

# 基于深度学习的番茄叶片病害图像识别技术

徐志望 谢侃侃 彭荣芳 吴克航 孙梦棋  
绍兴文理学院元培学院 浙江绍兴 312000

**摘要：**本文研究细粒度图像分类和弱监督学习在番茄病害检测中的应用，该研究一方面能够降低数据集对人工标注的依赖，另一方面能够更好的利用关键区域的特征表达能力，提高模型在番茄病害检测的准确度。本文提出了注意力指导下的图像增强和样本生成方法，不同于传统的翻转、平移、采集等图像扩充方法，是通过注意力裁剪和下降有选择的保留图像中的关键特征，从而合成新的高质量训练样本，在理论和方法上都具有创新性。

**关键词：**细粒度；弱监督；注意力裁剪

## 一、研究意义

番茄是我国重要的经济作物之一，因其营养丰富、易加工以及辅助治疗的功效，世界人民对番茄及其制品的需求不断增大，番茄的种植面积与种植规模也在逐年增加。然而在番茄的种植过程中，各种病害严重制约着番茄的生产，常见的有晚疫病、早疫病、叶霉病、花叶病毒病和斑枯病等，据统计番茄产业总产量的8%~10%的损失由病害造成。及早发现番茄作物叶片病害是保障番茄高产量与高品质的关键因素，但传统的人工鉴定方式依赖菜农或指导专家的经验，容易出现错诊、漏诊，导致农民错过防控病害的最佳时间。

## 二、国内外研究现状及发展动态分析

### （一）国外研究现状

随着机器学习的发展，外国研究者也将深度学习应用在植物病害检测，如MohamethF等人提出一种基于支持向量机（SVM）和卷积神经网络（CNN）的植物病害检测方法，通过CNN提取病害特征，所有得到的特征通过SVM进行分类，较于传统机器学习，大大提高了收集数据的速度，但准确度和速度还不够高。ArsenovicM等研究出一种基于两阶段体系结构的神经网络病害分类方法，对结合Plant village数据集及新建数据集共42种健康叶片和被感染叶片进行分类，自顶向下调制的Faster

R-CNN的表现优于其他策略，虽然速度较快，但准确率为93.67%相对不足。在图像处理方向，样本增强一直是研究者关注的热点。

### （二）国内研究现状

国内关于病害识别算法的研究主要针对传统或现有模型的性能提升，在实验中取得较高的识别精度。其中通过图像分割方法提升模型的识别能力是一个重要方向，如徐凯宏等采用分治中值滤波算法和分水岭算法进行病斑分割，对传统的BP算法进行优化，设计出一种诊断温室番茄病害的方法，训练所得模型收敛速度较快、识别效果较好，但识别种类较少。南京农业大学郭小清等<sup>[1]</sup>提出了基于改进Multi-Scale AlexNet的番茄叶病害图像识别方法，通过设置不同尺度卷积核来提取特征，并基于AlexNet的模型提升了多感受识别效果，该模型对番茄叶病害平均识别准确率达到92.7%，在田间应用的识别率达到89.2%。

目前国内外研究者采用机器学习在番茄病害识别取得了显著的成果，但研究还存在着特征提取准确率低，模型训练依赖大规模样本集，以及未分割目标病害区域导致的计算量大和冗余数据大等问题，番茄病害的自动识别技术还有巨大的提升空间。

## 三、研究内容

针对番茄病害诊断问题，足够的数据集规模和提取有辨别力的特征是提升分类效果的关键。在番茄病害数据集规模和特征提取能力有限的情况下，扩充数据集样本数量和优化关键部位特征提取机制成为了本项目的研究内容<sup>[2]</sup>。总体上可以概括成以下三点：1）面向类标签样本集的数据扩充增强技术研究；2）面向局部特征的多尺度注意力学习技术研究；3）番茄病害识别模型的构建和示范应用研究。

### 项目课题：

1. 绍兴市科技计划项目，基于深度学习的番茄叶片病害图像识别技术研究，课题编号：2023A12008
2. 2024年浙江省大学生科技创新活动计划（新苗人才计划），基于深度学习的番茄叶片病害图像识别技术研究，课题编号：2024R431A028

### (一) 面向类标签样本集的数据扩充增强技术研究

1. 通过弱监督学习生成注意力图来表示对象的辨别部分，提取更多有区别的的部分的特征。

2. 在注意力图的指导下增强图像，包括注意力裁剪和注意力下降，将生成的图像作为新的样本，进一步扩充与加强样本的数量与质量。

### (二) 面向局部特征的多尺度注意力学习技术研究

1. 研究通过注意力定位特征对象，在注意力图上预测特征对象的位置，来加强对对象的关注且过滤无效背景。所获得的特征对象图像不仅包含几乎对象的整个结构，而且包含更多的细节。

2. 研究一种注意力建议模块，获取信息丰富的零件区域，而不需要边界框或零件注释。通过获取不同尺度和具有更细粒度的特征的零件区域，来加强模型的分类

能力和鲁棒性。

### (三) 番茄病害识别模型的构建和应用示范研究

1. 将数据扩充网络与多尺度注意力学习网络进行融合，构建番茄病害识别模型，并在基准数据集对细粒度识别效果进行验证。

2. 将模型应用在番茄病害数据集，检验模型在标签级番茄病害图像的分类有效性，并通过实地采集番茄片图像数据进行示范应用。

## 四、研究方案

本文首先在图像数据增强部分通过弱监督学习方法来发现图像属性，运用注意力下降和注意力裁剪方法生成新的样本图像。其次在核心识别框架部分主要运用多尺度注意力学习方法来改进局部特征的可辨识性来提高特征性能<sup>[3]</sup>。具体研究方案如下图4.1所示。

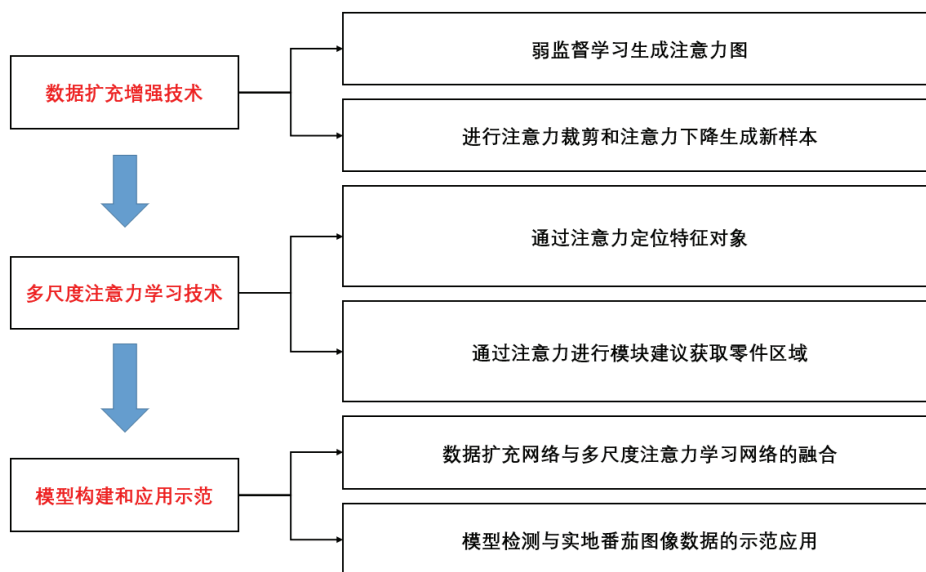


图4.1 研究方案

### (一) 弱监督注意力学习与数据增强

#### 1. 空间表示

首先预测物体的部分区域。为了贴近实际应用环境中图像的质量，训练和测试过程中，对象的人工位置注释（例如边界框或关键点）是不可用的。因此，本项目采用弱监督学习的方法，仅通过对象的类别注释（图片的标签信息）来预测对象的位置分布。

项目通过CNN来提取图像的特征，将 $F \in \mathbb{R}^{H \times W \times N}$ 作为特征图， $H$ 、 $W$ 和 $C$ 分别表示特征层的高度、宽度和通道数量<sup>[4]</sup>。目标图像区域的分布通过注意力图 $A \in \mathbb{R}^{H \times W \times M}$ 来代表，注意力图的获取来自特征图 $F$ ，用下列公式(1)表示：

$$A = f(F) = \bigcup_{k=1}^M A_k \# \quad (1)$$

其中 $f$ 是一个卷积函数。 $A_k \in \mathbb{R}^{H \times W}$ 代表目标物体的一个部分或视觉特征，例如鸟的头部、车的轮胎或者飞机的机翼、番茄病害的斑点等。 $M$ 是注意力图的数量。注意力图将被用来增强训练数据。

#### 2. 增强注意力图

随机数据增强较为低效，特别是当目标规模小和引入了较高背景噪声的时候。当有了注意力图，图像能被更高效的增强<sup>[5]</sup>。对于每个训练图像，随机选择一个的注意力图 $A_k$ 来引导数据增强过程，并将其规范化为 $k_n$ 增强映射图 $A_k^* \in \mathbb{R}^{H \times W}$ ，如公式(2)所示。

$$A_k^* = \frac{A_k - \min(A_k)}{\max(A_k) - \min(A_k)} \# \quad (2)$$

### 3. 注意力裁剪

基于增强注意力图  $A_k$ ，能够放大图像中响应值最高的区域并且提取更多详细的局部特征。通过设置元素  $A_k^*$  ( $i, j$ ) 获得裁剪掩膜  $C_k$ ，若  $A_k$  大于阈值  $\theta_c \in [0, 1]$ ，则  $C_k$  为 1；若  $A_k$  小于或等于阈值  $\theta_c \in [0, 1]$ ，则  $C_k$  为 0，如公式 (3) 所示。

$$C_k(i, j) = \begin{cases} 1, & \text{if } A_k^*(i, j) > \theta_c \\ 0, & \text{otherwise.} \end{cases} \# \quad (3)$$

### 4. 注意力下降

注意正则化损失监督的每个注意力图  $A_k^* \in \mathbb{R}^{H \times W}$  来表示相同的  $k_{th}$  目标部分，同时不同的注意力图可能集中在物体相似的部分。为了激励注意映射图表示多个不同目标部分，我们提出注意下降。我们获得了注意下降掩膜  $D_k$ ，通过设置元素  $A_k^*(i, j)$ ，若  $A_k^*$  大于阈值  $\theta_d \in [0, 1]$ ，则  $D_k$  为 0；若  $A_k^*$  小于或等于阈值  $\theta_d \in [0, 1]$ ，则  $D_k$  为 1，如公式 (4) 所示。

$$D_k(i, j) = \begin{cases} 0, & \text{if } A_k^*(i, j) > \theta_d \\ 1, & \text{otherwise.} \end{cases} \# \quad (4)$$

## (二) 核心识别框架优化方法

### 1. 注意力对象定位模块

一种特征提取方法是使用预先训练的模型来为细粒度图像检索任务提取图像特征，本项目的实现基于此方法，模型先通过预先训练提高它的定位性能，再经由 CNNs 特征图来生成目标位置坐标。

使用  $F \in \mathbb{R}^{C \times H \times W}$  来表示从输入图像  $X$  的最后一个卷积层输出  $C$  个通道空间大小为  $H \times W$  的特征图，其中  $f_i$  是对应通道的第  $i$  个特征映射。如公式 (5) 所示。

$$A = \sum_{i=0}^{c-1} f_i \# \quad (5)$$

激活图  $A$  可通过聚合特征图  $F$  获得。可以通过可视化发现深度神经网络的关注点从  $A$  中准确地定位目标区域。如公式 (6) 所示，其中  $\bar{a}$  是  $A$  的平均值。

$$\bar{a} = \frac{\sum_{x=0}^{W-1} \sum_{y=0}^{H-1} A(x, y)}{H \times W} \# \quad (6)$$

$\bar{a}$  作为阈值来确定  $A$  中该位置的元素是否属于目标物体， $(x, y)$  是  $H \times W$  激活映射中的特定位置。然后根据等式 (7)，从 ResNet-50 的最后一个卷积层  $conv_{5c}$  得到了一个大致 mask 图  $\tilde{M}_{conv_{5c}}$

$$\tilde{M}_{(x,y)} = \begin{cases} 1 & \text{if } A_{(x,y)} > \bar{a} \\ 0 & \text{otherwise} \end{cases} \# \quad (7)$$

我们发现目标通常位于  $\tilde{M}_{conv_{5c}}$  的最大连通分量中，因此使用包含最大连通区域的最小边界框作为目标定位的结果。为了提升识别准确率，我们还将使用训练集来训练 ResNet-50 以提高目标定位精度。根据位于  $Conv_{5c}$  的块前区域  $Conv_{5b}$ ，可以输出激活图  $\tilde{M}_{conv_{5b}}$ 。最后取  $\tilde{M}_{conv_{5c}}$  和  $\tilde{M}_{conv_{5b}}$  的交集后得到更准确的 mask  $M$ ，如公式 (8) 所示。

$$M = \tilde{M}_{conv_{5c}} \cap \tilde{M}_{conv_{5b}} \# \quad (8)$$

### 2. 注意力局部建议模块

虽然注意力对象定位模块可以实现更高的定位精度，但也有一些定位结果是目标物体的一部分。通过注意力局部建议模块和多尺度注意力学习可以提高模型在这种情况下鲁棒性。通过观察激活图，发现激活图中激活值高的区域往往是关键部分所在的区域。使用目标检测中的滑动窗来找到信息窗口作为局部图像。此外，通过全卷积网络实现了传统的滑动窗口方法，减少了计算量，就像 Overfeat 从上一分支输出的特征图中获取不同窗口的特征图一样。然后将通道维中每个窗口的激活图  $A_{\omega}$  聚合，并根据公式 (9) 得到其激活均值  $\bar{a}_{\omega}$ 。

$$\bar{a}_{\omega} = \frac{\sum_{x=0}^{W_{\omega}-1} \sum_{y=0}^{H_{\omega}-1} A_{\omega}(x, y)}{H_{\omega} \times W_{\omega}} \# \quad (9)$$

## 参考文献

- [1] 郭小清, 范涛杰, 舒欣. 基于改进 Multi-Scale AlexNet 的番茄叶部病害图像识别[J]. 农业工程学报, 2019, 35 (13): 162-169.
- [2] 樊杰. 基于深度残差网络的农作物叶片病症识别研究[D]. 内蒙古科技大学, 2020.
- [3] 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.
- [4] Dubey A, Gupta O, Raskar R, et al. Maximum-entropy fine-grained classification[J]. arXiv preprint arXiv:1809.05934, 2018.
- [5] Zheng H, Fu J, Zha Z J, et al. Learning deep bilinear transformation for fine-grained image representation[J]. arXiv preprint arXiv:1911.03621, 2019.