doi:10.6041/j.issn.1000-1298.2025.04.035

基于 TD－BlendMask 的复杂环境三七叶片病害实例分割方法

杨启良[1]　陈　成[1]　雷　炼[2]　周宁珊[1]　杨　玲[2]

（1.昆明理工大学现代农业工程学院，昆明 650500；2.昆明理工大学信息工程与自动化学院，昆明 650500）

摘要：针对三七叶片病害中灰霉病与疫病表型特征高度相似、炭疽病等病害病灶区域小且形态复杂导致的图像分割特征提取困难与识别精度不足问题，本文提出了 Transformer－DCNv2－BlendMask（TD－BlendMask）三七叶片多类别病害图像分割模型。首先，为了解决三七叶病害视觉相似问题，引入了 Transformer 编码器来捕获多种病害类别的长距离依赖性。其次，可变形卷积网络（DCNv2）通过引入偏移量，使其在分割各种复杂形状的病害方面具有更好的适应性。最后，与其他常用的实例分割模型（如 BoxInst、ConInst、SOLOv2、Mask R－CNN 和 YOLO v8－seg）在包含多类别疾病的三七叶片病害数据集上进行比较。实验结果表明，所提出模型精确度（AP）达到 86.14%，比基准模型高 3.17 个百分点，相比经典的 Mask R－CNN 模型高出 4.37 个百分点。在灰霉病、疫病和炭疽病类别上，分别提高 0.16、4.32、4.46 个百分点。因此，本文所提出的方法为在复杂环境中准确分割形状复杂且视觉高度相似的病害提供了有效解决方案，有助于实现病害准确量化。

关键词：三七叶片病害；实例分割；复杂田间环境；Transformer 编码器；可变形卷积

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2025)04-0375-12　　**OSID**：

TD－BlendMask-based Approach for Instance Segmentation of *Panax notoginseng* Leaf Diseases in Complex Environments

YANG Qiliang[1]　CHEN Cheng[1]　LEI Lian[2]　ZHOU Ningshan[1]　YANG Ling[2]

(1. *Faculty of Modern Agricultural Engineering*, *Kunming University of Science and Technology*, *Kunming* 650500, *China*
2. *Faculty of Information Engineering and Automation*, *Kunming University of Science and Technology*, *Kunming* 650500, *China*)

Abstract: Under practical *Panax notoginseng* cultivation scenarios, due to the visual similarities between diseases such as gray mold and plague, as well as small individual targets with complex and variable shapes of diseases such as anthracnose under natural conditions, current methods face a difficult problem of *P. notoginseng* leaf disease segmentation. A modified Transformer－DCNv2－BlendMask model for *P. notoginseng* leaf multicategory diseases image segmentation was proposed. To deal with the visual similarity problem and variable shape targets appeared on *P. notoginseng* leaf disease, a Transformer encoder to capture long-distance dependencies for multiple disease categories was introduced. And the deformable convolution networks v2 (DCNv2) showed a better adaptability of convolutional networks by enabling free-form deformation of the convolution to segment disease with various shape. The model and other instance segmentation models such as BoxInst, ConInst, SOLOv2, Mask R－CNN and YOLO v8－seg on the *P. notoginseng* leaf disease dataset, which contains multi-category diseases were compared. The results demonstrated the competitive performance of our model, achieving a average precision (AP) of 86.14%, outperforming the baseline BlendMask model by 3.17 percentage points and the previously best-performing Mask R－CNN by 4.37 percentage points. It also exceeded the baseline by 0.16 percentage points, 4.32 percentage points and 4.46 percentage points for the gray mold, plague and anthracnose categories, respectively. Thus, our method provides a robust solution for segmenting shape-variable and visually similar diseases in complex environments, helping to achieve accurate quantification of diseases.

Key words: *Panax notoginseng* leaf diseases; instance segmentation; complex field conditions; Transformer encoder; DCN

收稿日期：2024－03－04　修回日期：2024－06－04

基金项目：云南省科技厅重点项目（202201AS070034、202305AM070006）和云南省高校重点实验室项目（KKPS201923009）

作者简介：杨启良（1978—），男，教授，博士，主要从事农业工程研究，E-mail：yangqilianglovena@163.com

通信作者：杨玲（1994—），女，讲师，博士，主要从事计算机视觉研究，E-mail：yangling@kust.edu.cn

0 引言

三七(*Panax notoginseng*(BurK.) F. H. Chen)是一种药用植物,因其在心血管健康方面的广泛应用而闻名,其行业价值在 2020 年达到了 380 亿元人民币[1]。然而,根腐病、干叶病、炭疽病和灰霉病等病害对其产量造成了严重影响,并造成了严重的经济损失[2-3]。传统三七病害诊断主要依赖经验丰富的专家进行判断,效率较低,主观性较强,不适用于大规模的农业种植环境[4]。因此,迫切需要研究一种快速准确的分割方法来对三七叶片病害进行分割[5]。在此背景下,基于计算机视觉的叶片病害分割算法应运而生,通过对病害区域进行像素级的识别,实现对三七病害的精确量化,为三七病害的精准防治提供理论基础,进而阻止病害扩散、提高三七产量和减少农药施用。

近年来,随着深度学习在图像领域的发展,该方法也在农作物病害图像分割中得到了广泛应用。基于深度学习的图像分割算法包含语义分割、实例分割和全景分割,其性能和效率超越了传统的机器学习图像分割方法[6-8],由于具备更高的抗干扰性和更直接的端到端训练方式而取得显著分割效果。其中,语义分割通过对每个不同类别的图像像素进行分类,在准确分割植物叶病、土壤保护以及作物/杂草管理等任务中表现出优秀性能[9-13],语义分割神经网络类型包括全卷积网络(FCN)[14]、U-Net、金字塔场景解析网络(PSPNet)[15]和多版本的 DeepLab[16-18]。尽管语义分割技术在作物病害分割任务中取得了进步[19-21],但在处理包含多个病叶或重叠区域的复杂图像时,由于该类方法不能区分同种类别中的不同实例,其固有局限性变得明显。

为了克服这些局限性,实例分割技术通过融合目标检测与语义分割的优点,形成了具有对象级识别能力的解决方案。现有研究表明,基于卷积神经网络(CNN)的方法可以在不同光照条件下的田间图像中有效地检测和分割植物病变区域[22-23]。例如,改进的 Mask R-CNN 模型(ResNet-50 骨架)对小目标分割表现优异。这种改进不仅使模型在处理复杂背景时表现更为优异,还能更精准地分离出病害区域。此外,为了进一步应对田间图像中复杂背景干扰和样本类别不平衡的问题,引入 Focal Loss 通过动态调整损失权重,能够有效降低易分类样本对模型训练的影响,从而提升模型在少数类和难分类样本上的表现。尽管取得了一些进展,但由于 CNN 中每个卷积滤波器的感受野有限,难以关联图像中长距离空间信息,可能会导致特征丢失。

除此之外,上述方法在三七叶病害的实例分割中存在特定的挑战,包括:以往研究的图像采集是单片叶片,而田间条件下采集的多叶片多类病害图像含有大量的土壤背景和光照等噪声,且不同种类病害及不同严重程度的病害特征具有多样性和复杂性。特征相似的病害容易影响模型的先验框标签分类准确性,导致分割错误,如炭疽病等在自然条件下形状复杂多样和小目标的病害,使得特征的有效提取变得相当困难。虽然对病虫害、农作物等小目标识别提出了针对性的改进方法[24-26],但是对于其环境的目标物,无论是病虫还是大米等小目标其尺度变化较小,以往研究通常只关注单片叶片的图像采集,而忽略了田间条件下多叶片多类病害图像的复杂性。上述因素给病害分割带来了严重挑战。针对上述实例分割中的难点,WANG 等[27]针对实际种植场景中叶片和病害图像在复杂光照和重叠情况下的分割难题,提出了一种结合残差注意力机制、空洞空间金字塔池化和权重压缩损失的 RAAWC-UNet 模型,该方法有效提升了叶片和病害图像分割的精度。但是在三七病害数据集上,同一类病害有可能是一些细小的斑点,也可能会出现大面积坏死,并且形状复杂多样。不同病害类型之间存在特征相似性,如灰霉病和疫病。因此现有方法不能很好地应用于三七数据集。此外,不同种类的病害以及同一病害的不同严重程度,其特征都呈现出多样性和复杂性。特别是当病害特征相似时,很容易对模型的先验框标签分类造成干扰,从而影响分割的准确性。

为解决上述问题,本文提出一种基于 TD-BlendMask 的三七病害多类别实例分割方法。BlendMask 结合高层次的语义特征与位置敏感的低层次特征,使得模型不仅能捕捉到三七病害的全面语义信息,还能精确地分离和识别叶片上重叠或特征类似的病斑实例,实现疫病、灰霉病、炭疽病等多类病害的精准识别。针对三七病害特有的表现形式,提出一种结合卷积神经网络(CNN)和 Transformer 的混合模型 Transformer-DCNv2-BlendMask(TD-BlendMask)。针对图像中三七叶片与病害的形状和尺寸多变,且图像可能因为自然因素(如风、生长姿态等)或拍摄角度等因素导致植物部分的形态出现变形的问题,引入 DCNv2 动态调整其感受野,提高模型对不规则形状病害的定位和适应能力。在三七叶片病害数据集上对模型进行大量测试,并将 TD-BlendMask 与基准模型及其他先进的实例分割模型[28-32]进行对比,以验证本文方法的有效性。

1　材料与方法

1.1　数据采集与标注

三七叶片病害数据集采集于云南省红河州泸西县的昆明理工大学研究示范基地。本文使用相机手动拍摄了共计 1 886 幅图像，展示了三七叶片在多种光照条件下的形态及其不同的疾病表现形式。如图 1 所示，该数据集涵盖三七叶片及其 3 种常见病害：灰霉病、炭疽病和疫病。

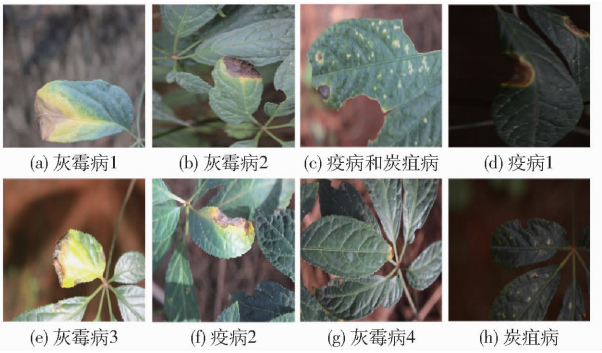


图 1　三七叶片病害示例图像

Fig. 1　Leaf diseases of *Panax notoginseng* under different light conditions

数据集的构建包含 3 个步骤：①数据裁剪和划分。从采集的数据中去除过于模糊且导致人为无法区分病害类型的图像，筛选出具有代表性的病害图像 300 幅，并将它们统一裁剪为分辨率 640 像素 × 640 像素图像。②数据标注。使用开源工具 LabelMe 对这些实例进行手动标注，包括三七叶片、灰霉病、炭疽病和疫病，如图 2 所示。这些实例的边界是通过多边形来明确描绘图像中三七叶片及各类病害的边界。③数据增强。为了避免网络训练中的过拟合问题和提高模型的泛化能力，采用旋转、翻转 2 种离线数据增强技术，来增强数据的多样性。本文将训练集 270 幅图像的样本数量增加至 1 345 幅，验证集为 270 幅图像，训练与验证比例为 8∶2。在训练集中 4 类标签具体实例数分别为：叶片 4 305 个、灰霉病 525 个、炭疽病 880 个、疫病 1 075 个。此外，在模型训练阶段，还利用 Detectron2 中的在线增强工具的随机亮度、对比度调整和加入高斯噪声等方法，进一步扩充数据集，从而提高模型泛化性能。

在本文的数据集中，三七叶片病害表现出 3 个富有挑战性的特性。首先叶子之间的重叠和遮挡对模型进行准确分割构成了障碍。其次，疫病和灰霉病颜色特征非常相似，容易分割错误。最后，炭疽病相较于叶子和其他疾病而言，其像素区域较小且形状复杂多样。此外，低照度条件下采集的图像也会造成叶片和病害分割困难。



图 2　数据集标注

Fig. 2　Manual labeling of *Panax notoginseng* leaves and diseases

1.2　总体框架

本文提出了一种结合 ResNet－101 的 BlendMask 模型。BlendMask[33] 是一个经典的实例分割方法，通过混合机制结合位置敏感的空间信息与语义特征来生成实例掩码。这一模型的优势包括能够准确划分对象边界、保留复杂的对象细节，并通过与各种检测框架的无缝整合支持可扩展的实例分割。尽管 BlendMask 在自然场景图像数据集上表现出色，但对于更为复杂的三七叶病害图像而言仍有不足。本文对 BlendMask 进行了一系列改进，使其更适合于三七叶片病害实例分割，其网络结构图如图 3 所示。

本文设计框架是一个紧密集成的端到端实例分割方法，它由 3 个主要部分构成：作为特征提取器的主干网络，负责识别疾病位置检测网络，以及结合空间细节和语义理解的掩码分支。主干网络提取全面的特征，而基于 FCOS 方法[34] 的检测头则专注于精确定位三七叶片和相关病害的位置。掩码分支将低层空间细节与高层语义理解相结合，以生成特定目标的实例掩码。具体来说，这个过程从主干网络提取多尺度的语义特征开始，然后将这些特征图传送到任务特定的分支，如 FCOS 检测头和分割头，以提高精度和效率。检测头将边界框（Bbox）分配给每个检测到的对象和实例框注意力模块，通过单层卷积层，预测每个检测头的主要关注区域，有助于编码对象的具体形状和方向等实例细节，从而实现更精细的分割。本文模型借鉴了其他基于区域建议的实例分割策略，如 DeeplabV3[18]，作为解码模块来预测得分图。最后，Blender 模块处理并输出最终的预测结果。

1.3　特征提取网络

采用 ResNet－101 模型，结合 Transformer encoder 模块和可变形卷积网络（Deformable convolutional neural network v2）。这些模块显著提升了网络在识别复杂模式和空间变化方面的能力。将 Transformer encoder 模块与 ResNet－101 模型结合，具体是将其融入在 ResNet 模型最后一层后，充分利用其处理全局数据依赖性的能力。Transformer

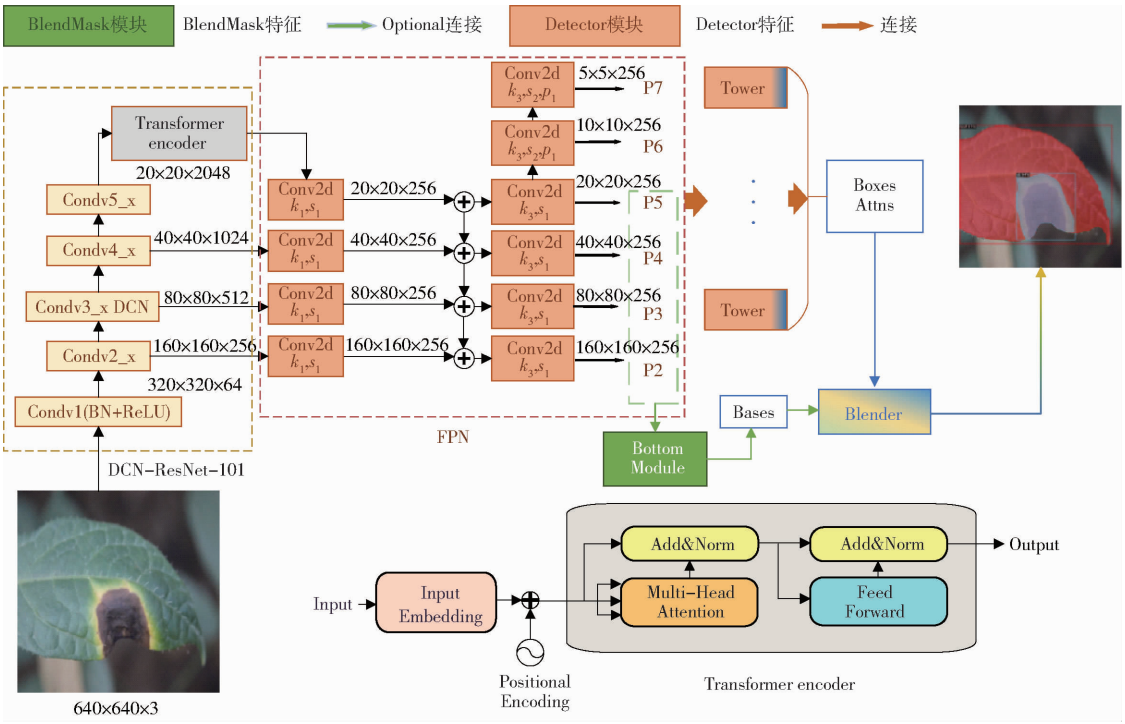


图 3 TD－BlendMask 总体框图

Fig. 3 Overall framework for TD－BlendMask

encoder 模块自注意力机制能够捕捉图像中的局部和全局依赖关系，这使得模型在处理三七病害图像时能够更好地理解和区分各种病害特征，例如叶片损伤、斑点或其他可见症状。同时，因为三七病害特征形状和尺寸可能不规则，将 ResNet 模型第 3 层 3×3 卷积块替换为 DCNv2，提高了模型对输入几何变化的适应性。DCNv2 通过对卷积核形状进行动态调整，可以更好地适应图像中各种病害形状和尺寸，骨干网络对比如表 1 所示。

表 1 改进的骨干网络对比

Tab. 1 DCN－ResNet－101 *vs* original ResNet－101

层名称	输出尺寸	模型	
		ResNet－101	DCN－ResNet－101
Conv1	112×112	7×7, 64, stride 2	
		3×3 max pool, stride 2	3×3 max pool, stride 2
Conv2_x	56×56	$\begin{pmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{pmatrix}\times3$	$\begin{pmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{pmatrix}\times3$
Conv3_x	28×28	$\begin{pmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{pmatrix}\times4$	$\begin{pmatrix}1\times1,128\\DCN,128\\1\times1,512\end{pmatrix}\times4$
Conv4_x	14×14	$\begin{pmatrix}1\times1,256\\3\times3,256\\1\times1,1\ 024\end{pmatrix}\times23$	$\begin{pmatrix}1\times1,256\\3\times3,256\\1\times1,1\ 024\end{pmatrix}\times23$
Conv5_x	7×7	$\begin{pmatrix}1\times1,512\\3\times3,512\\1\times1,2\ 048\end{pmatrix}\times3$	$\begin{pmatrix}1\times1,512\\3\times3,512\\1\times1,2\ 048\end{pmatrix}\times3$
	1×1	average pool, 1000-d fe, softmax	

1.4 BlendMask

由于三七病害数据的不同类别病害存在特征重叠的情况，特征相似的病害容易影响模型的先验框标签分类准确性，导致分割错误。BlendMask 通过融合更丰富的实例信息，并进行细致的位置敏感掩码预测，增强了基于区域建议的掩码方法。BlendMask 专注于编码 2 种不同类别的信息：①高层次语义特征，通常被称为语义掩码。②位置敏感的低层次特征。高层次的语义掩码通过其广阔的感受野，能够有效识别和区分出针对特定病害类型的像素。例如，它能区别出炭疽病的黑色斑点与叶片的其他部分，即使这些斑点在视觉上与背景或健康组织相近。低层次特征的位置敏感性对于三七病害特别重要，因为病害的斑点可能尺寸不一且形态各异。BlendMask 通过精确地捕捉这些特征的空间细节，帮助模型分辨即使是非常接近或部分重叠的斑点，确保了对每个病斑的准确识别和分割。BlendMask 的掩码分支包括 3 个主要模块：①底层模块专注于得分图的预测，解释底层的语义信息。②顶层模块负责预测实例的注意力，有助于辨别独特对象。③负责融合得分和注意力的 Blender 模块，将两层信息协调一致，以生成精确的实例掩码。

（1）底层模块：被称为基础 B 的底层模块专注于预测得分图。B 的维度为 $N\times K\times H/s\times W/s$，其中 N 代表批次大小，K 为基数的数量，s 为得分图的输出步长，H 和 W 分别为输入的高度和宽度。由于

其固有的灵活性和独立于全卷积网络(FCN)的特点，该模块特别擅长生成高分辨率的掩码。因此，BlendMask 能够产生具有更准确边缘的掩码，显示出其广泛的通用性和灵活性。

(2)顶层注意力模块：在顶层注意力部分，每个检测单元增加了一个额外的卷积层，专门用于预测注意力并有效地编码实例信息。对于每个分辨率为 $W_l \times H_l$ 的金字塔层，掩码系数呈现 $N \times K \times W_l \times H_l$ 的结构。$\boldsymbol{A}$ 代表与每个位置相对应的张量，其结构为 $N \times (K \times M \times M) \times W_l \times H_l$，其中 $M \times M$ 表示注意力的分辨率。$\boldsymbol{A}$ 通过一个具有 $K \times M \times M$ 输出通道的卷积来预测。在融合到后续模块前，进行后处理步骤，主要目的是锁定前 P 个框预测($P = \{p_d \in \mathbf{R}^4_{\geqslant 0} \mid d = 1,2,\cdots,D\}$)及其匹配的注意力($\boldsymbol{A} = \{a_d \in \mathbf{R}^{K \times M \times M} \mid d = 1,2,\cdots,D\}$)。变量 a'_d 是通过将 a_d 的尺寸从 $M \times M$ 插值到 $R \times R$ 得到的。注意力尺寸 M 通常比 R 小。接着，应用 softmax 函数结合 K 维度进行归一化，从而形成一组独特的得分图 s_d。

$$s_d = \text{softmax}(a'_d) \quad (\forall d \in \{1,2,\cdots,D\}) \tag{1}$$

其中

$$a'_d = \text{interpolate}_{M \times M \to R \times R}(a_d) \quad (\forall d \in \{1,2,\cdots,D\}) \tag{2}$$

式中　$\text{interpolate}_{M \times M \to R \times R}(\,)$——将原始尺寸为 $M \times M$ 的特征图通过插值调整到目标尺寸 $R \times R$

$\text{softmax}(\,)$——归一化函数，用于计算所有通道的 softmax 值，使得每个位置权值和为 1

(3)Blender 模块：Blender 模块充当整合工具，将实例级数据与密集的像素特征相结合。通过运用位置敏感基础的能力，Blender 模块精心打造最终的预测结果，这些预测在很大程度上依赖于注意力度量。接下来，Blender 模块把底层基础 B 与精心选择的顶层注意力 $\boldsymbol{A}$ 结合在一起，这个结合过程还包括边界框提议。运用 Mask R－CNN 中的核心组件 RoIPooler，每个提议 p_d 被精确地从基础中裁剪出来。这部分裁剪后的内容随后被调整尺寸以符合固定的 $R \times R$ 维度，形成特征图 r_d。在训练阶段，提议使用真实框，而在推理阶段则基于 FCOS 的预测。为了得出掩码逻辑 m_d，对 r_d 和 s_d 进行逐元素乘法计算。该乘积最终在 K 维度上进行求和，即

$$m_d = \sum_{k=1}^{K} s_d^k r_d^k \quad (\forall d \in \{1,2,\cdots,D\}) \tag{3}$$

$$r_d = \text{RoIPool}_{R \times R}(B, p_d) \quad (\forall d \in \{1,2,\cdots,D\}) \tag{4}$$

式中　$\text{RoIPool}_{R \times R}(\,)$——将输入的候选区域通过池化操作转换为固定尺寸 $R \times R$ 的特征图

B——基础特征图

p_d——第 d 个候选区域坐标信息

1.5　Transformer encoder

三七病害数据集中的病害，如炭疽病表现为小斑点，这些斑点广泛分布在叶片的各个部分。这些病害特征的多样性要求模型能够有效识别和区分局部细节与全局特征的复杂关系。为了解决这一问题，在 ResNet 模型架构的最终阶段引入 Transformer 编码器提取全局特征。通过这种方式，增强了网络在捕捉输入图像中全局上下文细节方面的能力。图像特征张量在通过 ResNet 模型的 res5 阶段后被调整，以适应 Transformer 的输入格式，然后传递至 Transformer 编码器。Transformer 作为一种革命性的深度学习架构，不仅在自然语言处理领域产生了巨大影响，而且在计算机视觉等领域也显示出巨大潜力。它的核心组件包括一个自注意力机制，这一机制允许序列中的每个输入元素与序列中的所有其他元素相关联，从而使模型能够捕捉序列内的各种依赖关系，有效捕捉远距离空间信息。在自注意力计算过程中，模型首先形成查询(Query)、键(Key)和值(Value)向量 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，然后通过点积注意力机制确定权重并集成输出。为了更好地捕获输入序列中的多个信息片段，Transformer 采用多个注意力头来同时进行自注意力计算，每个注意力头都有其独立的参数，以捕获不同方面的信息，其结构如图 4 所示。然后将各个头的输出连接起来，并通过线性层进行整合。尽管 Transformer 非常强大，但它缺乏固有的循环或卷积功能。为了弥补这一点，同时识别序列中固有的关系，位置向量被加入到输入嵌入中，赋予模型对序列中元素位置的敏感性。通过将位置向量添加到输入嵌入中，实现位置编码，让模型能够感知序列中元素的相对位置。Transformer 架构由多

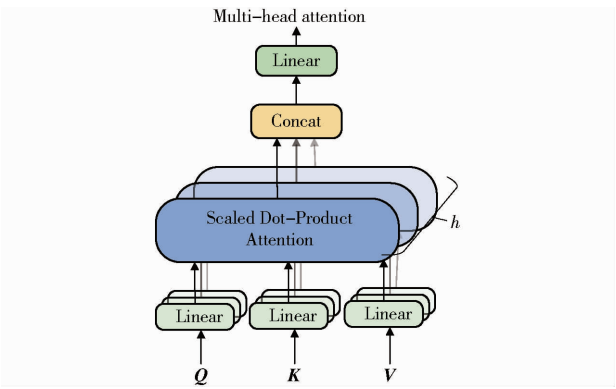


图 4　多头注意力机制结构

Fig. 4　Multi-head attention architecture

个编码器和解码器层叠加而成,每个编码器层包含一个多头注意子层和一个前馈神经网络(FFNN)子层,并采用残差连接和层级归一化。解码器层除了具有类似编码器的结构外,还增加了一个额外的多注意子层来关注编码器的输出。本文主要使用 Transformer 编码器部分来处理图像特征,捕捉全局上下文信息,从而提升 ResNet 模型表现。

1.6 可变形卷积

三七叶图像呈多样的叶子类型、形状和尺寸,使得对疾病进行准确分类变得相当复杂。此外,一些疾病在叶子上的表现相似,增加了分割难度。面对这些复杂性,需要一种能够有效识别不同叶病类的卷积神经网络(CNN),这样的网络需要充分利用多样化训练数据集。然而,考虑到 CNN 的固定几何结构,包括卷积核、采样位置和池化比率,存在一个实际的风险,即感受野保持不变,无法适应实例的不同病害尺度和轮廓。这种局限可能会限制 CNN 在特征提取方面的能力,从而导致其在叶病识别方面的准确性下降。因此可变形卷积应运而生,其结构如图 5 所示。

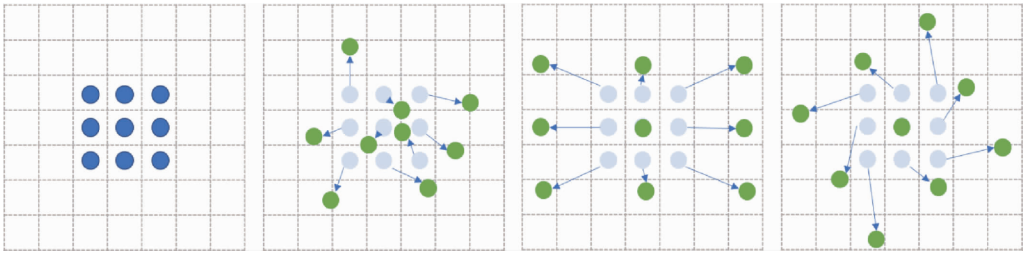


图 5 可变形卷积核

Fig. 5 Deformable convolutional network

在处理每个输入特征图时,本文在标准卷积过程中对几个不同位置的像素点 P_0 执行了 $y(P_0)$ 的计算。这个步骤重点关注像素点在各个位置的特征,以确保卷积能有效地识别和处理图像中的关键特征。

$$y(P_0)=\sum_{P_n\in R_f} w(P_n)x(P_0+P_n) \tag{5}$$

式中 $w()$——权重函数在偏移位置处权重

$x()$——输入在指定位置处的值

P_n——偏移距离

可变形卷积通过引入偏移量 ΔP_n 和权重 $\Delta m_n\in[0,l]$ 来改进计算过程。计算式为

$$y(P_0)=\sum_{P_n\in R_f} w(P_n)x(P_0+P_n)y(P_0)=\sum_{P_n\in R_f} w(P_n)x(P_0+P_n+\Delta P_n)\Delta m_n \tag{6}$$

三七病害数据集中的图像可能因为自然因素(风、生长姿态等)或拍摄角度导致植物部分的形态出现变形。DCNv2 模型通过其可变形卷积层和池化层,能够自适应这些几何变形,比如自动调整卷积核的形状来匹配叶片上病斑的实际形状和方向,从而更准确地捕捉和分析病害特征。同时,DCNv2 模型通过可变形层增强了网络对这些关键区域的感知能力,使得网络更倾向于聚焦于这些重要的感兴趣区域(RoI),而不是被背景或非病害区域干扰。DCNv2 模型在原始 DCN 模型基础上进行了改进。在 DCNv2 模型中,传统的可变形卷积操作被调制可变形卷积(MDC)所替代。MDC 模型不仅学习卷积核中每个元素的偏移量,还学习其调制标量,并且能在卷积过程中自适应调整每个元素的贡献。这使得模型能够更精准地聚焦于三七病害图像中关键的局部特征,如特定病斑或感染区域,即便这些区域的尺寸、形状和位置在不同图像之间有所不同。此外,DCNv2 模型对可变形 RoI 池化进行了改进,以适应调制学习,使得在池化过程中能够调整 RoI 中每个元素的贡献,从而使池化过程更加灵活和适应性强,能够帮助模型更好地理解和提取各种类别的病害特征,无论出现在图像的任何位置。

2 实验设计

2.1 实验环境

实验在一台配备 Intel Core i7 - 9700K CPU 和 NVIDIA GeForce RTX 3090 GPU 的计算机上运行,使用 Ubuntu 22.04.2 LTS 操作系统。在深度学习框架方面,本文选用了 Python 3.8 环境下的 PyTorch 1.8.2,并与 CUDA 11.1 集成。实验设置的批处理大小为 16,学习率为 0.025。另外,本文在 FCOS 检测头中应用了 GIoU 损失函数来优化模型。

2.2 评估指标

为了客观衡量模型在实例分割任务上的表现,采用平均精度(AP)、AP_{50} 和 AP_{75} 等指标来量化评估。其中 AP 表示不同目标类别上平均精度;AP_{50} 表示 IoU 阈值为 0.5 时平均精度;AP_{75} 表示 IoU 阈值为 0.75 时平均精度;AP_s 表示在小物体类别上平均精度;AP_m 表示在中等物体类别上平均精度;AP_l 表示在大物体类别上平均精度;AP_{leaf} 表示在叶片类

别上平均精度；$AP_{greymold}$表示在灰霉病类别上平均精度；$AP_{Anthracnose}$表示在炭疽病类别上平均精度；AP_{Plague}表示在疫病类别上平均精度。

3　实验结果与分析

为评估 TD－BlendMask 在三七叶片病害实例分割方面的性能，以三七叶片病害数据集为基础开展一系列实验。此过程中，TD－BlendMask 与其他实例分割方法，如 Mask R－CNN、BoxInst、ConInst、SOLOv2 和 YOLO v8－seg 进行了比较分析。同时，本文还探讨了不同模块对最终分割效果的影响，以进一步优化算法性能。

3.1　消融实验

TD－BlendMask 模型是在 BlendMask 的基础上改进而来的。为了验证本文中各个模块的有效性，主要进行了 2 个改进。首先，在骨干网络 ResNet－101 的第 5 层之后添加 Transformer encoder，并以此为基础，将 ResNet－101 第 3 层的 3×3 传统瓶颈块替换为更为灵活的可变形瓶颈块(DCN)。

TD－BlendMask 模型在三七叶片病害数据集上进行训练和测试。根据图 6 显示，训练过程中最初的 500 次迭代中损失值下降最快，此后损失值逐步减少。当训练迭代次数达到 8 000 时，损失曲线变得更加平缓，表明模型的准确度已达到稳定状态，训练过程也相应结束。TD－BlendMask 的损失函数相较于基准模型降低得更快，这体现了改进的有效性。

表 2 展示了在三七叶片病害数据集上的消融实验结果。实验结果表明，T－BlendMask 模型平均精度(AP)为 83.72%，与 BlendMask 模型相比高 0.75 个百分点，这说明 Transformer encoder 的融入有效提升了模型的分割能力。灰霉病和疫病 AP 分别提升 0.29 个百分点和 1.5 个百分点，表明将 Transformer encoder 集成于骨干网络之后，通过自注意力机制(Self－Attention)能够捕捉三七病害图像的全局依赖性，从而有效区分特征类似病害。炭疽病 AP 提升 0.75 个百分点，则说明由于 Transformer encoder 擅长处理复杂和细微的依赖关系，它在实例分割任务中适合处理病害图像中的炭疽病这种小尺寸或形态复杂的对象。相较于基准 BlendMask 模型，DCNv2－BlendMask 的 AP 从 82.97%增至 85.28%，在叶片、灰霉病、炭疽病、疫病分割 AP 上分别提升 2.13 个、0.96、2.27、3.90 个百分点。这表明通过将 ResNet－101 的第 3 层 3×3卷积瓶颈块替换为可变形瓶颈块，使得卷积操作能够动态调整其感受野，以适应三七病害图像内容的复杂边缘形状。DCNv2 的动态卷积核位置调整功能，还能有效应对三七叶片病害图像中各类病害尺度变化，这对于处理不同尺寸的同一类病害例如灰霉病至关重要。3 种病害在叶片上均可能出现不同尺寸，而 DCNv2 能够帮助模型在各个尺度上均保持高效的识别和分割能力，无需依赖额外尺度处理策略。

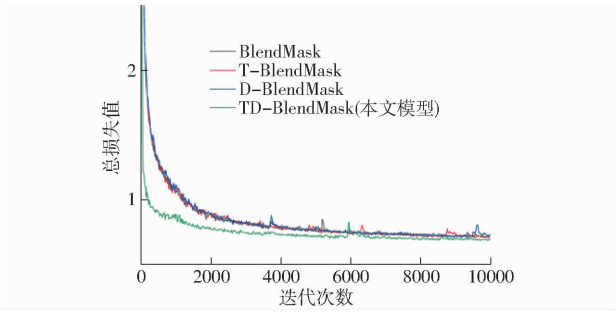


图 6　损失函数曲线

Fig. 6　Loss curve graph for model training

表 2　消融实验结果

Tab. 2　Ablation studies on all the proposed components at different IoUs and labels　%

模型	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l	AP_{leaf}	$AP_{greymold}$	$AP_{Anthracnose}$	AP_{Plague}
BlendMask	82.97	98.34	94.26	72.95	85.75	91.34	74.73	97.40	79.68	80.06
T－BlendMask	83.72	98.78	93.12	72.54	86.59	93.97	75.20	97.69	80.43	81.56
D－BlendMask	85.28	98.52	94.55	75.31	88.21	93.91	76.86	98.36	81.95	83.96
TD－BlendMask	86.14	98.56	95.94	77.76	88.05	93.61	78.47	97.56	84.14	84.38

总体而言，TD－BlendMask 的 AP 达到最高的 86.14%，比 BlendMask 模型提高 3.17 个百分点。TD－BlendMask 的 AP_{75}和 AP_s 也取得了最好的结果，相比 BlendMask 分别提高 1.68 个百分点和 4.81 个百分点。然而，TD－BlendMask 的 AP_{50}、AP_l 略低于 T－BlendMask，AP_m 和叶片标签略低于 D－BlendMask。这可能是因为 Transformer 和 DCNv2 主要提高了 TD－BlendMask 在小目标以及精细边缘分割方面的能力，而稍微降低了对大目标的分割能力，但这种轻微的损失是可以接受的。总的来说，BlendMask 的分割效果随着 Transformer encoder 和 DCNv2 的加入而提高。

图 7 中展现了在本文数据集上 BlendMask、T－BlendMask、D－BlendMask 和 TD－BlendMask 模

型的实例分割效果。在实例边缘处理和小目标分割方面，BlendMask 显示出了一些局限，但 TD－BlendMask 模型成功避免了 BlendMask 在这些方面的一些常见错误。从这些例子可以看出，本文所提出的 TD－BlendMask 在分割性能上相较于基准的 BlendMask 有了显著提高。

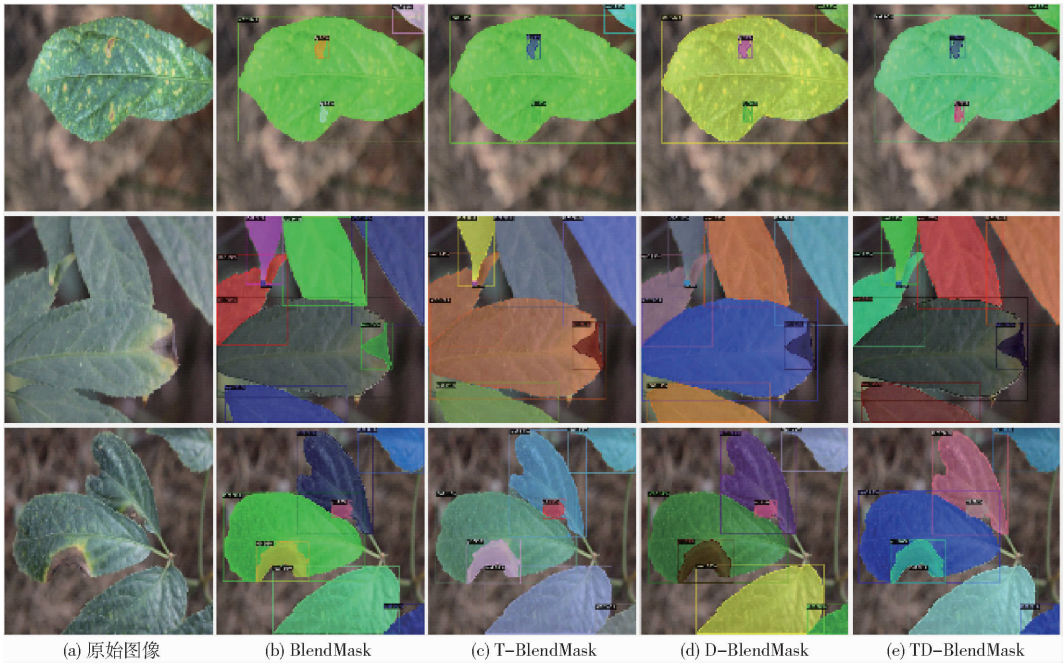


图 7 消融实验可视化结果

Fig. 7 Visualizing the instance segmentation result of ablation studies

3.2 对比实验

通过与 BoxInst、CondInst、SOLOv2、Mask R－CNN 和 YOLO v8－seg 实例分割模型比较实验，对比了不同模型在三七叶片病害数据集上的分割性能。为保证实验结果可比性和一致性，所有实验都在同一环境设置下进行。

表 3 展示了不同实例分割方法在三七叶片病害数据集上的性能表现，TD－BlendMask 算法在三七叶病害数据集上 AP 达到 86.14%，与单阶段模型 BoxInst、YOLO v8－seg、CondInst 和 SOLOv2 相比，分别提高 23.8、14.54、6.43、5.62 个百分点。与两阶段模型 Mask R－CNN 相比，则提高 4.37 个百分点。AP_{75}要求模型在更高的 IoU 阈值（即 0.75）下表现良好，模型需要更精确地预测三七病害目标的边界。虽然 Mask R－CNN 在生成掩模时相对精确，但其掩模精细度可能不足以在所有情况下满足高 IoU 的需求且其在处理病害与正常叶片之间的细微边界时可能不如 TD－BlendMask 模型。这是因为 Transformer 的全局视角和 DCNv2 的空间自适应特性可以更好地处理这些复杂的边界和细节问题。然而，Mask R－CNN 模型的 AP_{50}表现相比本文所提出方法高 0.08 个百分点，主要因为 AP_{50}衡量在较低 IoU 阈值（即 0.5）时模型的精度，此时主要考虑模型对三七病害目标大致位置的检测能力。由于 Mask R－CNN 在生成区域建议时较为精确，因此 AP_{50}表现出色。此外，SOLOv2 的 AP_l和灰霉病的 AP 达到最高，分别为 94.21% 和 97.96%，因为通过其独特的密集预测方法和空间编码机制，在大尺寸叶片实例的检测上表现出色。然而由于小尺寸病害实例的像素数量少，它们的类别和空间信息可能不足以使 SOLOv2 的算

表 3 对比实验结果

Tab. 3 Comparative experimental results %

模型	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l	AP_{leaf}	$AP_{greymold}$	$AP_{Anthracnose}$	AP_{Plague}
BoxInst	62.34	93.58	65.40	42.54	64.19	74.03	50.09	92.08	50.27	56.93
YOLO v8－seg	71.60						67.30	92.80	46.40	80.10
CondInst	79.71	97.55	91.61	65.58	82.82	91.84	68.16	96.90	76.12	77.65
SOLOv2	80.52	94.90	91.46	63.09	86.30	94.21	73.49	97.96	70.87	79.74
Mask R－CNN	81.77	98.64	95.16	72.97	83.77	88.90	76.53	96.09	77.41	77.04
TD－BlendMask	86.14	98.56	95.94	77.76	88.05	93.61	78.47	97.56	84.14	84.38

法有效地识别和分割，其表现不如本文所提出方法。TD－BlendMask 之所以优异，主要归功于其结合了低级细节信息与高级语义信息，并通过 Transformer encoder 模块整合上下文信息以及可变形卷积对复杂形状病害目标的特征提取，这充分证实了强大的主干网络和分割模型对于实现高质量图像分割的必要性。因此，在三七叶病害数据集上，TD－BlendMask 是一种准确的分割模型。

图 8 展现了 TD－BlendMask 模型及其他一些经典实例分割模型在本文数据集上的分割效果。从图像中可以清晰看出，其他方法在分割实例边缘和小目标时存在不足。通过观察图 8 可以发现，TD－BlendMask 模型能有效分割出不同疾病标签的大多数样本图像，而 SOLOv2 在所有样本图像的正确分割上却未能达标。SOLOv2 分割精度不高的主要原因有两点：其在准确勾勒目标边界上不足；它倾向于将单一实例误判为多个实例。相反，TD－BlendMask 凭借两大优势有效克服了这些问题，即结合低层细节信息和高层语义信息，以及先进行检测后进行分割的处理流程，与数据标注步骤保持一致。TD－BlendMask 通过结合来自不同方向的注意力，显著提高了类似目标和形状复杂目标的分割效果。图 9 展示了投射到原始图像上的完整基点和关注点。前 2 个基点（红色和蓝色）检测叶片或病害右上角和左下角的点。第 3 个基点（黄色）会激活更有可能位于叶片或病害上的点。第 4 个基点（绿色）只在叶片或病害的边界上激活。位置敏感特征可以帮助分离重叠的叶片或病害实例。与此相比，本文提出的方法在这两方面的表现更佳，凸显了其明显的优势。

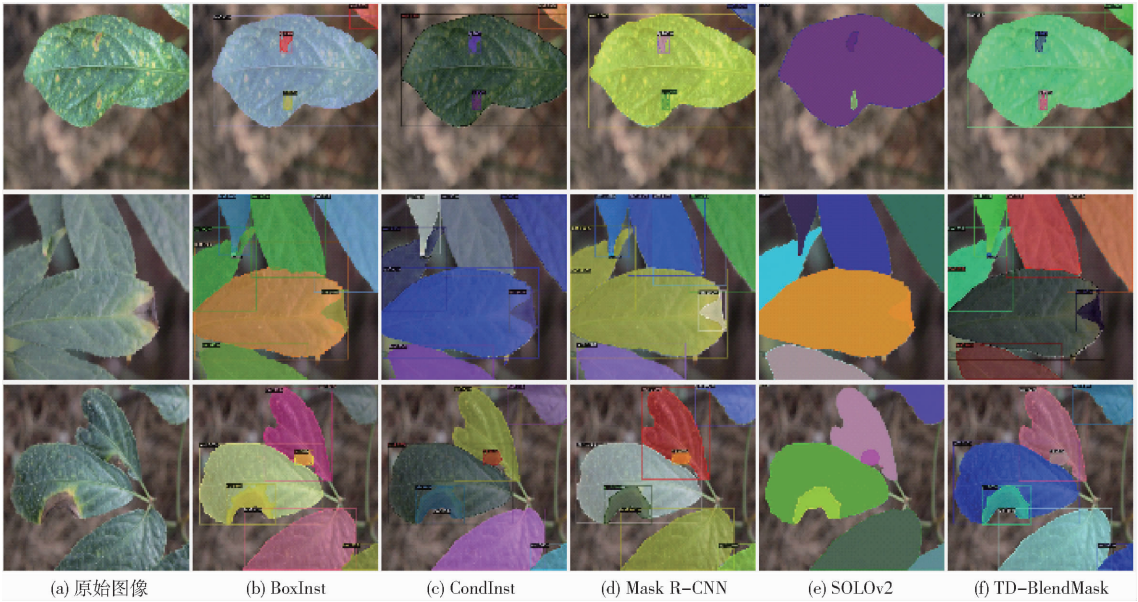


图 8　对比实验结果

Fig. 8　Comparison outputs of different methods for *Panax notoginseng* leaf disease image segmentation

尽管 TD－BlendMask 在大多数情况下表现出色，但如图 8f 所示，在某些图像样本中，它也表现出了分割的不足。观察图中第 2 排中间的情形，可以看到上面的叶子被中间的叶子遮盖，导致下方较小的叶尖未能被有效分割。其他几种模型也遇到了类似的问题。当顶部叶子被中间的叶子完全遮住时，算法难以判断这个小目标叶尖到底属于哪个实例，因此产生了分割上的错误。

3.3　在 plant and disease identification 公共数据集上的对比实验

为验证所提出的 TD－BlendMask 算法模型的泛化性，在 plant and disease identification[35] 多种植物和病害公开数据集上，开展与其他模型的比较实验。该数据集包含在实验室光照下共计 29 类病害标签的多种植物单叶片图像，共计 7 181 幅图像，训练集、验证集和测试集比例为 7∶2∶1，图像数量分别为 4 990、1 464、727 幅。数据集中部分图像样本如图 10 所示。

由表 4 可知，本文提出的 TD－BlendMask 模型的 AP、AP_{75} 和 AP_l 达到最佳，与 BlendMask 模型相比，分别提高 4.56、5.36、4.53 个百分点。上述结果表明，通过在 BlendMask 模型中融合 Transformer encoder 和 DCNv2，可有效解决模型在分割形状复杂多变以及复杂背景中的病害叶片实例的挑战。Mask R－CNN 作为一种二阶段的实例分割，在生成区域建议时较为精确，因此 AP_{50}、AP_s 和 AP_m 相比本文方法高 0.18、4.79、0.38 个百分点。由于区域建议的存在，其在小目标和中目标上表现较好，

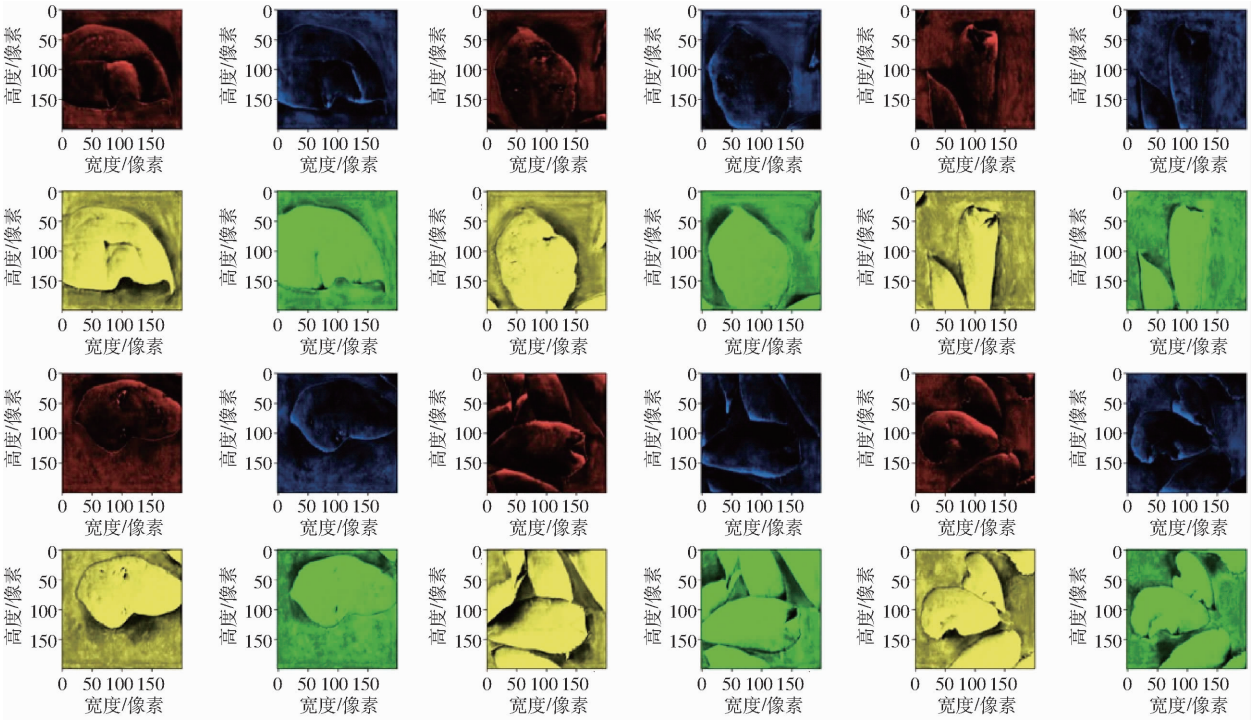


图 9 TD－BlendMask 模型混合注意力可视化

Fig. 9 Visualizing the instance segmentation result of the attention mechanism

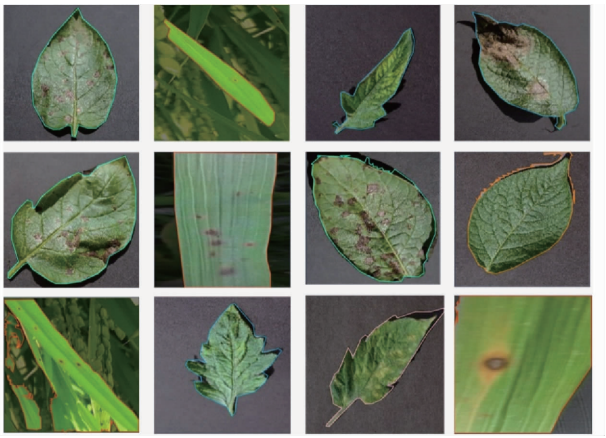


图 10 Plant and disease identification 公共数据集部分图像样本

Fig. 10 Plant and disease identification datasets

表 4 Plant and disease identification 数据集对比实验结果

Tab. 4 Results of comparison experiments on plant and disease identification datasets %

模型	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
BlendMask	65.52	74.65	67.56	0.23	4.60	66.91
CondInst	65.64	75.64	68.84	2.73	5.05	67.49
Mask R－CNN	66.53	78.33	71.94	5.14	6.13	67.56
BoxInst	67.44	74.82	70.07	0.51	5.74	69.92
SOLOv2	67.87	76.58	70.80	0.00	5.92	69.71
TD－BlendMask（本文模型）	70.08	78.15	72.92	0.35	5.75	71.44

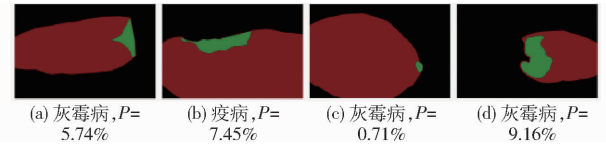
但因为其在面对复杂边缘的大目标上比本文所提出方法低 3.88 个百分点，其总体表现不如本文所提出的 TD－BlendMask 模型。因此，本文所提出的算法改进具有泛用性，可为其他作物病害图像分割提供参考。

3.4 三七叶片病害量化实验

三七病害的严重程度估算是制定预防和控制策略的依据之一。采用一种量化方法来评估三七叶片的病害程度 P，该方法基于同一叶片上病斑面积与总叶面积的百分比。具体而言，本文为计算病斑面积 S_{Disease}（用绿色表示）与叶片总面积 S_{Leaf}（红色部分代表健康组织，面积为 S_{health}；绿色部分代表病斑组织）的百分比。

$$P=\frac{S_{\text{Disease}}}{S_{\text{Leaf}}}\times 100\%=\frac{S_{\text{Disease}}}{S_{\text{health}}+S_{\text{Disease}}}\times 100\% \quad (7)$$

为了更直观地展示量化结果，图 11 展示了部分三七叶片病斑面积占比的计算结果。通过这种方法，能够定量地分析叶片上病害的分布情况，为进一步的病情指数计算提供重要的参考依据。叶片病害面积占比的准确计算，不仅有助于监测病害的发展动态，还能够为农业管理者提供科学的决策支持，从而更有效地实施防治措施。



(a) 灰霉病，P= 5.74%　(b) 疫病，P= 7.45%　(c) 灰霉病，P= 0.71%　(d) 灰霉病，P= 9.16%

图 11 三七叶片不同病斑面积占比

Fig. 11 Proportion of different diseased spots on leaves of *Panax notoginseng*

4　结束语

针对三七叶片病害中相似目标和形状多变的小目标分割的困难问题，本文提出了 TD－BlendMask 实例分割模型，旨在提高现有模型在三七叶片病害分割下的性能。研究结果显示，TD－BlendMask 模型在三七叶片病害分割数据集上表现出色，其平均精度（AP）达到 86.14%，在 plant and disease identification 公共数据集上达到 70.08% 的平均精度（AP），这凸显了其在处理在复杂背量下特征类似、形状复杂多变等叶片或病害目标的困难问题方面的巨大潜力。目前的研究初步证明了 TD－BlendMask 在有限数据集上超越现有模型的能力。

参　考　文　献

[1] XU C, WANG W, WANG B, et al. Analytical methods and biological activities of *Panax notoginseng* saponins: recent trends [J]. Journal of Ethnopharmacology, 2019, 236: 443－465.

[2] ZHANG Y, ZHENG Y, XIA P, et al. Impact of continuous *Panax notoginseng* plantation on soil microbial and biochemical properties[J]. Scientific Reports, 2019, 9(1): 13205.

[3] LI J, BAO Y, WANG Z, et al. Research progress in diseases of *Panax notoginseng*[J]. Physiological and Molecular Plant Pathology, 2022, 121: 101878.

[4] BONKRA A, PATHAK S, KAUR A, et al. Exploring the trend of recognizing apple leaf disease detection through machine learning: a comprehensive analysis using bibliometric techniques[J]. Artificial Intelligence Review, 2024, 57(2): 21.

[5] LEI L, YANG Q, YANG L, et al. Deep learning implementation of image segmentation in agricultural applications: a comprehensive review[J]. Artificial Intelligence Review, 2024, 57(6): 149.

[6] MA J, DU K, ZHANG L, et al. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing[J]. Computers and Electronics in Agriculture, 2017, 142: 110－117.

[7] 李翠玲，李余康，谭昊然，等. 基于 K-means 聚类和 RF 算法的葡萄霜霉病检测分级方法[J]. 农业机械学报，2022，53(5)：225－236.
LI Cuiling, LI Yukang, TAN Haoran, et al. Grading detection method of grape downy mildew based on K-means clustering and random forest algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 225－236. (in Chinese)

[8] SINGH V, MISRA A K. Detection of plant leaf diseases using image segmentation and soft computing techniques[J]. Information Processing in Agriculture, 2017, 4(1): 41－49.

[9] PICON A, EGUSKIZA I, GALAN P, et al. Crop-conditional semantic segmentation for efficient agricultural disease assessment [J]. Artificial Intelligence in Agriculture, 2025, 15(1): 79－87.

[10] 李凯雨，朱昕怡，马浚诚，等. 基于混合扩张卷积和注意力的黄瓜病害严重度估算方法[J]. 农业机械学报，2023，54(2)：231－239.
LI Kaiyu, ZHU Xinyi, MA Juncheng, et al. Estimation method of leaf disease severity of cucumber based on mixed dilated convolution and attention mechanism[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 231－239. (in Chinese)

[11] 邓阳君，王旭，龙陈锋，等. 基于 VCDM－UNet 的水稻叶瘟病斑分割和病害程度分级[J]. 农业工程学报，2024，40(23)：190－198.
DENG Yangjun, WANG Xu, LONG Chenfeng, et al. Segmenting and grading the blast disease of rice leaves using VCDM－UNet[J]. Transactions of the CSAE, 2024, 40(23): 190－198. (in Chinese)

[12] 钟昌源，胡泽林，李淼，等. 基于分组注意力模块的实时农作物病害叶片语义分割模型[J]. 农业工程学报，2021，37(4)：208－215.
ZHONG Changyuan, HU Zelin, LI Miao, et al. Real-time semantic segmentation model for crop disease leaves using group attention module[J]. Transactions of the CSAE, 2021. 37(4): 208－215. (in Chinese)

[13] KITZLER F, BARTA N, NEUGSCHWANDTNER R W, et al. WE3DS: an RGB－Dimage dataset for semantic segmentation in agriculture[J]. Sensors, 2023, 23(5): 2713.

[14] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3431－3440.

[15] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230－6239.

[16] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834－848.

[17] WANG Z, YANG L, WANG R, et al. WE－DeepLabV3＋: a lightweight segmentation model for *Panax notoginseng* leaf diseases[J]. Computers and Electronics in Agriculture, 2024, 227: 109612.

[18] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder - Decoder with atrous separable convolution for semantic image segmentation[C]// European Conference on Computer Vision, 2018: 11211.

[19] 胡晓波, 许桃胜, 黄伟, 等. 交互式双分支特征融合的草莓病害程度快速诊断方法[J]. 农业机械学报, 2023,54(11): 225 - 235.
HU Xiaobo, XU Taosheng, HUANG Wei, et al. Interactive bilateral feature fusion network for real-time strawberry disease diagnosis[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(11):225 - 235. (in Chinese)

[20] ZHANG S, ZHANG C. Modified U - Net for plant diseased leaf image segmentation[J]. Computers and Electronics in Agriculture, 2023, 204: 107511.

[21] ZHU S, MA W, LU J, et al. A novel approach for apple leaf disease image segmentation in complex scenes based on two-stage DeepLabv3 + with adaptive loss[J]. Computers and Electronics in Agriculture, 2023, 204: 107539.

[22] TASSIS L M, TOZZI DE SOUZA J E, KROHLING R A. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images[J]. Computers and Electronics in Agriculture, 2021, 186: 106191.

[23] YAO N, NI F, WU M, et al. Deep learning-based segmentation of peach diseases using convolutional neural network[J]. Frontiers in Plant Science, 2022, 13: 876357.

[24] DAI F, WANG F, YANGD, et al. Detection method of citrus psyllids with field high-definition camera based on improved cascade region-based convolution neural networks[J]. Front. Plant Sci., 2022, 12: 816272.

[25] XU X, GENG Q, GAO F, et al. Segmentation and counting of wheat spike grains based on deep learning and textural feature [J]. Plant Methods, 2023,19: 77.

[26] LI S, LI B, LI J,et al. Semantic segmentation algorithm of rice small target based on deep learning[J]. Agriculture,2022; 12(8): 1232.

[27] WANG J, JIA J, ZHANGY,et al. RAAWC - UNet: an apple leaf and disease segmentation method based on residual attention and atrous spatial pyramid pooling improved UNet with weight compression loss[J]. Front. Plant Sci., 2024, 15: 1305358.

[28] TIAN Z, SHEN C, WANG X, et al. Boxinst: high-performance instance segmentation with box annotations[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 5439 - 5448.

[29] TIAN Z, SHEN C, CHEN H. Conditional convolutions for instance segmentation[C]// ECCV, 2020: 282 - 298.

[30] WANG X, ZHANG R, KONG T, et al. SOLOv2: dynamic and fast instance segmentation[C]// Neural Information Processing Systems, 2020: 17721 - 17732.

[31] HE K, GKIOXARI G, DOLLAR P, et al. Mask R - CNN[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2980 - 2988.

[32] VARGHESE R, SAMBATH M. Yolov8: a novel object detection algorithm with enhanced performance and robustness[C]// 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024: 1 - 6.

[33] CHEN H, SUN K, TIAN Z, et al. BlendMask: top-down meets bottom-up for instance segmentation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 8570 - 8578.

[34] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9626 - 9635.

[35] SINGH D, JAIN N, JAIN P, et al. PlantDoc: a dataset for visual plant disease detection[C]// Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, 2020: 249 - 253.