

多尺度自注意力特征融合的茶叶病害检测方法

孙艳歌¹ 吴飞¹ 姚建峰¹ 周棋赢² 沈剑波³

(1. 信阳师范大学计算机与信息技术学院, 信阳 464000; 2. 河南省茶树生物学重点实验室, 信阳 464000;
3. 北京市农林科学院智能装备技术研究中心, 北京 100097)

摘要: 针对茶叶病害检测面临的病害尺度多变、病害密集与遮挡等诸多问题,提出了一种多尺度自注意力茶叶病害检测方法(Multi-scale guided self-attention network, MSGSN)。该方法首先采用基于VGG16的多尺度特征提取模块,以获取茶叶病害图像在不同尺度下的局部细节特征,例如纹理和边缘等,从而有效表达多尺度的局部特征。其次,通过自注意力模块捕获茶叶图像中像素之间的全局依赖关系,实现病害图像全局信息与局部特征之间的有效交互。最后,采用通道注意力机制对多尺度特征进行加权融合,提升了模型对病害多尺度特征的表征能力,使其更加关注关键特征,从而提高了病害检测的准确性。实验结果表明,融合多尺度自注意力的茶叶病害检测方法在背景复杂、病害尺度多变等场景下具有更好的检测效果,平均精度均值达到92.15%。该方法可为茶叶病害的智能诊断提供参考依据。

关键词: 茶叶; 病害检测; 目标检测; 多尺度; 自注意力机制; 特征融合

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2023)12-0308-08

OSID:



Tea Disease Detection Method with Multi-scale Self-attention Feature Fusion

SUN Yange¹ WU Fei¹ YAO Jianfeng¹ ZHOU Qiyang² SHEN Jianbo³

(1. College of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China

2. Henan Key Laboratory of Tea Plant Biology, Xinyang 464000, China

3. Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China)

Abstract: Accurate detection of tea diseases is crucial for a high yield and quality of tea, thereby increasing production and minimizing economic losses. However, tea disease detection faces several challenges, such as variations in disease scales and densely occluded disease areas. To tackle these challenges, a novel method for detecting tea diseases called multi-scale guided self-attention network (MSGSN) was introduced, which incorporated multi-scale guided self-attention. The MSGSN method utilized a VGG16-based module for extracting multi-scale features to capture local details like texture and edges in tea disease images across multiple scales, effectively expressing the local multi-scale features. Subsequently, the self-attention module captured global dependencies among pixels in the tea leaf image, enabling effective interaction between global information and the disease image's local features. Finally, the channel attention mechanism was employed to weight, fuse, and prioritize the multi-scale features, thereby enhancing the model's ability to characterize the multi-scale features of the disease and improving disease detection accuracy. Experimental results demonstrated the MSGSN method's superior detection performance in complex backgrounds and varying disease scales, achieving an accuracy rate of 92.15%. This method served as a valuable reference for the intelligent diagnosis of tea diseases. In addition, the method can provide a scientific basis for the prevention and control of tea diseases and help farmers take timely and effective control measures. At the same time, the method can also provide technical support for the development of the tea industry.

Key words: tea; disease detection; object detection; multi-scale; self-attention mechanism; feature fusion

收稿日期: 2023-09-14 修回日期: 2023-10-14

基金项目: 河南省自然科学基金青年项目(222300420274)、河南省自然科学基金面上项目(232300421167)、河南省研究生课程思政示范课程项目(YJS2023SZ23)和信阳师范大学研究生科研创新基金项目(2021KYJJ56)

作者简介: 孙艳歌(1982—),女,副教授,博士,主要从事计算机视觉研究,E-mail: ygsun1982@126.com

0 引言

茶叶作为我国传统的经济作物之一,具有重要的经济和文化价值。然而,茶叶的生长过程容易受到病害的影响,直接影响茶叶产量和品质,严重降低了茶叶的经济价值^[1]。传统的人工检测方法需要耗费大量的人力和时间成本,检测结果还依赖于检测人员的专业知识,导致检测效率低、成本高^[2]。此外,茶园环境复杂,茶叶背景包括杂草、树枝和土壤等,加上病害尺度多变以及病害密集与遮挡等客观因素,进一步增加了精确检测的难度。因此,迫切需要研究一种快速、准确的方法,在茶叶病害发展早期阶段准确检测茶叶病害。这样的方法将有助于及时采取必要的控制措施,阻止病害蔓延,保护茶树健康,进而推动茶产业可持续发展。

传统机器学习模型如决策树、支持向量机(SVM)、K-means 聚类、随机森林等需要针对不同类型的病害人工设计特征提取方法,然后进行训练和检测。这个过程非常复杂且容易受到光线、拍摄角度和背景等环境因素的干扰。而且,模型泛化能力较弱,不能有效地应用于其他类型的检测任务^[3]。近年来,随着深度学习在目标检测领域的应用发展,该方法也被应用于茶叶病害检测^[4-7]。其中,卷积神经网络(Convolutional neural network, CNN)^[8]凭借其强大的特征提取能力,在茶叶病害检测中能够自动学习病害图像的形状、颜色、纹理等多层次、高阶特征。相比于传统方法需要人工设计特征的不足, CNN 在病害检测方面具有明显优势,并已取得了许多研究成果^[9-11]。LEE 等^[12]使用卷积层等基础网络提取特征以检测茶叶病害,其在对 4 种虫害和 3 种病害的检测中获得 77.5% 的准确率。另外,孙道宗等^[13]使用深度可分离卷积来检测自然环境下的 3 种常见茶叶病害,取得了较高的识别精度。谢军等^[14]利用卷积神经网络模型对 5 种茶叶病害进行检测,解决了人工观察茶叶病害耗时冗长的问题。HU 等^[15]则通过改进的深度卷积生成对抗网络(C-DCGAN)克服了样本不足的挑战。此外, CHEN 等^[16]提出了名为 LeafNet 的卷积神经网络模型,可自动从图像中提取茶叶病害的特征。XUE 等^[17]在 YOLO v5 基础上提出了一种茶叶病害检测模型 YOLO-Tea,旨在利用卷积神经网络自动提取茶叶病害特征,并进一步进行病害检测。然而,在处理复杂茶叶病害检测场景时,卷积神经网络生成的特征图虽然能有效表达局部特征,但无法捕捉到长距离像素间的全局关联信息^[18]。这限制了传统 CNN 方法对茶叶病害图像整体结构和全局特征的

准确理解,特别是在处理病害尺度多变和背景复杂的情况下,传统 CNN 方法可能无法准确地理解茶叶病害的整体情况。

与 CNN 相比,Transformer 模型通过自注意力机制赋予了特征提取器更强大的全局信息交互能力,使其能够快速建立全局感受野,从而实现更准确的场景理解^[19-20]。这一优势有效地弥补了 CNN 在处理茶叶病害检测时存在的不足。然而,在处理图像数据时,传统的 Transformer 模型通常将图像分割成一系列小块,并分别处理这些小块。这种方法使用全局自注意力,使得每个位置的特征能够与其他位置进行交互。然而,这种方式的计算复杂度高和内存开销大,无法捕捉到长距离的依赖关系^[19-20]。为了解决这个问题,CSwin Transformer 模型^[21]在 Transformer 基础上引入了交叉交叉窗口的概念,将注意力范围限定在一个窗口内,即交叉窗口自注意力。这种方法限制了注意力范围在一个窗口内,从而降低了计算复杂度,并能有效处理长距离的依赖关系。

基于上述分析,本文将卷积神经网络(CNN)与 CSwin Transformer 结合,提出一种融合多尺度自注意力茶叶病害检测网络(Multi-scale guided self-attention network, MSGSN),以期实现茶叶病害信息的多尺度特征融合,提高检测效果。

1 多尺度自注意力特征融合病害检测算法

1.1 网络整体结构

融合多尺度自注意力茶叶病害检测网络模型基于 YOLO v5^[22] 网络模型,并结合了卷积和 CSwin Transformer 多尺度自注意力机制,旨在兼顾病害图像局部和全局特征。网络结构如图 1 所示,包含 3 个关键功能模块:多尺度特征提取模块(Multi-scale feature extraction module, MFEM)、尺度自注意力模块(Scale self-attention module, SSAM)和尺度感知特征融合模块(Scale-aware feature fusion, SFF)。

该算法综合运用多尺度卷积块和自注意力机制的融合,用于提取丰富多样的特征信息。具体步骤如下:

(1)多尺度特征提取:采用 VGG16 作为多尺度卷积块骨干网络(Backbone),对茶叶病害图像进行特征提取。这些卷积块能够捕捉不同尺度和局部范围的茶叶病害信息。

(2)自注意力机制:将多尺度的特征图并行输入 CSwin Transformer 尺度自注意力模块(Neck),实现全局和局部信息的交互。这个机制使模型能自适应地调整每个尺度特征的重要性权重,以更好地捕

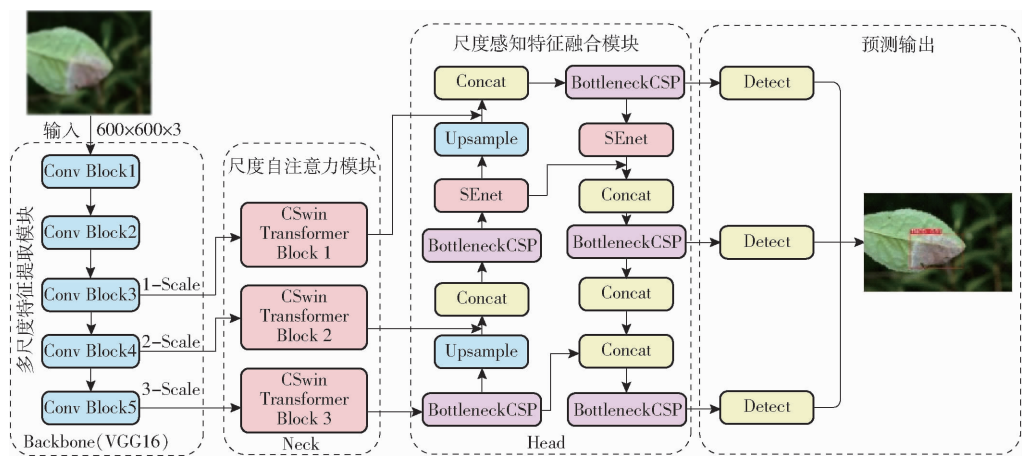


图 1 多尺度自注意力特征融合病害检测网络架构

Fig. 1 Framework of multi-scale guided self-attention network

提茶叶病害的特征。

(3)尺度感知特征融合:尺度感知特征融合模块(Head)应用通道注意力机制,对茶叶图像不同尺度的特征进行加权融合。能够更好地结合茶叶病害在不同尺度上的特征信息,提高定位和识别的准确性。

(4)预测模块:此模块能够利用之前提取的特征信息,对复杂背景下和不同尺度的茶叶病害特征进行高效定位和识别。

综合考虑了茶叶病害的尺度变化和局部特征,利用多尺度卷积块、自注意力机制和尺度感知特征融合模块的融合方法,能够提取丰富的特征信息,并更准确地定位和识别茶叶病害。该方法应对茶叶病虫害的尺度多变、病害密集和遮挡等复杂情况表现出良好性能。

1.2 多尺度特征提取模块

传统的图像特征提取方法通常只能关注到局部或全局的单一信息,无法全面反映图像的多样性和复杂性^[23]。近年来,基于深度学习的方法逐渐成为图像特征提取的主流趋势。其中,视觉Transformer^[24]将图像分割为具有位置嵌入的Token序列,利用自注意力模块提取参数化向量作为视觉表示,并广泛应用于图像特征提取任务中。然而,由

于每个块对应的Token序列尺度是固定的,无法捕捉不同尺度的特征结构。在茶叶病害检测中,由于尺度的变化,该方法无法有效地适应。此外,自注意力机制更专注于全局信息的获取,忽略了局部特征细节,这在一定程度上降低了茶叶病害图像中复杂背景和前景之间的可辨别性,限制了其在病害特征提取领域的应用。

本文采用了并行多尺度的茶叶病害特征提取模块。具体做法是先缩放原始图像,然后将不同尺度的图像同时输入到自注意力机制模块中,利用多头自注意力机制获取各个尺度下的全局上下文信息,并行多尺度Token序列原理如图2所示。以VGG16特征提取网络为例,如图1所示,保留了最后3个卷积块,将病害图像分辨率缩小为原图像的1/2、1/4、1/8,形成多尺度特征结构。通过保留卷积块的输出构成多尺度特征图,弥补了视觉Transformer局部特征表示不足的问题,同时实现多尺度结构。

为了确保大尺度的特征图包含充分的空间信息,移除最后一个卷积层级的池化操作。值得注意的是,当图像划分为固定尺寸时,使用多尺度特征图作为视觉Transformer的输入可以间接改变Token序列的长度。这样可以使每个Token包含更多的空间局部信息,从而显著增强了Transformer的归纳偏置

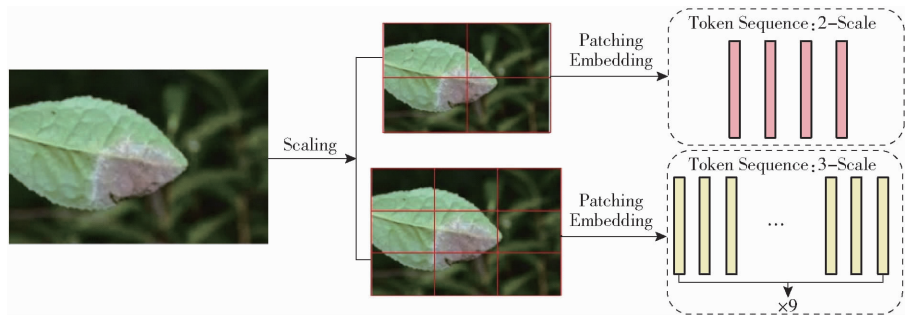


图 2 并行多尺度Token序列原理

Fig. 2 Parallel multi-scale Token sequences principle

能力。对多尺度特征提取模块定义为

$$x_i = C(x_0) \quad (1)$$

其中, $x_0 \in \mathbf{R}^{H \times W \times 3}$ 表示给定的病害图像, $x_i \in \mathbf{R}^{\frac{H}{2^i} \times \frac{W}{2^i} \times C_i}$ 代表病害图像与每个卷积块的输出特征图, i 表示尺度序号, H 和 W 分别表示图像的高度和宽度。

1.3 尺度自注意力模块

尺度自注意力模块结构如图 3a 所示。首先,将

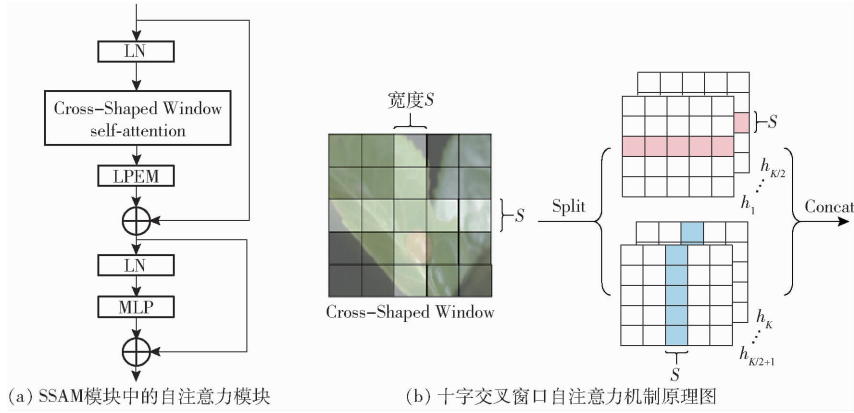


图 3 CSwin Transformer 模块

Fig. 3 CSwin Transformer block

多头的自注意力模型, CSwin Transformer 模块将一半的头数用于横向自注意力, 另一半用于纵向自注意力。最后将这 2 个特征拼接起来, 如图 3b 所示。假设网络有 K 个头, 其中 1 到 $K/2$ 用于横向自注意力计算, $K/2 + 1$ 到 K 用于纵向自注意力计算。

具体而言, 输入序列划分成宽度相同且不重叠的 S 条带。对于每个条带 $\mathbf{X}^i = (X^1, X^2, \dots, X^M) \in \mathbf{R}^{(S \times D) \times C}$, 其包含 $S \times D$ 个 Token, D 表示 Token 的特征维度, S 为条带的宽度, M 为将特征均匀划分为不重叠等宽的条带数量。多头注意力机制以条带为单位进行计算, Queries、Keys 和 Values 的维度设为 d_m , 计算过程为

$$Y_k^i = \text{Attention}(\mathbf{X}^i \mathbf{W}_k^Q, \mathbf{X}^i \mathbf{W}_k^K, \mathbf{X}^i \mathbf{W}_k^V) \quad (2)$$

式中, $\mathbf{W}_k^Q, \mathbf{W}_k^K, \mathbf{W}_k^V$ 分别表示第 k 个多头注意力的 Q, K, V 的投影矩阵, Attention 为注意力。

其次, 通过线性变换将输入条带映射为 Queries、Keys 和 Values, 并计算 Queries 和 Keys 的点积, 得到注意力矩阵。将注意力矩阵与 Values 相乘并加权, 得到垂直和水平方向的注意力, 计算公式为

$$E_k = \begin{cases} H - \text{Attention}(\mathbf{X}) & (k = 1, 2, \dots, K/2) \\ V - \text{Attention}(\mathbf{X}) & (k = K/2 + 1, K/2 + 2, \dots, K) \end{cases} \quad (3)$$

式中, $V - \text{Attention}(\mathbf{X})$ 、 $H - \text{Attention}(\mathbf{X})$ 分别表示垂直自注意力、水平自注意力, E_k 表示将多头注意力分为垂直自注意力和水平自注意力并行处理。

多尺度特征转换为 Transformer 的二维标准输入形式 $x_i \in \mathbf{R}^{N_i \times D_i}$, 其中 N_i 表示序列长度, D_i 表示向量的维度。该过程通过一个尺寸为 $K \times K$ 、步长为 K 的滑动窗口实现。其次, 利用多头注意力机制将 x_i 映射到多维特征子空间中。

为了降低计算复杂度, 在尺度自注意力模块中采用了 CSwin Transformer。其核心思想是通过并行的横向自注意力和纵向自注意力来实现。对于一个

最后, 将垂直和水平的注意力并行输出并拼接, 形成十字交叉窗口自注意力机制 CSWinattention 。将自注意力的结果映射到目标输出维度, 以提取更准确的特征信息。计算公式为

$$\text{CSWinattention}(\mathbf{X}) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_K) \mathbf{W}^0 \quad (4)$$

式中, Concat 为垂直和水平注意力拼接操作。 $\mathbf{W}^0$ 表示常用的投影矩阵, $\text{head}_1 \sim \text{head}_k$ 表示多头注意力机制中的不同头。上述操作使得模型能够更好地捕捉图像或特征图中的细节和结构信息, 从而增强模型的感知能力和表达能力。

1.4 尺度感知特征融合模块

多尺度特征融合模块利用通道注意力机制将多个尺度的特征融合, 增强了整体特征的表达能力。如图 4 所示, 融合模块将特征图进行上采样 (Upsample)、拼接 (Concat)、卷积 (BottleneckCSP) 和加权融合 (SEnet) 等操作。通过上采样, 使不同尺度的特征具有相同的尺寸; 拼接操作将不同尺度的特征图合并, 以获得更丰富的特征表示; 卷积操作进一步提取特征信息; 加权融合操作引入通道注意力机制, 对多尺度特征进行加权融合, 以消除冗余信息。其中, 通道注意力机制通过全局平均池化进行计算, 定义式为

$$\mathbf{z} = \mathbf{F}_{sq} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \mathbf{X}(i, j) \quad (5)$$

式中 $\mathbf{z}$ ——压缩后的特征向量

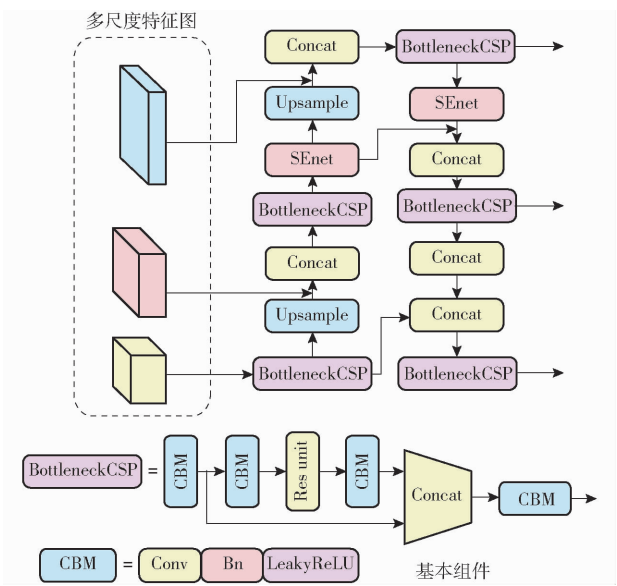


图 4 多尺度特征融合模块

Fig. 4 Multi-scale feature fusion module

特征通道间的相关性通过学习参数 w 来建模,从而减少通道个数以降低计算量,具体计算式为

$$s = F_{ex}(z) = \sigma(w_2 \delta(w_1 z)) \tag{6}$$

式中, s 为各通道的自适应权重, w_2 和 w_1 表示全连接层的参数, δ 为激活函数 Sigmoid, σ 为激活函数 ReLU。将通道自适应权重 s 与 z 相结合,为每个现有通道分配新的自适应权重,计算式为

$$X_c = F_{scale}(X, s) = Xs \tag{7}$$

式中 X_c ——经过增强的第 c 个特征图

1.5 预测模块

预测模块通过特征融合网络的输出和并行分支的协同作用,用于高效定位和识别不同尺度茶叶病害信息。同时,通过 IoU 分支提升预测结果的准确性。该模块包含 1×1 卷积层和并行分支,每个分支都有 2 个 3×3 卷积层,如图 5 所示(图中 C 为类别数)。其中,最上方 2 个 3×3 卷积用于分类任务,最下方 2 个 3×3 卷积用于回归任务。为了提高预测准确性,还额外添加了一个 IoU 分支。

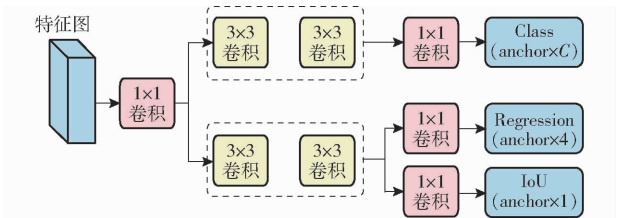


图 5 预测模块

Fig. 5 Prediction module

1.6 损失函数

本文引入 SIoU^[25] 作为损失函数,该函数考虑了预测框和真实框之间的交集大小、角度和距离信息,

能更好地处理茶叶病害尺度多变、病害密集等复杂因素。总损失函数为

$$L = W_{\text{box}} L_{\text{box}} + W_{\text{cls}} L_{\text{cls}} \tag{8}$$

其中
$$L_{\text{box}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \tag{9}$$

- 式中
- IoU ——交并比
 - Δ ——距离损失
 - Ω ——形状损失
 - W_{box} ——框损失模型体积
 - L_{box} ——回归损失
 - W_{cls} ——分类损失模型体积
 - L_{cls} ——焦点损失

2 实验

2.1 数据集及实验设置

实验使用的茶叶病害数据集来源于安徽农业大学江朝晖教授课题组^[14],共包含 776 个分辨率为 906 像素 $\times$ 600 像素的病害图像,涵盖了茶网饼病、茶圆赤星病、茶云纹叶枯病、茶饼病、茶红锈藻病、茶藻斑病共 6 种常见茶叶病害。部分示例如图 6 所示。原始样本使用并未提供茶叶病害目标的边界框信息。为了适用于病害检测任务,使用 Labelimg 图像标注工具,在每个样本图像上,由专业人员标注茶叶病害目标的边界框。此过程为训练和评估茶叶病害检测模型提供了标准数据集,使得模型能够自动地识别和定位茶叶病害。同时,为了提高模型的泛化能力,采用数据增强方法对数据集进行扩充。数据增强的目的是增加原始图像的多样性,使得模型能够更好地适应未处理过的图像预测任务。在本实验中,使用多种数据增强方法,包括旋转、添加噪声和曝光。通过应用这些方法,生成 7 630 个增强后的样本。



图 6 茶叶病害叶片示例图像

Fig. 6 Sample images of diseased tea leaves

统计和排序目标掩膜相对于图像的像素比例:约 20% 的目标尺度小于 0.020 7,而 34% 的目标尺度大于 0.345。这表明数据集中目标的尺度跨度非常大,存在着显著差异。

在实验中,基于 Python 3.8 和 PyTorch 1.7.0、CUDA 11.0 框架实现了算法,并在一台配备 RTX 3090GPU 的服务器上进行了训练。对增强后的样本进行处理时,按比例 8:2 划分数据为训练集和验证集。这种数据划分方式有助于保持数据的均衡,并在测试阶段评估模型性能,以获得更可靠的结果。

2.2 评价指标

使用平均精度均值 (Mean average precision, mAP) 作为主要评估指标。mAP 通过计算准确率-召回率曲线下的面积来度量模型的性能。用精确率 (Precision) 和召回率 (Recall) 进一步分析模型的性能。精确率是指正确检测到的目标框数与所有预测的目标框数的比值,反映模型的准确性。召回率是指正确检测到的目标框数量与实际目标框数量比值,反映模型的完整性。

2.3 不同主干网络对比

本文对 VGG16、ResNet-101 和 Darknet53 等不同主干网络进行了测试和比较,结果如表 1 所示。由表 1 可知,Darknet53 作为主干网络的模型在精确率、召回率和平均精度均值方面表现最佳。相比 VGG16,ResNet-101 精确率提高 2.14 个百分点,召回率提高 4.06 个百分点,mAP 提高 0.81 个百分点。

表 1 不同骨干网间对比

Tab.1 Comparison among different backbone networks			
%			
主干网络	精确率	召回率	平均精度均值
VGG16	90.02	89.45	92.15
ResNet-101	92.16	93.51	92.96
Darknet53	92.94	94.49	93.71

Darknet53 的网络结构深,包含多个卷积层和池化层。这种深度的结构赋予模型强大的表达能力,但也增加了模型复杂度和训练难度。另外,Darknet53 中的一些操作 (如 Dropout、Batch Normalization 等) 虽能提高模型泛化能力和稳定性,但也增加了模型复杂度。相比于 ResNet-50,ResNet-101 是更深的深度残差网络,需要更多的计算资源和训练时间。

相对于其他网络结构,VGG16 在茶叶病害检测任务中表现出色,且具有简单、易于理解和解释的特点。已有在 ImageNet 上训练的 VGG16 模型公开发布,也展现了在其他数据集上 Fine-tuning 的良好适应能力。因此,本文选择 VGG16 作为主干网络,并采用构建多尺度特征表示的通用方法来解决尺度问题。

2.4 经典目标检测算法性能对比

为了验证 MSGSN 模型对茶叶病害的检测能

力,本研究选择 6 种主流目标检测方法进行对比试验:SSD^[26]、Fast R-CNN^[27]、EfficientDet^[28]、YOLO v5s^[22]、YOLO-Tea^[17] 和 AX-RetinaNet^[29]。SSD 方法采用多尺度检测;Fast R-CNN 方法采用 VGG16 作为主干网络;EfficientDet 方法采用双向特征融合提高检测精度;YOLO v5s 方法是使用 IoU Loss 损失提高检测精度;YOLO-Tea 方法在 YOLO v5 中集成了自注意力模块;AX-RetinaNet 方法采用 ResNet 作为主干网用于检测自然场景下的茶叶病害。在训练过程中,所有网络使用相同的训练集,并通过测试集图像对模型性能进行综合评估。不同算法检测结果如表 2 所示。

表 2 茶叶病害检测算法对比

Tab.2 Comparison of STOA methods		
算法	年份	平均精度均值/%
SSD	2016	87.52
Fast R-CNN	2015	88.93
EfficientDet	2020	89.09
YOLO v5s	2020	89.31
YOLO-Tea	2023	89.82
AX-RetinaNet	2022	92.93
MSGSN	2023	92.15

由表 2 可得,本文改进算法在检测精度方面有明显优势。相比使用 IoU Loss 的 YOLO v5s 算法,本文算法实现了整体检测精度的提升,这是因为 MSGSN 采用 SIoU 作为损失函数可以更好地处理茶叶病害尺度多变问题。与最优秀的方法相比,所提出的模型仍存在一定差距,这是因为 AX-RetinaNet 网络模型采用 ResNet 作为主干网络来提取目标特征信息。EfficientDet 算法的检测精度高于 Fast R-CNN 和 SSD,这可能是由于 EfficientDet 算法采用双向特征融合模块,可以提高检测精度。与 SSD 算法相比,本文算法平均精度略有提高。

综上所述,本文方法通过融合 CNN 和自注意力机制的特性,在融合全局和局部信息方面取得了显著改进。通过使用 CNN 提取多尺度特征图,能够获取更丰富的空间局部信息,并结合自注意力机制,全面考虑了全局信息的影响,使全局与局部信息能够有效地交互。这种模型设计充分考虑了人类视觉机制,能够有效地处理复杂背景和尺度变化的病害图像问题。引入自注意力机制使得模型能够自动选择和聚焦于图像中最相关的特征,从而显著提高了病害检测的准确性。

2.5 消融实验

为验证本文提出的网络模型的有效性,使用

VGG16 作为骨干网,并逐步增加各网络模块来验证网络结构的合理性和有效性,结果如表 3 所示。

表 3 各模块消融实验结果

Tab.3 Results of ablation experiments by module

模块	平均精度均值	%
MFEM	89.26	
MFEM + SSAM	91.32	
MFEM + SSAM + SFF(本文模型)	92.15	

相较于仅使用 MFEM 模块, MFEM + SSAM 模块显著提高了性能,mAP 提高 2.06 个百分点。这是因为 MFEM + SSAM 模块利用多头自注意力机制在多尺度特征图上捕获全局上下文信息,实现了全局与局部信息的交互,并通过加权处理病斑颜色和叶片边缘等特征来提高检测的精确度。进一步引入尺度感知特征融合模块 SFF,并利用通道注意力机制对不同尺度的特征进行加权融合。这样就能更好地结合茶叶病害在不同尺度上的特征信息,从而提高检测准确性。相对于 MFEM + SSAM,mAP 提升 0.83 个百分点,说明 SFF 模块的引入可以提高网络的检测精度。

2.6 不同尺度病害情况下检测结果分析

为进一步说明本模型在不同尺度变化下的性能,选择了有代表性的病害图像进行检测,结果见图 7~9。图 7 和图 8 的局部放大图可更清晰展示茶叶病害的形态、尺度和茶叶病害周围的背景信息。从图 7 可以看出,MSGSN 能高效地检测不同尺度的病害。这得益于 MSGSN 采用基于 VGG16 的多尺度特征提取模块,能获取茶叶病害图像在不同尺度下的局部细节特征,例如纹理和边缘等。这使得模型能够更好地适应不同尺度的茶饼病害。从图 8 可以看出,在病害复杂背景下,本文算法可同时检测出多种尺度的病害。这得益于尺度自注意力模块捕获茶叶图像中像素间全局依赖关系,实现全局与局部特征之间的有效交互。这有助于模型在提取局部特征的基础上,进一步整合全局信息,提高检测的准确性。从图 9 可以看出,在不同尺度病害下,左边图像置信度为 0.81,其 TRRAD 是茶红锈藻病,右边图像置信度为 0.93,其 TCD 是茶饼病,本文算法都达到了很高的置信度。



图 7 尺度变化的病害检测结果

Fig.7 Detection results of scale variations

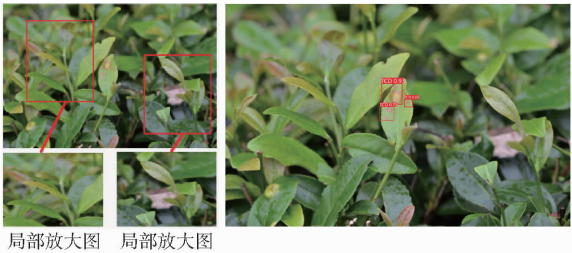


图 8 复杂背景下不同尺度变化检测结果

Fig.8 Detection results of scale variations in complex backgrounds



图 9 不同尺度病害的检测结果对比

Fig.9 Comparison of detection results for different scales of disease

3 结束语

为解决田间环境下茶叶病害背景复杂、病害尺度多变等因素引起的检测困难问题,本文提出了一种融合多尺度自注意力的茶叶病害检测方法。该方法通过将卷积神经网络与 CSwin Transformer 融合,构建了自注意力机制的多尺度网络结构,实现了全局与局部之间的交互。实验对比经典的目标检测方法表明,融合多尺度自注意力的茶叶病害检测方法的平均精度均值达到 92.15%。不同尺度病害情况下检测结果表明,在背景复杂、病害尺度多变等场景下具有更好的检测效果。因此,该模型可为茶叶病害智能诊断提供参考依据,利于更准确地检测和识别茶叶病害,及时采取控制措施,保护茶树的健康,提升茶叶的品质和产量。

参 考 文 献

[1] ABADE A, FERREIRA P A, DE BARROS VIDAL F. Plant diseases recognition on images using convolutional neural networks: a systematic review[J]. Computers and Electronics in Agriculture, 2021, 185:106125.

[2] 李子茂, 徐杰, 郑禄, 等. 基于改进 DenseNet 的茶叶病害小样本识别方法[J]. 农业工程学报, 2022, 38(10):182-190.

LI Zimao, XU Jie, ZHENG Lu, et al. Small sample recognition method of tea disease based on improved DenseNet[J]. Transactions of the CSAE, 2022, 38(10): 182-190. (in Chinese)

- [3] BHAVSAR N, TAYADE A, GHODMARE H, et al. A review on image processing algorithm for foliage target detection and classification[J]. *International Journal of Engineering Research and Reviews*, 2022, 10(4): 52 – 61.
- [4] 丁永军, 张晶晶, 李民赞. 基于卷积胶囊网络的百合病害识别研究[J]. *农业机械学报*, 2020, 51(12): 246 – 251, 331. DING Yongjun, ZHANG Jingjing, LI Minzan. Disease detection of lily based on convolutional capsule network [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(12): 246 – 251, 331. (in Chinese)
- [5] 刘阳, 高国琴. 采用改进的 SqueezeNet 模型识别多类叶片病害[J]. *农业工程学报*, 2021, 37(2): 187 – 195. LIU Yang, GAO Guoqin. Identification of multiple leaf diseases using improved SqueezeNet model[J]. *Transactions of the CSAE*, 2021, 37(2): 187 – 195. (in Chinese)
- [6] 苏仕芳, 乔焰, 饶元. 基于迁移学习的葡萄叶片病害识别及移动端应用[J]. *农业工程学报*, 2021, 37(10): 127 – 134. SU Shifang, QIAO Yan, RAO Yuan. Recognition of grape leaf diseases and mobile application based on transfer learning[J]. *Transactions of the CSAE*, 2021, 37(10): 127 – 134. (in Chinese)
- [7] 侯俊铭, 姚恩超, 朱红杰. 基于卷积神经网络的蓖麻种子损伤分类研究[J]. *农业机械学报*, 2020, 51(增刊 1): 440 – 449. HOU Junming, YAO Enchao, ZHU Hongjie. Classification of castor seed damage based on convolutional neural network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(Supp. 1): 440 – 449. (in Chinese)
- [8] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770 – 778.
- [9] 何自芬, 黄俊璇, 刘强, 等. 基于非对称混洗卷积神经网络的苹果叶部病害分割[J]. *农业机械学报*, 2021, 52(8): 221 – 230. HE Zifen, HUANG Junxuan, LIU Qiang, et al. High precision identification of apple leaf diseases based on asymmetric shuffle convolution [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(8): 221 – 230. (in Chinese)
- [10] 余小东, 杨孟辑, 张海清, 等. 基于迁移学习的农作物病虫害检测方法研究与应用[J]. *农业机械学报*, 2020, 51(10): 252 – 258. YU Xiaodong, YANG Mengji, ZHANG Haiqing, et al. Research and application of crop diseases detection method based on transfer learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(10): 252 – 258. (in Chinese)
- [11] 刘斌, 徐皓玮, 李承泽, 等. 基于快照集成卷积神经网络的苹果叶部病害程度识别[J]. *农业机械学报*, 2022, 53(6): 286 – 294. LIU Bin, XU Haowei, LI Chengze, et al. Apple leaf disease identification method based on snapshot ensemble CNN [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(6): 286 – 294. (in Chinese)
- [12] LEE S H, LIN S R, CHEN S F. Identification of tea foliar diseases and pest damage under practical field conditions using a convolutional neural network[J]. *Plant Pathology*, 2020, 69(9): 1731 – 1739.
- [13] 孙道宗, 刘欢, 刘锦源, 等. 基于改进 YOLOv4 模型的茶叶病害识别[J]. *西北农林科技大学学报(自然科学版)*, 2023, 51(9): 145 – 154. SUN Daozong, LIU Huan, LIU Jinyuan, et al. Recognition of tea diseases based on improved YOLOv4 mode[J]. *Journal of Northwest A&F University(Natural Science Edition)*, 2023, 51(9): 145 – 154. (in Chinese)
- [14] 谢军, 江朝晖, 李博, 等. 基于二次迁移模型的小样本茶树病害识别[J]. *江苏农业科学*, 2021, 49(6): 176 – 182. XIE Jun, JIANG Chaohui, LI Bo, et al. Image recognition of tea plant disease small samples based on secondary migration model[J]. *Jiangsu Agricultural Sciences*, 2021, 49(6): 176 – 182. (in Chinese)
- [15] 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.
- [16] CHEN J, LIU Q, GAO L. Visual tea leaf disease recognition using a convolutional neural network model[J]. *Symmetry*, 2019, 11(3): 343.
- [17] XUE Z, XU R, BAI D, et al. YOLO – Tea: a tea disease detection model improved by YOLOv5[J]. *Forests*, 2023, 14(2): 415.
- [18] LIN J, BAI D, XU R, et al. TSBA – YOLO: an improved tea diseases detection model based on attention mechanisms and feature fusion[J]. *Forests*, 2023, 14(3): 619.
- [19] GEHRING J, AULI M, GRANGIER D, et al. Convolutional sequence to sequence learning[C] // *International Conference on Machine Learning*, 2017: 1243 – 1252.
- [20] YANG X. An overview of the attention mechanisms in computer vision[J]. *Journal of Physics: Conference Series*, 2020, 1693(1): 012173.
- [21] DONG X, BAO J, CHEN D, et al. CSwin Transformer: a general vision transformer backbone with cross-shaped windows[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 12124 – 12134.
- [22] WU W, LIU H, LI L, et al. Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image[J]. *PloS One*, 2021, 16(10): e0259283.
- [23] NIU Z, ZHONG G, YU H. A review on the attention mechanism of deep learning[J]. *Neurocomputing*, 2021, 452: 48 – 62.
- [24] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey[J]. *ACM Computing Surveys (CSUR)*, 2022, 54(10): 1 – 41.
- [25] GEVORGYAN Z. Siou loss; more powerful learning for bounding box regression[J]. *arXiv Preprint*, arXiv:2205.12740, 2022.
- [26] KUMAR A, SRIVASTAVA S. Object detection system based on convolution neural networks using single shot multi-box detector[J]. *Procedia Computer Science*, 2020, 171: 2610 – 2617.
- [27] GIRSHICK R. Fast R – CNN[C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440 – 1448.
- [28] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10781 – 10790.
- [29] BAO W, FAN T, HU G, et al. Detection and identification of tea leaf diseases based on AX – RetinaNet[J]. *Scientific Reports*, 2022, 12(1): 1 – 16.