺蠖体积小和形态多变的影响。注意力模块还使用不同大小的卷积核,自适应调节感受野大小,以适应大小、形状不同的茶尺蠖图像,有效提取茶尺蠖的全局信息。

注意力模块由分割、融合和选择 3 个部分组成。Split 部分将输入的特征图 x 分别与 3×3 卷积核、膨胀率为 3 的 3×3 卷积核和 5×5 的深度可分离卷积核进行卷积,得到特征图 \hat{U} 、 \hat{V} 和 \hat{W} 。融合部分将经过分割得到的特征图 \hat{U} 、 \hat{V} 和 \hat{W} 进行叠加、全局平均池化、全连接和 softmax 后,输出两个矩阵 a 和 b 。选择部分利用 a 和 b 分矩阵对 \hat{U} 和 \hat{W} 进行加权得到 \hat{Y} 和 \hat{Z} ,利用全连接层的输出 Z 对 U 进行加权得到 \bar{Y} ,叠加 \hat{Y} 、 \hat{Z} 和 \bar{Y} 得到注意力模块的输出 Y 。

1.2.4 Focal loss 损失函数

由于野外拍摄的茶尺蠖图像的背景复杂,背景目标的类别多,为了减小前景和背景的不平衡对检测结果的影响,本研究将交叉熵损失函数替换为 Focal loss 损失函数 [12]。

Focal loss 损失函数是在交叉熵损失函数基础上进行改进的,交叉熵损失函数 L 为

$$L = -y \ln y' - (1 - y) \ln(1 - y') = \begin{cases} -\ln y', & y = 1 \\ -\ln(1 - y'), & y = 0 \end{cases} \quad (2)$$

式中 y 为样本类别的真值; y' 是经过激活函数得到的输出,在 0~1 之间,表示预测前景目标的概率。因此对前景目标而言,预测概率越小,损失值越大。而对于背景目标而言,预测概率越小,损失值越小。如果存在大量的背

景目标, 则损失函数迭代缓慢且可能无法优化至最优。Focal loss 损失函数 (L_n) 在交叉熵损失函数的基础上加了因子 $\gamma > 0$, 降低易分类目标的损失, 使得网络更关注于困难的、错分的目标, 减小前景和背景的类不平衡对检测结果的影响。

$$L_n = \begin{cases} -(1-y')^\gamma \ln y', & y=1 \\ -y'^\gamma \ln(1-y'), & y=0 \end{cases} \quad (3)$$

1.2.5 茶尺蠖检测网络

本研究所用的茶尺蠖检测网络以图 2 中 YOLOv5s 作为基线网络, 使用卷积核组替换掉骨干网络中所有的 3×3 卷积核, 以增强网络对茶尺蠖的特征提取能力, 降低复杂背景的干扰。在 C3 模块中添加可以自适应调节感受野大小的注意力模块, 以关注不同尺度的茶尺蠖, 提取茶尺蠖完整的信息, 减少茶尺蠖大小和形状不对检测结果的影响。使用 Focal loss 作为损失函数, 以此减少前景和背景类别不平衡带来的影响。改进的卷积模块、C3 模块结构如图 3 所示。

1.2.6 参数设置与精度评价

茶尺蠖检测网络用 pytorch 框架搭建, 在 NVIDIA Tesla V100 服务器上进行训练和测试。图像大小调整为 640×640 像素, 批量大小 (batch size) 设置为 16, 学习率 (learning rate) 为 0.01, 训练步数为 1 000。采用 Adam 优化, 动量 (momentum) 设为 0.9, 权重衰减 (weight decay) 为 0.000 5。

本研究使用精度 (Precision)、召回率 (Recall)、平均精度均值 (mean Average Precision, mAP) 和检测速度作为茶尺蠖检测结果的定量评价指标。设置检测正确的标准为交并比 (Intersection-over-Union, IoU) ≥ 0.5 , 其中 IoU 是预测边界框和真实边界框之间的重叠率。

2 结果与分析

2.1 与传统机器学习方法对比

将所提方法与 SVM、BP 神经网络、AdaBoost 等经典机器学习方法进行对比, 试验过程设置如下: 增广后的训练集和验证集图像被调整成 $1\,000 \times 1\,000$ 大小, 每张图像按照长宽 10 等分分成 100×100 大小的 100 张图像。为了保证背景图像和害虫图像数量的均衡, 从 10 等分后的图像中挑选出合适的背景图像 2 000 张, 害虫图像 2 000 张, 共 4 000 张图像进行训练, 利用训练好的模型对 3 400 张 (即 34 张测试集图像按照长宽 10 等分) 测试图像进行测试。SVM 模型的核函数选择 RBF 核 (SVM 模型中的参数), 惩罚系数 C 设置为 200; BP 算法的学习率为 0.01, 批量大小为 16, 权重衰减为 0.000 5, 动量为 0.9, 迭代次数为 1 000 次; AdaBoost 算法的弱分类器数量设置为 20。不同方法的检测结果如表 2 所示。图 4 显示了不同方法的一些检测结果示例, 图中的小框为模型检测出的目标, 大框为根据检测结果手动标注正确检测出的完整茶尺蠖目标。从表 2 和图 4 可以看出, 由于样本中存在茶尺蠖被遮挡、大小形状不一和复杂背景等问题, 所以利用人工提取特征的 SVM、BP 和 Adaboost 检测方法难以精确检测出茶尺蠖。3 种方法的漏检测数分别为

63、49 和 40, 并且有较多的误检测。而本文所提方法仅有 9 个漏检测, 检测效果优于 SVM、BP 和 Adaboost 传统检测方法。

表 2 本文方法与经典机器学习方法的检测结果比较

Table 2 Comparison of detection results between the proposed method and classical machine learning methods

方法 Methods	TP	FP	FN	召回率 Recall R	精度 Precision P	平均精度均值 Mean Average Precision mAP/%
SVM	88	79	63	0.58	0.52	52.34
BP	102	36	49	0.67	0.73	65.46
AdaBoost	111	61	40	0.73	0.64	69.88
本文方法 Proposed method	142	5	9	0.94	0.96	92.89

注: TP、FP 和 FN 分别表示茶尺蠖被正确检测、背景目标被错误地检测为茶尺蠖和茶尺蠖被误检测为背景目标的数量。SVM 为支持向量机; BP 代表反向传播方法。

Note: TP, FP and FN represented the number of *Ectropis oblique* were correctly detected, background targets incorrectly detected as *Ectropis oblique*, and the number of *Ectropis oblique* incorrectly detected as background targets, respectively. SVM is Support Vector Machine; BP is Back Propagation method.

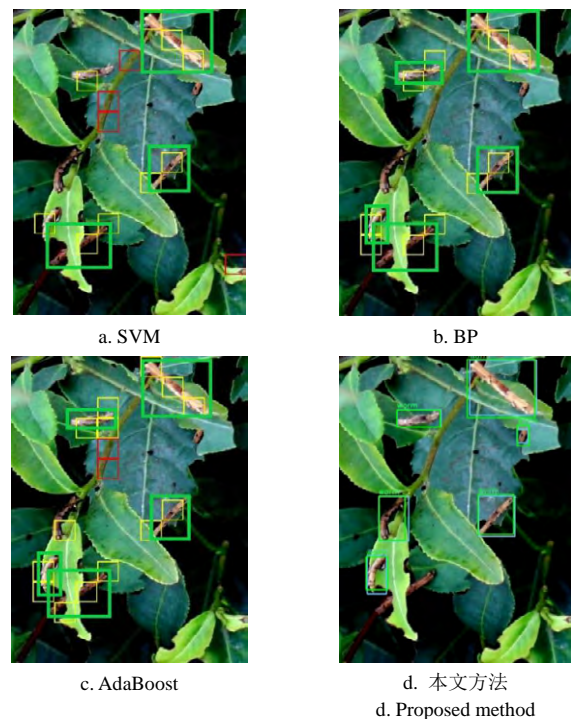


图 4 本文方法和经典机器学习方法的检测结果示例

Fig.4 Examples of detection results of the proposed method and the classical machine learning methods

2.2 与常用目标检测网络的对比

将所提方法与 SSD、Faster-RCNN、RetinaNet、YOLO 系列等常用目标检测网络方法进行对比。不同的目标检测网络都使用预训练模型进行训练, 网络的参数设置如表 3 所示。不同网络方法的检测结果如表 4 所示。从表 4 可以看出, 在不使用数据增广和对比度预处理的条件下, Faster-RCNN 方法的漏检测数为 15, 而误检测数达 86, 这是由于数据集的背景复杂, Faster-RCNN 作为二阶段目标检测网络, 会产生大量的误检测, 导致检测精度不高。由于数据集中茶尺蠖形态多样、被遮挡, SSD 和 RetinaNet 方法的漏检测数为 39 和 21, 可以看出 SSD 和 RetinaNet 方法的检测精度不高。YOLO 系列方法也存在漏检测和

误检测多的问题。所提方法有较少的漏检测和误检测数， P 、 R 、 mAP 指标值高于上述常用目标检测网络方法，其中 mAP 比 SSD、Faster RCNN、YOLOv4 分别高 17.18 个百分点、6.52 个百分点和 4.78 个百分点。如图 5 所示为不同检测网络的检测结果，各网络的茶尺蠖检测框和

真实框比较，当交并比 $IoU \geq 0.5$ ，则认定模型检测正确，反正则认定模型检测错误。与表 4 和图 5 可以看出，在使用数据增广和对比度预处理的条件下，不同网络方法的检测精度都有所提高，但本文方法仍优于上述常用目标检测网络方法。

表 3 训练不同检测网络的参数设置
Table 3 Parameter settings for different detection networks

方法 Method	批量大小 Batch size	学习率 Learning rate	训练步数 Epoch	优化器 Optimizer	动量 Momentum	权重衰减 Weight decay
SSD	16	0.001	1 000	Adam	0.9	0.000 5
Faster-RCNN	16	0.001	1 000	Adam	0.9	0.000 5
RetinaNet	16	0.0001	1 000	Adam	0.9	0.000 5
YOLOv3	16	0.001	1 000	Adam	0.9	0.000 5
YOLOv4	16	0.001	1 000	Adam	0.9	0.000 5
YOLOv5s	16	0.01	1 000	Adam	0.9	0.000 5
本文方法 Proposed method	16	0.01	1 000	Adam	0.9	0.000 5

表 4 不同网络方法的检测结果
Table 4 The detection result of different methods

方法 Methods	无增广无预处理 No augmentation and no preprocessing						增广+预处理 Augmentation and preprocessing					
	TP	FP	FN	R	P	$mAP/\%$	TP	FP	FN	R	P	$mAP/\%$
SSD	112	10	39	0.74	0.91	73.06	116	6	35	0.76	0.95	75.71
Faster-RCNN	136	86	15	0.90	0.61	84.37	138	78	13	0.91	0.63	86.37
RetinaNet	130	19	21	0.86	0.87	84.92	132	11	19	0.87	0.92	86.79
YOLOv3	131	26	20	0.86	0.83	85.15	133	17	18	0.88	0.88	87.42
YOLOv4	132	23	19	0.87	0.85	85.94	135	15	16	0.89	0.90	88.11
YOLOv5s	133	27	18	0.88	0.83	86.45	136	19	15	0.90	0.87	88.68
本文方法 Proposed method	138	17	13	0.91	0.89	90.02	142	5	9	0.94	0.96	92.89

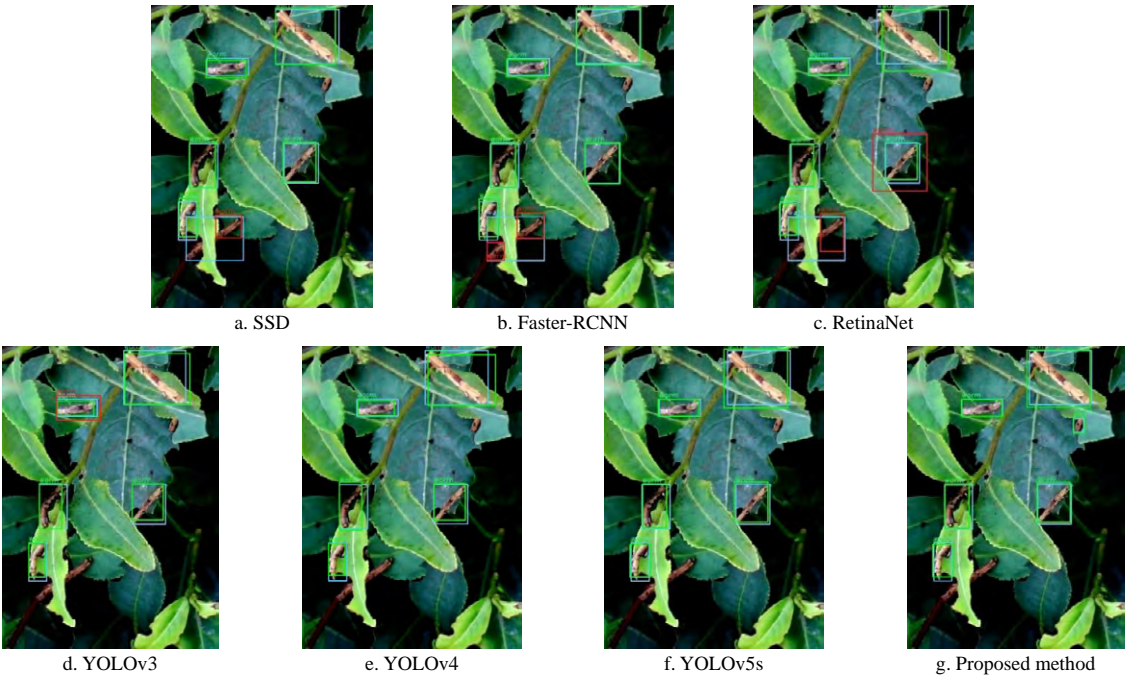


图 5 本文方法和常用目标检测网络方法的检测结果示例

Fig.5 Examples of detection results of the proposed method and common target detection network methods

2.3 消融试验

消融试验结果如表 5 所示。从表 5 可以看出，对数据进行增广和预处理，利用卷积核组增强对茶尺蠖的特征提取，通用注意力模块关注茶尺蠖整体，并使用 Focal loss 损失函数减少前景和背景的类不平衡，提升了模型的检测精度、召回率和 mAP 。

由 2.3 节和表 5 可知，对数据进行增广和预处理，可提高网络泛化性，小幅度提升网络检测效果。上述条件

下，在 YOLOv5s 网络中分别加入卷积核组和注意力模块，可以看出，都对召回率提升明显，这是因为数据集背景复杂和茶尺蠖形态变化大，卷积核组和注意力模块有效抑制它们的影响。最后通过替换交叉熵损失函数为 Focal loss 时，降低了因前景和背景类别不平衡导致的误检测，因而所提方法具有较少的误检测和漏检测，获得了好的检测结果。与基线网络相比，该方法的平均精度均值提高了 6.44 个百分点。

表 5 消融试验结果
Table 5 The results of Ablation experiments

基线网络 Baseline network	增广 Augmentation	预处理 Preprocessing	卷积核组 Convolution kernel group	注意力模块 Attention block	Focal loss	<i>R</i>	<i>P</i>	mAP/%
YOLOv5s						0.88	0.85	86.45
YOLOv5s	√					0.89	0.86	87.83
YOLOv5s	√	√				0.90	0.87	88.68
YOLOv5s	√	√	√			0.92	0.90	90.77
YOLOv5s	√	√		√		0.92	0.87	90.69
YOLOv5s	√	√	√	√		0.94	0.91	92.75
YOLOv5s	√	√	√	√	√	0.94	0.96	92.89

图 6 给出了数据增广前后本文所用模型的训练和验证损失曲线。由图 6a 可知, 当数据无增广时, 训练损失和验证损失存在较大差异, 且验证损失波动较大, 说明模型在训练集上表现较好, 但在验证集上表现较差, 模型训练出现过拟合; 由图 6b 可知, 当数据增广后, 训练损失和测试损失差异极小, 且验证损失波动较小, 说明模型在训练集和验证集上表现都较好。

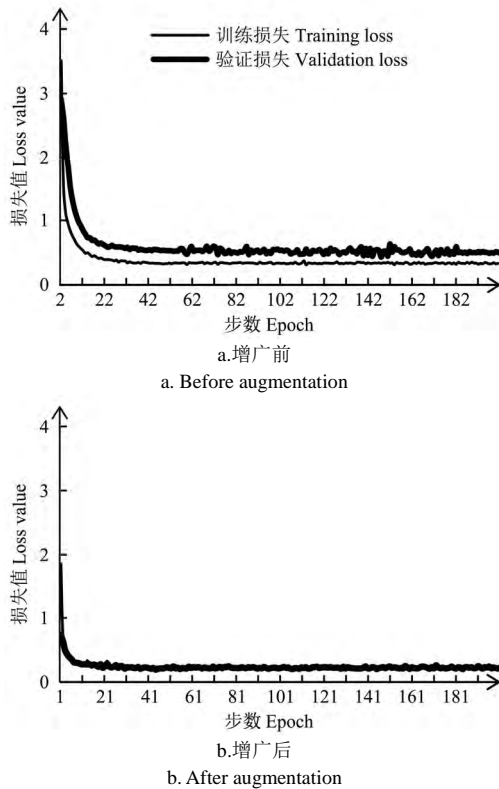


图 6 训练和验证损失曲线

Fig.6 Loss curves of training and validation

2.4 特征可视化

当图像输入网络提取特征时, 特征提取网络越深, 得到的特征图越抽象。为了直观展现卷积核组的效果,

本研究对骨干中第一个 C3×9 模块最后一层的特征图进行可视化, 如图 7 所示, 从图中可以看出, 基线网络使用卷积核组后获取的特征图中茶尺蠖目标变得更加明显, 说明了卷积核组增强了网络对茶尺蠖目标的特征提取能力。

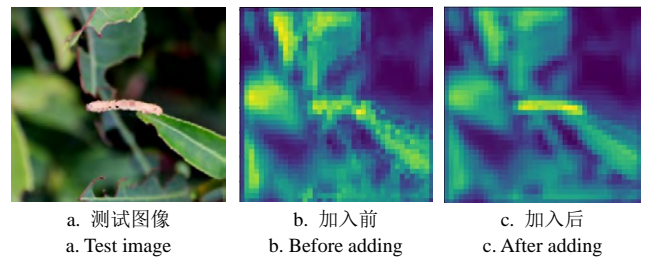


图 7 测试图像和加入卷积核组前后特征图可视化

Fig.7 Test image and feature map visualization before and after adding convolution kernel group

2.5 数据量可靠性分析

虽然增广后的数据集数量依然不是很多, 但有研究者也使用较少的数据集进行试验, 如 Cheng 等^[21]对 10 类害虫进行识别, 每类害虫 55 张图像, 共 550 张图像, 其中每类中 40 张图像用于训练, 分类准确率达到 98.67%; Chen 等^[34]使用 382 张图像作为训练集, 55 张图像作为测试集对麦田蚜虫进行识别, 可达到 96.4%的准确率; Hu 等^[35]使用 398 张茶叶枯萎病进行轻度和重度严重程度估计, 并将训练集增广到 700 张, Faster RCNN 用于检测病叶, VGG16 用于严重程度分析, 取得了较好的效果。

为了验证本研究小样本数据量的可靠性, 又裁剪 153 张茶尺蠖图像, 按照 6:1:3 分成训练集、验证集和测试集, 将训练集和测试集增广后加入本研究的数据集, 获得 636 张训练图像、104 张验证图像和 80 张测试图像, 其中测试图像中有 336 个茶尺蠖样本, 使用本文方法对两种数据集进行试验, 试验结果如表 6 所示, 从表 6 可以看出, 当数据量增长 1.36 倍时, 检测结果的 mAP 仅增长 0.09 个百分点, 说明本文小样本数据量满足本文方法要求。

表 6 不同增广数据集试验结果

Table 6 Experimental results of different augmented data sets

数据集 Dataset	训练集 Training set		验证集 Validation set		测试集 Test set		<i>R</i>	<i>P</i>	mAP/%
	图像数 Number of images	茶尺蠖数 Number of <i>Ectropis oblique</i>	图像数 Number of images	茶尺蠖数 Number of <i>Ectropis oblique</i>	图像数 Number of images	茶尺蠖数 Number of <i>Ectropis oblique</i>			
A	268	1 120	44	184	34	151	0.94	0.96	92.89
B	636	2 596	104	418	80	336	0.94	0.96	92.98

3 结 论

由于自然场景中获取的茶尺蠖图像的背景复杂、茶尺蠖大小形状不一且存在部分遮挡等问题, 现有方法很难获得较高的检测精度, 本研究提出了一种以 YOLOv5s 为基线网络的茶尺蠖检测方法。

1) 为了减少复杂背景的影响, 本研究将骨干网络中的 3×3 卷积核替换为卷积核组, 通过特征可视化分析可知, 卷积核组增强网络的特征提取能力; 为了减少茶尺蠖体积小和形态多变的影响, 本研究引入注意力模块, 通过试验结果分析可知, 注意力可以有效提取茶尺蠖的全局信息; 为了减小前景和背景类不平衡对检测结果的影响, 本研究将交叉熵损失函数替换为 Focal loss 函数, 试验结果表明, Focal loss 损失函数更进一步减少背景的影响。

2) 改进的 YOLOv5s 模型对茶尺蠖的检测可以达到 92.89% 的平均精度均值 (mean Average Precision, mAP), 高于经典的机器学习方法和常用的深度学习方法。

[参 考 文 献]

- [1] Li L, Li T, Jiang Y, et al. Alteration of local and systemic amino acids metabolism for the inducible defense in tea plant (*Camellia sinensis*) in response to leaf herbivory by *Ectropis oblique*[J]. Archives of Biochemistry and Biophysics, 2020, 683: 108301.
- [2] Wang Y N, Tang L, Hou Y, et al. Differential transcriptome analysis of leaves of tea plant (*Camellia sinensis*) provides comprehensive insights into the defense responses to *Ectropis oblique* attack using RNA-Seq[J]. Functional & Integrative Genomics, 2016, 16(4): 383-398.
- [3] Hu G, Wu H, Zhang Y, et al. A low shot learning method for tea leaf's disease identification[J]. Computers and Electronics in Agriculture, 2019, 163: 104852.
- [4] Hu G, Yang X, Zhang Y, et al. Identification of tea leaf diseases by using an improved deep convolutional neural network[J]. Sustainable Computing: Informatics and Systems, 2019, 24: 100353.
- [5] Kasinathan T, Singaraju D, Uyyala S R. Insect classification and detection in field crops using modern machine learning techniques[J]. Information Processing in Agriculture, 2021, 8(3): 446-457.
- [6] Ebrahimi M A, Khoshtaghaza M H, Minaei S, et al. Vision-based pest detection based on SVM classification method[J]. Computers and Electronics in Agriculture, 2017, 137: 52-58.
- [7] Qing Y A, Xian D, Liu Q, et al. Automated counting of rice planthoppers in paddy fields based on image processing[J]. Journal of Integrative Agriculture, 2014, 13(8): 1736-1745.
- [8] 潘春华, 肖德琴, 林探宇, 等. 基于 SVM 和区域生长结合算法的南方主要蔬菜害虫分类识别[J]. 农业工程学报, 2018, 34(8): 192-199.
Pan Chunhua, Xiao Deqin, Lin Tanyu, et al. Classification and recognition for major vegetable pests in Southern China using SVM and region growing algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(8): 192-199. (in Chinese with English abstract)
- [9] Long D, Yan H, Hu H, et al. Research on Image Location Technology of Crop Diseases and Pests Based on Haar-Adaboost[C]//2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS). Hunan: IEEE, 2019: 163-165.
- [10] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [11] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [12] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Hawaii: IEEE, 2017: 2980-2988.
- [13] 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. Columbus: IEEE, 2014: 580-587.
- [14] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [15] 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: 91-99.
- [16] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263-7271.
- [17] Redmon J, Farhadi A. YOLOv3: An incremental improvement[Z]. [2021-05-21], <https://arxiv.org/abs/1804.02767>.
- [18] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[Z]. <https://arxiv.org/abs/2004.10934>.
- [19] 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. Las Vegas: IEEE, 2016: 779-788.
- [20] Wang Q J, Zhang S Y, Dong S F, et al. Pest24: A large-scale very small object data set of agricultural pests for multi-target detection[J]. Computers and Electronics in Agriculture, 2020, 175: 105585.
- [21] Cheng X, Zhang Y, Chen Y, et al. Pest identification via deep residual learning in complex background[J]. Computers and Electronics in Agriculture, 2017, 141: 351-356.
- [22] Yue Y, Cheng X, Zhang D, et al. Deep recursive super resolution network with Laplacian Pyramid for better agricultural pest surveillance and detection[J]. Computers and Electronics in Agriculture, 2018, 150: 26-32.
- [23] Muppala C, Guruviah V. Detection of leaf folder and yellow stemborer moths in the paddy field using deep neural network with search and rescue optimization[J]. Information Processing in Agriculture, 2021, 8(2): 350-358.
- [24] Wang F, Wang R, Xie C, et al. Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition[J]. Computers and Electronics in Agriculture, 2020, 169: 105222.
- [25] Lyu Z, Jin H, Zhen T, et al. Small object recognition algorithm of grain pests based on SSD feature fusion[J]. IEEE Access, 2021, 9: 43202-43213.
- [26] 王林惠, 兰玉彬, 刘志壮, 等. 便携式柑橘虫害实时检测系统的研制与试验[J]. 农业工程学报, 2021, 37(9): 282-288.
Wang Linhui, Lan Yubin, Liu Zhizhuang, et al. Development and experiment of the portable real-time detection system for

- citrus pests[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2007, 23(6): 175-180. (in Chinese with English abstract)
- [27] 张博, 张苗辉, 陈运忠. 基于空间金字塔池化和深度卷积神经网络的作物害虫识别[J]. 农业工程学报, 2019, 35(19): 209-215.
- Zhang Bo, Zhang Miaohui, Chen Yunzhong, et al. Crop pest identification based on spatial pyramid pooling and deep convolution neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35 (19): 209-215. (in Chinese with English abstract)
- [28] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.
- [29] Shi R, Li T, Yamaguchi Y. An attribution-based pruning method for real-time mango detection with YOLO network[J]. Computers and Electronics in Agriculture, 2020, 169: 105214.
- [30] Wu D, Lv S, Jiang M, et al. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments[J]. Computers and Electronics in Agriculture, 2020, 178: 105742.
- [31] Suo R, Gao F, Zhou Z, et al. Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking[J]. Computers and Electronics in Agriculture, 2021, 182: 106052.
- [32] 王金鹏, 高凯, 姜洪喆, 等. 基于改进的轻量化卷积神经网络火龙果检测方法[J]. 农业工程学报, 2020, 36(20): 218-225.
- Wang Jinpeng, Gao Kai, Jiang Hongzhe, et al. Method for detecting dragon fruit based on improved lightweight convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(20): 218-225. (in Chinese with English abstract)
- [33] Li X, Wang W, Hu X, et al. Selective kernel networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 510-519.
- [34] Chen P, Li W L, Yao S J, et al. Recognition and counting of wheat mites in wheat fields by a three-step deep learning method[J]. Neurocomputing, 2021, 437: 21-30.
- [35] Hu G, Wang H, Zhang Y, et al. Detection and severity analysis of tea leaf blight based on deep learning[J]. Computers & Electrical Engineering, 2021, 90: 107023.

Detection of *Ectropis oblique* in complex background images using improved YOLOv5

Hu Gensheng, Wu Jitian, Bao Wenxia^{*}, Zeng Weihui

(National Engineering Research Center for Agro-ecological Big Data Analysis and Application, Anhui University, Hefei 230601, China)

Abstract: Diseases and pests have posed a great threat to the yield and quality of tea in recent years. Among them, the *Ectropis oblique* is one of the most common pests in tea growth. A traditional detection has normally used the appearance of the pests, such as the color, morphology, and texture. But, these are more sensitive to the environments, particularly to the complex background, where the pests appear. A rapid and accurate detection cannot be realized, because: 1) The training samples are taken in different scales, while the pest is normally small in size; 2) The pest with the changeable shape and color may be shielded to obscure during imaging; 3) The color and texture of the pest can be similar to the tree branches and dead leaves of tea. Therefore, it is very necessary to identify and recognize the pest in a complex background in tea production. In this study, a rapid and accurate detection was proposed for the *Ectropis oblique* in complex background images using the improved YOLOv5 deep learning. Definitely, the YOLOv5 was taken as the baseline network. A labeling operation was first used to manually label the pest samples in the training and validation images. The data was then enhanced using the flipping, and contrast enhancement, particularly that the Gaussian noise was added to prevent data from overfitting. Meanwhile, the contrast of the test image was adjusted to reduce the influence of complex backgrounds, such as the tea pole on the detection of the scorpion. A convolution kernel group was also used to enhance the feature extraction without increasing the computation load. Furthermore, an attention module was utilized to adaptively adjust the receptive field, thereby enhancing the feature representation, according to the size and shape of the *Ectropis oblique*. More importantly, a Focal Loss function was used to reduce the impact of class imbalances between foreground and background during detection. The experimental results show that the convolution kernel group was effectively reduced the interference of complex background to the detection of tea geometrid. The attention module also presented an excellent performance to reduce the missed detection, due to the varying sizes and shapes of targets. Specifically, the best detection was achieved for the images with a complex background, where 0.94 recall, 0.96 precision, and 92.89% mean average precision. The improved accuracy increased by 6.44 percentage points, compared with the original YOLOv5. Moreover, there were 17.18 percentage points higher than the SSD, 6.52 percentage points higher than the Faster-RCNN, and 4.78 percentage points higher than the YOLOv4, compared with the SSD, Faster-RCNN, and YOLOv4. Consequently, the improved YOLOv5 can be widely expected to realize the intelligent monitoring of *ectropis oblique* pests in the precise pesticide application for the higher yield and quality of tea.

Keywords: agriculture; algorithm; object detection; deep learning; convolution kernel group; attention module; *Ectropis oblique*