

基于注意力机制的改进 UNet 草莓病害语义分割^①



毛万菁, 阮炬全, 刘 朔

(武汉轻工大学 数学与计算机学院, 武汉 430048)

通信作者: 刘 朔, E-mail: 874477154@qq.com

摘 要: 针对当前传统农作物病害语义分割方法精度不高、鲁棒性差等问题, 本文提出了基于注意力机制的改进 UNet 草莓病害语义分割模型. 首先, 在编码器中加入 CNN-Transformer 混合结构, 增强全局信息与局部细节信息的特征提取能力. 其次, 在解码器中将 dual up-sample 模块替换传统上采样, 提高特征提取能力与分割精度. 再使用 hard-swish 激活函数代替 ReLU 激活函数, 更加平滑的曲线有助于提高泛化性和非线性特征提取能力, 防止梯度消失. 最后, 通过使用结合交叉熵 Dice 损失函数, 加强模型对分割结果的约束, 进一步提升分割精度. 实验采用了由 7 种草莓病害 2 500 张图像组成的数据集, 在复杂背景下对草莓病害进行分割, 语义分割像素精度达到 92.56%, 平均交并比达到 84.97%. 实验结果表明, 本文的改进 UNet 在草莓病害语义分割方面, 能实现更好的分割效果, 优于大多数分割模型.

关键词: 语义分割; 深度学习; 草莓病害; 注意力机制; UNet

引用格式: 毛万菁, 阮炬全, 刘朔. 基于注意力机制的改进 UNet 草莓病害语义分割. 计算机系统应用, 2023, 32(6): 251-259. <http://www.c-s-a.org.cn/1003-3254/9165.html>

Improved UNet Based on Attention Mechanism in Semantic Segmentation of Strawberry Diseases

MAO Wan-Jing, RUAN Ju-Quan, LIU Shuo

(School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430048, China)

Abstract: The existing traditional semantic segmentation methods of crop diseases have low accuracy and poor robustness. In order to address these problems, an improved UNet semantic segmentation model of strawberry diseases based on an attention mechanism is proposed. Firstly, a CNN-Transformer hybrid structure is added to the encoder to improve the feature extraction ability of global information and local detail information. Secondly, the traditional up-sampling is replaced by a dual up-sample module in the decoder to enhance the feature extraction ability and segmentation accuracy. Thirdly, the hard-swish activation function is employed to replace the ReLU activation function, and the smoother curve helps to improve generalization and nonlinear feature extraction ability and prevent gradient disappearance. Finally, the segmentation accuracy is further improved by using a combined cross-entropy Dice loss function to strengthen the model's constraints on the segmentation results. A dataset consisting of 2 500 images of seven strawberry diseases is used to segment strawberry diseases in a complex background. The semantic segmentation pixel accuracy reaches 92.56%, and the average cross-merge ratio reaches 84.97%. The experimental results show that the improved UNet in this study can achieve better segmentation results and outperform most segmentation models in the semantic segmentation of strawberry diseases.

Key words: semantic segmentation; deep learning (DL); strawberry disease; attention mechanism; UNet

^① 收稿时间: 2022-11-10; 修改时间: 2023-02-03; 采用时间: 2023-02-20; csa 在线出版时间: 2023-04-23

CNKI 网络首发时间: 2023-04-24

国家在“十三五”规划中已经明确提出,在农业领域中应用基础以及软件科学研究亟待强化,通过信息技术来促进农业技术的发展.为了使农业生产更加智能化和现代化,杰出的学者把图像分割技术运用到农业领域中,解决了病虫害防治和诊断不够及时的问题^[1].农作物的病害需要在早期识别出来,便于实施治疗措施,减少农作物的经济损失.因此,高精度且低成本的农作物病害诊断方法是农作物生产阶段不可或缺的需求.

近年来农作物图像病害分割常用技术主要有基于传统方法和基于深度学习方法的分割.传统语义分割方法是通过图像的纹理、颜色、几何形状以及空间结构信息来分割目标,把语义相同的像素划分到一个区域中,并且每个区域之间不存在交集^[2].传统语义分割方法主要可以分为基于阈值、基于聚类、基于边缘的分割和基于区域等类型^[3].但是传统方法的计算复杂度较高、耗时较长,不能有效地处理大规模图像,以及有噪声、多种物体、复杂背景等的复杂图像,因而分割结果准确率较低,适用范围受限.随着深度神经网络的出现,图像分割技术打破了传统模式进入了快速发展阶段.

Long 等人^[4]提出了全卷积网络 FCN (fully convolutional networks),将编码器、解码器引入了图像分割领域. Ronneberger 等人^[5]提出了全卷积网络改进得到的 UNet,将编码器和解码器阶段特征图拼接,形成梯形结构,让每个阶段的解码器都能学习编码器池化中丢失的特征. Zhao 等人^[6]提出了 PSPNet,该网络采用了金字塔池化模块 (pyramid pooling module),显著提高了全局信息特征提取的能力. Chen 等人^[7]提出了 DeepLabV3,使用空洞卷积提取特征,采用 ASPP 模块,扩大感受野,加强了特征提取能力. Vaswani 等人^[8]提出 Transformer 自注意力 (self-attention, SA) 机制,重点提取特征图中关注对象的特征信息,同时减少无用信息,提高特征提取效率. Chen 等人^[9]等人提出了 TransUNet (merit both Transformers and UNet),通过将 Transformer 融入 UNet 网络,以获得图像全局联系,实现特征图像多尺度预测与监督,兼具 Transformer 与 UNet 两者的优点.

综上所述,基于深度学习的语义分割模型在场景分割领域取得了良好的效果.然而以上方法应用于农作物病害语义分割场景时,易受季节、光照、环境等影响,且在复杂背景中进行准确分割具有极大挑战性.在编码过程中,缺乏局部信息交互,对图像边缘与几何形状等特征提取能力较弱.针对以上问题,本文提出

了一种改进 UNet 草莓病害语义分割网络,应用于草莓病害数据集.如图 1 所示,本文针对草莓角斑病叶、叶斑病、白粉病叶的叶片病斑区域进行分割,对花朵花腐病的腐烂部分进行分割,以及对炭疽病、灰霉病、果实白粉病的果实病变区域进行分割,根据不同的严重程度对叶片和果实的部分或整体区域进行标注与分割.在训练阶段,引入了交叉熵 Dice 损失函数,加强模型对分割结果的约束,进一步提升分割精度.在编码器中,加入 CNN-Transformer 混合结构,增强全局信息与局部细节信息的特征提取能力.在解码器部分,引入了 dual up-sample 模块,提高特征提取能力与分割精度.使用了 hard-swish 激活函数,提高泛化性和非线性特征提取能力,防止梯度消失.

1 数据来源

本文实验中草莓病害图像样本源于 Afzaal 等人^[10]发表的 strawberry disease detection 公开数据集.该数据集由 JBNU 计算机科学与工程部人工智能实验室收集,由 2500 张来自真实的田地、温室中收集的图像组成,引入多种挑战,如复杂背景、不同光照条件等.该数据集包含的 7 种草莓病害如图 1 所示.

2 改进的 UNet 模型

本节介绍了本文基于注意力机制的改进 UNet 草莓病害语义分割模型网络架构与模块详情,以及本文使用的 CNN-Transformer 混合结构、dual up-sample 模块、hard-swish 激活函数与交叉熵 Dice 损失函数.改进的 UNet 网络模型结构如图 2 所示.

2.1 CNN-Transformer 混合结构

传统的 CNN 卷积具有较好的局部感知能力,具备局部空间信息,不同的卷积核可具备不同的感受野,但缺点是无法综合全局信息,池化层会丢失部分特征信息,缺乏局部与整体的关联性.而 Transformer 网络具有 self-attention 结构,更注重全局信息的特征提取,但缺点是对细节信息的提取能力较差.

因此本文在编码器部分采用了 CNN-Transformer 混合结构, CNN 用作特征提取器,通过卷积操作提取特征学习,利用 CNN 获取详细高分辨率空间信息.再使用 Transformer 将 self-attention 机制引入编码器设计,通过 Transformer 的内部自注意力改善了卷积运算无法建立远距离模型的局限性.



图1 7种草莓病害图像

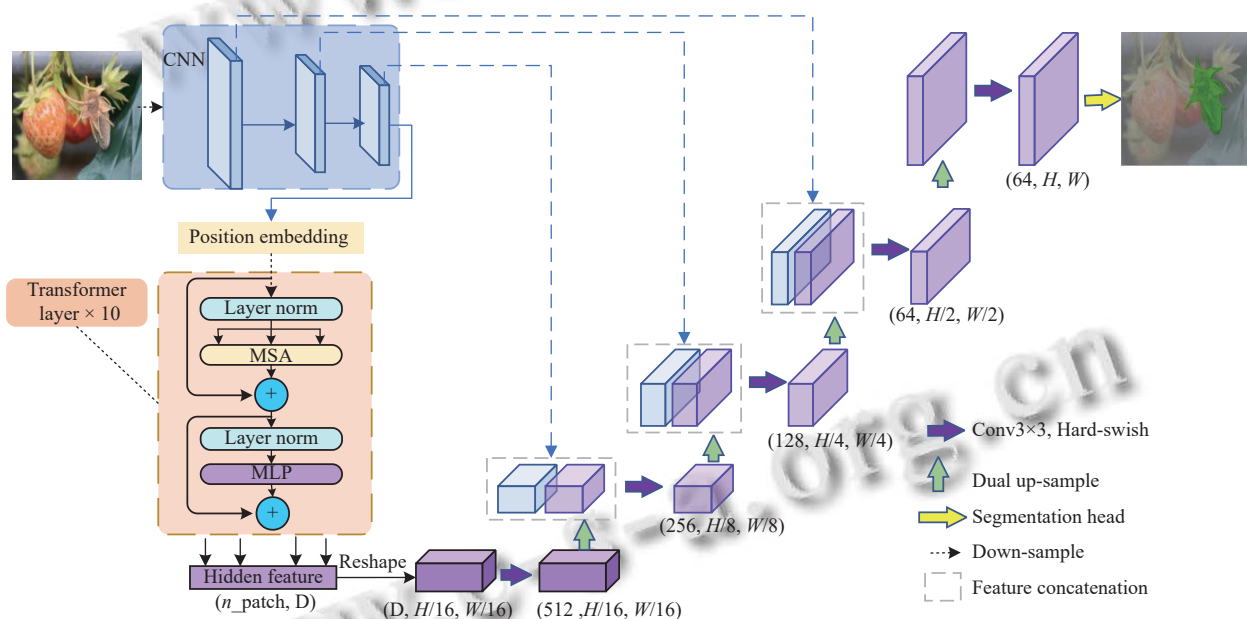


图2 改进的UNet网络结构图

首先将原图输入 CNN 进行特征提取,进行了3层卷积下采样,特征图相较原图缩小,变为原图的 1/2、1/4 与 1/8。然后将下采样后的图像输入 embedding 层。首先将图片分成一堆 patches,按给定大小划分后通过线性映射将每个 patch 映射到一维向量中输出一个向量序列,即二维矩阵,再输入 Transformer 层中循环 10 次。

具有全局自注意力的 Transformer 模块基于 vision Transformer^[11],包含 layer norm、多头自注意力模块

(multi-head self-attention, MSA)、多层感知机 (multi-layer perceptron, MLP)。其多头自注意力模块的运算公式为式 (1)。

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

2.2 Dual up-sample 模块

UNet 网络属于编码解码模型的框架,通过卷积操作提取特征学习,特征图相较原图像缩小,因此需要在 Decoder 解码端进行上采样恢复图像大小。普通的反卷

积 (deconv) 上采样会导致棋盘效应, 反最大池化 (max unpooling) 和亚像素卷积层 (pixel shuffle)^[12] 上采样方法又会破坏特征的连续性, 而更基础的双线性插值上采样 (bilinear upsampling) 又没有可学习的参数, 效果较差. 亚像素卷积层 (pixel shuffle) 的主要功能是将低分辨率特征图像, 通过多通道的卷积和重组来获得高分辨率的特征图, 有效改善棋盘效应. 双线性插值上采样相较于最邻近法结果更加光滑, 且速度快, 计算量小.

文献 [13] 采用双线性插值与反卷积构建上采样块, 以减少池化阶段下采样的负面影响, 运用于 OCR 研究中. 文献 [14] 中提出了双上采样块结构, 以防止棋盘效应, 运用于图像去噪任务中. 本研究将 dual up-sample 模块^[14] 代替 UNet 原本的上采样方案, 将亚像素卷积层 (pixel shuffle) 与双线性插值上采样 (bilinear upsampling) 两者融合, 并加入了 hard-swish 激活函数, 提升上采样时特征提取能力与图像边缘分割精度, 以防止棋盘效应, 弥补 Transformer 带来的特征分辨率的损失. 本文的 dual up-sample 模块如图 3 所示, 分别包含两个卷积层、hard-swish 层、亚像素卷积层、双线性插值, 由双通道构成的上采样结构.

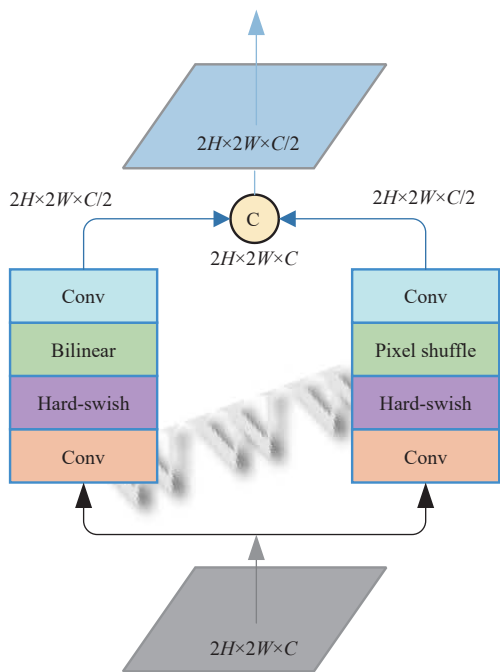


图 3 Dual up-sample 模块

2.3 Hard-swish 激活函数

传统神经网络中, 激活函数的选择对模型训练效率与性能有着显著影响, 一般使用最广泛的激活函数

是整流线性单元 (ReLU)^[15]. Ramachandran 等人^[16] 提出了 swish 激活函数, 在深层网络中 swish 的表现比 ReLU 更好. swish 结构简单, 与 ReLU 结构相似, 但是其缺点在于计算量较大.

$$hard-swish(x) = x \cdot \frac{ReLU(x+3)}{6} \quad (2)$$

本文采用了 Howard 等人^[17] 在 MobileNet V3 中提出的 hard-swish 激活函数, 具体公式为式 (2), 相较于 swish 有更好的数值稳定性和计算速度, 在实验中使用 hard-swish 非线性激活函数在准确率上无显著差异, 且在部署至嵌入式移动设备来看存在着多重的优点, hard-swish 与其导数 (hard-swish'(x)) 的函数图像如图 4 所示.

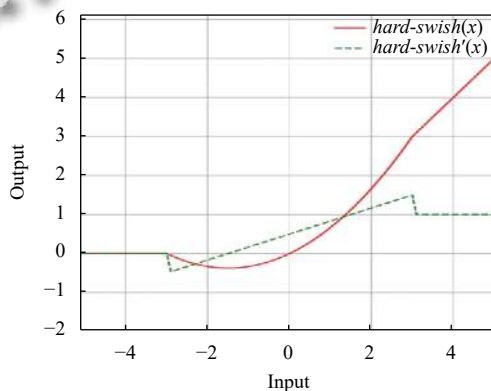


图 4 hard-swish 与其导数的函数图像

在实际应用中, 可将 hard-swish 激活函数实现成分段功能来减少内存访问的次数, 使等待时间大幅减少, 具体公式为式 (3).

$$hard-swish(x) = \begin{cases} 0, & \text{if } x \leq -3 \\ x, & \text{if } -3 < x < 3 \\ x(x+3)/6, & \text{otherwise} \end{cases} \quad (3)$$

2.4 交叉熵 Dice 混合损失函数

在神经网络训练阶段, 利用损失函数计算出每次迭代结果与实际值的差距, 以引导后续训练朝正确方向发展. 对损失函数的改进主要针对的是类不平衡的问题, 因此本文采用了融合交叉熵损失函数和 Dice 损失函数^[18] 结合的混合损失函数, 结合了交叉熵的稳定性和类不平衡, 并且不影响 Dice loss 的特性. 本文采用的混合损失函数有比 Dice loss 更优秀的稳定性, 且较交叉熵能更好地解决类不平衡的问题. 交叉熵损失函数和 Dice 损失函数的公式分别如式 (4) 与式 (5) 所示.

$$Loss_{cross} = \frac{1}{N} \sum_i - \sum_{c=1}^C y_i \lg(p_i) \quad (4)$$

交叉熵损失函数用来评价图像数据进行分割时像素点分类所造成的损失情况,它可以度量相同随机变量下两种不同概率分布之间的差值大小,数值越小说明该模型预测结果越理想.其中, C 为标签, y_i 指类别是否为 i ,若是则 $y_i = 1$,不是则 $y_i = 0$. p_i 是指样本 i 属于类别 C 的概率.

$$Loss_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

Dice 损失函数用于对预测分割图像和实际分割图像进行相似度度量损失评价,取值范围为 $[0, 1]$.其中 $|X \cap Y|$ 表示真实图片与预测的分割图像的交集, $|X|$ 与 $|Y|$ 表示各自的元素个数.

改进 UNet 模型的总损失函数为 $Loss_{Total}$, 公式如式 (6) 所示.

$$Loss_{Total} = \frac{Loss_{cross} + Loss_{Dice}}{2} \quad (6)$$

3 实验过程与对比

实验主要以本研究的改进 UNet 网络模型在草莓病害数据集上的训练过程以及预测结果来衡量本文模型的有效性,并采用语义分割的常用评价指标与其他主流网络模型进行对比.

3.1 评价指标

语义分割常用的评价指标有像素准确率 (pixel accuracy, PA)、类别平均像素准确率 (mean pixel accuracy, MPA)、平均交并比 (mean intersection over union, $MIoU$) 等.本研究主要采用的类别平均像素准确率与平均交并比作为评价指标.为方便说明评价指标的计算公式,设像素点数量为 k , P_{ii} 代表属于 i 类且预测为 i 类的像素点总数 (TP), P_{jj} 代表实际为 j 类而预测为 j 类的数量 (TN), P_{ij} 代表本属于 i 类却被预测为 j 类的像素点总数 (FP), P_{ji} 则表示属于 j 类却被预测为 i 类的像素点总数 (FN).

像素准确率 (PA) 表示测试目标正确像素点数与全部测试区域总像素之比,公式如下:

$$PA = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \quad (7)$$

类别平均像素准确率 (MPA) 就是一个改进的像素准确率 PA 计算公式,它针对测试目标进行 PA 值计算,进而计算全部 PA 值的均值,其公式如下:

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij}} \quad (8)$$

平均交并比 ($MIoU$) 是语义分割的评价度量标准,其通过计算测试目标合集中真实区域和预测区域之间的交集和并集之比,公式如下:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ji} + \sum_{j=0}^k P_{ji} + P_{ii}} \quad (9)$$

准确率 P (precision) 公式如下:

$$P = \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + P_{ii}} \quad (10)$$

召回率 R (recall) 公式如下:

$$R = \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ji} + P_{ii}} \quad (11)$$

$F1$ 公式如下:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (12)$$

3.2 训练过程及结果

3.2.1 实验平台

本文实验采用深度学习框架 PyTorch, Window 11 操作系统, NVIDIA GeForce RTX 3080 显卡, 运行内存为 32 GB. 详细配置如表 1 所示.

表 1 实验配置

名称	配置信息
操作系统	Windows 11
实验环境	Python 3.6.13
深度学习框架	PyTorch 1.7.0
CUDA	CUDA 11.0
显卡GPU	NVIDIA GeForce RTX 3080 (12 GB)
CPU	Intel(R) Core(TM) i7-12700
内存容量	32 GB

3.2.2 消融实验

本文选取了 UNet 作为基准网络. 如表 2 所示, 为验证本文所设计的改进 UNet 网络模型的可行性与有效性, 在相同实验条件下, 采用以下 5 种消融实验方案: 1) 基础 UNet 网络; 2) 在方案 1 的基础上加入 CNN-Transformer 混合结构; 3) 在方案 1 的基础上加入 dual up-sample (DU) 模块; 4) 在方案 2 和方案 3 的基础上

加入 hard-swish (HS) 激活函数; 5) 在方案 4 的基础上加入交叉熵 Dice 混合损失函数对模型进行训练。

表 2 改进的 UNet 模型消融实验对比

编号	改进策略	MPA (%)	MIoU (%)
1	UNet	84.97	75.84
2	UNet+CNN-Transformer	88.46	79.52
3	UNet+DU	86.87	77.35
4	UNet+CNN-Transformer+DU+HS	91.12	82.59
5	UNet+CNN-Transformer+DU+HS (交叉熵Dice混合损失函数)	92.56	84.97

在上述 5 种消融实验方案结果中, 比较方案 1 与方案 2 可知, 加入 CNN-Transformer 混合结构对模型准确率与平均交并比有显著提升, 分别提升了 3.49% 与 3.68%。比较方案 1 与方案 3 可知, 加入 dual up-sample 模块对模型平均像素准确率与平均交并比有小幅提升, 分别提升了 1.9% 与 1.51%。验证了 CNN-Transformer 混合结构与 dual up-sample 模块对模型性能提升的作用。由方案 1 与方案 4 对比可知, 同时加入 CNN-Transformer 混合结构、dual up-sample 模块以及 hard-swish 激活函数, 平均像素准确率与平均交并比达到了 91.12% 与 82.59%, 效果良好。由方案 4 与方案 5 可知, 在方案 4 的基础上加入交叉熵 Dice 混合损失函数对模型性能有进一步提升。由此可知, 本研究的网络训练策略与模型改进方案是有效可行的。

3.2.3 模型训练

本研究利用改进的 UNet 网络模型对 strawberry disease detection 数据集进行训练得到训练模型。下列表 3 中列出了该数据集图像的标签类别及其对应的训练集、测试集、验证集对应的数量, 图像数据未扩增。

为衡量本文模型的有效性, 考虑到试验设备性能及预训练训练效果, 实验每批次样本数 (batch size) 设定为 16, 初始学习速率设置为 0.01, 迭代次数 (epoch) 设置为 100 次。通过预训练, 提高了模型的训练速度并有效地加强了网络的拟合, 达到了提高有限数据集上病斑分割准确度的目的。本文最终平均交并比 (MIoU) 达到 84.97%。

将预训练完成的模型在 strawberry disease detection 数据集训练, 每 5 轮迭代次数记录训练结果, 计算平均交并比 (MIoU)。由图 5 可知, 随着训练迭代次数的不断增加, 模型的分割准确率不断上升。当迭代次数达到 50 次时, 平均交并比 (MIoU) 稳定在 80% 以上, 浮动变化在 2 个百分点以内。模型分割准确率较高, 鲁棒性较好。

表 3 草莓数据集信息

疾病类别	训练集	测试集	验证集
草莓角斑病	245	43	147
炭疽病	52	12	33
花腐病	117	29	62
灰霉病	255	77	145
叶斑病	382	71	162
果实白粉病	80	12	43
白粉病叶	319	63	151
总数	1450	307	743

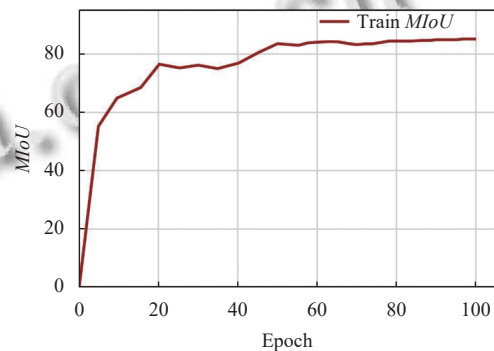


图 5 模型 MIoU 变化曲线

3.3 实验对比

为进一步衡量本文模型的各方面性能与效果, 本研究复现了目前主流的语义分割模型: UNet、PSPNet、DeepLabV3、TransUNet。UNet 采用编码-解码框架, 使用 VGG 模型^[19] 作为特征提取网络。PSPNet 与 DeepLabV3 采用 MobileNet^[20] 作为特征提取网络, 在保证性能的同时能够降低计算量。TransUNet 与本文的改进 UNet 模型都采用 ResNet50^[21] 作为特征提取网络。

在相同实验参数配置下, 都采用 strawberry disease detection 数据集, 将本文的改进 UNet 模型其他 4 种模型进行比较, 实验对比指标包括精准率、 R 值、 $F1$ -score、MPA、MIoU 及训练时间, 如表 4、表 5、表 6 所示。

由表 4 和表 5 数据可知, 各模型下的 7 种病害分割效果差距较大, 且相同模型对不同病害的语义分割效果也存在一定差距。主要由于部分病害较为相似, 如叶片类病害, 模型可准确分割但存在语义类别判断错误的情况, 导致部分病害精确率不高。本文的改进 UNet 模型与其他模型相对比, 其各类病害的语义分割精准率与 recall 值大部分优于其他模型, 部分病害识别效果稍差于 TransUNet, 整体效果较好, 可有效识别与分割各类病害。

表4 各模型下7种病害精准率

病害类型	UNet	PSPNet	DeepLabV3	TransUNet	本文改进UNet
草莓角斑病 (angular leafspot)	0.85	0.84	0.77	0.93	0.95
炭疽病 (anthracnose fruit rot)	0.82	0.81	0.84	0.90	0.89
花腐病 (blossom blight)	0.81	0.80	0.76	0.88	0.90
灰霉病 (gray mold)	0.80	0.77	0.80	0.89	0.92
叶斑病 (leaf spot)	0.83	0.85	0.81	0.94	0.94
果实白粉病 (powdery mildew fruit)	0.79	0.76	0.77	0.90	0.88
白粉病叶 (powdery mildew leaf)	0.78	0.75	0.73	0.86	0.91

表5 各模型下7种病害 Recall 值

病害类型	UNet	PSPNet	DeepLabV3	TransUNet	本文改进UNet
草莓角斑病 (angular leafspot)	0.75	0.73	0.71	0.84	0.85
炭疽病 (anthracnose fruit rot)	0.72	0.71	0.69	0.80	0.82
花腐病 (blossom blight)	0.72	0.70	0.67	0.79	0.80
灰霉病 (gray mold)	0.76	0.78	0.72	0.86	0.84
叶斑病 (leaf spot)	0.80	0.77	0.74	0.79	0.81
果实白粉病 (powdery mildew fruit)	0.74	0.72	0.69	0.81	0.81
白粉病叶 (powdery mildew leaf)	0.72	0.74	0.70	0.82	0.83

表6 各模型类别平均对比结果

模型	MPA (%)	MIoU (%)	F1-score (%)	训练时间 (h)
UNet	84.97	75.84	85.54	4.27
PSPNet	83.16	75.27	84.74	3.84
DeepLabV3	82.55	74.82	83.48	4.07
TransUNet	90.53	82.36	89.64	4.73
本文的改进UNet模型	92.56	84.97	91.48	4.31

由表6数据可知,从评价指标上看,五种模型类别平均像素准确率分别为84.97%、83.16%、82.55%、90.53%、92.56%。其中,DeepLabV3在整体上看效果较差,准确率偏低,在该数据集上的分割效果并不理想。其次PSPNet网络分割效果稍好,且训练时间最短。

UNet整体效果较好,编解码架构在草莓病害语义

分割中得到较好的应用。加入了CNN-Transformer模块的TransUNet与UNet相比,类别平均像素准确率提高了5.56%,平均交并比提高了6.52%。而本文的改进UNet模型效果最佳,类别平均像素准确率为92.56%,平均交并比为84.97%,分别比TransUNet提升了约2.03%和2.61%,则表明被正确分割的像素比例增加,正确分割率提升。

从训练时间上来看,PSPNet模型速度最快,其次是DeepLabV3、UNet模型,分别为3.84 h、4.07 h、4.27 h。而TransUNet模型训练时间较慢是因为加入了注意力机制,导致增加了一定参数与计算量。本研究为均衡注意力机制的影响,采用hard-swish激活函数减少内存访问的次数,引入交叉熵Dice混合损失函数,降低训练时间,加速模型收敛,最终训练时间为4.31 h。

图6为不同模型架构的病害语义分割效果图,其中分为7种草莓病害类别,分别用不同颜色进行分割标注。由图可知,UNet与PSPNet网络模型的病害分割效果相差不大,均能分割出病害的大致区域,但部分区域仍存在分割不精准现象。DeepLabV3网络模型通过空洞卷积扩大感受野,但该方法对病害边缘分割效果较差,且双线性上采样过程中细节丢失过多,导致分割结果不理想。TransUNet网络模型与前几类网络模型相比较,分割效果较好,准确率高,但部分图像边缘分割仍存在问题。而本文针对上述问题,通过引入CNN-Transformer结构与dual up-sample模块等改进,有效改善了边缘分割的问题,分割效果更佳,提升了分割精度,其预测的分割结果,可有效运用于草莓病害的检测。

4 结论

本研究利用基于UNet网络语义分割方法,并结合CNN-Transformer模型,引入dual up-sample模块并采用hard-swish激活函数与交叉熵Dice混合损失函数,得到一个改进的UNet草莓病害语义分割模型。为验证本文模型,在包含2500张图像的草莓病害数据集上进行训练,对7类草莓病害进行语义分割,通过大量试验并与其他网络模型对比,语义分割的MPA值达到了92.56%,MIoU值达到了84.97%。试验结果表明,本文的改进UNet草莓病害语义分割模型训练稳定性好,泛化能力较强,鲁棒性较好。与其他主流网络模型相比,分割效果更好。

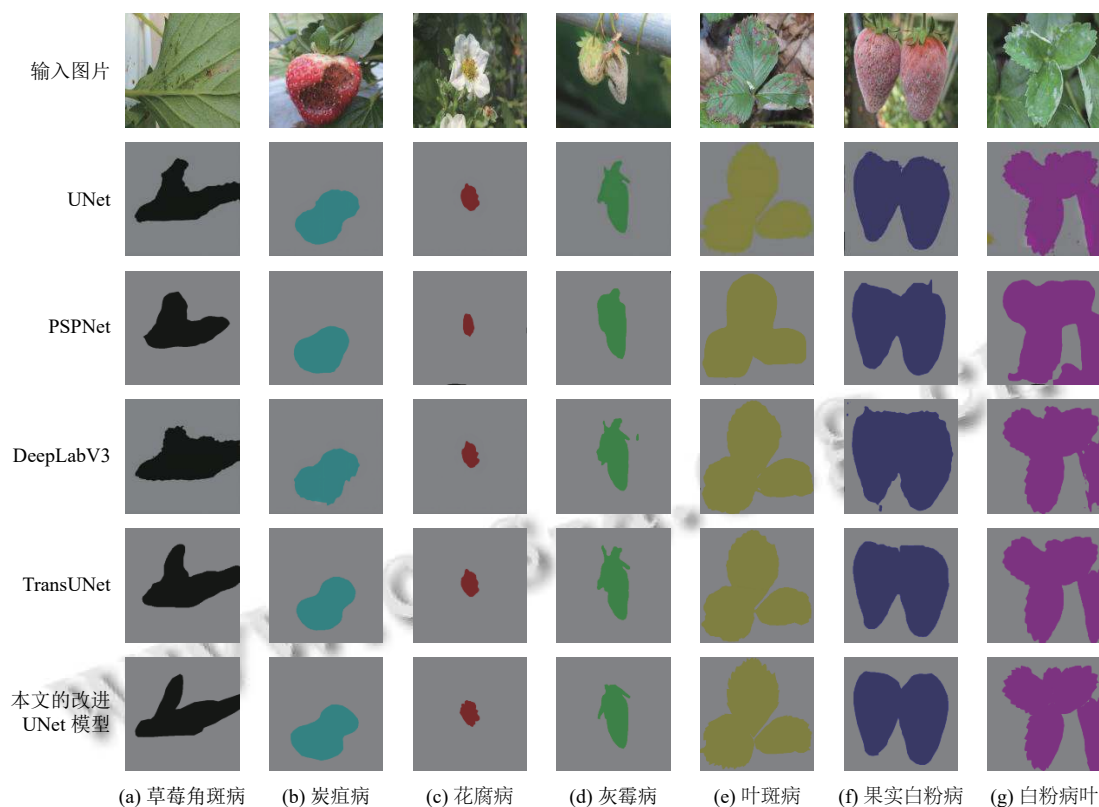


图6 不同语义分割方法效果对比图

然而, 本文基于注意力机制的改进 UNet 的语义分割算法仍然有一定局限性. 其计算量大, 训练时间较长, 参数量较大, 并不适用于移动端或者嵌入式设备中. 未来将进一步研究轻量级网络的语义分割模型, 更加便于移动端的对语义分割模型的使用.

参考文献

- 赵小虎, 李晓. 基于语义分割的农作物病害识别综述研究. 现代计算机, 2021, 27(18): 122–126. [doi: 10.3969/j.issn.1007-1423.2021.18.024]
- 王雪, 尹来武, 郭鑫鑫. 室外多变光照条件下农田绿色作物的图像分割方法. 吉林大学学报(理学版), 2018, 56(5): 1213–1218. [doi: 10.13413/j.cnki.jdxblxb.2018.05.28]
- 宋西平, 李国琴, 罗陆锋, 等. 基于 HSI 色彩空间与 FFCM 聚类的葡萄图像分割. 农机化研究, 2015, 37(10): 40–44. [doi: 10.3969/j.issn.1003-188X.2015.10.009]
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.
- Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230–6239.
- Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.
- Afzaal U, Bhattarai B, Pandeya YR, *et al.* An instance segmentation model for strawberry diseases based on mask R-CNN. Sensors, 2021, 21(19): 6565. [doi: 10.3390/s21196565]

- 11 Shi WZ, Caballero J, Huszár F, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1874–1883.
- 12 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929, 2020.
- 13 Tang ZM, Jiang WM, Zhang Z, *et al.* DenseNet with up-sampling block for recognizing texts in images. Neural Computing and Applications, 2020, 32(11): 7553–7561. [doi: [10.1007/s00521-019-04285-8](https://doi.org/10.1007/s00521-019-04285-8)]
- 14 Fan CM, Liu TJ, Liu KH. SUNet: Swin Transformer UNet for image denoising. Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS). Austin: IEEE, 2022. 2333–2337.
- 15 Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: JMLR.org, 2011. 315–323.
- 16 Ramachandran P, Zoph B, Le QV. Searching for activation functions. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018.
- 17 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.
- 18 Li X, Sun X, Meng Y, *et al.* Dice loss for data-imbalanced NLP tasks. arXiv:1911.02855, 2019.
- 19 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
- 20 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 21 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.

(校对责编: 牛欣悦)