

基于语义分割与实例分割的玉米茎秆截面参数测量方法

陈燕<sup>1,2</sup> 李想<sup>1</sup> 曹勉<sup>1</sup> 胡小春<sup>3</sup> 王令强<sup>4</sup>

(1. 广西大学计算机与电子信息学院, 南宁 530004;  
2. 广西多媒体通信与网络技术重点实验室, 南宁 530004;  
3. 广西财经学院大数据与人工智能学院, 南宁 530007; 4. 广西大学农学院, 南宁 530004)

**摘要:** 茎秆微观结构与其力学性能密切相关,影响作物的抗倒伏性能。但作物茎秆微观表型参数难以通过人工方式获取,因此急需自动化的测量方法。本研究以玉米为材料,通过光学显微镜获得玉米茎秆横截面切片图像,基于深度学习架构融合 ResNet 和 Unet 构建语义分割 Res-Unet 网络模型,对截面表皮、周皮和髓区 3 个功能区域进行分割;针对维管束数量多、面积小、密度大的特点,以 EfficientDet 作为基础网络架构,根据维管束尺寸小的特性,减少双向特征图金字塔(BiFPN)的层数,达到提高推理速度、减少显存占用量的目的,同时添加掩膜分割分支,构造新的网络 Eifff-BiFPN 实现对维管束的分割。实验结果表明,功能区域分割的平均 DICE 达到 88.17%;维管束分割的 AP<sub>50</sub> 和 AP<sub>50:70</sub> 分别达到 88.78% 和 72.80%。根据分割结果,可以获得玉米茎秆截面尺寸、各功能区域尺寸和维管束数量、面积等微观结构参数。本文方法具有精确性、实时性和可用性,可用于玉米茎秆微观结构参数的自动化测定,为作物抗倒伏研究提供技术基础。

**关键词:** 玉米茎秆; 截面参数; 语义分割; 实例分割; 显微结构

中图分类号: TP391.41      文献标识码: A      文章编号: 1000-1298(2023)06-0214-09      OSID: 

Measurement of Maize Stem Cross Section Parameters Based on Semantic Segmentation and Instance Segmentation

CHEN Yan<sup>1,2</sup> LI Xiang<sup>1</sup> CAO Mian<sup>1</sup> HU Xiaochun<sup>3</sup> WANG Lingqiang<sup>4</sup>

(1. College of Computer and Electronic Information, Guangxi University, Nanning 530004, China  
2. Guangxi Key Laboratory of Multimedia Communications Network Technology, Nanning 530004, China  
3. School of Bigdata and Artificial Intelligence, Guangxi University of Finance and Economics, Nanning 530007, China  
4. College of Agriculture, Guangxi University, Nanning 530004, China)

**Abstract:** Stem microstructure is closely related to its mechanical properties and affects lodging resistance in crops. But crop microphenotypic parameters are difficult to obtain manually. Therefore, automated measurement methods are urgently needed. The lack of measurement methods for high-throughput vascular bundle parameters seriously restricts the in-depth study. Based on the deep learning architecture, ResNet and Unet network were merged to construct the semantic segmentation model Res-Unet to segment function zones in maize stem cross section. In view of the small area, large number and dense distribution of vascular bundle in maize stem cross section, EfficientDet was used as the basic network architecture. According to the characteristics of small size of vascular bundles, the number of layers of BiFPN was reduced to improve reasoning speed and reduce the occupation of video memory. Mask segmentation branches were added to construct a new network Eifff-BiFPN to segment vascular bundles. The results showed that the DICE of each function zone could reach an average DICE of 88.17%, and the vascular bundle segmentation task could reach 88.78% and 72.80% on AP<sub>50</sub> and AP<sub>50:70</sub>, respectively. Therefore, the proposed method was accurate, real-time and available, which can be used for automatic determination of microstructure parameters of maize stem, and a technical basis was established for the study of crop lodging resistance. According to the segmentation results, the cross-

收稿日期: 2022-09-29    修回日期: 2022-12-04  
基金项目: 广西科学研究与技术开发计划项目(桂科 AA20302002-3)和广西自然科学基金项目(2020GXNSFAA159090)  
作者简介: 陈燕(1975—),女,教授,博士,主要从事智能算法与数据科学及应用研究,E-mail: cy@gxu.edu.cn  
通信作者: 胡小春(1974—),男,副教授,主要从事智能算法及应用、大数据与高性能计算研究,E-mail: huxch999@163.com

sectional size of corn stem, the size of each functional area, the number and area of vascular bundles and other microstructure parameters can be obtained.

**Key words:** maize stem; section parameters; semantic segmentation; instance segmentation; microstructure

# 0 引言

维管束为植物体输送水分、无机盐和有机物质,在“库-源-流”系统中扮演“流”的重要角色<sup>[1]</sup>。玉米茎秆倒伏与茎秆微观表型,如截面面积、截面不同区域的面积占比、维管束数量和各维管束面积等性状密切相关<sup>[2]</sup>。机械组织比例等表型性状影响茎秆的机械强度<sup>[3]</sup>,从而影响作物的抗倒伏性能。由于大规模获取茎秆截面微表型参数工作量大、效率低、测量指标有限,作物的倒伏研究主要局限在茎秆的形态学性状和力学指标,而对茎秆的组织解剖学特征研究则较少<sup>[4]</sup>。

图像分割就是把图像分成若干个特定的、具有独特性质的区域并提出感兴趣目标的技术。图像分割方法是对图像中属于特定类别的像素进行分类的过程。通过对图像的分割,以充分理解图像中的内容,便于对图像各部分的关联性进行分析。传统的图像分割方法基于灰度值的不连续和相似的性质,易受限于特定的图像特征而欠缺泛化能力;基于深度学习的图像分割技术是利用卷积神经网络(Convolutional neural network, CNN)强大的特征提取能力来理解图像中每个像素所代表的真实物体<sup>[5]</sup>,从而具有较好的稳健性和适应性。基于深度学习的图像分割有语义分割与实例分割,皆可在像素级别区分物体轮廓。TERAMOTO 等<sup>[6]</sup>用基于 U-net 的语义分割方法识别沟槽剖面图内的水稻根系分布。FETTER 等<sup>[7]</sup>提出一种使用深度卷积神经网络的系统来识别和统计银杏显微图像中的气孔,识别准确率达到 98.1%,提供了针对作物微观表型智能化研究范例。刘文波等<sup>[8]</sup>提出改进 SOLO v2 的实例分割方法对番茄叶部病害进行分割和识别,可保证算法的实时性和准确性。熊俊涛等<sup>[9]</sup>用 Mask R-CNN 对大豆叶片进行实例分割,平均分割准确率达到 88.3%。文献[10]列举了深度学习在作物表型研究中具有代表性的工作,为作物表型的智能化识别提供了稳健的解决方案。

基于深度学习方法在特征提取方面的优势明显,无需对特征进行设计就可以实现对图像特征的提取,但目前应用于作物微表型识别或组织解剖学特征的研究还很少。目前对作物茎秆微表型结构参数的获取和分析普遍采用显微镜<sup>[11]</sup>或作物成像专

用电子计算机断层扫描(Computed tomography, CT)<sup>[12]</sup>获得茎秆截面微表型图像,再利用传统图像分割方法对图像进行处理。徐胜勇等<sup>[13]</sup>使用基于 OpenCV 的滤波、边缘检测等操作,对各个组织结构进行独立的阈值分割、拼接等操作,以此为基础测量各种参数,并将各种参数、性状特征进行统计分析。赵欢等<sup>[14]</sup>利用基于图形学的区带表型解析方法,可获取茎秆不同节间相关表型的多项指标。但此类方法对图像目标与背景色差弱或亮度差异小,存在气泡、过曝或曝光不足时,所获性状参数易出现较大误差。目前仅有文献[15]和文献[16]是基于 CNN 对作物微表型识别进行研究。WU 等<sup>[15]</sup>利用作物 CT 成像并基于 SegNet 架构的语义分割水稻茎秆微表型,用于提取水稻茎秆性状特征参数。陈燕等<sup>[16]</sup>基于 U-net 架构的语义分割对普通光学显微镜成像的小麦茎秆截面的组织结构进行分割和量化,获得小麦维管束和功能区域的表型参数,为获取小麦茎秆的微表型参数提供了较为准确的方法。

虽然语义分割和实例分割都可在像素级别对物体进行分割,但是语义分割只能分割一类对象,导致无法区分同一类别之间的不同对象;而实例分割可针对特定物体的像素进行分类、直接分割出不同个体,从而可区分不同的实例个体。玉米茎秆截面的维管束分布密集,后期需要获得每个维管束的质心、面积性状参数等,除了要识别出维管束,还需区分他们的不同个体。文献[16]提出的语义分割方法可以识别截面的维管束,但相邻较近的个体易被识别为同一个维管束实例。

目前较为成熟的实例分割方法以 Mask R-CNN<sup>[17]</sup>为代表,其分割目标通常为自然场景物体,适用于检测数量少、物体形状尺寸差异较大的目标。但玉米茎秆截面微表型参数的数量大、面积小且分布密集,已有的实例分割方法并不直接适用。而且,以 Mask R-CNN 为基础的网络普遍存在架构设置不适用、推理速度较慢、占用显存较多的问题<sup>[17]</sup>。因此,本文选用具有复合扩张能力的 EfficientDet<sup>[18]</sup>作为基础网络架构,增加掩膜分支以实现实例分割,通过减少输出特征图的尺度数量构造新的实例分割网络 Eif-BiFPN,以提高推理速度、减少显存的占用,从而实现对维管束的分割。文献[16]对截面功能区域的分割是以 MobileNet 作为骨干网络,但文中

对小麦的截面只分为 2 个功能区域,识别难度较低,适合选用较为轻量型的 MobileNet 网络。由于玉米茎秆截面需要分成表皮、周皮和髓区 3 个区域,而且不同功能区域之间还具有易混淆性,如果直接将文献[16]的功能区域分割网络用于玉米茎秆截面的功能区域分割,其准确率较低。因此本文选用提取特征能力更强的 ResNet<sup>[19]</sup> 作为骨干网络,并与 Unet<sup>[20]</sup> 融合成为 Res - Unet 网络模型用于玉米茎秆截面的功能区域分割。

1 数据集获取与处理

1.1 数据集来源

数据集包含 116 份不同的玉米种质材料,由广西农业科学院玉米研究所和广西大学提供,品种有美玉 27、天桂糯 932、福华甜、CML161、CML171、Gui39722、Guizhao18421、PH6WC 和昌 7-2 等,可代表温带和亚热带地区主要的玉米种质资源。各玉米品种在试验田的种植密度每公顷约 75 000 株,行长 3 m,行宽 0.65 m,每行种植 15 株,水肥管理同大田生产。

玉米节间茎秆徒手切片,厚度 0.2 ~ 0.5 mm,用 5% 间苯三酚(乙醇与水的体积比为 95:5)和浓盐酸染色,染色时间 2 min,steREO Discovery. V20 体视显微镜拍照,放大倍数为 6.7 ~ 15,图像存储格式为 TIF,分辨率为 1 790 像素 × 1 370 像素。共获取图像 180 幅,其中 113 幅来自不同种质材料的抽雄期茎秆自顶向下最后 3 个节间,其余 67 幅来自剩余 3 种样本材料处于不同生长期的茎秆。因此,数据集中的图像具有茎秆解剖特征的多样性。

1.2 图像标注

使用 Labelme 工具对每幅图内维管束的轮廓和功能区域的轮廓进行标注,将玉米茎秆截面分为表皮区、周皮区、髓区 3 个区域,得到的标注图样例如图 1 所示。为了全面地验证模型效果,数据集按独立同分布的原则划分,选择 156 幅作为训练集,其余 24 幅样本作为验证集。

2 网络结构设计

本文用于玉米截面分割的网络模型由两部分组

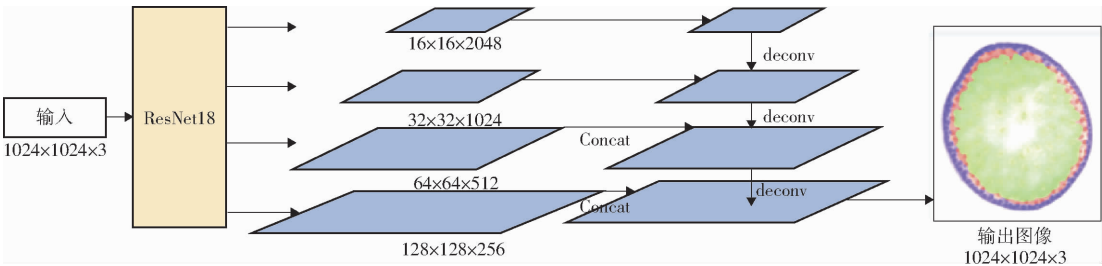


图 2 功能区域分割网络模型

Fig. 2 Architecture of function zone segmentation network

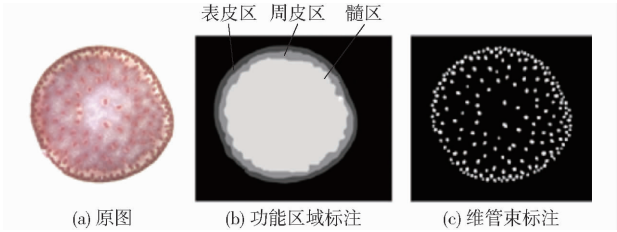


图 1 原图和标注图示例

Fig. 1 Example of original image and annotation

成,第 1 部分用于功能区域分割,以 ResNet 作为基准网络,并与 Unet 融合成 Res - Unet 网络,用于检测、分割表皮、周皮和髓区 3 个功能区域;第 2 部分用于维管束的分割,以 EfficientNet 为基准网络,改变 BiFPN 的连接方式和层数,增加掩膜输出分支,用于检测、分割每个维管束实例。

2.1 功能区域分割网络模型

功能区域分割用语义分割方法。选用 ResNet 作为骨干网络,并与 Unet 融合成为 Res - Unet 网络模型,用于分割截面的表皮区、周皮区和髓区。功能区分割网络模型 Res - Unet 由解码器和编码器两部分组成,网络结构如图 2 所示。输入玉米茎秆截面样本的彩色图像,经过图像预处理和图像增强操作之后,通过编码器 ResNet 对输入图像进行下采样,获取 4 种尺度分别为 16 × 16 × 2 048、32 × 32 × 1 024、64 × 64 × 512、128 × 128 × 256 的特征图,经过特征融合与拼接后,输出所有功能区域分割后的彩色掩膜。

编码器的基础卷积块由残差模块组成,使用恒等映射使卷积层在输入特征的基础上学习新的特征,再提取图像的语义特征,如轮廓、边缘、颜色等信息。解码器对编码器生成的特征图进行拼接,再利用浅层网络中的语义信息辅助位置信息对图像进行分割,并将特征精确定位并映射到图像上。为了减少冗余计算量,根据分割结果将样本图裁剪为该玉米茎秆截面的最小外接矩形截图,作为第 2 部分模型的输入。

2.2 维管束分割网络模型

维管束分割采用实例分割方法,选用具有复合扩张能力的 EfficientNet<sup>[21]</sup> 作为基准网络,在双向特

征金字塔 BiFPN<sup>[18]</sup> 结构中减少输出特征图的尺度数量,通过改变内部的连接方式以高效提取图像特征,同时添加基于锚框的边界框输出分支 (Box head) 和类别输出分支 (Class head) 两部分,然后由检测框 (Proposal boxes) 输出到掩膜输出分支 (Mask

head), 构建基于目标检测的实例分割网络 Eifff-BiFPN, 具体结构如图 3 所示。将图像输入基准网络后,通过 BiFPN 获得 3 种不同尺寸的特征图,经过检测输出、分类输出与分割输出分支结构得到最后的维管束分割结果。

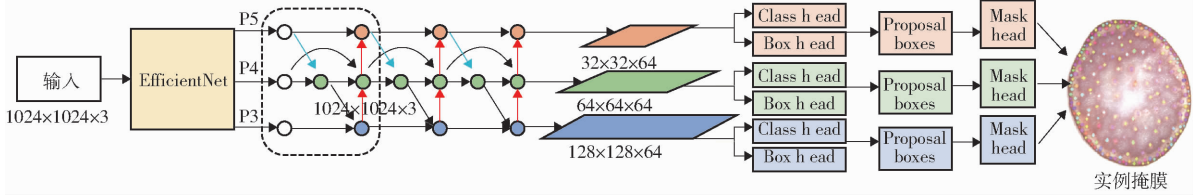


图 3 维管束分割网络模型

Fig. 3 Architecture of vascular bundles segmentation network

2.2.1 基准网络

深度学习算法大都通过扩大网络规模来提升网络泛化能力,提高预测精度。扩大卷积网络结构只对网络的深度、宽度和分辨率参数进行微调。为了在有限的计算资源内获得更高的精度,同时也为了减少调参的工作量,EfficientNet 从网络的深度  $d$ 、宽度  $w$  (通道数) 和输入图像的尺寸  $r$  (分辨率) 3 个维度上进行复合扩张。网络深度、宽度和分辨率的缩放存在如下关系

$$\begin{cases} d = d_{\text{ratio}}^{\phi} \\ w = w_{\text{ratio}}^{\phi} \\ r = r_{\text{ratio}}^{\phi} \\ \text{s. t.} \begin{cases} d_{\text{ratio}} w_{\text{ratio}}^2 r_{\text{ratio}}^2 \approx 2 \\ d_{\text{ratio}} \geq 1 \\ w_{\text{ratio}} \geq 1 \\ r_{\text{ratio}} \geq 1 \end{cases} \end{cases} \quad (1)$$

式中  $d_{\text{ratio}}$  ——网络深度缩放系数  
 $w_{\text{ratio}}$  ——宽度缩放系数  
 $r_{\text{ratio}}$  ——分辨率缩放系数  
 $\phi$  ——缩放系数

设定  $\phi = 1$ , 通过网格搜索得到  $d_{\text{ratio}}$ 、 $w_{\text{ratio}}$  和  $r_{\text{ratio}}$  的最优解,通过相应的  $d$ 、 $w$  和  $r$  构造了最小网络规模的最优模型 B0<sup>[19]</sup>。 $\phi$  越大,网络的 3 个维度扩张越多,模型消耗的资源也越大,模型的精确度也越高, $\phi$  取值为 1 ~ 7 的整数。

2.2.2 改进的 BiFPN 结构

特征图金字塔 (FPN)<sup>[22]</sup> 可对经过骨干网络后得到的不同尺寸的特征图进行多尺度加权的特征提取与融合。BiFPN 为双向特征图金字塔结构,除保持 FPN 自上而下和自下而上的连接外,增加了特征图之间的跨层和跳跃连接。与 FPN 相比,BiFPN 的特征提取能力更强,网络参数更少,运行速度却更快。

文献[19]采用 5 层 BiFPN 结构,先将 P5 进行下采样生成 P6 和 P7,再将 P3 ~ P7 共 5 层的特征图进行融合。但由于 P6 和 P7 的尺寸只占原输入的 1/64 与 1/128,分辨率太小,在下采样过程中容易丢失小目标的特征信息,P6 和 P7 两层不适用于针对维管束这类小目标的检测任务,因此本文改变了原来的 BiFPN 结构及其内部的连接方式,只保留 3 层的 BiFPN 结构。具体结构如图 4 所示,仅将 P3 ~ P5 共 3 层的特征图进行融合,使网络更好地学习到小目标信息,提升维管束边缘细节的分割效果。其中蓝色箭头和红色箭头为跨层连接,弧线形箭头为跳跃连接。灰色虚线框是一个基本单元,左边白色一列为各层的输入,中间一列为跨层连接,右边一列获得输入的双向连接和同层的跳跃连接,同时右边一列也作为下一个单元的输入。此单元重复堆叠 3 次即构成一个完整的 BiFPN 结构,并输出 3 幅尺寸为 32 × 32 × 64、64 × 64 × 64、128 × 128 × 64 的特征图。

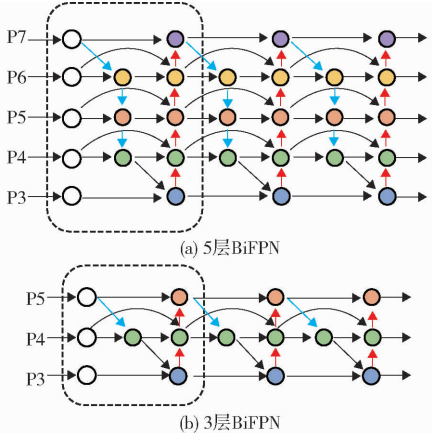


图 4 多层 BiFPN 结构对比

Fig. 4 Comparison of BiFPN structure with multiple-layers

2.2.3 检测输出分支结构

检测输出分支采用基于锚定框的检测方法,包括对应边界框输出头 (Box head) 和类别输出头 (Class head) 两部分 (图 3)。以每个像素为中心生

成 9 个锚框,锚框采用大、中、小 3 个不同的尺寸,分别为 0.5、1、2;3 个比例的策略分别为  $1$ 、 $2^{1/3}$ 、 $2^{2/3}$ 。如特征图的下采样率为 8,尺寸为  $32 \times 32$ ,则该特征图共生 9  $\times$   $32 \times 32$  个锚框。

经过 BiFPN 输出的每一层特征图分别输入到对应边界框输出头 (Box head) 和类别输出头 (Class head) 进行处理。Box head 输出格式为  $[b_s, a \times 4, h, w]$ , Class head 输出格式为  $[b_s, a, h, w]$ , 其中  $b_s$  是批量大小,  $a$  是特征图中一个像素点锚定框的数量,  $h$  是特征图的高。Box head 的输出为每个锚框对标注框的 4 个顶点回归值, Class head 的输出则为每个锚框内是否包含维管束的分类结果。当判定锚框包含真实值时,将与标注边界框重合 50% 以上的锚框标记为正例。

2.2.4 分割输出分支结构

经过检测分支后,可获得每个维管束的检测框 (Proposal boxes) (图 3), 然后通过分割输出分支 (Mask head) 分割出框内的维管束掩膜。分割输出分支由 3 层卷积组成,将输入的特征图通道压缩至 3 层,输出检测框区域像素的预测结果。在训练过程中,使用标注框提取出相应特征图内的像素,使用 RoIAlign<sup>[17]</sup> 将其全部统一对齐到固定尺寸之后,得到每个检测框内的维管束轮廓掩膜。在推理过程中,选出置信度较高的检测框,将框内的特征图像素送入掩膜输出分支,得到框内维管束轮廓掩膜,再将 这些维管束掩膜根据检测框的位置还原回原图像,即可得到图像的维管束实例分割掩膜图像。

3 实验

实验硬件环境为 B365M - POWER、32 GB 内存、NVIDIA GeForce RTX 3090 显卡、24 GB 显存、操作系统为 Ubuntu 7.5.0,编程语言为 Python 3.8,深度学习框架为 Pytorch 1.7.0。

3.1 数据增强

由于数据集中样本数量较少,所以在训练阶段需要对原始数据集进行图像增强处理,以提高训练数据集内样本的多样性,增强网络的泛化性能。在功能区域分割部分,使用随机翻转、高斯模糊、亮度对比度变化、自适应直方图均衡等图像增强操作。图像增强处理后再做统一的归一化操作,目的是移除图像中相同部分,凸显图像特征。归一化具体操作将图像 RGB 三通道内所有像素各减去固定的均值,并除以标准差,将所有像素标准化处理。本文使用的均值和标准差均为 ImageNet 统计的均值和标准差,分别为  $[0.485, 0.456, 0.406]$  和  $[0.229, 0.224, 0.225]$ 。

3.2 模型训练

骨干网络和特征提取部分的网络权重可通过迁移学习将已在大规模数据集 ImageNet 上训练好的网络权重迁移过来,以提高网络的收敛速度。整个网络模型的下游解码器和输出分支部分,需要在训练过程中对权重进行微调。

模型训练中,功能区域分割网络的训练最大迭代次数设为 100,批量大小为 2,图像缩放尺寸为  $512 \times 512$ ;维管束分割网络训练的最大迭代次数为 200,批量大小为 1,图像缩放尺寸为  $1\,024 \times 1\,024$  像素。

在预置锚框的设置上,采用的 3 种不同比例尺寸为  $2^0$ 、 $2^{1/3}$ 、 $2^{2/3}$ ,3 种不同长宽比为  $1:1$ 、 $1.25:0.8$ 、 $0.8:1.25$ ,即在特征图的每一个像素上设置 9 个不同尺寸、不同长宽比的锚框。在 3 层 BiFPN 中,锚框最大边长为 95.2 像素,最小边长为 9.6 像素。

3.3 损失函数与优化器

损失函数用于度量预测值和真实值的差距,可衡量模型预测的好坏,在深度学习模型中起到决定网络优化的作用。由于整个玉米茎秆横截面的分割由 2 个相对独立的网络模型组成,对维管束分割是由检测网络架构增加掩膜分支构成,因此功能区域分割的 Res - Unet 需要一个损失函数,维管束分割网络中不同的输出分支需要不同的损失函数。

(1) 功能区域分割网络使用交叉熵损失函数  $L_{ce}$   
$$L_{ce}(y_i, p_i) = y_i \lg p_i + (1 - y_i) \lg (1 - p_i) \quad (2)$$
式中  $y_i$ 、 $p_i$ ——第  $i$  个样本的真实值和预测值

(2) 在维管束分割网络中, Class、Box 和 Mask 3 个头分支分别使用不同的损失函数,以评价每个头分支的损失,最后再用 3 个损失函数的加权和作为维管束分割网络总的损失,以评价整个维管束分割网络的性能。

Class head 选用适合于密集型物体检测任务的焦点损失函数 focal loss<sup>[22]</sup>,表示为

$$L_{class} = \begin{cases} \alpha(1 - p_i)^\gamma \lg p_i & (y_i = 1) \\ p_i^\gamma \lg (1 - p_i) & (y_i = 0) \end{cases} \quad (3)$$

式中  $\alpha$ 、 $\gamma$ ——超参数,  $\alpha \in (0, \infty)$   
 $L_{class}$ ——焦点损失函数

调整参数  $\alpha$  以解决前景框和背景框数量不均衡的问题;取  $\gamma \in (0, \infty)$ ,调整  $\gamma$  以提高难区分实例的损失权重、降低简单样本的损失权重。

Box head 选用带参数的回归损失函数 Huber loss<sup>[23]</sup>,表示为

$$L_{box} = \begin{cases} \frac{1}{2}(y_i - p_i)^2 & (|y_i - p_i| \leq \delta) \\ \delta |y_i - p_i| - \frac{1}{2}\delta^2 & (|y_i - p_i| > \delta) \end{cases} \quad (4)$$



式中  $L_{box}$ ——回归损失函数  
 $\delta$ ——阈值参数,表示真实值与预测值的偏差  
当  $\delta \sim 0$  时,Huber loss 会趋向于平均平方误差 MAE;当  $\delta \sim \infty$ ,Huber loss 会趋向于平均绝对误差 MSE。因此,Huber loss 结合了 MSE 和 MAE 的优点,相比于最小二乘的线性回归,可降低对异常点的敏感性。

Mask head 选用二元交叉熵损失函数 BCE loss,表示为

$$L_{mask} = -[y_i \lg p_i + (1 - y_i) \lg (1 - p_i)] \quad (5)$$

在式(2)的基础上,使用独热编码计算出所有框对所有类别进行二分类概率的均值。

上述 3 个输出头分支的损失加权和作为维管束分割网络的总体损失  $L_v$ ,表示为

$$L_v = W_{class} L_{class} + W_{box} L_{box} + W_{mask} L_{mask} \quad (6)$$

式中  $W_{class}$ 、 $W_{box}$ 、 $W_{mask}$ ——Class、Box 和 Mask 输出头的权重,取 1、800、0.2

此外,选用对超参数不敏感的 AdamW 模型优化器,学习率动态调整器为 ReduceLROnPlateau,使每个参数都具有动态学习率,初始学习率设置为  $10^{-3}$ ,当验证集损失值在 10 个迭代周期内未下降时,将全局学习率缩小为原来的 1/2。

3.4 网络模型评价指标

3.4.1 DICE 系数

骰子系数(DICE coefficient)是一种集合相似度量指标,用于表示 2 个样本的相似程度<sup>[24]</sup>。取值范围为 0~1,其中 1 表示预测结果和实际结果完全重合,0 表示预测结果和实际结果完全不相交,因此 DICE 系数越接近 1 表示网络的性能越好。本文用 DICE 系数从玉米茎秆截面功能区域在像素级别上评价网络模型的分割性能。

$$D_{ICE} = \frac{2X_{TP}}{2X_{TP} + X_{FP} + X_{FN}} \times 100\% \quad (7)$$

式中  $D_{ICE}$ ——DICE 系数,%  
 $X_{TP}$ ——预测为正例的正例像素数  
 $X_{FP}$ ——预测为负例的正例像素数  
 $X_{FN}$ ——预测为负例的负例像素数

3.4.2 平均精度

平均精度(AP)表示精确率-召回率(Precision-recall)曲线与坐标轴所围面积。

由于使用单个指标评价目标检测模型有一定的局限性,因此使用 AP 作为目标检测和实例分割的评价指标。给定一个阈值  $t$ (一般取 0.5~0.9)用于表示预测框与真实框的重复置信度。如果交并比(IOUS)大于  $t$ ,则预测样本为正例。例如  $t$  取值为 0.7,则  $AP_{70}$  的度量值是表示预测框与真实框的重

叠区域大于两种区域并集面积的 70% 及以上,才能被判定为正例。因此  $t$  越大,对应的 AP 就越小,评价指标就越严格,预测样本被认为是正例就越困难。一般情况下, $t$  越小,对应的 AP 就越大。

3.4.3 平均绝对误差

平均绝对误差(MAPE)是衡量预测准确性的统计指标。MAPE 可避免误差相互抵消,能准确反映实际预测误差的大小。

3.5 实验结果与分析

由于功能区域和维管束的分割分别由两个独立的网络完成,因此在对实验结果进行分析时也分别对 2 个不同的网络模型及其性能进行评价分析。

使用本文方法对玉米茎秆截面功能区域和维管束的分割效果示例如图 5 所示。图 5 展示了 4 种具有代表性的不同品种或不同生长期的玉米茎秆截面分割结果。其中图 5b 为对功能区域分割的结果,红色为表皮区域,绿色为周皮区域,蓝色为髓区区域;图 5c 为对维管束分割的结果,图中不同颜色的掩膜代表不同的维管束实例。

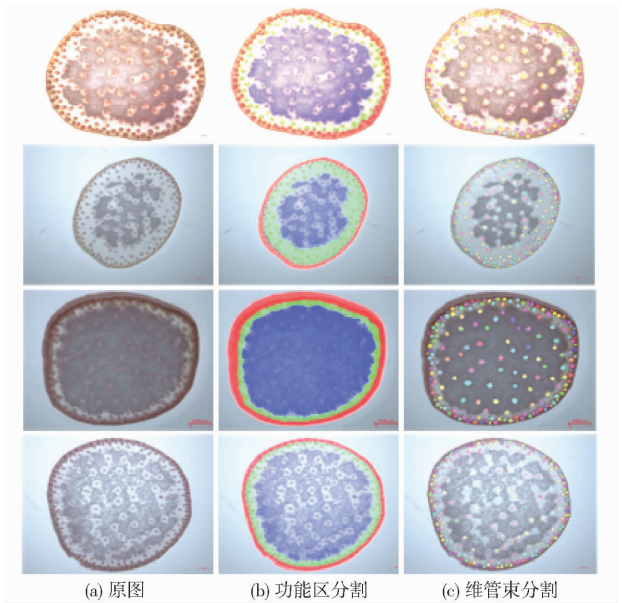


图 5 功能区分割与维管束分割结果  
Fig. 5 Results of functional zone segmentation and vascular bundles segmentation

3.5.1 功能区域分割网络性能分析

先分析本文所设计的用于功能区域分割网络 Res-Unet 的性能,再用 DICE 系数作为功能区域分割性能的评价指标,分析对表皮、周皮和髓区 3 个功能区域的分割性能。

3.5.1.1 网络模型性能分析

主要分析针对骨干网络 ResNet 不同层数的构造对各功能区域分割的 DICE 系数,以选择合适的层数搭建实用的 Res-Unet 网络。

基于不同层数的 Res - Unet 模型对功能区域分割的 DICE 系数如表 1 所示。由表 1 可看出,当骨干网络为 ResNet18 时对功能区域分割效果最好,3 个区域的 DICE 都达到 84% 以上,平均 DICE 达到 88.17%,其中髓区区域的 DICE 达到 93.61%。其中,对表皮区域分割准确率略低的原因可能是部分切片较厚,用显微镜拍摄图像的表皮边缘出现阴影。此外,随着骨干网络层数增加,表 1 中各区域的 DICE 没有明显提升,这说明随着网络模型的深度增加,模型出现了过拟合现象。

表 1 不同骨干网络下的 DICE 对比  
Tab.1 Comparison of DICE under different backbones

| %         |       |       |       |       |
|-----------|-------|-------|-------|-------|
| 骨干网络      | 表皮区域  | 周皮区域  | 髓区区域  | 平均值   |
| ResNet18  | 84.62 | 86.29 | 93.61 | 88.17 |
| ResNet34  | 84.23 | 86.40 | 93.58 | 88.07 |
| ResNet50  | 82.98 | 83.69 | 92.60 | 86.42 |
| ResNet101 | 83.95 | 85.38 | 93.25 | 87.53 |

3.5.1.2 与其他文献方法对比分析

与文献[16]方法进行对比,结果如表 2 所示。由表 2 可得,本文方法在功能区域上的分割性能整体优于文献[16]的方法。

表 2 DICE 对比  
Tab.2 Comparison of DICE

| %      |       |       |       |       |
|--------|-------|-------|-------|-------|
| 方法     | 表皮区域  | 周皮区域  | 髓区区域  | 平均值   |
| 本文方法   | 84.62 | 86.29 | 93.61 | 88.17 |
| 文献[16] | 82.80 | 56.48 | 78.34 | 72.53 |

根据表 3 可知,本文方法所获取的参数在 MAPE 上与文献[16]相比有明显降低,在功能区域平均面积上降低 38.232 个百分点。由于文献[16]中有关功能区域划分不一致,表皮区域面积虽然具有一定误差,但尚在可接受范围内,但是对于周皮区域面积与髓区区域面积的获取上几乎呈现出无法识别的状态,与人工标注值误差巨大。

表 3 功能区分割网络的平均绝对误差对比  
Tab.3 Comparison of MAPE in function area segmentation network

| %      |        |        |        |        |
|--------|--------|--------|--------|--------|
| 网络模型   | 表皮面积   | 周皮面积   | 髓区面积   | 平均面积   |
| 本文方法   | 16.068 | 14.549 | 11.729 | 14.115 |
| 文献[16] | 24.227 | 80.116 | 52.698 | 52.347 |

3.5.2 维管束分割网络性能分析

先分析本文所设计的用于维管束分割网络 Eifff - BiFPN 的性能,再用 AP 作为评价指标分析维管束分割的性能。

3.5.2.1 网络模型性能分析

首先分析维管束分割在 EfficientNet 不同级别缩放系数的性能,再分析在不同层数 BiFPN 的性能,以选择最合适的结构用于 Eifff - BiFPN 网络分割维管束。

基于 3 层 BiFPN 使用不同缩放系数的 EfficientNet 训练,维管束分割 AP 性能如表 4 所示。由表 4 可知,系数为 B4 的分割效果最好。

表 4 不同骨干网络的维管束分割网络的 AP (使用 3 层 BiFPN)

| %                 |                  |                  |                  |                  |                  |                     |
|-------------------|------------------|------------------|------------------|------------------|------------------|---------------------|
| 骨干网络              | AP <sub>50</sub> | AP <sub>60</sub> | AP <sub>70</sub> | AP <sub>80</sub> | AP <sub>90</sub> | AP <sub>50:90</sub> |
| EfficientNet - B1 | 86.25            | 73.33            | 44.51            | 10.28            | 0.25             | 42.92               |
| EfficientNet - B2 | 87.46            | 75.71            | 49.73            | 13.01            | 0.29             | 45.24               |
| EfficientNet - B3 | 87.75            | 76.86            | 50.62            | 12.90            | 0.31             | 45.68               |
| EfficientNet - B4 | 88.78            | 77.21            | 52.46            | 14.97            | 0.39             | 46.76               |
| EfficientNet - B5 | 86.82            | 75.58            | 46.65            | 10.17            | 0.25             | 43.89               |
| EfficientNet - B6 | 84.20            | 70.75            | 44.89            | 11.17            | 0.20             | 42.24               |

基于 EfficientNet - B4,使用不同层数的 BiFPN 做训练,维管束分割 AP 性能如表 5 所示。由表 5 可知,3 层 BiFPN 的特征提取效果最好。

表 5 不同 BiFPN 结构的维管束分割网络的 AP (以 EfficientNet - B4 为骨干网络)

| %        |                  |                  |                  |                  |                  |                     |
|----------|------------------|------------------|------------------|------------------|------------------|---------------------|
| BiFPN 层数 | AP <sub>50</sub> | AP <sub>60</sub> | AP <sub>70</sub> | AP <sub>80</sub> | AP <sub>90</sub> | AP <sub>50:90</sub> |
| 3        | 88.78            | 77.21            | 52.46            | 14.97            | 0.39             | 46.76               |
| 4        | 88.90            | 77.01            | 50.95            | 13.45            | 0.33             | 46.13               |
| 5        | 87.29            | 76.06            | 50.37            | 14.13            | 0.33             | 45.64               |

骨干网络缩放系数越小、BiFPN 特征图层级越少,网络越容易因为缺少充分利用数据而出现欠拟合的现象;骨干网络缩放系数越大、BiFPN 特征图层级越多,网络越容易因为训练数据中的细节和噪声而出现过拟合的现象。通过对比实验分析,骨干网络为 EfficientNet - B4、使用 3 层 BiFPN 的网络能够获得最好的分割结果。

3.5.2.2 网络模型预测结果可视化分析

为直观体现维管束分割任务的效果,利用验证集的 24 个样本的维管束数量和维管束平均面积分析预测结果和原始标注的拟合程度,如图 6 所示。从图 6a 可见,两个数值回归到最小二乘法拟合直线上。图 6b 为测试集中每个样本的维管束平均标注面积和预测面积的曲线,从图中可见,

大部分样本的预测值与标注值较为接近,部分甚至几乎一致,但也有少部分差异较大,其中差异最大达到 $0.0197\text{ mm}^2$ ,这与人工标注原始图像存在一定误差有关。

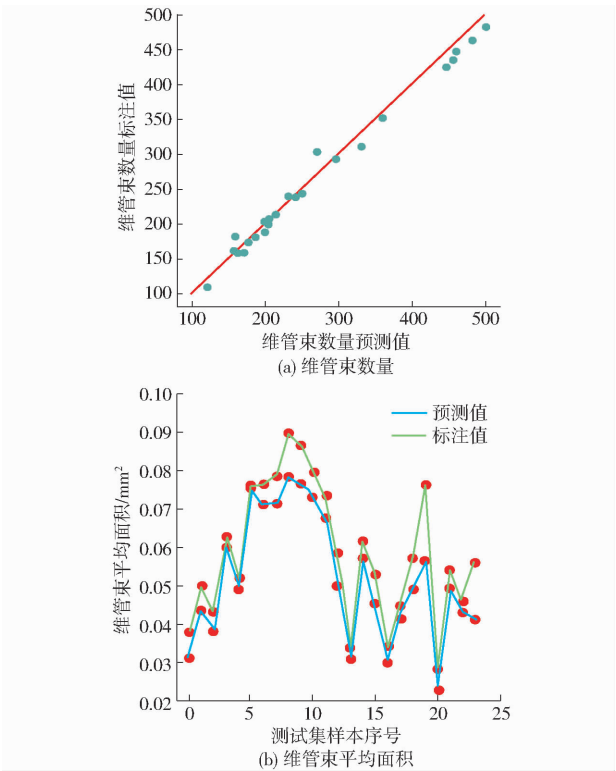


图 6 维管束数量与面积的预测结果

Fig. 6 Prediction results of vascular bundle number and area

3.5.2.3 与其他文献方法对比分析

选择在维管束数量、维管束面积方面与文献[16]的方法对比。由于玉米的茎秆截面微表型形态比小麦复杂,文献[16]网络模型无法对玉米茎秆截面微表型特征进行良好的训练和拟合。根据表 6 可知,本文方法所获取的参数在 MAPE 上与文献[16]相比有明显降低,在维管束数量和面积上分

表 6 维管束分割网络的平均绝对误差对比

| Tab. 6 Comparison of MAPE in vascular bundle segmentation network |        |        | % |
|-------------------------------------------------------------------|--------|--------|---|
| 方法                                                                | 维管束数量  | 维管束面积  |   |
| 本文方法                                                              | 12.316 | 21.596 |   |
| 文献[16]                                                            | 28.584 | 23.364 |   |

别降低 16.268、1.768 个百分点。  
从表 7 可见,在  $AP_{50:70}$  上本文方法更优,其中只有  $AP_{70}$  值略低。

表 7 AP 值对比

| Tab. 7 Comparison of AP |           |           |           | %            |
|-------------------------|-----------|-----------|-----------|--------------|
| 方法                      | $AP_{50}$ | $AP_{60}$ | $AP_{70}$ | $AP_{50:70}$ |
| 本文方法                    | 88.78     | 77.21     | 52.46     | 72.80        |
| Mask R-CNN              | 87.71     | 74.35     | 41.15     | 67.70        |

在维管束的分割任务上,与应用最广泛的实例分割网络 Mask R-CNN 模型进行对比,比较的指标除了 AP 之外,还对存储内存占用量和推理内存占用量两方面进行比较,结果如表 8 所示。

表 8 所需内存对比

| Tab. 8 Comparison of memory |            |            |
|-----------------------------|------------|------------|
| 方法                          | 存储内存占用量/MB | 推理内存占用量/GB |
| 本文方法                        | 78         | 13         |
| Mask R-CNN                  | 170        | > 24       |

在存储内存占用量和推理内存占用量方面,本文方法均比 Mask R-CNN 所需要的内存少。其中,Mask R-CNN 推理单幅图像占用的显存就超过 24 GB,因此必须使用压缩图像算法才能完成维管束的分割任务。而本文方法在不使用压缩算法处理的情况下,单幅图像推理占用显存仅为 13 GB,节省 50%。因此,即使是在推理阶段,Mask R-CNN 方法的实用性也较差。

4 结论

- (1) 在功能区域分割任务中,平均 DICE 达到 88.17%;在维管束实例分割任务中,  $AP_{50}$  达到 88.78%,  $AP_{50:70}$  达到 72.80%。
- (2) 与常用的实例分割网络 Mask R-CNN 相比,本文方法在不同阈值下的 AP 指标表现更佳,且不需压缩算法即可完成推理,所需存储内存占用量更少,获取参数的平均绝对误差更低。
- (3) 与文献[16]相比,本文方法的功能区域分割网络更适用于玉米茎秆,DICE 系数更高;在维管束数量和面积上分别降低 16.268、1.768 个百分点,在功能区域平均面积上降低 38.232 个百分点。

参 考 文 献

[1] 赵雪,周顺利. 玉米抗茎倒伏能力相关性状与评价研究进展[J]. 作物学报, 2022, 48(1):15-26.  
ZHAO Xue, ZHOU Shunli. Research progress on traits and assessment methods of stalk lodging resistance in maize[J]. Acta Agronomica Sinica, 2022, 48(1):15-26. (in Chinese)

[2] 王庭杰,张亮,韩琼,等. 玉米茎秆细胞壁和组织构建对抗压强度的影响[J]. 植物科学学报, 2015, 33(1):109-115.  
WANG Tingjie, ZHANG Liang, HAN Qiong, et al. Effects of stalk cell wall and tissue on the compressive strength of maize [J]. Plant Science Journal, 2015, 33(1):109-115. (in Chinese)

[3] 王夏青,宋伟,张如养,等. 玉米茎秆抗倒伏遗传的研究进展[J]. 中国农业科学, 2021, 54(11):2261-2272.  
WANG Xiaqing, SONG Wei, ZHANG Ruyang, et al. Genetic research advances on maize stalk lodging resistance[J]. Scientia



- Agricultura Sinica, 2021, 54(11):2261–2272. (in Chinese)
- [4] 周海宇,江禹奉,杨明冲,等. 玉米茎秆强度与维管束数目及纤维含量的关系和评价[J]. 植物遗传资源学报,2022, 23(6):1636–1643.  
ZHOU Haiyu, JIANG Yufeng, YANG Mingchong, et al. Relationship and evaluation of stalk strength, vascular bundle and fiber content in maize[J]. Journal of Plant Genetic Resources, 2022, 23(6):1636–1643. (in Chinese)
- [5] GHOSH S, DAS N, DAS I, et al. Understanding deep learning techniques for image segmentation: ACPUB27New York, NY, USA, 10.1145/3329784[P]. 2019–08–30.
- [6] TERAMOTO S, UGA Y. A deep learning-based phenotypic analysis of rice root distribution from field images[J]. Plant Phenomics, 2020, 2(1):3194308.
- [7] FETTER K C, EBERHARDT S, BARCLAY R S, et al. Stomatacounter: a neural network for automatic stomata identification and counting[J]. New Phytologist, 2019, 223(3):1671–1681.
- [8] 刘文波,叶涛,李顺. 基于改进 SOLO v2 的番茄叶部病害检测方法[J]. 农业机械学报, 2021, 52(8):213–220.  
LIU Wenbo, YE Tao, LI Qi. Tomato leaf disease detection method based on improved SOLO v2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(8):213–220. (in Chinese)
- [9] 熊俊涛,戴森鑫,区炯洪,等. 基于深度学习的大豆生长期叶片缺素症状检测方法[J]. 农业机械学报, 2020, 51(1):195–202.  
XIONG Juntao, DAI Senxin, OU Jionghong, et al. Leaf deficiency symptoms detection method of soybean based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(1):195–202. (in Chinese)
- [10] 岑海燕,朱月明,孙大伟,等. 深度学习在植物表型研究中的应用现状与展望[J]. 农业工程学报,2020,36(9):1–16.  
CEN Haiyan, ZHU Yueming, SUN Dawei, et al. Current status and future perspective of the application of deep learning in plant phenotype research[J]. Transactions of the CSAE, 2020, 36(9):1–16. (in Chinese)
- [11] 徐胜勇,段宏兵,李东臣,等. 小麦茎秆截面参数显微图像测量系统[J]. 农业机械学报, 2017, 48(7):46–52.  
XU Shengyong, DUAN Hongbing, LI Dongchen, et al. Measurement system for parameters of wheat stem section based on microimage processing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(7):46–52. (in Chinese)
- [12] ZHANG Ying, DU Jianjun, WANG Jinglu, et al. High-throughput micro-phenotyping measurements applied to assess stalk lodging in maize (*Zea mays* L.)[J]. Biological Research, 2018, 51(1):40.
- [13] 徐胜勇,彭程里,陈可,等. 基于扇环形区域图像分割的小麦秸秆截面参数测量方法[J]. 农业机械学报, 2018, 49(4):53–59.  
XU Shengyong, PENG Chengli, CHEN Ke, et al. Measurement method of wheat stalks cross section parameters based on sector ring region image segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(4):53–59. (in Chinese)
- [14] 赵欢,王璟璐,张颖,等. 基于 Micro-CT 的玉米自交系茎秆维管束表型高通量获取与精准解析[J]. 江苏农业科学, 2022, 50(12):201–211.
- [15] WU Di, WU Dan, FENG Hui, et al. A deep learning-integrated Micro-CT image analysis pipeline for quantifying rice lodging resistance-related traits[J]. Plant Communications, 2021, 2(2):100165.
- [16] 陈燕,朱成宇,胡小春,等. 基于深度学习的小麦茎秆截面参数检测[J]. 农业机械学报, 2021, 52(7):169–176.  
CHEN Yan, ZHU Chengyu, HU Xiaochun, et al. Detection of wheat stem section parameters based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7):169–176. (in Chinese)
- [17] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2):386–397.
- [18] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020.
- [19] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [21] TAN M, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980–2988.
- [23] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017:2117–2125.
- [24] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference 3D Vision (3DV). IEEE, 2016.
- [25] HUBER P J. Robust statistics (second edition)[M]. New Jersey: John Wiley & Sons, 2009.