

曾晏林, 贺壹婷, 蔺 瑶, 等. 基于 BCE - YOLOv5 的苹果叶部病害检测方法[J]. 江苏农业科学, 2023, 51(15): 155 - 163.  
doi:10.15889/j.issn.1002-1302.2023.15.022

# 基于 BCE - YOLOv5 的苹果叶部病害检测方法

曾晏林, 贺壹婷, 蔺 瑶, 费加杰, 黎 强, 杨 毅

(云南农业大学大数据学院, 云南昆明 650201)

**摘要:**针对自然环境中, 人工目视解译苹果叶部病害耗时耗力、人为主观因素强的问题。本研究提出了一种融合自注意力机制和 Transformer 模块的目标检测算法——BCE - YOLOv5, 实现对自然环境下对苹果叶片病虫害的自动识别与检测。该算法首先使用 BotNet、ConvNeXt 模块分别替换 Backbone 网络和 Neck 网络的 CSP 结构, 增加自注意力机制对目标的特征提取能力。通过将改进的 CBAM 引入 YOLOv5 的特征融合网络之后, 使注意力机制对特征融合信息更加地关注。最后, 用  $\alpha$  - IoU 损失函数替换 IoU 损失函数, 使得网络在模型训练过程中收敛的更加稳定。BCE - YOLOv5 算法在传统算法 YOLOv5 基础上平均精准率均值提升了 2.9 百分点, 并且改进后的算法的模型大小和计算量较传统算法分别减小了 0.2 M 和 0.9 GFLOPs。平均精度均值比 YOLOv4s、YOLOv6s、YOLOx - s 和 YOLOv7 模型分别高 2.5、1.3、3.5、2.2 百分点。该方法能快速准确识别苹果叶部病害, 为苹果种植过程中提供智能化管理做参考。

**关键词:** 苹果; 叶片病害; 识别; 注意力机制; YOLOv5; BotNet; ConvNeXt; CBAM;  $\alpha$  - IoU

**中图分类号:** TP391.41      **文献标志码:** A      **文章编号:** 1002 - 1302(2023)15 - 0155 - 09

苹果种植在我国已有 2 000 多年的发展历程, 目前栽培面积正逐渐增加。在国家宏观经济政策的扶持下, 中国传统苹果行业正逐渐地向现代化方向发展, 我国已逐步发展为全球的苹果行业强国。但同时, 在苹果栽培行业高速发展的背后, 虫害防控问题始终是困惑果农的主要难点所在。

随着深度学习的崛起和计算机计算能力的增强, 采用智慧农业物联网系统对农作物病虫害进行智能诊断, 不仅省时省力, 而且科学精准, 从动态采集农作物的病虫害信息, 到精准识别, 提出有针对性的防治决策, 全过程基本实现自动化、科学化, 降低农业成本。

近年来, 国内外学者提出许多基于传统机器学习算法来解决针对苹果叶部病害的识别。刘斌等研究了在移动端建立苹果叶部病虫害轻量级识别模型<sup>[1]</sup>; 龙满生等利用改进的 Alex Net 网络对棉花叶部病虫害进行识别<sup>[2]</sup>; 于洪涛等提出的 VGG - F 采用学习率的动态调整实现苹果病害的识别, 该方法在复杂情况下识别准确率较差<sup>[3]</sup>; 王云露等基于

Faster R - CNN 对苹果叶部病害进行识别, 其算法时间效率不高<sup>[4]</sup>; 田靓靓等利用 SSD 网络对苹果叶片病害进行检测<sup>[5]</sup>; Di 等建立了基于改进卷积神经网络的苹果叶部病害检测方法<sup>[6]</sup>; Tian 等基于 YOLOv3 深度学习方法对果园苹果病斑进行检测<sup>[7-8]</sup>; Li 等利用 YOLOv5 对苹果早期叶部病害在移动终端进行识别<sup>[9]</sup>; Wang 等通过改进 YOLOv5 对苹果病害的精度和检测速度来提高识别率<sup>[10-12]</sup>; 晁晓菲等基于轻量级改进的 YOLO 对苹果叶部病害进行检测<sup>[13-14]</sup>; 在轻量级检测上 Polder 等在深度学习上利用智能相机进行苹果和葡萄的早期病害检测<sup>[15]</sup>; Liu 等基于深度卷积神经网络对苹果叶部病害进行识别<sup>[16]</sup>。虽然上述算法对苹果病害的检测过程做了一定优化, 但是在检测能力方面仍旧存在不足, 例如检测速度较慢、检测条件要求较高等问题。在基准模型 YOLOv5 的基础上, 本研究提出一种基于 YOLOv5 的目标检测算法——BCE - YOLOv5。针对苹果叶片的病害检测任务, 为了提高对苹果叶片病害检测精度和检测速率, 以当前发现较严重的病害(斑叶病、灰斑病、铁锈病)为研究对象。

## 1 YOLOv5 算法简述

YOLOv5 结构由 4 个部分组成, 分别为输入端、骨干网络部分、多尺度特征融合部分、预测端, 其网络结构如图 1 所示。

收稿日期: 2023 - 01 - 04

基金项目: 云南省重大科技专项(编号: 202002AE09001002)。

作者简介: 曾晏林(1993—), 男, 四川广安人, 硕士研究生, 主要从事计算机视觉方面研究。E - mail: 786823791@qq.com。

通信作者: 杨 毅, 教授, 主要从事农业信息化方面研究。E - mail: yyang66@126.com。

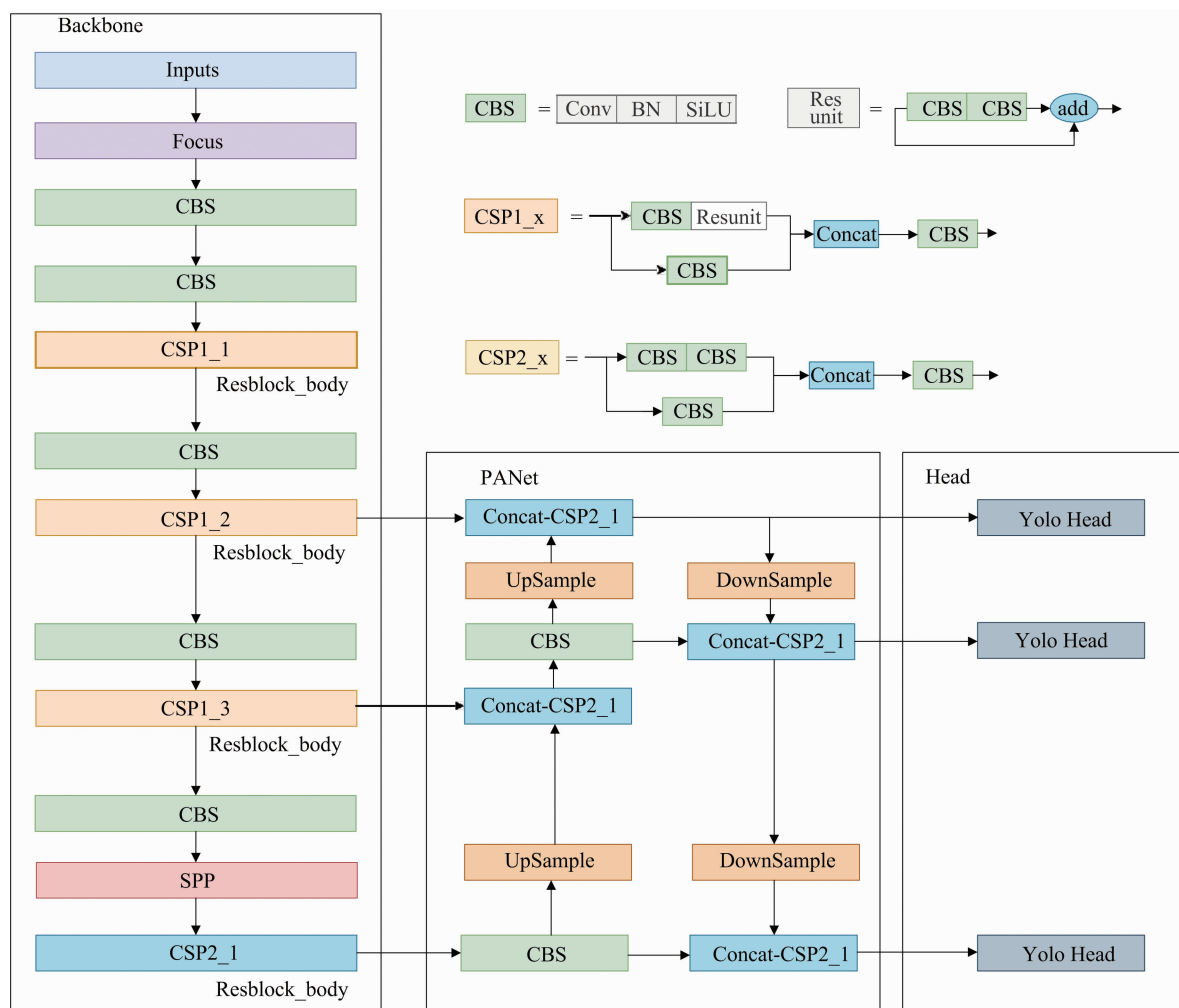


图1 YOLOv5 网络结构图

在输入端, YOLOv5 可以通过 Mosaic、自适应锚框算法、自适应图像压缩等信号增强手段, 对输入信号进行预处理。Mosaic 数据增强技术通过对多张数据以随机压缩、随机编辑、随机排布等的方法重新拼接到一起, 使之进一步丰富了数据的样本, 同时提高了系统的泛化性能。YOLOv5 默认使用 3 个锚框对应检测不同大小的目, 其锚框的参数范围为  $[10, 13, 16, 30, 33, 23]$   $[30, 61, 62, 45, 59, 119]$   $[116, 90, 156, 198, 373, 326]$  分别对应尺度为  $80 \times 80$ 、 $40 \times 40$ 、 $20 \times 20$  的特征图, 自适应锚框是预测框值与实际框值之差, 并利用数据迭代方法获得最佳的锚框值。而自适应图像压缩则是将数据集中在各个尺寸的特征图像, 统一缩至相应的尺寸。

骨干网络 (Backbone) 由切片结构 Focus、卷积模块 CBS、瓶颈层 CSP (cross stage partial network) 模块以及空间金字塔池化 (spatial pyramid pooling, SPP) 组成。Focus 模块采用隔列切片操作, 把高分

分辨率特征图拆分成多个低分辨率的特征图, 然后再拼接在一起。CBS 基本模块由 Conv2d + Batch Normalization + SiLu 组成。CSP 残差结构可以增加深层网络中层间反向传播的梯度值, 避免了深层网络导致的梯度消失, 加快了推理速度。SPP 模块采用一个卷积层和 3 个多尺度最大池化进行多尺度融合。

多尺度特征融合 (Neck) 主要使用自上而下增强语义特征提取的特征金字塔网络 (feature pyramid networks, FPN) 结构, 使用自底向上增强定位特征的路径聚合网络 (path aggregation network, PAN) 结构, 实现不同尺寸目标特征信息的融合, 解决了多尺度问题。

预测端头部网络 (Head) 采用 3 种损失函数分别计算分类、定位和置信度损失, 通过非极大抑制 (non maximum suppression, NMS) 去除检测框重叠并提高多目标的检测能力。

## 2 YOLOv5 算法改进

### 2.1 改进策略

植物叶片病虫害检测任务中容易存在背景干扰,待检测病斑被遮挡、尺寸小等问题。本研究在 YOLOv5 模型的基础上提出 4 点改进,首先改进骨干网络 Backbone 部分,使用改进后的 SPPF 替换原有的 SPP 层,并用 BoTNet<sup>[17]</sup> 结构替换相应的 CSP 结构,增加有效信息与位置感知关联。再引入改进的卷积块注意力模块 (convolution block attention

module, CBAM) 提高了神经网络识别特征区域的能力,并选择有效的位置添加到 YOLOv5 网络中,使模型更加精准地定位和识别感兴趣的目标特征。再次用 ConvNeXt<sup>[18-19]</sup> 结构对 Neck 中的特征金字塔部分的 CSPLayer 层进行改进,增加对语义特征的提取能力,提高特征提取的精准度;最后在预测部分改进损失函数,用  $\alpha$ -IoU 替换现有的 IoU 损失计算方式。本研究改进的网络模型 BCE-YOLOv5 的结构如图 2 所示。

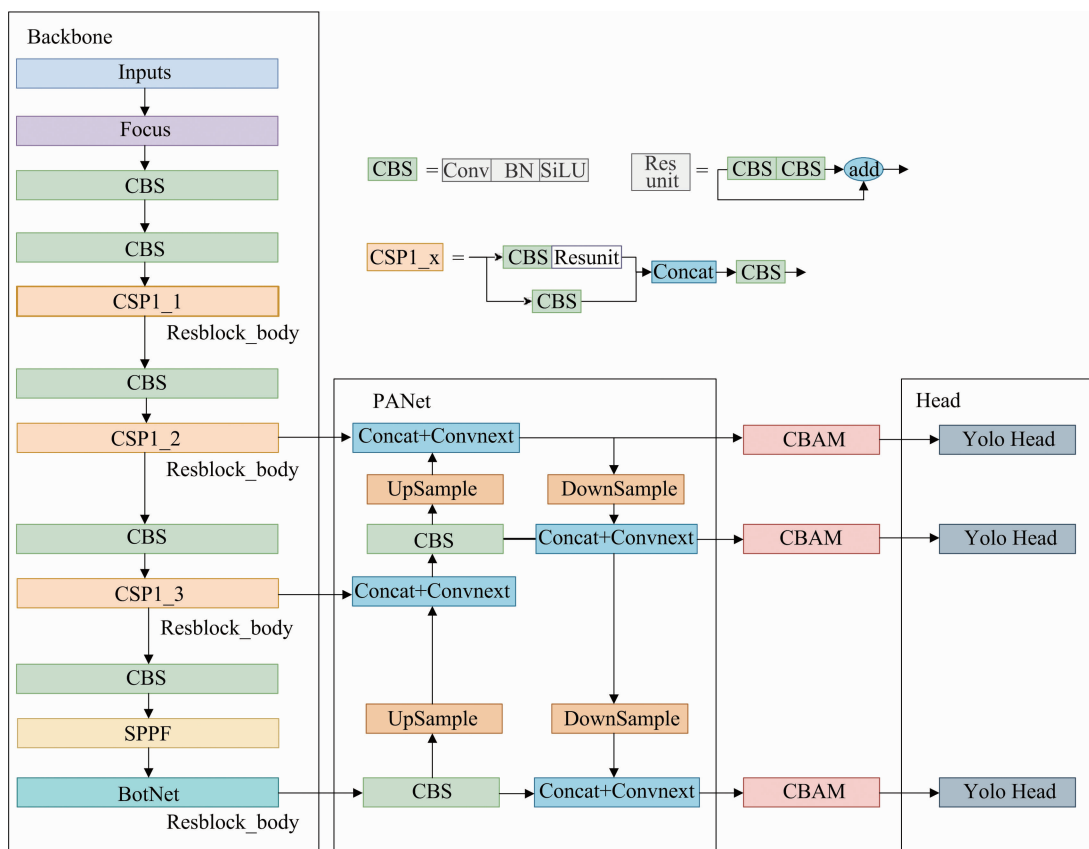


图2 BCE-YOLOv5 总体框架图

### 2.2 SPP 结构替换

空间金字塔池化能将任意大小的图片转换成网络需要的固定大小特征向量,减少对候选区域的重复卷积计算,减少计算冗余。SPP 的结构首先经过一个卷积再分别经过 3 个不同大小的最大池化后再做拼接,最后再经过一个卷积,这能够在一定程度上解决目标多尺度的问题。其结构如图 3 所示。

通过把 SPP 换成了 SPPF,其作用是一样的,但后者效率更高。SPP 需要指定 3 个最大池化 (maxpool),池化核大小分别为  $5 \times 5$ 、 $9 \times 9$ 、 $13 \times 13$ ,而 SPPF 只需要指定 3 个大小为  $5 \times 5$  的 maxpool。

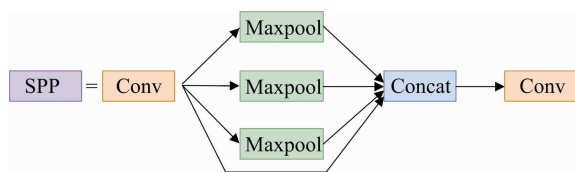


图3 SPP 结构

在结构上串行 2 个  $5 \times 5$  大小的 maxpool 层和一个  $9 \times 9$  大小的 maxpool 层计算结果是一样的,同样的串行 3 个 maxpool 层和一个  $13 \times 13$  大小的 maxpool 层计算结果是一样的。但是 SPPF 在计算量上要比 SPP 小,并且效率上比 SPP 要更快。其结构如图 4 所示。

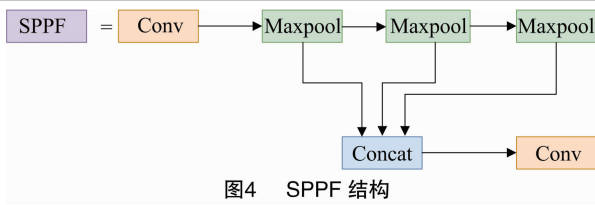


图4 SPPF 结构

### 2.3 BoTNet 结构

BotNet 结构是一种基于 Transformer 的网络结构,它利用了卷积和自注意力的混合网络,以提升模型的性能,BotNet 如图 5 所示。通过结合 CNN + Transformer<sup>[20-21]</sup> 的组合方式提出一种 Bottleneck Transformer 来代替 ResNet Bottleneck,即仅在 ResNet 框架的最后 3 个 bottleneck blocks 中使用全局多头自注意力(multi-head selfattention, MHSA)替换  $3 \times 3$  空间卷积。

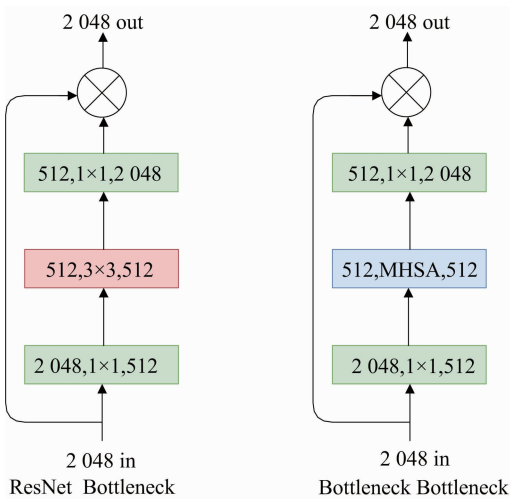


图5 ResNet 和 BotNet 网络结构

MHSA 层利用全局注意力对特征图中的信息进行捕获,增加全局特征之间的联系,可以有效地处理大图像,通过卷积进行空间下采样,让注意力能在小分辨率图像上获取特征信息。在 MHSA 模块中对  $Q$  (查询向量)、 $K$  (键向量)、 $V$  (值向量) 进行自注意力变换,对过程重复  $h$  次( $h$  为 MHSA 的头数)的结果进行拼接,最后进行线性变换得到 MHSA 模块的最终输出。

在 YOLOv5 的整体改进中,在最后一个阶段中的 3 个 Bottleneck 使用了 MHSA 模块,因为自注意力机制在训练过程中需要消耗巨大的计算量,而在 Backbone 的最后 CSP 模块加入的时候特征图的大小较小,从而可以使得整个网络的计算量减少。

### 2.4 注意力机制

在输入的病虫害叶片图像中,除了叶片的病

斑,经常还伴有噪音,随着卷积的深度增加会产生大量的冗余信息,从而丢失部分有用信息,导致目标检测精度降低。为此,本研究提出了一种改进的 CBAM<sup>[22]</sup>,并选择有效的网络位置添加到 YOLOv5 网络中,这使得网络能够更准确地定位和识别感兴趣的区域。

CBAM 是一种即插即用并且使用方便的注意力机制,其由通道域和空间域组成。通道域关注不同通道上特征点的权重,利用权重对不同通道进行乘法运算,以增强网络对关键通道域的注意力。首先,特征图进入通道域,然后在宽度和高度上进行全局最大池化(global max pooling, GAP)和全局平均池化(global average pooling, GMP),并使用多层感知器(multi-layer perception, MLP)计算不同通道的注意力权重,接下来使用 Sigmoid 函数对这些权重进行归一化处理,最后通过每个通道的权重乘以原始输入特征图,以获得更准确的结果。

具体公式如下所示。

$$M_c(F) = \sigma \{ MLP [ AvgPool(F) ] + MLP [ MaxPool(F) ] \} = \sigma \{ W_1 [ W_0(F_{avg}^c) ] + W_1 [ W_0(F_{max}^c) ] \} \quad (1)$$

在公式(1)中,  $F_{avg}^c$  和  $F_{max}^c$  分别表示平均池化和最大池化后的特征图,  $\sigma$  是 Sigmoid 激活函数,  $W_1$  和  $W_2$  表示 2 层全连接层的权重。

根据公式(1)得到的感兴趣的通道特征  $M_c$  与原始图像特征相乘,并送至空间域模块。空间域模块关注的是目标特征在图像上的位置信息,并将空间特征做对应的变换,从而将关键信息提取出来。通道注意特征  $M_c$  与原图像特征  $F$  相乘,得到一个新的特征图  $F_s$ ,公式如下所示。

$$F_s = M_c \otimes F \quad (2)$$

空间域经过全局最大池化和全局平均池化,得到新的特征图  $F_s$  的宽度和高度,使特征维度从  $H \times W$  转化为  $1 \times 1$ ,然后通过卷积核进行  $7 \times 7$  卷积,再通过 Sigmoid 激活函数进行归一化,最后与通道注意力输出特征图进行合并,公式如下所示。

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

在公式(3)中,  $f^{7 \times 7}$  表示  $7 \times 7$  的卷积层。

通过公式(2)的  $F_s$  和公式(3)的  $M_s$  做乘积运算,再与原图像特征图  $F$  相加,经过 ReLU 激活函数,得到特征图  $M_F$ ,公式如下所示。

$$M_F(F) = \sigma [ (M_s \otimes F_s) \oplus F ] \quad (4)$$

式中:⊕表示特征合并操作;⊗表示逐元素相乘;σ表示激活函数。

上述 CBAM 中的通道域采用全连接层,压缩通道特征的空间维度以降低模型的复杂性,通道特征被映射到一个较低维度的空间并再次映射。在降维中不可避免地导致一些通道域的感兴趣区域信息丢失,对于复杂的特征图,不可能获得完整的特征信息。因此,为了减少上述通道域感兴趣信息的丢失,以及提高病变识别网络检测的准确性,本研究引入改进的 CBAM,其中原通道域使用改进通道注意力网络 (efficient channel attention network, ECA-Net) 进行替换,保留原空间域。

ECA-Net<sup>[23-24]</sup> 模块使用一种不降维的局部跨通道交互策略,有效避免了降维对于通道注意力学习效果的影响。对每个特征图求均值,从一个  $H \times W \times C$  变成  $1 \times 1 \times C$ ,通过大小为  $K$  的一维卷积来有效实现,其中卷积核大小为  $K$  代表了局部跨信道交互的覆盖率。通过公式调整  $K$  值大小,公式如下所示。

$$K = \varphi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (5)$$

在公式(5)中: $C$ 表示通道数; $b$ 设置为1; $\gamma$ 设置为2; $\lfloor t \rfloor_{\text{odd}}$ 表示 $t$ 的最接近奇数。

本研究通过把 ECA-Net 模块里面原图像特征图与经过 Sigmoid 函数处理的特征图的乘积处理去掉,得到新的 ECA-Net 应用于 CBAM 中的通道域,改进后的 CBAM<sup>[25]</sup> 结构图如图 6 所示。

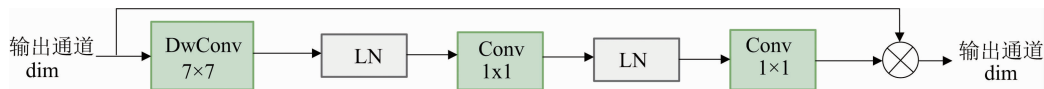


图6 改进后 CBAM 的结构示意图

## 2.5 ConvNeXt 结构

ConvNeXt 结构于 2022 年被提出,该网络结合了近年来卷积神经网络方面提出的各项主要改进,并将其融合<sup>[18]</sup>。ConvNeXt 网络结构如图 7 所示,卷积块的输入数据为通道数维度 (dim) 的特征图。首先经过卷积块大小为  $7 \times 7$  的深度卷积层,随后通过 LN 层进行归一化后再经过卷积块大小为  $1 \times 1$  的卷积层,得到通道数变为原输入通道的 4 倍。再经过 GELU 激活函数赋予非线性,在此之后经过卷积块大小为  $1 \times 1$  输出通道变回原来的维度。最后输入数据和输出数据通过一个残差连接进行拼接,得到最终的结果。

图7 ConvNeXt 网络结构

在深度卷积神经网络中,图像的语义信息得到了增强,但随着图像的不断卷积,详细的特征被减少。相反,浅层的图像细节表现能力强,但是语义信息表达较弱。YOLOv5 采用的特征金字塔结构来进行特征融合,FPN 结构是自顶向下传递高层图像的语义信息,但是在传递过程中会丢失位置细节信息。为使网络更加适合不同大小的病变检测,在网络 Neck 部分原有的 CSP 结构上进行改进,替换 CSPLayer 层并使用 ConvNeXt 层。

## 2.6 改进损失函数

传统的 YOLOv5 采用  $\text{Loss}_{\text{GIOU}}$  作为回归损失的损失函数,当实际框与预测框水平或垂直重合或 2

个框为包含关系时,此时  $\text{GIOU}$  值相同,无法区分相对位置关系,从而  $\text{Loss}_{\text{GIOU}}$  退化成  $\text{Loss}_{\text{IoU}}$ ,而  $\text{Loss}_{\text{IoU}}$  不能解决实际框与预测框不重合的问题。

为了解决这个问题,本研究引入  $\text{Loss}_{\alpha-\text{IoU}}$  用于改进原损失函数<sup>[26]</sup>, $\text{Loss}_{\alpha-\text{IoU}}$  在保留原损失函数的性质时,同时可以更好地关注目标,提高了检测框回归的准确性。 $\text{Loss}_{\alpha-\text{IoU}}$  在原有的  $\text{Loss}_{\text{IoU}}$  上加上一个幂指数,以  $\alpha$  为幂,其公式如下所示。

$$\text{Loss}_{\alpha-\text{IoU}} = \frac{1 - \text{IoU}^\alpha}{\alpha}, \alpha > 0. \quad (6)$$

在不增加额外的参数和训练时间下,通过对  $\alpha - \text{IoU}$  中的参数  $\alpha$  进行调制,可以很容易地扩展到



更复杂的具有多个 IoU 或惩罚项的损失函数。

本研究采用效果较好的  $\text{Loss}_{\text{CloU}}$  改进为  $\text{Loss}_{\alpha-\text{CloU}}$ , 其公式如下所示。

$$\begin{aligned} \text{Loss}_{\alpha-\text{CloU}} &= 1 - \text{IoU} + \frac{\rho^{2\alpha}(b, b^{gt})}{c^{2\alpha}} + (\beta v)^\alpha; \\ v &= \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2; \\ \beta &= \frac{v}{(1 - \text{IoU}) + v} \end{aligned} \quad (7)$$

公式(7)中:  $b$ 、 $b^{gt}$  分别表示预测框和真实框的中心线距离;  $\rho^{2\alpha}$  代表的是在 2 个中心点之间的欧式间隔;  $c$  是能容纳预测框和真实框的最小封闭区间

的对角线长度;  $\frac{w^{gt}}{h^{gt}}$  表示目标框的宽高比;  $\frac{w}{h}$  表示预测框各自的宽高比。

### 3 试验与分析

#### 3.1 试验数据集

苹果叶部病虫害收集了云南省昭通市邵阳区守望乡苹果种植和百度飞浆公开数据集共 5 106 张图片, 共分为 3 类常见病害: 斑点落叶病 (alternaria boltch)、灰斑病 (grey\_spot)、锈病 (rust) 作为试验数据集, 图 8 是部分数据集图像, 其中有较大的病斑图像, 有较小的病斑, 还有阴影的病斑。

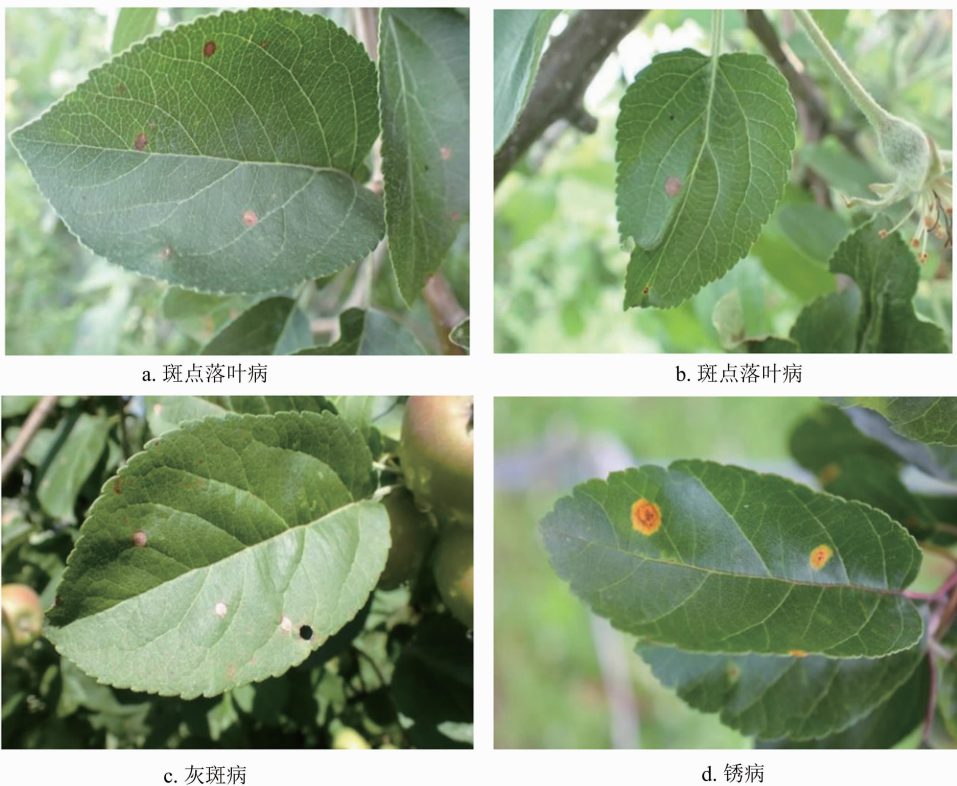


图8 苹果叶部病虫害数据集部分图像

#### 3.2 预训练

本研究训练使用的操作系统为 Windows 10, CPU 型号为 AMD Ryzen7 5800H, GPU 型号为 RTX 3060, 深度学习框架为 Pytorch 1. 10. 0 以及 CUDA 11.4 版本框架。由于试验数据的样本数量不多, 因此在训练过程中采用迁移学习策略来解决数据匮乏的问题, 利用 YOLOv5 官方提供的预训练权重作为训练初始化参数。

#### 3.3 评价标准

在目标检测领域, 评价网络性能的指标是平均精准率均值 (mean average precision, mAP)、模型大

小以及计算量 (GFLOPS)。mAP 评价指标是对精准率 (precision, P)、召回率 (recall, R) 的综合考量, 可以评估模型的有效性。

精准率是指正确被预测为某类病斑目标占全部预测为该正确病斑目标的比例, 召回率是指正确被预测为某类病斑目标占实际上全部为该病斑目标的比例。

模型大小能够表明模型对硬件内存的需求, 对于外界复杂环境且硬件设备受限的情况下则需要轻量级模型; 计算量反映了模型对硬件设备的计算单元的要求, 同样对于硬件设备受限的情况下, 计

算量越小越好。

### 3.4 结果与对比

为了直观地体现改进后的 YOLOv5 算法的有效性,如图 9 所示,通过对比 4 组图片,分别使用传统的 YOLOv5s 和本研究改进的 BCE-YOLOv5 算法对测试图片进行检测,从对比图可以看出传统的 YOLOv5 检测的精度普遍较低,而改进后的 YOLOv5 检测精度有所提升。通过第 3 组对比图片可以看出,传统的 YOLOv5 存在漏检,而改进后的算法对叶片灰斑病细节更加注重,改善了对背景模糊的漏检。通过第 4 组对比图片可知,在大图片小目标叶片背景下病斑难以区分的情况下,传统的 YOLOv5 存在错检,而改进后的算法具有较好的识别能力。



a. YOLOv5 检测效果

b. BCE-YOLOv5 检测效果

图9 YOLOv5 改进前后结果对比

### 3.5 消融试验

为了检验本研究中改进后的 CBAM 是否对训练有效,试验在 YOLOv5 颈部网络 Neck 后分别引入原始 ECA 注意力模块、原始 CBAM 以及改进后的 CBAM,本研究将改进的 CBAM 表示为 ECBAM。其试验对比结果如表 1 所示,在 YOLOv5 传统算法的相同位置上引入注意力机制其 mAP 均有提升,其中

ECA 和 CBAM 注意力模块分别贡献了 0.2 个百分点和 1.6 个百分点的提升,而 ECBAM 提升了 1.8 百分点。这表明改进后的 ECBAM 在通道域上不压缩感兴趣特征,而是对通道域进行全局池化然后再进行一维卷积,从而有效避免降维带来的通道特征丢失,提高 ECBAM 在聚合网络中的整体能力。结果显示,改进后的 ECBAM 能够有效提升网络性能。

表 1 不同注意力机制在 YOLOv5 中的对比

算法	AP(%)			mAP (%)
	斑点落叶病	灰斑病	锈病	
YOLOv5	84.6	87.5	95.6	89.3
YOLOv5-ECA	82.9	88.1	97.4	89.5
YOLOv5-CBAM	86.5	88.3	98.0	90.9
YOLOv5-ECBAM	86.4	88.7	98.4	91.1

为了验证本研究提出的改进模型的有效性,在训练参数和数据集相同的情况下对改进的 YOLOv5 网络进行对比试验,结果如表 2 所示,表中 Backbone 改进、Neck 改进、CBAM 改进分别表示在传统 YOLOv5 网络中加入改进方案。由表 2 可以分析得到,首先采用改进的 Backbone 后为模型提升了 1.4 百分点,这使得模型检测的准确性更高,并在一定程度上缓解了由于背景复杂的模糊图像造成的漏检或误检问题;其次,加入 CBAM 后,模型比基准 mAP 提升了 2.5 百分点,说明添加 CBAM 模块可以使网络更加注意病斑,使得整个模型的性能提升;再在原 CBAM 基础之上加以改进,得到新的注意力机制 ECBAM,在模型中加入 ECBAM 后获得比原 CBAM 提升了 0.4 百分点,由此可以看出改进后的注意力机制较原注意力机制更能注重小目标;使用上述改进方法最终为模型的 mAP 比基线模型性能提高了 2.9 百分点的增益,证明了 3 种改进方法的有效性。

同时,图 10 清楚地显示了 YOLOv5 的学习过程中损失曲线在改进前后的变化。其中  $\alpha$ -CIoU 损失曲线在 GIoU 损失曲线的下方,说明  $\alpha$ -CIoU 损失值更低,同时在出现拐点之后  $\alpha$ -CIoU 损失值在 10 个 Epoch 后趋于平稳,而 GIoU 损失曲线在 85 个 Epoch 后和  $\alpha$ -CIoU 曲线基本重合,说明  $\alpha$ -CIoU 损失在整个训练过程中网络可以更快、更平稳地收敛。

### 3.6 对比试验

对比本研究 BCE-YOLOv5 目标检测算法相比其他算法的优越性,本研究与各种先进的目标检测算法

表 2 改进策略的消融试验

骨干网络改进	颈部网络改进	CBAM	ECBAM	AP( % )				涨点 (百分点)
				斑点落叶病	灰斑病	锈病	mAP	
-	-	-	-	84.6	87.5	95.6	89.2	
✓	-	-	-	84.6	89.9	97.4	90.6	1.4
✓	✓	-	-	86.9	88.7	98.4	91.3	2.1
✓	✓	✓	-	85.6	91.1	98.6	91.7	2.5
✓	✓	-	✓	88.0	89.5	98.7	92.1	2.9

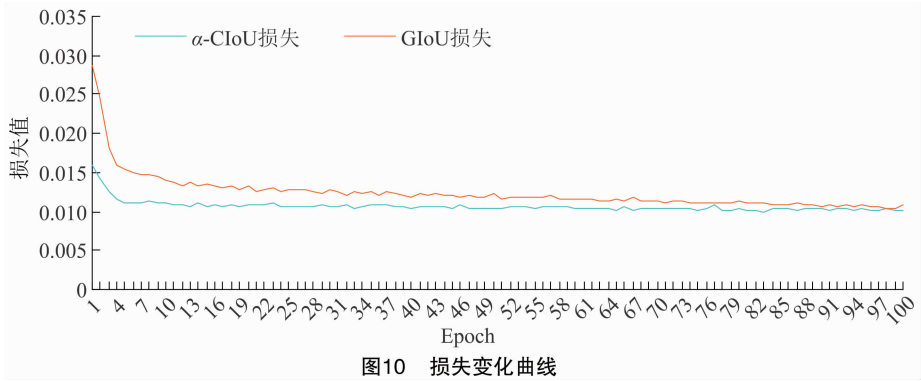


图10 损失变化曲线

(YOLOv4s<sup>[27]</sup>、YOLOxs<sup>[28]</sup>、YOLOv6s<sup>[29]</sup>、YOLOv7s<sup>[30]</sup>) 进行对比试验。由表 3 可以看出,改进后的 YOLOv5s 算法在检测精度上达到 92.1%, 优于其他算法,比 YOLOv4s、YOLOxs、YOLOv6s 和 YOLOv7s 分别高出 2.5、3.5、1.3、2.2 百分点;BCE – YOLOv5 模型大小分别比其低 230.5、58.3、24.6、61.3 M;在计算量上 BCE – YOLOv5 相比 YOLOv5s、YOLOxs、YOLOv6s 和 YOLOv7s 算法分别低 0.9、29.0、11.7、88.3 GFLOPs。本研究改进的 BCE – YOLOv5 算法在模型大小、检测精度及计算量上均优于其他类算法,证明其改进算法的有效性。说明本研究的 BCE – YOLOv5 算法适合后续在边缘设备部署。

表 3 各种算法对比

算法	模型大小 (Mb)	检测精度 (%)	计算量 (GFLOPs)
YOLOv4s	244.0	89.6	
YOLOv5s	13.7	89.2	16.0
YOLOv6s	38.1	90.8	44.1
YOLOxs	71.8	88.6	26.8
YOLOv7s	74.8	89.9	103.4
BCE – YOLOv5	13.5	92.1	15.1

4 结论

针对苹果叶部病虫害检测的识别任务,本研究

提出一种基于 YOLOv5 改进模型的苹果叶检测方法。首先将 BotNet 和 ConvNeXt 分别与原始 YOLOv5 的骨干网络和特征融合网络融合,利用 Transformer 的上下文理解能力,提高网络对图像特征的识别;其次使用 CBAM 注意力机制引入到 YOLOv5 多尺度特征融合模块之后,能够有效的提升对目标的检测精度;最后使用  $\alpha$  – IoU 损失函数代替原来的损失函数,使网络可以更快、更平稳地收敛,并改善了复杂图像的漏检和小目标识别效果差的问题。在消融试验对比中,本研究改进的 BCE – YOLOv5 与原 YOLOv5 相比,其 mAP 提高了 2.9 百分点,具有较高的精准率。本研究改进的 BCE – YOLOv5 网络模型具有检测速度快、检测准确率高、计算量和模型小等特点。

为了达到推广应用的效果,今后还需开展以下研究:(1)加大对苹果叶部病害分类的研究;(2)进一步扩充数据,增加病害数据以及种类;(3)进一步优化模型,提高模型的训练效果在复杂背景下的识别水平;(4)研究在不同阶段对苹果叶部病害的识别能力,从而提高模型的泛化能力。

参考文献:

[1] 刘 斌,贾润昌,朱先语,等. 面向移动端的苹果叶部病虫害轻量级识别模型[J]. 农业工程学报,2022,38(6):130 – 139.



- [2] 龙满生, 欧阳春娟, 刘欢, 等. 基于卷积神经网络与迁移学习的油茶病害图像识别[J]. 农业工程学报, 2018, 34(18): 194–201.
- [3] 于洪涛, 袁明新, 王琪, 等. 基于 VGG-F 动态学习模型的苹果病虫害识别[J]. 科学技术与工程, 2019, 19(32): 249–253.
- [4] 王云露, 吴杰芳, 兰鹏, 等. 基于改进 Faster R-CNN 的苹果叶部病害识别方法[J]. 林业工程学报, 2022, 7(1): 153–159.
- [5] 田靓靓. 基于 SSD 网络的苹果叶片病害检测方法研究[D]. 杨凌: 西北农林科技大学, 2022.
- [6] Di J, Li Q. A method of detecting apple leaf diseases based on improved convolutional neural network[J]. PLoS One, 2022, 17(2): e0262629.
- [7] Tian Y, Yang G, Wang Z, et al. Detection of apple lesions in orchards based on deep learning methods of cyclegan and yolov3-dense[J]. Journal of Sensors, 2019, 2019.
- [8] Mathew M P, Mahesh T Y. Determining the region of apple leaf affected by disease using YOLO V3 [C]//2021 International Conference on Communication, Control and Information Sciences (ICCISc). IduKKi, India, 2021: 1–4.
- [9] Li J, Zhu X, Jia R, et al. Apple-YOLO: a novel mobile terminal detector based on YOLOv5 for early apple leaf diseases [C]//2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC). Los Alamitos, CA, USA, 2022: 352–361.
- [10] Wang Y L, Sun F G, Wang Z J, et al. Apple leaf disease identification method based on improved YoloV5 [M]//Sun J D, Wang Y, Huo M Y, et al. Signal and information processing, networking and computers. Singapore: Springer, 2023: 1246–1252.
- [11] Kuznetsova A, Maleva T, Soloviev V. YOLOv5 versus YOLOv3 for apple detection [M]//Kravets A G, Bolshakov A A, Shcherbakov M. Cyber-physical systems: modelling and intelligent control. Cham: Springer, 2021: 349–358.
- [12] Chen Z, Wu R, Lin Y, et al. Plant disease recognition model based on improved YOLOv5[J]. Agronomy, 2022, 12(2): 365.
- [13] 晁晓菲, 池敬柯, 张继伟, 等. 基于 PSA-YOLO 网络的苹果叶片病斑检测[J]. 农业机械学报, 2022, 53(8): 329–336.
- [14] 邸洁, 曲建华. 基于 Tiny-YOLO 的苹果叶部病害检测[J]. 山东师范大学学报(自然科学版), 2020, 35(1): 78–83.
- [15] Polder G, Blok P M, van Daalen J T, et al. Early disease detection in apple and grape using deep learning on a smart-camera [C]//Proceedings of the European Conference on Agricultural Engineering. EurAgEng, Cranfield, 2021: 51–56.
- [16] Liu B, Zhang Y, He D J, et al. Identification of apple leaf diseases based on deep convolutional neural networks[J]. Symmetry, 2017, 10(1): 11.
- [17] Srinivas A, Lin T Y, Parmar N, et al. Bottleneck transformers for visual recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Kuala Lumpur, 2021: 16519–16529.
- [18] Liu Z, Mao H Z, Wu C Y, et al. A convnet for the 2020s [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 11976–11986.
- [19] 黄文博, 黄钰翔, 姚远, 等. 融合注意力的 ConvNeXt 视网膜病变自动分级[J]. 光学精密工程, 2022, 30(17): 2147–2154.
- [20] Zhang Z, Lu X, Cao G, et al. ViT-YOLO: Transformer-based YOLO for object detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2799–2808.
- [21] Liu Y, He G, Wang Z, et al. NRT-YOLO: Improved YOLOv5 based on nested residual transformer for tiny remote sensing object detection[J]. Sensors, 2022, 22(13): 4953.
- [22] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, 2018: 3–19.
- [23] Wang Q L, Wu B G, Zhu P F, et al. Eca-net: Efficient channel attention for deep convolutional neural networks [C]//2020 IEEE Conference on Computer Vision and Pattern Recognition. Seattle, 2020.
- [24] Wang Q, Wu B, Zhu P, et al. Supplementary material for ‘ECA-Net: efficient channel attention for deep convolutional neural networks [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA. 2020: 13–19.
- [25] Li X, Xia H, Lu L. ECA-CBAM: Classification of diabetic retinopathy: classification of diabetic retinopathy by cross-combined attention mechanism [C]//2022 the 6th International Conference on Innovation in Artificial Intelligence (ICIAI). 2022: 78–82.
- [26] He J B, Erfani S M, Ma X J, et al. Alpha-IoU: a family of power intersection over union losses for bounding box regression [J]. Advances in Neural Information Processing Systems, 2021, 34: 20230–20242.
- [27] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection [EB/OL]. [2022-12-12]. <https://arxiv.org/abs/2004.10934>.
- [28] Ge Z, Liu S T, Wang F, et al. Yolox: Exceeding yolo series in 2021 [EB/OL]. [2022-12-12]. <https://arxiv.org/abs/2107.08430>.
- [29] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications [EB/OL]. [2022-12-12]. <https://arxiv.org/abs/2209.02976>.
- [30] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [EB/OL]. [2022-12-12]. <http://arXiv.ORG/ABS/2207.02696>.