doi:10.6041/j.issn.1000-1298.2021.12.018

基于区域语义和边缘信息融合的作物苗期植株分割模型

廖　娟[1,2]　陈民慧[1]　张　锴[1]　邹　禹[3]　张　顺[1]　朱德泉[1,2]

(1. 安徽农业大学工学院, 合肥 230036; 2. 安徽农业大学智慧农业研究院, 合肥 230036;
3. 安徽省农业科学院水稻研究所, 合肥 230031)

摘要: 为在自然环境下准确分割作物苗期植株,实现苗期植株定位及其表型自动化测量,本文提出一种融合目标区域语义和边缘信息的作物苗期植株分割网络模型。以 U-Net 网络构建主干网络,基于侧边深度监督机制,引导主干网络在提取特征时能感知植株边缘信息;利用空间空洞特征金字塔构建特征融合模块,融合主干网络和边缘感知模块提取的特征,融合后的特征图具有足够的细节信息和更强的语义信息;联合边缘感知的损失与特征融合的损失,构建联合损失函数,用于整体网络优化。实验结果表明,本文模型对不同数据集的作物植株的语义分割像素准确率高达 0.962,平均交并比达到 0.932;与 U-Net、SegNet、PSPNet、DeepLabV3 模型相比,本文模型在不同数据集上平均交并比最高提升 0.07,对自然环境下作物苗期植株具有良好的分割效果和泛化能力,可为植株定位、对靶喷药、长势识别等应用提供重要依据。

关键词: 作物苗期; 植株分割; U-Net 网络; 区域语义信息; 边缘感知模块; 特征融合模块

中图分类号: TP391.41; S513　　**文献标识码**: A　　**文章编号**: 1000-1298(2021)12-0171-11　　**OSID**:

Segmentation of Crop Plant Seedlings Based on Regional Semantic and Edge Information Fusion

LIAO Juan[1,2]　CHEN Minhui[1]　ZHANG Kai[1]　ZOU Yu[3]　ZHANG Shun[1]　ZHU Dequan[1,2]

(1. *School of Engineering*, *Anhui Agricultural University*, *Hefei* 230036, *China*
2. *Institute of Intelligent Agriculture*, *Anhui Agricultural University*, *Hefei* 230036, *China*
3. *Rice Research Institute*, *Anhui Academy of Agricultural Sciences*, *Hefei* 230031, *China*)

Abstract: To segment crop plant seedlings accurately in natural environment, a segmentation network model based on regional semantic and edge information was presented. Firstly, the U-Net network was used as the backbone network, and the side depth supervision mechanism was used to guide the backbone network to perceive the plant edge information when extracting features. Then, based on atrous spatial pyramid pooling, the feature fusion module was built to fuse the semantic information in the backbone network and the edge information extracted by the edge perception module. The fused feature map would have enough detail information and strong semantic information. Besides, combined with the loss of edge perception and the loss of feature fusion, the joint loss function was defined for the overall network optimization. The experimental results showed that the proposed model can achieve the pixel accuracy of 0.962 and the mean intersection over union of 0.932. Compared with the U-Net, SegNet, PSPNet and DeepLabV3 models, the mean intersection over union of the used model was about 0.07 higher. Therefore, the proposed model can achieve good segmentation effect and generalization ability for crop plant seedlings in natural environment, which can provide important basis for plant location, target spraying, growth recognition and other applications.

Key words: plant seedling; segmentation; U-Net network; regional semantic information; edge perception module; feature fusion module

收稿日期: 2021-07-13　修回日期: 2021-09-28

基金项目: 智慧农业研究院开放基金项目(IAR2021A02)、安徽省自然科学基金项目(2108085MC96、1808085ME158)和安徽省研发计划项目(202004a06020016、202004a06020061)

作者简介: 廖娟(1986—),女,讲师,博士,主要从事机器视觉研究,E-mail: liaojuan308@163.com

通信作者: 朱德泉(1969—),男,教授,博士,主要从事智能农业装备技术研究,E-mail: zhudequan@ahau.edu.cn

0 引言

随着农业自动化、智能化技术的不断发展,机器视觉技术已广泛应用于作物生长信息获取、病虫害识别、农药精准喷施、田间智能除草以及农业机械视觉导航等领域[1]。而作物植株的准确分割是机器视觉技术的关键步骤,但不同作物的外观(如形状、尺寸、姿态等)存在较大差异,且田间自然光照环境复杂,这些因素给作物植株图像分割带来了一定困难。

传统的植株分割方法多通过提取目标特定图像中颜色[2]、纹理[3]及几何形状[4]等浅层特征来区分作物植株与背景,但浅层特征通常缺乏表达高层语义的能力,在表征作物植株的特定信息方面具有局限性,从而导致传统分割方法性能下降。相比传统方法,深度学习技术凭借其卓越的特征学习能力,可在图像语义分割中取得更好的分割结果,现已广泛应用于农业领域[5]。段凌凤等[6]提出了基于全卷积神经网络 SegNet 的大田稻穗分割算法,通过网络训练自动学习和提取稻穗图像中深度特征,实现稻穗分割。MILIOTO 等[7]构建了 14 通道的输入图像,提出了一种深度编-解码的语义分割网络,能够实时分割作物和杂草。韩振浩等[8]基于全卷积神经网络 U-Net[9],实现了果园图像中果树与道路的像素级平稳分割。

基于深度学习的语义分割方法在作物植株分割应用中优势明显,但其分割效果取决于网络在训练过程中能否学习到丰富的有用特征,其中,高层的语义特征包含更多抽象的语义信息,低层的语义特征包含较多的边界纹理等信息[10]。然而,当前作物植株语义分割研究多用高层特征实现像素分类,忽略了低层特征,导致分割结果缺乏精确的边缘信息。而边缘作为分割目标的一个重要特征,能有效表征不同区域间语义信息。为获取精细边缘,XIE 等[11]提出整体嵌套边缘检测网络(Holistically-nested edge detection,HED),通过构建多层边缘深度监督,实现了利用卷积神经网络提取边缘。CHENG 等[12]提出了一种融合边缘与区域语义信息的遥感图像分割模型,通过单层的边缘监督,增强主干网络对边缘信息的感知,有效提高了小尺寸物体的分割精度。TAKIKAWA 等[13]构建了一种并行卷积神经网络的语义分割模型,明确将边缘信息作为单独的处理分支,采用边缘标签监督支路的优化,以并行的方式处理信息,能够对目标边界产生更清晰的预测。但现有的结合边缘信息的语义分割网络缺少针对作物苗期植株分割的研究,且边缘信息对于作物生长信息获取、生长检测等是不可或缺的数据依据。

针对作物苗期植株语义分割中边缘信息缺失问题,本文提出一种同时关注目标植株语义信息和边缘信息的作物苗期植株分割网络模型,联合边缘检测与区域语义分割任务,实现作物苗期图像中植株与背景的精确分割。以 U-Net 为主干框架,基于 HED 中侧边深度监督的思想构建边缘感知模块(Edge awareness module,EAM),并弱化边缘深度监督带来的监督冲突,引导网络在提取植株区域的语义分割特征时,能够关注植株的边缘信息;基于空间空洞特征金字塔(Atrous spatial pyramid pooling,ASPP)结构[14]设计特征融合模块(Feature fusion module,FFM),融合主干网络中作物植株区域的语义分割特征和边缘感知模块中边缘特征,以充分利用边缘感知模块获取的丰富的边缘语义信息,并利用联合损失优化整体网络,最终实现具有精细边缘的作物植株分割。

1 数据集构建

1.1 数据来源

采用 3 个作物植株数据集开展实验,数据集 1 为自制的田间水稻秧苗图像数据集,图像于 2018 年 6 月、2019 年 5 月及 2020 年 6 月采集自安徽农业大学郭河实验基地和安徽省农业科学院全椒实验基地,水稻品种为双季早籼稻,采集设备为 Canon EOS 850D 型相机,拍摄高度为 0.5 ~ 1.5 m,拍摄垂直角度为 0° ~ 60°,取多个拍摄点,在不同光照条件下共采集图像样本 560 幅,图像分辨率为3 456 像素 × 2 304 像素。数据集 2 和数据集 3 分别为公开的 Plant Phenotyping Dataset[15] 和 Crop/Weed Field Image Datase[16],其中,数据集 2 共有原始图像 350 幅,图像分辨率为2 448 像素 ×2 048 像素,主要包含3 种不同形态的植株图像:叶少的 Y 形拟南芥幼苗 135 幅、长叶柄莲座形态的拟南芥植株 145 幅和紧密的莲座形态的烟草植株 70 幅。数据集3 共有原始图像 60 幅,图像分辨率为1 296 像素 ×966 像素。图 1 为3 个数据集中部分样本示例。为减少后续网络训练所占的内存和显存,对数据集图像进行压缩处理,处理后图像分辨率为 416 像素 ×416 像素。

1.2 数据集构建与标注

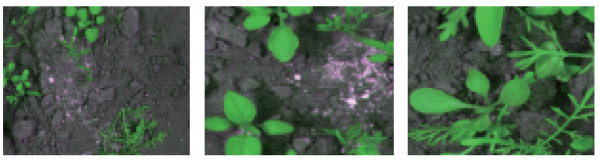
为增加数据集的多样性,增强模型的泛化能力和鲁棒性,采用 5 种图像增强技术:亮度随机减弱或增强、图像对比度随机增强、色度动态变化 30%、添加高斯噪声和图像随机旋转(0°、90°、180°、270°)。对数据增强后的图像进行筛选,剔除植株信息严重缺失的图像,筛选后的数据集 1 扩充至 1 400 幅,数



(a) 不同光照秧苗



(b) 叶少的Y形拟南芥幼苗、长叶柄的莲座形态的拟南芥植株和紧密的莲座形态的烟草植株



(c) 不同拍摄距离的胡萝卜植株

图 1　作物植株数据集部分样本

Fig. 1　Samples of plant data set

据集 2 增加到 1 424 幅，数据集 3 增加到 549 幅。

将数据增强后的 3 个数据样本按照 7∶2∶1分为训练集、验证集和测试集，对应的统计表如表 1 所示。使用 Labelme 工具对数据集 1 分别进行手动标注水稻秧苗区域像素和边缘像素，标注标签分别是背景和植株，标注后生成 8 位 PNG 灰度图，黑色为背景，白色为植株。而数据集 2、3 已公开了相应的语义分割标注图，仅使用 Labelme 标注其边缘像素。原图和可视化标注结果如图 2 所示。

表 1　数据分类统计表

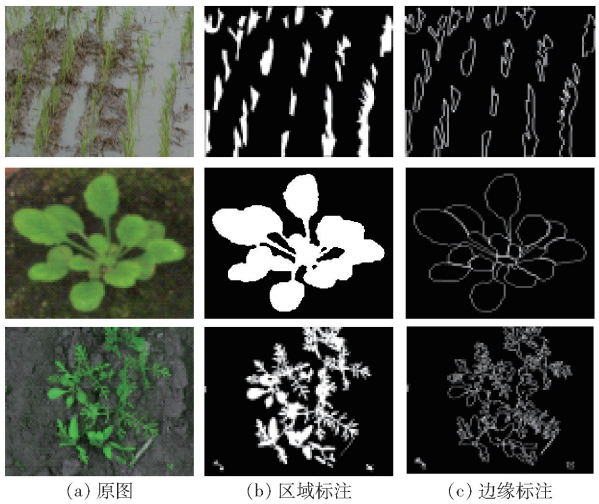
Tab. 1　Data classification statistics

数据集	总数/幅	训练集		验证集		测试集	
		原始	扩增	原始	扩增	原始	扩增
1	1 400	392	588	112	168	56	84
2	1 424	245	755	70	213	35	106
3	549	42	345	12	95	6	49

2　植株分割网络结构设计

2.1　网络结构概述

语义分割网络 U－Net 整体呈“U”形，通过跳跃连接的方式将特征图进行维度拼接，能够保留更多的位置和特征信息，在小样本数据集上分割性能优于其他网络结构，且适合处理小尺寸物体的语义分割任务[17]。为此，本文在 U－Net 网络模型基础上搭建一个能共同学习目标的区域语义和边缘特征的作物苗期植株分割网络，网络整体结构如图 3 所示，主要包括主干网络、边缘感知模块和特征融合模块 3 部分，其中，主干网络主要用于提取作物植株区域的语义分割特征；边缘感知模块(EAM)通过引入侧面输出，并采用图像的边缘标签对其进行深度监督，



(a) 原图　(b) 区域标注　(c) 边缘标注

图 2　原图和标注图

Fig. 2　Original images and annotated images

以提取作物植株的边缘特征；特征融合模块(FFM)主要用于融合植株区域的语义分割特征和边缘特征，提高网络检测精度。

2.2　主干网络构建

U－Net 采用了编-解码结构，编码器主要用于提取图像特征，经过 4 次池化，获取的特征尺寸逐渐缩小为输入图像尺寸的 1/16。但原始 U－Net 网络的参数较多，直接添加其他模块易造成网络参数量过大，网络训练困难[18]，运行时间长，为实现田间作物苗期植株特征提取的低延迟，本文将深度可分离卷积(Depthwise separable convolution)[19]与 U－Net 网络结构相结合。

如图 4 所示，深度可分离卷积将标准卷积操作分解为逐深度卷积和逐点 1×1 卷积，通过逐深度卷积对每一个输入通道进行单个卷积核的卷积运算，以保持输入特征图与输出特征图的通道数一致，然后用 1×1 逐点卷积组合不同深度卷积的输出，实现通道数改变。深度可分离卷积与标准卷积的计算量比例公式定义为

$$\frac{D_k^2 M + MO}{D_k^2 MO} = \frac{1}{O} + \frac{1}{D_k^2} \tag{1}$$

式中　D_k——深度可分离卷积的卷积核尺寸

M——输入特征通道数

O——输出特征通道数

相比于标准卷积层将所有卷积核都与每一个输入特征通道进行计算，深度可分离卷积可有效地减少计算量和模型训练参数。使用计算量小的深度可分离卷积层代替编码器的标准卷积层，可实现 U－Net 的轻量化[20]。引入深度可分离卷积主干网络的编码器结构如表 2 中 Stage1～Stage5 所示。

解码器部分的结构与原始 U－Net 相同，由表 2 中 4 个阶段(Stage6～Stage9)组成，每个阶段包含 1

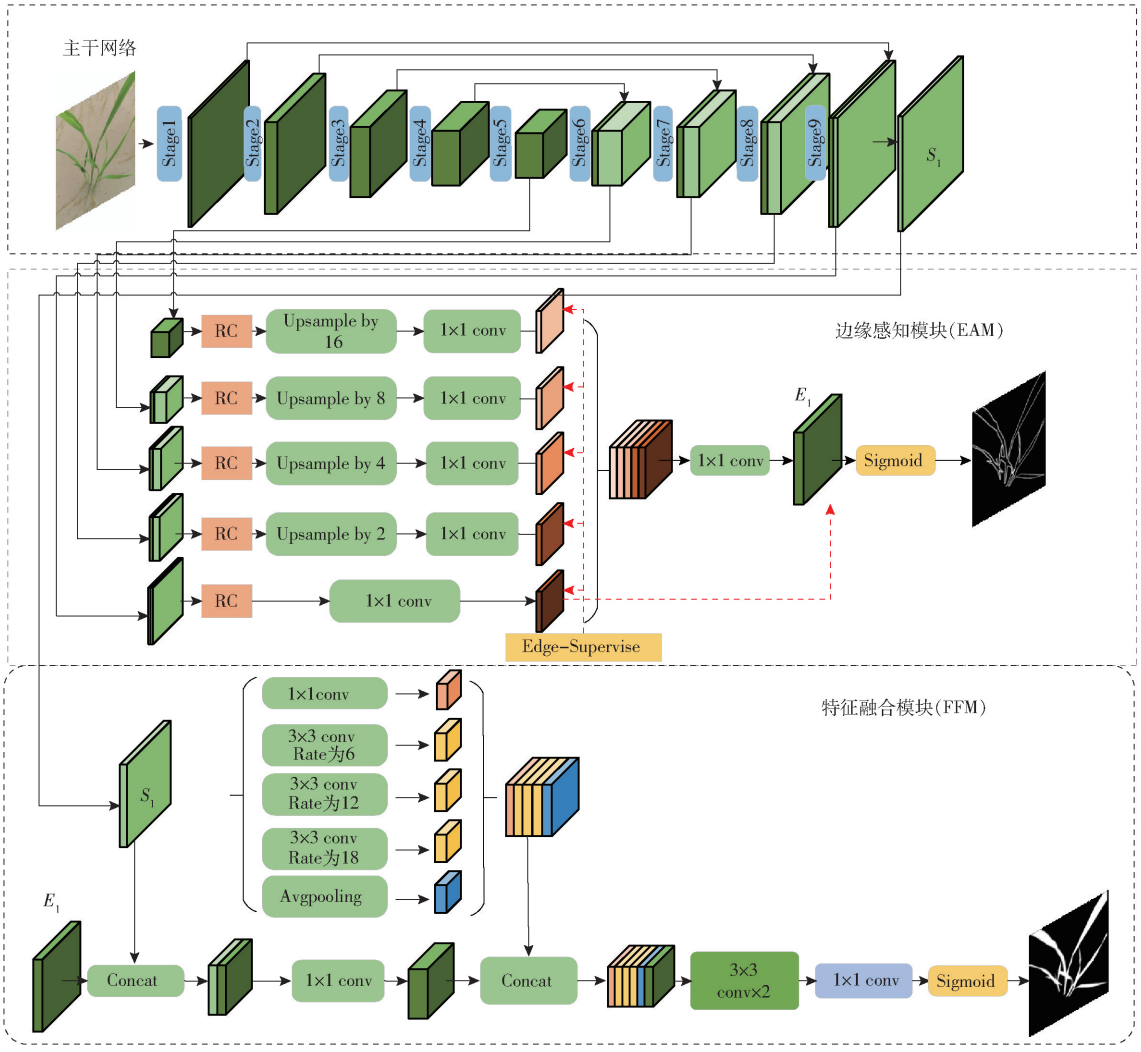


图 3 网络整体结构

Fig. 3 Structure of network

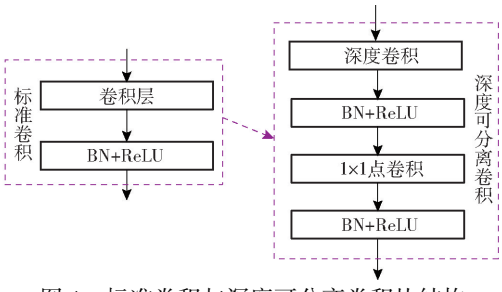


图 4 标准卷积与深度可分离卷积块结构

Fig. 4 Convolution and depthwise separable convolution

层上采样层、特征融合层和 2 层 3×3 的卷积,其中特征融合层分别融合上采样后的特征层和编码器中具有相应尺寸的特征层。为防止过拟合和抑制梯度消失,解码器中各激活层(Rectified linear unit, ReLU)[21]前引入批归一化层(Batch normalization, BN)[22],最后输出尺寸为 416 像素×416 像素×32 通道的语义分割特征,记作 S_1。

2.3 边缘感知模块

针对区域语义分割网络易导致分割后植株边缘粗糙及边缘信息丢失的问题,本研究基于 HED 的深度监督策略构造边缘感知模块(EAM),在语义分割的基础上,联合边缘分割任务,并利用边缘标签同区域语义标签共同监督网络优化。但 2 种不同的标签监督同时作用到网络中,会引起监督冲突,即边缘标签的监督引导主干网络通过捕获图像区域间的不连续性来定位精细的边缘,而植株的语义分割标签对网络的监督将会引导网络通过识别目标类别的外观变化获取抽象的高级语义信息,2 种不同的引导信息将导致网络提取有用特征时发生冲突。

为弱化监督冲突,LIU 等[23]提出通过构建卷积模块延长每一阶段的辅助监督到边缘分割网络的反向传播距离,可缓冲 2 种不同梯度信号对边缘分割网络的监督冲突。但由本文的主干网络结构可知,U-Net 的编码器主要用于特征提取,解码器用于特征融合和恢复图像尺寸,对于目标特征的提取,编码器占主导地位[9]。若将边缘感知模块直接构建在编码器部分,边缘监督会干扰编码器提取语义特征。为此,本研究在主干网络的解码器部分构建边缘感知模块,利用解码器作为缓冲部分,减弱边缘监督对

表 2　主干网络卷积层结构

Tab. 2　Convolutional layer structure of backbone network

阶段	卷积块类型	卷积核尺寸	步长	输入尺寸
Stage1	Conv2D	3×3	1	416×416×3
	Depthwise_Sep_Conv	3×3,1×1	1	416×416×32
Stage2	Depthwise_Sep_Conv	3×3,1×1	2	416×416×32
	Depthwise_Sep_Conv	3×3,1×1	1	208×208×64
Stage3	Depthwise_Sep_Conv	3×3,1×1	2	208×208×64
	Depthwise_Sep_Conv	3×3,1×1	1	104×104×128
Stage4	Depthwise_Sep_Conv	3×3,1×1	2	104×104×128
	Depthwise_Sep_Conv	3×3,1×1	1	52×52×256
Stage5	Depthwise_Sep_Conv	3×3,1×1	2	52×52×256
	Depthwise_Sep_Conv	3×3,1×1	1	26×26×512
	Depthwise_Sep_Conv	3×3,1×1	1	26×26×512
	Depthwise_Sep_Conv	3×3,1×1	1	26×26×512
	Depthwise_Sep_Conv	3×3,1×1	1	26×26×512
	Depthwise_Sep_Conv	3×3,1×1	1	26×26×512
Stage6	Upsampling2D	2×2		26×26×512
	Conv2D + BN + ReLU	3×3	1	52×52×512
	Concat			52×52×256 52×52×256
	Conv2D + BN + ReLU	3×3	1	52×52×512
Stage7	Upsampling2D	2×2		52×52×256
	Conv2D + BN + ReLU	3×3	1	104×104×256
	Concat			104×104×128 104×104×128
	Conv2D + BN + ReLU	3×3	1	104×104×256
Stage8	Upsampling2D	2×2		104×104×128
	Conv2D + BN + ReLU	3×3	1	208×208×128
	Concat			208×208×64 208×208×64
	Conv2D + BN + ReLU	3×3	1	208×208×128
Stage9	Upsampling2D	2×2		208×208×64
	Conv2D + BN + ReLU	3×3	1	416×416×64
	Concat			416×416×32 416×416×32
	Conv2D + BN + ReLU	3×3	1	416×416×64

编码器特征提取的冲突。边缘感知模块(EAM)如图 3 所示,EAM 包含 5 个边缘特征提取分支,每个分支由一个残差卷积块(Residual convolution, RC)[24]、上采样层和 1×1 卷积层构成,其中 RC 结构如图 5 所示。EAM 对主干网络解码器的每个阶段均引出侧层输出,通过每个边缘特征提取分支获取侧层边缘特征,并利用边缘标签对侧层边缘特征进行监督,且每个分支的卷积层延长了边缘监督到解码器的梯度信号反向传播路径,起到缓解边缘监督对解码器网络优化的冲突作用。由反向传播机制可知,每个分支的输出层参数通过权值误差传播路径的反向传播来更新,每一阶段的侧层深度监督使得每一级的侧边网络层分别在不同的尺度上提取边缘特征。因此,多层的深度监督将获取不同尺度的边缘特征,将每个分支提取的侧层边缘特征经过融合层堆叠融合,经过 1×1 卷积恢复通道数获得边缘感知输出特征,记作 E_1,同时采用边缘标签进行深度监督,经过 Sigmoid 层输出尺寸为 416 像素×416 像素×2 通道的边缘预测图。

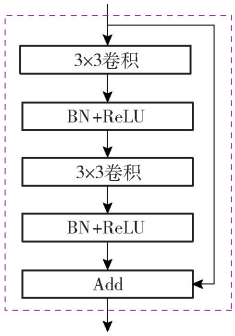


图 5　残差卷积块

Fig. 5　Residual convolution block

2.4　特征融合模块

边缘感知模块可提取丰富的边缘特征,如只在训练过程中依靠边缘感知模块来修正主干网络提取的植株区域语义特征,会遗失边缘感知模块提取的边缘特征信息。为充分利用边缘感知模块提取的边缘特征,进一步获取边缘信息完整的植株区域分割特征,提高分割效果,本文构建空间特征金字塔结构(ASPP)的特征融合模块(FFM),融合主干网络输出的植株语义区域特征和边缘感知模块输出的边缘特征,其结构见图 3。

在 ASPP 中使用扩张率为 1、6、12 和 18 的空洞卷积和平均池化操作,空洞卷积根据扩张率对卷积核的大小进行空洞填充扩张,增大卷积感受野的同时不引入额外的参数。例如,扩张率为 2 的 3×3 卷积可以达到与 5×5 卷积相同的感受野。不同扩张率的空洞卷积操作可以获取不同分辨率下的特征,可实现在多个尺度上捕获目标语义信息和上下文的语义信息,记获取包含上下文的多尺度特征为 F_1。为弥补空洞卷积中丢失的语义信息,将获取的多尺度特征与 S_1、E_1 进行堆叠融合,并通过 2 层 3×3 的卷积操作细化特征,使用 1×1 卷积恢复特征通道数,获取的特征尺寸为 416 像素×416 像素×2 通道,经过 Sigmoid 函数输出分割预测图 P,其融合过程表达式为

$$P=\sigma(\mathrm{Conv}(S_1\oplus E_1)\oplus F_1) \tag{2}$$

其中
$$F_1=\mathrm{ASPP}(S_1)$$

式中　ASPP——空间空洞卷积函数

⊕——特征按通道堆叠

Conv(·)——卷积运算函数

σ(·)——Sigmoid 函数

2.5 多任务训练损失

本文训练过程中采用联合损失函数完成网络参数优化,损失函数定义为特征融合模块(FFM)损失与边缘感知模块(EAM)损失的联合损失,表达式为

$$l_{\text{Loss}} = l_{\text{seg}} + l_{\text{edg}} \tag{3}$$

式中 l_{seg}——FFM 预测结果的损失值

l_{edg}——EAM 预测结果的损失值

l_{Loss}——联合损失函数的损失值

l_{seg}采用二分类交叉熵损失函数(Binary cross entropy,BCE)和 MIOU 损失函数的混合损失函数进行计算,定义为

$$l_{\text{seg}} = l_{\text{bce}} + l_{\text{miou}} \tag{4}$$

式中 l_{bce}——二分类交叉熵损失函数的损失值

l_{miou}——MIOU 损失函数的损失值

二分类交叉熵损失函数 l_{bce}是二分类分割中应用最广泛的损失,计算式为

$$l_{\text{bce}} = -\frac{1}{N}\sum_{i,j}\left[y_{i,j}\ln\hat{y}_{i,j} + (1-y_{i,j})\ln(1-\hat{y}_{i,j})\right] \tag{5}$$

式中 N——总样本个数

$\hat{y}_{i,j}$——语义分割预测图(i,j)处的像素值

$y_{i,j}$——语义分割标签图(i,j)处的像素值

(i,j)——图像位置坐标

其中$(i,j)\in(W,H)$,W、H 分别是图像的宽度、高度;$\hat{y}_{i,j}\in(0,1)$,$y_{i,j}\in\{0,1\}$,语义分割标签图中背景像素值为 0,作物像素值为 1。

MIOU 损失可以衡量 2 个集合的相似度,已有研究[25]证明,当分割前景与背景相差较大时,MIOU 的损失函数可取得较好的效果。而作物苗期植株图像中,背景所占区域较大,前景植株所占区域较少,导致前景像素与背景像素的占比失衡,无法获取良好的分割结果。因此,采用 MIOU 的损失函数可有效改善背景与前景植株所占像素不平衡带来的训练困难。MIOU 的损失函数定义为

$$l_{\text{miou}} = 1 - \sum_{i,j}\frac{y_{i,j}\hat{y}_{i,j}}{y_{i,j}+\hat{y}_{i,j}-y_{i,j}\hat{y}_{i,j}} \tag{6}$$

边缘感知模块的损失函数定义为 EAM 的 5 个侧层输出和最终 EAM 输出的平均值。

$$l_{\text{edg}} = \frac{1}{K}\sum_{k=1}^{K} l^{(k)} \tag{7}$$

式中 K——输出层总数

$l^{(k)}$——第 k 层侧层输出的损失值

由边缘感知模块结构知 $K=6$,包括 5 个侧层输出和 1 个边缘感知输出的融合层输出。

由于边缘感知模块源于 HED 边缘检测网络,可采用 HED 网络中加权交叉熵函数定义 $l^{(k)}$。但植株图像中边缘像素占比较小,加权交叉熵损失函数未考虑到边缘像素的邻域信息。文献[26]表明结构相似性损失 SSIM(Structural similarity,SSIM)可赋予边缘更高的权重,在训练初始时会使得沿着边缘的损失值最大,以有助于网络训练时更多地集中于边缘部分的优化。因此,本文在构建边缘感知模块损失函数时引入结构相似性损失 SSIM,采用 7×7 滑动窗口从边缘预测图和边缘标签中选取对应图像块,记作 $\hat{x}=\{\hat{x}_r:r=1,2,\cdots,M^2\}$ 和 $x=\{x_r:r=1,2,\cdots,M^2\}$,则结构相似性损失函数定义为

$$l_{\text{ssim}}^{(k)} = 1 - \frac{(2\mu_x\hat{\mu}_x+0.01^2)(2\sigma_{x\hat{x}}+0.03^2)}{(\mu_x^2+\hat{\mu}_x^2+0.01^2)(\sigma_x^2+\sigma_{\hat{x}}^2+0.03^2)} \tag{8}$$

式中 $l_{\text{ssim}}^{(k)}$——第 k 层侧层输出的结构相似性损失值

μ_x、$\hat{\mu}_x$——x、$\hat{x}$ 的平均值

σ_x、$\sigma_{\hat{x}}$——x、$\hat{x}$ 的标准偏差

$\sigma_{x\hat{x}}$——x 与 $\hat{x}$ 的协方差

第 k 层侧层输出的损失值计算式为

$$l^{(k)} = l_{\text{hed}}^{(k)} + l_{\text{ssim}}^{(k)} \tag{9}$$

其中

$$l_{\text{hed}}^{(k)} = -\beta\sum_{j\in Y_+} b_j\ln a_j - (1-\beta)\sum_{j\in Y_-}(1-b_j)\ln(1-a_j) \tag{10}$$

$$\beta = \frac{|Y_-|}{|Y_-|+|Y_+|} \tag{11}$$

式中 $l_{\text{hed}}^{(k)}$——第 k 层侧层输出加权交叉熵损失函数损失值

β——边缘标签图中边缘像素与非边缘像素的比值

Y_-——边缘标签图的边缘像素集合

Y_+——边缘标签图中的非边缘像素集合

a_j——边缘预测图在像素 j 处的像素值

b_j——边缘标签图在像素 j 处的像素值

3 实验与分析

3.1 实验环境与参数设置

实验设备选用 Intel(R)-Xeon(R) CPU E5-2699 v3, 2.30 GHz, GPU 选用 NVIDIA Quadro P2000,显存 5 GB。软件环境为 Windows 64 位操作系统,使用 Python 作为编程语言,使用 Tensorflow 作为深度学习框架。

网络训练中采用分批次(batch)训练方法,将训练集和验证集分成多个批次训练,所有训练集图像在网络模型中完成遍历计算即为一次迭代(epoch)。网络模型初始化采用加载预训练的权重初始化主干网络。初始学习率为 0.001,采用 Adam 算法优化,计算每个权重参数的自适应学习率。

3.2　评价指标

本文采用的评价标准主要包括像素准确率(Pixel accuracy,PA)、平均交并比(Mean intersection over union,MIOU)、精准率(Precision,P)和 F1 值(F1)。其中,PA 定义为预测类别正确的像素数占总像素数的比例,MIOU 为预测的结果和真实值的交集和并集之比,P 表示为预测分割目标类别正确的像素准确率,F1 值(F1)为综合考虑召回率和精准率的平衡指标,同时兼顾分割的精准程度和完整性,且召回率可以反映真实标签为分割目标的像素被正确分割的比例,衡量分割区域的完整性。

3.3　实验结果分析

3.3.1　不同分割模型性能对比分析

基于测试集样本数据分别对本文网络与 U－Net、SegNet、PSPNet、DeepLabV3 进行性能测试,表 3 为 5 种网络分割性能参数对比结果。由表 3 可知,本文设计的植株分割网络分割性能最优,对数据集 1,其各项指标较 U－Net、SegNet、PSPNet、DeepLabV3 均有提高,其中,MIOU 较 U－Net 提高了 0.019;对数据集 2,本文网络在 MIOU 评价指标上达到 0.932,F1 值为 0.978,相较于目前最为先进的 DeepLabV3 网络模型,分别提高了 0.041、0.042;对数据集 3,图像中目标植株较多且细小,图像的细节信息更复杂,本文网络较 U－Net 网络,MIOU 提高了 0.07,PA 提高了 0.123。可见,本文网络中边缘感知模块和特征融合模块的设计,能够促使网络在训练过程中学习到更多的细节信息,获取较好的植株分割性能。此外,本文基于深度可分离卷积对主干网络进行了轻量化设计,本文网络参数共 9.6×10^6 个,而原始 U－Net 的网络参数为 2.42×10^7 个,且对单帧图像的平均检测速度比本文网络低 32%,表明本文网络具有参数规模小、实时性高的特点,适合应用于硬件系统资源有限的智能农业设备。

表 3　不同模型作物植株分割比较

Tab.3　Performance comparison of different segmentation models

方法	数据集 1				数据集 2				数据集 3			
	PA	MIOU	P	F1 值	PA	MIOU	P	F1 值	PA	MIOU	P	F1 值
DeepLabV3	0.963	0.700	0.934	0.920	0.931	0.891	0.950	0.936	0.886	0.836	0.983	0.944
PSPNet	0.962	0.723	0.937	0.496	0.824	0.808	0.944	0.854	0.817	0.721	0.973	0.910
SegNet	0.945	0.725	0.938	0.956	0.899	0.759	0.970	0.866	0.653	0.504	0.935	0.664
U－Net	0.964	0.733	0.941	0.962	0.951	0.912	0.986	0.964	0.783	0.823	0.955	0.972
本文模型	0.975	0.752	0.943	0.964	0.962	0.932	0.991	0.978	0.906	0.893	0.991	0.968

图 6 为不同网络模型的分割效果,前 5 幅测试图来源于数据集 1、2、3,后 2 幅玉米苗期图不同于训练集,用于验证网络的泛化性。由图 6 可知,U－Net 网络虽可分割出较全的植株区域,但易将背景像素错误分类为植株像素,不利于后续的植株定位、生长检测等。SegNet 存在严重的边缘信息缺失,且对于细小区域的像素如植株茎部位难以准确分割,尤其是对包含细节信息较多的数据集 3 的分割效果较差,无法分割出完整植株。DeepLabV3 和 PSPNet 的分割结果存在明显的边缘粗糙和边缘信息缺失问题。而本文网络在 3 个数据集及玉米苗期图像都可正确地分割出植株,且植株边缘信息能较好地保留,表明网络中边缘标签监督能在网络训练过程中增加植株边缘的约束条件,具有更好的高层语义和低层边缘特征提取能力,避免背景误分割为植株。

3.3.2　EAM 与 FFM 对网络分割性能的影响

为了验证本文设计的植株分割网络中 EAM、FFM 的有效性,基于测试集样本数据分别对主干网络、主干网络＋EAM 及主干网络＋EAM＋FFM 3 种结构进行性能测试,表 4 为 3 种结构性能参数对比结果。由表 4 可知,主干网络＋EAM 和主干网络＋EAM＋FFM 在 3 个数据集上各项评价指标均优于主干网络,其中主干网络＋EAM＋FFM 结构在 3 个数据集上获得的 MIOU 相对于主干网络分别提升了 0.024、0.029、0.047。表明将边缘检测任务与语义分割任务进行联合训练,边缘感知模块对语义分割网络产生正向的效果,能够增加分割网络对边缘信息的感知,有效改善语义分割的结果。

为了更直观展示 FFM 模块对边缘信息与区域语义信息的融合作用,图 7 给出了主干网络＋EAM 和主干网络＋EAM＋FFM 的分割效果图,其中,图 7e 为边缘叠加图,右下角为图中标记框的放大显示,图中绿色、蓝色、红色分别代表标签边缘、主干网络＋EAM 分割结果边缘及主干网络＋EAM＋FFM 分割结果边缘。对比边缘线可见,红色边缘线比蓝色线更加逼近绿色边缘线,结合表 4 中的数据,主干网络＋EAM＋FFM 的分割性能较主干网络＋EAM 有进一步提升,表明特征融合模块 FFM 可充分利用边缘感知模块提取的边缘特征信息来弥补主干网络

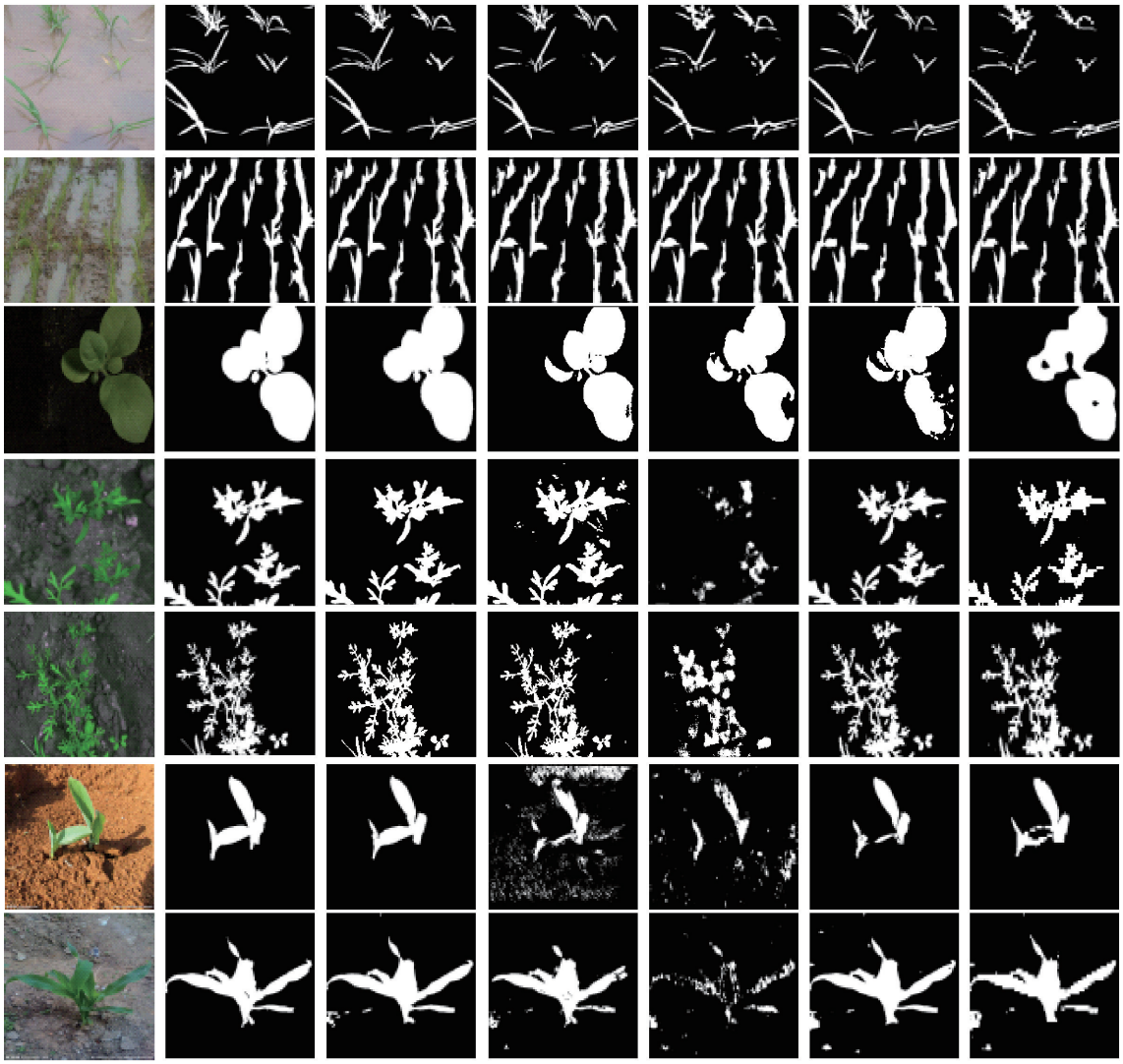


图 6 植株分割结果对比

Fig. 6 Comparisons of plant segmentation results

表 4 不同结构对测试图像的分割性能比较

Tab. 4 Performance comparison of different modules on testing images

方法	数据集 1				数据集 2				数据集 3			
	PA	MIOU	P	F1 值	PA	MIOU	P	F1 值	PA	MIOU	P	F1 值
主干网络	0.963	0.728	0.940	0.948	0.936	0.903	0.953	0.938	0.796	0.846	0.950	0.966
主干网络 + EAM	0.971	0.739	0.942	0.964	0.957	0.921	0.988	0.972	0.860	0.867	0.989	0.970
主干网络 + EAM + FFM	0.975	0.752	0.943	0.964	0.962	0.932	0.991	0.978	0.906	0.893	0.991	0.968

提取的特征中遗失的细节边缘信息,获取更完整的边缘信息,从而得到具有精细边缘的分割结果,提高分割精度。

3.3.3 损失函数对网络分割性能的影响

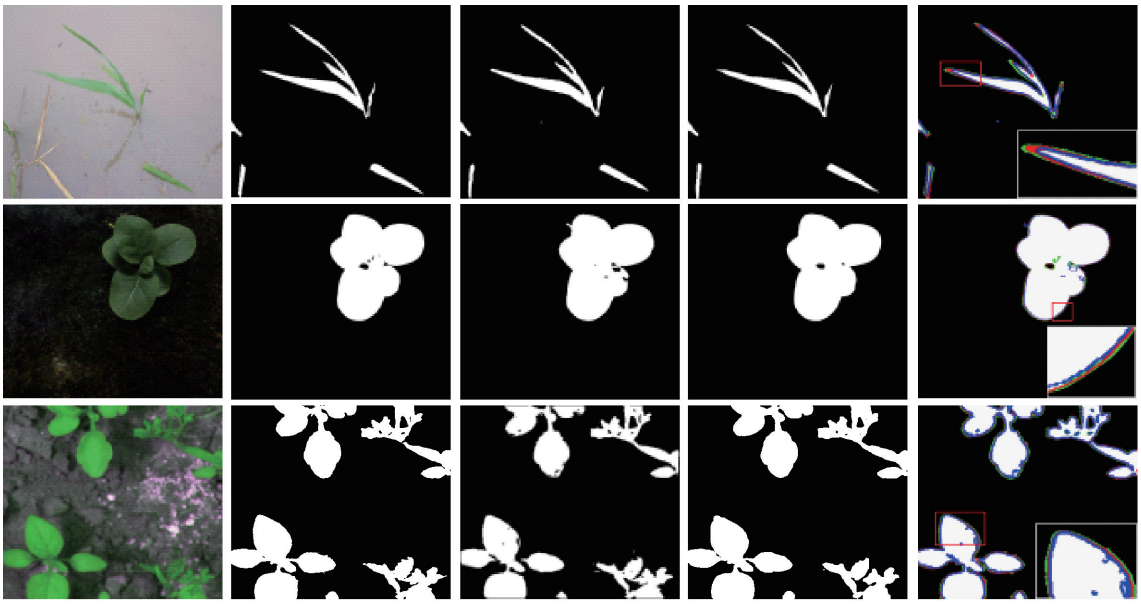
网络训练过程中,损失函数通过对预测结果和标签数据产生的误差反向传播指导网络参数学习。为验证本文设计的联合损失函数的有效性,对网络的损失函数进行相关的消融实验。联合损失函数分别定义为

$$l_{\text{Loss1}} = l_{\text{bce}} + l_{\text{hed}} \tag{12}$$

$$l_{\text{Loss2}} = l_{\text{bce}} + l_{\text{hed}} + l_{\text{ssim}} \tag{13}$$

$$l_{\text{Loss3}} = l_{\text{bce}} + l_{\text{miou}} + l_{\text{hed}} + l_{\text{ssim}} \tag{14}$$

其中,l_{Loss3}为本文设计的联合损失函数。以 3.2 节中提及的评价指标,对比本文设计的分割网络在 3 种损失函数下各数据集的分割性能如表 5 所示。由表 5 可见,在 3 个数据集上,l_{Loss2}的 MIOU 比 l_{Loss1}高 0.007 ~ 0.014,证实将 SSIM 损失函数项引入边缘感知模块,边缘感知模块可获取更丰富的边缘特征,引导主干网络关注植株的边缘信息,提高网络整体的分割效果;进一步引入 MIOU 损失函数后,在数据集 1



(a) 原图　(b) 标签　(c) 主干网络+EAM　(d) 主干网络+EAM+FFM　(e) 边缘叠加图

图 7　不同结构分割结果对比图

Fig. 7　Comparisons of segmentation results of different modules

表 5　不同 Loss 函数对测试图像的分割性能比较

Tab. 5　Performance comparison of different Loss functions on testing images

联合损失函数	数据集 1				数据集 2				数据集 3			
	PA	MIOU	P	F1 值	PA	MIOU	P	F1 值	PA	MIOU	P	F1 值
l_{Loss1}	0.973	0.741	0.938	0.956	0.962	0.898	0.986	0.952	0.863	0.864	0.983	0.961
l_{Loss2}	0.977	0.748	0.945	0.966	0.958	0.912	0.987	0.964	0.874	0.872	0.990	0.964
l_{Loss3}	0.975	0.752	0.943	0.964	0.962	0.932	0.991	0.978	0.906	0.893	0.991	0.968

上，F1 值、P、PA 虽因人工标注误差存在而无提升，但 MIOU 提升 0.004，且对于数据集 2、3，在各评价指标上均有提升，最高提升 0.021，表明在 SSIM、MIOU 损失函数项的共同作用下 l_{Loss3} 能有效提升网络分割性能。

3.3.4　不同算法的边缘检测性能对比

为衡量本文构建的网络模型在提取苗期植株边缘上的性能，将本文构建的网络模型获取的边缘结果与 HED 边缘检测模型和 Canny 算子获取的边缘结果进行对比，其中，量化评价指标为 F1 值、P 和平均绝对误差(Mean absolute error，MAE)。MAE 表示边缘预测图与边缘标签图之间的绝对像素误差，定义式为

$$E_{MAE}=\frac{1}{HW}\sum_{j=1}^{H}\sum_{i=1}^{W}|S(i,j)-G(i,j)| \quad (15)$$

式中　E_{MAE}——绝对像素误差

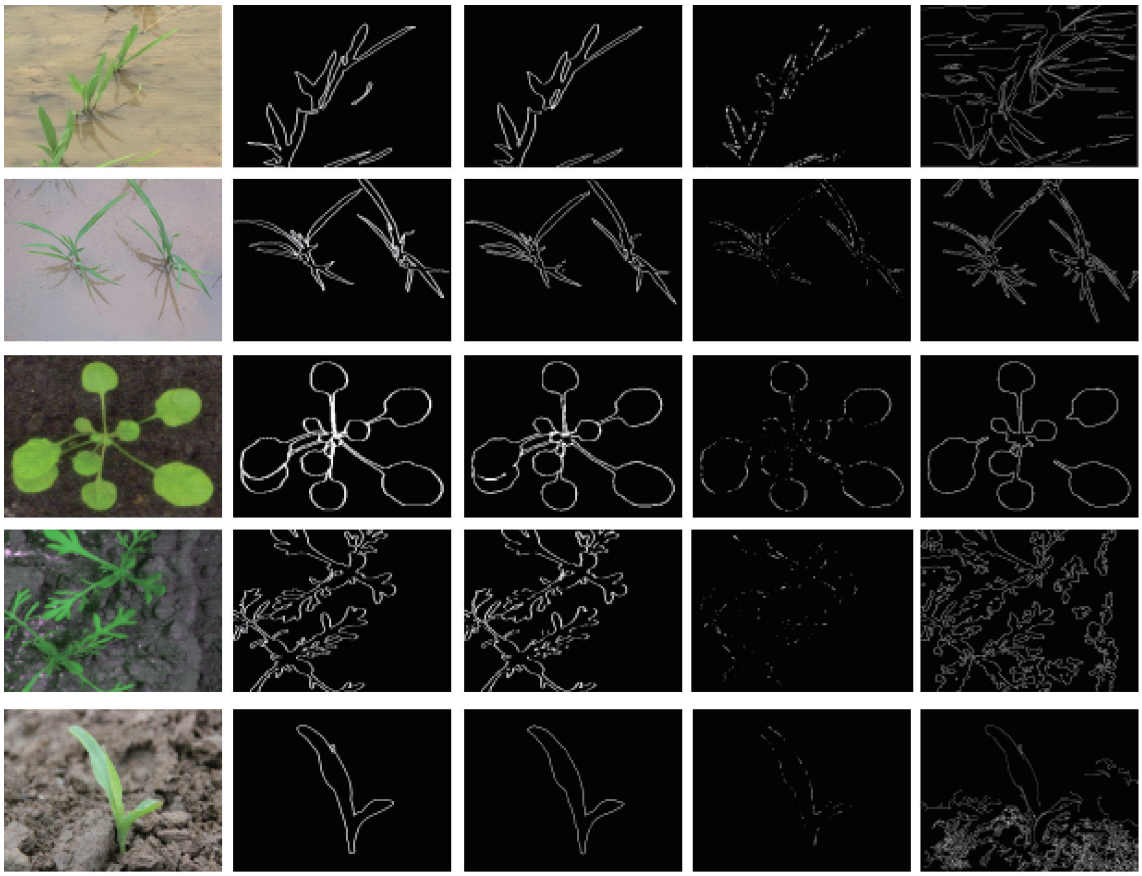
$S(i,j)$——边缘预测图在像素坐标(i,j)处的像素值

$G(i,j)$——边缘标签图在像素坐标(i,j)处的像素值

图 8 为本文网络及 HED 和 Canny 算子在测试集及玉米苗期植株的边缘分割结果。可见，对于仅有边缘检测任务的 HED 网络，由于缺少区域语义信息的补充，检测的植株边缘断裂，造成边缘信息缺失。Canny 算子基于浅层特征区分植株边缘和非植株边缘，易将非植株边缘误检为植株边缘，且边缘细化效果差。而本文网络对 3 个测试集均可获取准确的植株边缘，且边缘细化效果较好，如表 6 所示，本文网络对 3 个测试集的边缘检测精准率均在 0.9 以上，且 MAE 低于 HED 和 Canny 算子，表明边缘检测任务与语义分割任务的联合训练，能够有效提高作物苗期植株边缘细化效果，弥补区域语义分割中的边缘缺失问题。此外，本文网络对未经训练的玉米苗期植株也可获取完整的边缘，具有良好的泛化性。

4　结论

(1)以 U－Net 构建主干网络，通过侧边深度监督策略构建具有缓冲单元的边缘感知模块，缓冲两种监督冲突，并增强网络特征提取过程中对植株边缘的感知能力；基于空间特征金字塔结构构建特征融合模块，充分融合区域语义特征和边缘特征，融合后的特征图中具有更强的植株区域语义信息和浅层细节信息，有效提高了作物植株分割精度和边缘细化效果。



(a) 原图 (b) 边缘标签 (c) 本文网络 (d) HED (e) Canny算子

图 8 植株边缘分割结果

Fig. 8 Segmentation results of plant edge

表 6 不同算法的边缘分割性能比较

Tab. 6 Performance comparison of different methods on edge segmentation

方法	数据集 1			数据集 2			数据集 3		
	F1 值	P	MAE	F1 值	P	MAE	F1 值	P	MAE
Canny 算子	0.537	0.703	0.330	0.568	0.953	0.141	0.265	0.941	0.159
HED	0.388	0.889	0.569	0.368	0.818	0.083	0.787	0.912	0.112
本文网络	0.908	0.926	0.060	0.864	0.964	0.052	0.954	0.948	0.059

(2)将边缘感知模块损失函数和特征融合模块损失函数联合,构建联合损失函数,用于整体网络参数的学习优化,进一步提高网络分割精度。与当前主流的分割模型 U - Net、SegNet、PSPNet、DeepLabV3 对比,对不同数据集的作物植株的语义分割像素准确率高达 0.962,平均交并比达到 0.932,具有良好的植株分割效果和泛化能力。在后续的研究中将对缓冲边缘感知模块对主干网络的监督冲突做进一步研究,以及研究网络的轻量化,在保证网络模型分割精度基础上提高其效率。

参 考 文 献

[1] 杨长辉,刘艳平,王毅,等. 自然环境下柑橘采摘机器人识别定位系统研究[J/OL]. 农业机械学报, 2019, 50(12): 14 - 22,72.
YANG Changhui, LIU Yanping, WANG Yi, et al. Research and experiment on recognition and location system for citrus picking robot in natural environment[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 14 - 22,72. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20191202&flag = 1. DOI:10.6041/j.issn.1000-1298.2019.12.022. (in Chinese)

[2] RIEHLE D, REISER D, GRIEPENTROG H W. Robust index-based semantic plant/background segmentation for RGB-images [J]. Computers and Electronics in Agriculture, 2020, 169: 105201.

[3] 伍艳莲,赵力,姜海燕,等. 基于改进均值漂移算法的绿色作物图像分割方法[J]. 农业工程学报,2014,30(24): 161 - 167.
WU Yanlian, ZHAO Li, JIANG Haiyan, et al. Image segmentation method for green crops using improved mean shift [J]. Transactions of the CSAE, 2014, 30(24): 161 - 167. (in Chinese)

[4] 王侨,孟志军,付卫强,等. 基于机器视觉的玉米苗期多条作物行线检测算法[J/OL]. 农业机械学报, 2021,52(4):208 - 220.
WANG Qiao, MENG Zhijun, FU Weiqiang, et al. Detection algorithm of multiple crop row lines based on machine vision in maize seedling stage[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 208 - 220. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20210422&flag = 1. DOI:10.6041/j.issn.1000-1298.2021.04.022. (in Chinese)
[5] KAMILARIS A, PRENAAFET-BOLDÚ F X. Deep learning in agriculture: a survey[J]. Computers and Electronics in Agriculture, 2018, 147: 70 - 90.
[6] 段凌凤,熊雄,刘谦,等. 基于深度全卷积神经网络的大田稻穗分割[J]. 农业工程学报,2018,34(12):202 - 209.
DUAN Lingfeng, XIONG Xiong, LIU Qian, et al. Field rice panicles segmentation based on deep full convolutional neural network[J]. Transactions of the CSAE, 2018, 34(12): 202 - 209. (in Chinese)
[7] MILIOTO A, LOTTES P, STACHNISS C. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs[C]//International Conference on Robotics and Automation, 2018: 2229 - 2235.
[8] 韩振浩,李佳,苑严伟,等. 基于 U - Net 网络的果园视觉导航路径识别方法[J/OL]. 农业机械学报, 2021, 52(1): 30 - 39.
HAN Zhenhao, LI Jia, YUAN Yanwei, et al. Path recognition of orchard visual navigation based on U - Net[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1): 30 - 39. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20210104&flag = 1. DOI:10.6041/j.issn. 1000-1298. 2021.01.004. (in Chinese)
[9] OLAF R, PHILIPP F, THOMAS B. U - Net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention, 2015: 234 - 241.
[10] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[J]. Lecture Notes in Computer Science, 2014,8689: 818 - 833.
[11] XIE S, TU Z. Holistically-nested edge detection[C]//International Conference on Computer Vision,2017:3 - 18.
[12] CHENG D, MENG G, XIANG S, et al. FusionNet: edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(12): 5769 - 5783.
[13] TAKIKAWA T, ACUNA D, JAMPANI V, et al. Gated - SCNN: gated shape CNNs for semantic segmentation[C]// International Conference on Computer Vision, 2019: 5228 - 5237.
[14] CHEN L, ZHU Y, PAPANDEROU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision, 2018: 801 - 818.
[15] MINERVINI M, FISCHBACH A, SCHARR H, et al. Finely-grained annotated datasets for image-based plant phenotyping [J]. Pattern Recognition Letters, 2016, 81: 80 - 89.
[16] HAUG S, OSTERMANN J. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks [C]//European Conference on Computer Vision, 2014: 105 - 116.
[17] HE K, SUN J. Convolutional neural networks at constrained time cost[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 5353 - 5360.
[18] 陈进,韩梦娜,练毅,等. 基于 U - Net 模型的含杂水稻籽粒图像分割[J]. 农业工程学报, 2020, 36(10): 174 - 180.
CHEN Jin, HAN Mengna, LIAN Yi, et al. Segmentation of impurity rice grain images based on U - Net model [J]. Transactions of the CSAE, 2020, 36(10): 174 - 180. (in Chinese)
[19] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//International Conference on Computer Vision and Pattern Recognition, 2018: 4510 - 4520.
[20] 孟庆宽,张漫,杨晓霞,等. 基于轻量卷积结合特征信息融合的玉米幼苗与杂草识别[J/OL]. 农业机械学报,2020, 51(12):238 - 245.
MENG Qingkuan, ZHANG Man, YANG Xiaoxia, et al. Recognition of maize seedling and weed based on light weight convolution and feature fusion[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(12): 238 - 245. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20201226&flag = 1. DOI:10.6041/j.issn. 1000-1298.2020.12.026. (in Chinese)
[21] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines[C]//International Conference on Machine Learning, 2010:807 - 814.
[22] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning, 2015:448 - 456.
[23] LIU Y, CHENG M, FAN D, et al. Semantic edge detection with diverse deep supervision[J]. arXiv e-prints, 2018: 1804 - 2864.
[24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//International Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.
[25] RAHMAN M A, WANG Y. Optimizing intersection-over-union in deep neural networks for image segmentation[C]// International Symposium on Visual Computing, 2016: 234 - 244.
[26] QIN X, ZHANG Z, HUANG C, et al. BASNet: boundary-aware salient object detection[C]//International Conference on Computer Vision and Pattern Recognition, 2019: 7471 - 7481.