

多通道深度可分离卷积模型实时识别复杂背景下甜菜与杂草

孙 俊, 谭文军, 武小红, 沈继锋, 芦 兵, 戴春霞

(江苏大学电气信息工程学院, 镇江 212013)

摘 要: 针对实际复杂田间环境下杂草与作物识别精度低和实时性差的问题, 为减少弱光环境对分割识别效果的影响, 实现甜菜与杂草的实时精确分割识别, 该文首先将可见光图像进行对比度增强, 再将近红外与可见光图像融合为 4 通道图像; 将深度可分离卷积以及残差块构成分割识别模型的卷积层, 减少模型参数量及计算量, 构建编码与解码结构并融合底层特征, 细化分割边界。以分割识别精度、参数量以及运行效率为评价指标, 通过设置不同宽度系数以及输入图像分辨率选出最优模型。试验结果表明: 本文模型的平均交并比达到 87.58%, 平均像素准确率为 99.19%, 帧频可达 42.064 帧/s, 参数量仅为 525 763, 具有较高分割识别精度和较好实时性。该方法有效实现了甜菜与杂草的精确实时识别, 可为后续机器人精确除草提供理论参考。

关键词: 作物; 图像分割; 卷积神经网络; 深度学习; 甜菜; 杂草; 实时

doi: 10.11975/j.issn.1002-6819.2019.12.022

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2019)-12-0184-07

孙 俊, 谭文军, 武小红, 沈继锋, 芦 兵, 戴春霞. 多通道深度可分离卷积模型实时识别复杂背景下甜菜与杂草[J]. 农业工程学报, 2019, 35(12): 184—190. doi: 10.11975/j.issn.1002-6819.2019.12.022 <http://www.tcsae.org>
Sun Jun, Tan Wenjun, Wu Xiaohong, Shen Jifeng, Lu Bing, Dai Chunxia. Real-time recognition of sugar beet and weeds in complex backgrounds using multi-channel depth-wise separable convolution model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(12): 184—190. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2019.12.022 <http://www.tcsae.org>

0 引 言

杂草是影响作物生长的重要因素之一, 会与作物争肥争水, 增加病虫害的传播, 故及时去除杂草对于保证作物高产具有重要意义。传统除草一般是人工除草, 通过人力挖除或者喷药实现除草, 但需要耗费大量人力。另外喷药产生的农药残留不仅对人类健康产生巨大威胁, 而且容易破坏生态环境^[1-2]。随着绿色农业的概念提出, 为了提高作物产量、减少环境污染, 实现智能化除草, 解决杂草与作物实时精确识别问题就显得尤为重要^[3-6]。

机器视觉在农业中应用广泛^[7-8], 目前大多数杂草与作物识别研究也是基于机器视觉技术, 通常包含特征提取与模式识别两个步骤。何东健等^[9]提取植物叶片形状、纹理等特征并提出了一种多特征融合的方法, 最后通过支持向量机 (support vector machine, SVM) 建模识别杂草。赵川源等^[10]将多光谱图像转换为 Lab 颜色空间, 提取叶片多种特征, 并采用 C4.5 算法对杂草进行识别。Pulido 等^[11]计算图像灰度共生矩阵 (gray-level co-occurrence matrix, GLCM), 然后使用支持向量机分类器区分杂草与蔬菜。以上文献均是先提取特征然后通过

分类器实现杂草识别, 虽然在特定环境下取得较好效果, 但是实际杂草识别中, 识别效果容易受光照、遮挡、以及图像采集质量等因素的影响, 所以提高杂草识别方法的鲁棒性在传统机器视觉中仍然是一个难以解决的问题。

深度学习^[12]是机器学习的一个分支, 其中的卷积神经网络 (convolutional neural network, CNN) 成为了图像处理中一个研究热点。近些年, 卷积神经网络在图像识别^[13-16]、语义分割^[17-18]、目标检测^[19-20]中取得重大突破, 由于其对图像特征表征能力强, 所以越来越多学者将卷积神经网络应用在农业领域。王璨等^[21]以幼苗期玉米及杂草为研究对象, 构建卷积神经网络模型, 提取图像高斯金字塔多尺度特征, 并结合超像素分割避免杂草识别中重叠问题。孙俊等^[22]提出了一种空洞卷积与全局池化相结合的多尺度特征融合卷积模型, 有效的识别多种杂草。Tang 等^[23]采用 K-means 学习的图像特征, 并将其参数替换卷积神经网络的随机初始化参数, 最后通过微调模型参数的方法提高杂草识别精度。上述文献均是采用基于分类思想实现杂草识别, 为了更精确定位杂草与作物位置, Milioto 等^[24]采用基于语义分割的模型实现端到端的图像像素点分类, 提取传统图像特征并按通道进行像素级叠加从而构建 14 通道的输入图像, 获得了较好分割识别效果。

以上文献表明将卷积神经网络模型应用于杂草识别是可行的, 且能避免传统机器视觉方法中特征提取过程带来的主观性, 提高了杂草识别精度。但是, 由于实际环境下除草机器设备配置低, 因此如何在有限硬件环境配置下节省资源消耗, 达到实时精确的识别效果是需要解决的一个关键问题。为解决上述问题, 本文以甜菜为

收稿日期: 2018-10-21 修订日期: 2019-03-06

基金项目: 国家自然科学基金资助项目 (No.31471413); 江苏高校优势学科建设工程资助项目 PAPD (苏政办发 2011 6 号); 江苏省六大人才高峰资助项目 (ZBZZ-019)。

作者简介: 孙 俊, 博士, 教授, 博士生导师。研究方向为计算机技术在农业工程中的应用。Email: sun2000jun@ujs.edu.cn

中国农业工程学会会员: 孙俊 (E041200652S)

对象,通过获取田间甜菜的近红外以及可见光图像,构建了 4 通道输入图像,并提出了一种编码与解码型轻量级卷积神经网络结构,通过加入深度可分离卷积减少模型参数量与计算量。本文最后通过试验研究模型参数优化选择,为后续智能除草机的研制提供理论依据。

1 试验数据

本文采用 2016 年德国波恩大学采集的甜菜与杂草图片作为数据集^[25],该数据集在德国波恩的农场采集,数据集涵盖了不同生长阶段的甜菜图片。采集图片的机器人上配备 JAI AD-130 GE 摄像头,可以同时获取 RGB 图片和近红外 (near infrared, NIR) 图片。由于像素级别的图片标注难度大,故该数据集中带标签的图片较少,共 283 张。数据集样例如图 1 所示,图片包含甜菜和其他各类杂草,其中标签的红色区域表示甜菜,其他颜色区域表示杂草。

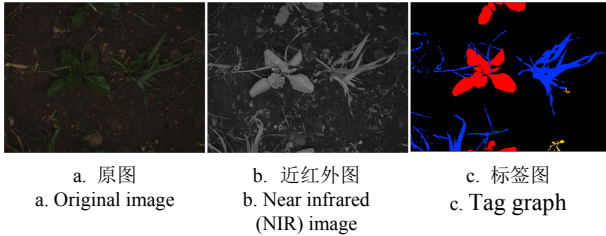


图 1 数据集示例

Fig.1 Examples of dataset

深度学习是数据驱动的自动特征学习方法,通常深度学习语义分割模型的输入图片是 RGB 三通道图片。由于已标注图片数量过少,训练一个鲁棒性较好的深度学习模型难度较大,一方面,浅层次模型难以拟合复杂背景图片,另一方面,设计深层模型对数据量要求高,容易发生拟合现象。针对上述问题,Milioto 等^[24]在论文中提出增加额外特征图作为卷积神经网络模型的输入,其试验结果证明了特征融合的方法确实可以提升模型的精度。本文参考该论文的方法,与其不同的是,本文先对原始 RGB 图像进行对比度增强,然后将图像增强后的图片与 NIR 图片进行融合,构成 4 通道输入图像,避免原始图片特征冗余和减少计算量。随机选取 226 张图片作为训练集,剩余 57 张作为测试集。

2 甜菜与杂草分割识别模型

2.1 深度可分离卷积

传统卷积神经网络采用大尺寸卷积核,如 AlexNet 中使用 11×11 尺寸的卷积核,虽然得到的感受野较大,但使模型的参数量也大大增加。GoogLeNet 模型中使用多个 3×3 的小尺寸卷积核级联结构,在保持对原图感受野相同的同时使模型的参数量大量减少,并增加了网络的深度,增强对图像特征的非线性表达能力。由于传统卷积核作用于输入特征图的每个通道,计算量仍然太大,Howard 等^[26]使用深度可分离卷积构建卷积神经网络使其在移动端也能高效运行。深度可分离卷积的思想是将传统的卷积操作分为 2 步:首先,对输入特征图的每一个通道进行一对一的 2 维卷积用于减少参数计算,该过程

被称为 depth-wise 卷积;然后用 1×1 大小卷积核继续做传统卷积 (3 维卷积) 操作从而组合各通道特征,也称 point-wise 卷积。深度可分离卷积的结构如图 2 所示。假设输入特征图大小为 $D_f \times D_f$,通道数为 M ,卷积核尺寸为 $D_k \times D_k$,共 N 个,则传统卷积与深度可分离卷积的计算量分别如式 (1) 和式 (2) 所示。

$$D_f \times D_f \times M \times N \times D_k \times D_k \quad (1)$$

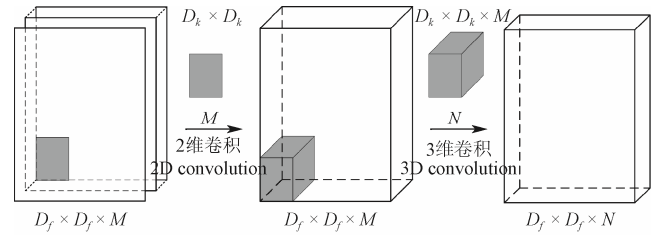
$$D_f \times D_f \times M \times D_k \times D_k + D_f \times D_f \times M \times N \quad (2)$$

故深度可分离卷积与传统卷积的计算量比为

$$\frac{D_f \times D_f \times M \times D_k \times D_k + D_f \times D_f \times M \times N}{D_f \times D_f \times M \times N \times D_k \times D_k} \quad (3)$$

$$= \frac{1}{N} + \frac{1}{D_k \times D_k}$$

可见,深度可分离卷积计算量的减少与输出通道 N 和使用的卷积核尺寸 $D_k \times D_k$ 有关。实际中,深度可分离卷积一般使用 3×3 大小卷积核,如果输出通道为 64,通过式 (3) 可计算得到深度可分离卷积的计算量仅为传统卷积参数计算量的 0.126 倍。



注: $D_k \times D_k$ 表示卷积核的大小; $D_f \times D_f$ 表示特征图大小; M 与 N 为通道数。

Note: $D_k \times D_k$ is the size of the convolution kernel; $D_f \times D_f$ is the size of feature map; M and N are the number of channels.

图 2 深度可分离卷积结构图

Fig.2 Structure of depth-wise separable convolution

用参数 r 表示 depth-wise 卷积的输出与输入通道数的倍数关系,上述深度可分离卷积结构在进行 depth-wise 卷积时输出通道数与输入通道数相同,即上述结构的 r 为 1。由于在输入通道数较少的情况下,模型特征图对图像特征的表达能力较弱,因此本文将在 depth-wise 卷积时对输入通道数进行扩张,以增强模型非线性映射能力,提高模型精度。

2.2 分割模型结构

编码与解码器结构在 SegNet^[27]之后被广泛应用在语义分割模型中,编码结构用于提取图像深层特征,随着层次变深,特征图尺寸逐渐变小,特征也越抽象。解码结构负责将深层的特征图重新采样至原图尺寸,通过对大量图片的学习,得到原图与标签之间的映射关系。

本文模型将输入图片统一缩放至同一尺寸,输入层通道数根据图片预处理后的通道数确定,假设输入数据是 RGB 图像与近红外图像 NIR 的融合,则输入通道为 4。

模型下采样层参考 E-net 结构^[28],使用最大池化层和步长为 2 的卷积层对输入特征图进行下采样,其中卷积核大小为 3×3 。杂草分割模型结构如图 3 所示,每经过 1 次下采样,特征图尺寸缩小至原来的一半。深度可分离

卷积的步长为 1, 故输入与输出的特征图尺寸相同, 并通过残差块连接。由于目标图像包含各个生长阶段的作物与杂草, 其覆盖的像素范围不同, 大感受野对于覆盖像素范围较大的物体分割效果好, 同样分割小目标则需要较小感受野。为了实现对覆盖像素范围小的作物或杂草精确分割, 上采样层使用反卷积, 并在解码结构中将感受野较小的浅层特征与上采样层的输出特征融合, 然后级联深度可分离卷积层, 本文共采用 3 次上采样将低分辨率特征图逐渐还原为原图尺寸。

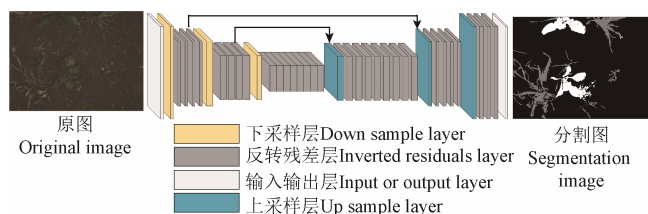


图 3 杂草分割模型结构图

Fig.3 Structure of weed segmentation model

以输入图像尺寸 384×384 为例, 图 3 所示模型结构的各层具体参数如表 1 所示。其中 r 是指宽度系数, Depth-wise residuals layer 指深度可分离卷积组成的残差块。Downsample layer 为下采样层, 由最大池化层 (Max pooling) 和普通卷积层 (Conv) 构成, 步长 (stride) 均为 2, 其中最大池化层的池化核大小为 2×2 , 卷积层的卷积核大小为 3×3 。下采样层的输出由池化层和卷积层的特征图叠加构成, 以下采样层 1 为例, 其输出通道 32 是由池化核为 2×2 的 3 通道和卷积核为 3×3 的 29 通道叠加构成。Upsample layer 为上采样层, 使用步长为 2、大小为 2×2 的转置卷积。

表 1 杂草分割模型参数表

Table 1 Table of weed segmentation model parameter

层 Layer	输出尺寸 Output size	模型参数 Model parameters
下采样层 1	$192 \times 192 \times 32$	$2 \times 2, 3$, 最大池化层, 步长为 2 $3 \times 3, 29$, 普通卷积层, 步长为 2
残差块 1 ($\times 4$)	$192 \times 192 \times 32$	$\begin{bmatrix} 3 \times 3, 32 * r \\ 1 \times 1, 32 \end{bmatrix} \times 4$
下采样层 2	$96 \times 96 \times 64$	$2 \times 2, 32$, 最大池化层, 步长为 2 $3 \times 3, 32$, 普通卷积层, 步长为 2
残差块 2 ($\times 4$)	$96 \times 96 \times 64$	$\begin{bmatrix} 3 \times 3, 64 * r \\ 1 \times 1, 64 \end{bmatrix} \times 4$
下采样层 3	$48 \times 48 \times 128$	$2 \times 2, 64$, 最大池化层, 步长为 2 $3 \times 3, 64$, 普通卷积层, 步长为 2
残差块 3 ($\times 8$)	$48 \times 48 \times 128$	$\begin{bmatrix} 3 \times 3, 128 * r \\ 1 \times 1, 128 \end{bmatrix} \times 8$
上采样层 1	$96 \times 96 \times 64$	$2 \times 2, 64$, 转置卷积层, 步长为 2,
残差块 4 ($\times 8$)	$96 \times 96 \times 64$	$\begin{bmatrix} 3 \times 3, 64 * r \\ 1 \times 1, 64 \end{bmatrix} \times 8$
上采样层 2	$192 \times 192 \times 32$	$2 \times 2, 32$, 转置卷积层, 步长为 2
残差块 5 ($\times 4$)	$192 \times 192 \times 32$	$\begin{bmatrix} 3 \times 3, 32 * r \\ 1 \times 1, 32 \end{bmatrix} \times 4$
上采样层 3	$384 \times 384 \times 32$	$2 \times 2, 32$, 转置卷积层, 步长为 2
残差块 6 ($\times 4$)	384×384	$\begin{bmatrix} 3 \times 3, 32 * r \\ 1 \times 1, 32 \end{bmatrix} \times 4$

注: $\times 4$ 、 $\times 8$ 指该层在模型中通过级联的方式重复级联 4 次与 8 次; r 是指宽度系数; * 表示乘。下同。

Note: $\times 4$ and $\times 8$ indicate that the layer is cascaded 4 times and 8 times in the model respectively; r is the width factor; * indicates multiplication. The same below.

3 模型训练

3.1 试验平台

本文试验均在 Ubuntu 16.04 LTS 64 位系统环境下运行, 采用 Bonnet^[29]及 Tensorflow 深度学习框架。计算机硬件配置为 16 GB 内存, 搭载 Intel® Core™ i7-7700K CPU 和 GTX1080Ti 显卡。

3.2 试验参数设置

采用带动量的 Adam 算法优化梯度下降, 其中一阶矩估计的指数衰减率 β_1 为 0.9, 二阶矩估计的指数衰减率 β_2 为 0.99, ε 取值 10^{-8} 。初始学习率设置为 0.001, batch size 为 5, 每迭代 50 个 epoch 学习率下降为原来的二分之一。采用 L2 和 dropout 正则化方法防止模型过拟合, 其中 L2 正则化系数 decay 为 10^{-4} , dropout 比率为 0.3。本文设置迭代 300 个 epoch 结束训练。由于图片集中作物、杂草、背景所占图像像素比例差异大, 类别不平衡, 为避免模型对背景信息过拟合, 使用带权值 ω_j 的损失函数 $L(y, \hat{y})$, 定义如下:

$$\begin{cases} L(y, \hat{y}) = -\sum_{i=1}^P \sum_{j=1}^C \omega_j y_{ij} \log(\hat{y}) \\ \omega_j = \frac{\bar{f}}{f_j} \\ f_j = \frac{P_j}{P}, j = 1, \dots, C \end{cases} \quad (4)$$

式 (4) 中, y 与 $\hat{y}$ 分别是像素的实际标签和模型预测的输出值, C 表示类别总数; y_{ij} 取值为 1 或 0, 为 1 时表示第 i 个像素的实际类别为第 j 类, 否则为 0; ω_j 为第 j 类的权重, f_j 为图片中第 j 类别像素数 P_j 占总像素 P 的比值, $\bar{f}$ 表示 f_j 的中位数。可见当模型对像素所占比例较小的杂草和作物误分时, 计算得到的损失值更大, 更利于作物与杂草图像的分割。

4 结果与分析

本文从模型的分割精度、参数量及效率 3 个方面进行模型性能评价, 分割精度主要使用平均像素准确率 (mean pixel accuracy)、平均交并比 (mean intersection over union, mIOU) 评估模型, 效率方面主要计算处理视频图像的帧数。最后权衡分割精度、参数量和效率选取最优模型, 并与文献[24]方法进行对比。

4.1 图像分辨率对分割精度的影响

图像特征与图像分辨率之间具有一定联系, 通常分辨率越高的图像, 其表征的特征也越丰富。研究表明^[26]保留较大图像分辨率对图像识别和图像检测具有重要意义。MobileNet 使用分辨率因子控制输入图像分辨率, YOLOv3^[30]采用 320、416、608 三种分辨率图像训练目标检测模型, 两者的研究结果均表明使用较高分辨率图像能提升识别和检测的效果。本文使用的数据集中原始图像分辨率为 1296×966 , 为平衡分割精度与计算量之间的关系, 将原始图像使用双线性插值法分别缩放为

384×384、512×384、512×512 三种分辨率，分别训练模型。不同输入图像分辨率下模型分割精度如表 2 所示。

表 2 不同图像分辨率下分割精度的比较
Table 2 Comparison of segmentation accuracy under different image resolutions

输入图片类型 Input picture type	输入图片尺寸 Size of input picture/pixel	交并比 Intersection union ratio /%				平均像素准确率 Mean pixel accuracy/%
		平均交并比 Mean intersection union ratio	土壤 Soil	杂草 Weed	作物 Crop	
RGB	384×384	80.09	98.62	53.99	87.67	98.59
	512×384	81.18	98.74	55.80	89.01	98.70
	512×512	83.16	98.82	60.92	89.75	98.83
RGB+NIR	384×384	85.96	99.14	67.84	90.9	99.05
	512×384	87.04	99.18	70.40	91.53	99.11
	512×512	87.26	99.19	70.68	91.92	99.17

表 2 所示模型结构采用的宽度系数均为 1，可见，在 3 种分辨率下，采用 3 通道的 RGB 图像和 4 通道 RGB+NIR 图像分别训练的模型分割精度随分辨率增大而提升。在输入通道为 4 时，分辨率从 384×384 调整为 512×384 之后，平均交并比从 85.96% 提升至 87.26%，当输入图像分辨率为 512×512 时，土壤、杂草、作物的交并比均取得最好效果，分别为 99.19%、70.68% 和 91.92%。此外，采用 4 通道 RGB+NIR 图像训练的模型分割精度优于直接采用普通 RGB 图像训练的模型，说明近红外 NIR 图像特征可以在一定程度上弥补普通 RGB 图像的缺陷，更加适应暗光环境。

4.2 宽度系数对分割精度的影响

宽度系数可以调整 depth-wise 卷积的输出通道，在输入特征图通道较少时，图像特征已经被压缩，此时若再使用相同通道的 depth-wise 卷积则会使模型的特征表征能力减弱。针对上述问题，本文研究了不同宽度系数对模型分割精度的影响，使用 1、2、4 三种宽度系数，共构建了 9 个模型，并与文献[24]中方法对比，结果见表 3。

表 3 不同宽度系数下分割精度的比较
Table 3 Comparison of segmentation precision under different width factor

方法 Methods	输入图片尺寸 Size of input picture/pixel	r	平均交并比 Mean intersection union ratio/%		平均像素准确率 Mean pixel accuracy /%	
			RGB	RGB+NIR	RGB	RGB+NIR
文献 [24]	512×384	-	73.68	78.96	90.34	94.12
		1	80.09	85.96	98.59	99.05
	384×384	2	82.38	86.21	98.74	99.08
		4	81.48	86.21	98.71	99.07
本文 方法	512×384	1	81.18	87.04	98.70	99.11
		2	83.56	87.58	98.88	99.19
		4	83.09	87.43	98.80	99.17
	512×512	1	83.16	87.26	98.83	99.17
		2	84.35	87.64	98.92	99.21
		4	83.62	87.53	98.90	99.20

如表 3，在相同输入图像分辨率情况下，使用宽度系数为 2 或 4 的模型分割精度均优于使用宽度系数为 1 的模型。当输入图像分辨率为 512×384 时，文献[24]中模型的平均交并比与平均像素准确率最高值分别为 78.96% 与 94.12%，而本文模型的平均交并比与平均像素准确率最高值分别能达到 87.58% 与 99.19%，分别提高 10.92% 和 5.39%，分割精度明显提升。

4.3 模型参数量以及运行效率比较

模型参数量与运行效率是衡量深度学习模型的一个重要指标，由于嵌入式设备软硬件环境配置较低，基于该环境下的模型实时推理仍然是现有研究的难点问题。本文计算了 2 种方法中模型的参数量并在测试集上测试了各模型的运行效率，其中运行效率通过帧频（在 1 s 内模型可处理的图片数）表示，具体见表 4。

表 4 模型参数量以及运行效率比较
Table 4 Comparison of model parameters and operation efficiency

方法 Methods	输入图片尺寸 Size of input picture/ pixel	r	参数量 Number of parameters		帧频 Frame rates 帧/s	
			RGB	RGB+NIR	RGB	RGB+NIR
文献[24]	512×384	-	1 037 950	1 038 171	32.637	32.524
		1	313 446	313 667	86.225	86.105
		2	525 542	525 763	61.163	60.232
本文 方法	384×384	4	949 734	949 955	40.824	40.327
		1	313 446	313 667	61.906	60.732
		2	525 542	525 763	42.187	42.064
	512×384	4	949 734	949 955	29.849	28.233
		1	313 446	313 667	45.814	44.701
		2	525 542	525 763	31.876	30.897
	512×512	4	949 734	949 955	21.669	21.537

模型参数量与使用的卷积核尺寸以及输入输出通道数有关，而与输入图像尺寸无关，故表 4 中 3 种输入图像分辨率下使用相同宽度系数的模型参数量相同。随着宽度系数的增加，模型参数量也大幅度增加，其中使用宽度系数为 4 的模型参数量约为使用宽度系数为 1 的 3 倍。模型的计算量与参数量不同，计算量与输入图像尺寸有关，图像尺寸越大，所需要的乘积操作也越多，故表 4 中帧频随着输入图像尺寸变大而逐渐降低。本文使用的 RGB+NIR 图像为 4 通道，相比使用原始 RGB 图像训练的模型，输入层为 4 通道模型的参数量与计算量稍有增加，其增加量仅在第一层下采样层，后面层参数与 RGB 图像训练的模型相同。综合比较不同分辨率输入、不同宽度系数下所有模型的分割精度、参数量和效率，本文最优模型为输入尺寸为 512×384 的 4 通道输入模型，该模型的宽度系数为 2，其分割平均交并比达到 87.58%，平均像素准确率为 99.19%，参数量仅为 525 763，帧频可达 42.064 帧/s。相对于文献[24]中的模型，本文最优模型参数量缩减为该模型的 50.64%，平均交并比与平均像素准确率分别提升了 8.62% 和 5.07%，能达到实时精确分割作物与杂草图像的效果。

4.4 作物与杂草图像分割效果

依据上述试验讨论结果,选择本文最优模型,随机抽取测试集中 3 张图片作为样例图片,获取测试集上杂草识别效果图。图 4 为甜菜与杂草分割识别效果图,第一列是原始图像,中间列是本文模型的预测结果图,最后一列是实际标签图。可以明显的看到本文模型可以精确区分杂草与甜菜,对小目标具有较好的识别效果,本文模型能较好保留甜菜与杂草的边缘细节信息,可实现图像完整分割。另外,对于杂草与甜菜重叠的情况具有较好适应性。

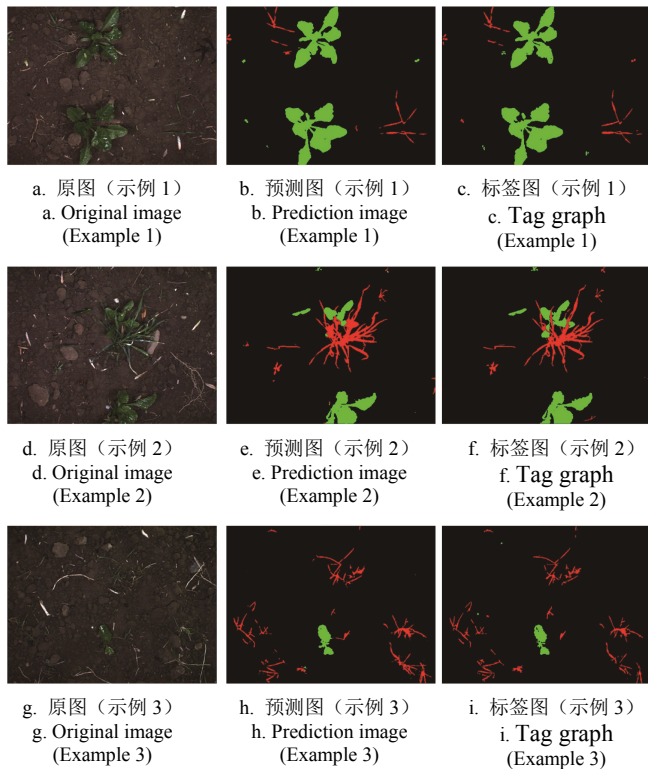


图 4 分割效果图

Fig.4 Effects of segmentation

5 结 论

本文提出了一种基于深度可分离卷积的卷积神经网络模型用于甜菜与杂草图像分割识别。获取田间甜菜的近红外和可见光图像,对原始可见光图像进行对比度增强,然后与近红外图像叠加,构成 4 通道图像。本文模型采用深度可分离卷积块减少模型参数量与计算量,利用残差块避免模型梯度消失和梯度弥散问题,并采用编码与解码结构以及融合底层特征实现甜菜与杂草图像实时分割识别。综合考虑分割精度、参数量以及运行效率 3 个方面,本文最优模型分割的平均交并比可达 87.58%,平均像素准确率为 99.19%,帧频可达 42.064 帧/s,具有较高分割识别精度和较好实时性,可为后续的智能除草机的研制提供理论依据。

[参 考 文 献]

[1] 齐月,李俊生,闫冰,等. 化学除草剂对农田生态系统野生植物多样性的影响[J]. 生物多样性, 2016, 24(2): 228—236.

Qi Yue, Li Junsheng, Yan Bing, et al. Impact of herbicides on wild plant diversity in agro-ecosystems: A review[J]. Biodiversity Science, 2016, 24(2): 228—236. (in Chinese with English abstract)

- [2] 杨靛,刘小娟,郭玉双. 农药残留快速检测技术研究进展[J]. 黑龙江农业科学, 2012(10): 150—153.
- Yang Liang, Liu Xiao Juan, Guo Yu Shuang. Progress of rapid detecting technique of pesticide residues[J]. Heilongjiang Agricultural Sciences, 2012(10): 150—153. (in Chinese with English abstract)
- [3] 侯学贵,陈勇,郭伟斌. 除草机器人田间机器视觉导航[J]. 农业机械学报, 2008, 39(3): 106—108.
- Hou Xuegui, Chen Yong, Guo Weibin. Machine vision-based navigation for a weeding robot[J]. Transactions of The Chinese Society For Agricultural Machinery, 2008, 39(3): 106—108. (in Chinese with English abstract)
- [4] 李碧青,朱强,郑仕勇,等. 杂草自动识别除草机器人设计——基于嵌入式 Web 和 ZigBee 网关[J]. 农机化研究, 2017, 39(1): 217—221.
- Li Biqing, Zhu Qiang, Zheng Shiyong, et al. Design for weeding robot based on embedded web and zigBee gateway[J]. Journal of Agricultural Mechanization Research, 2017, 39(1): 217—221. (in Chinese with English abstract)
- [5] 李谦,蔡晓华. 机器视觉在除草机器人中的应用[J]. 农机化研究, 2014(7): 204—206.
- Li Qian, Cai Xiaohua. Application of machine vision in weeding robot[J]. Journal of Agricultural Mechanization Research, 2014(7): 204—206. (in Chinese with English abstract)
- [6] 张文莉,陈树人,褚德宏. 除草机器人研究现状与趋势[J]. 农业装备技术, 2015, 41(2): 6—10.
- Zhang Wenli, Chen Shuren, Zhu Dehong. Research review on field weeding robot[J]. Agricultural Equipment & Technology, 2015, 41(2): 6—10. (in Chinese with English abstract)
- [7] 孙俊,芦兵,毛罕平. 基于双目识别技术的复杂背景中果实识别试验[J]. 江苏大学学报: 自然科学版, 2011, 32(4): 423—427.
- Sun Jun, Lu Bing, Mao Hanping. Fruits recognition in complex background using binocular stereovision[J]. Journal of Jiangsu University: Natural Science Edition, 2011, 32(4): 423—427. (in Chinese with English abstract)
- [8] 司永胜,乔军,刘刚,等. 基于机器视觉的苹果识别和形状特征提取[J]. 农业机械学报, 2009, 40(8): 161—165.
- Si Yongsheng, Qiao Jun, Liu Gang, et al. Recognition and shape features extraction of apples based on machine vision[J]. Transactions of The Chinese Society For

- Agricultural Machinery, 2009, 40(8): 161—165. (in Chinese with English abstract)
- [9] 何东健, 乔永亮, 李攀, 等. 基于 SVM-DS 多特征融合的杂草识别[J]. 农业机械学报, 2013, 44(2): 182—187.
He Dongjian, Qiao Yongliang, Li Pan, et al. Weed recognition based on SVM-DS multi-feature fusion[J]. Transactions of The Chinese Society For Agricultural Machinery, 2013, 44(2): 182—187. (in Chinese with English abstract)
- [10] 赵川源, 何东健, 乔永亮. 基于多光谱图像和数据挖掘的多特征杂草识别方法[J]. 农业工程学报, 2013, 29(2): 192—198.
Zhao Chuanyuan, He Dongjian, Qiao Yongliang. Identification method of multi-feature weed based on multi-spectral images and data mining[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(2): 192—198. (in Chinese with English abstract)
- [11] Pulido C, Solaque L, Velasco N. Weed recognition by SVM texture feature classification in outdoor vegetable crop images[J]. Ingeniería e Investigación, 2017, 37(1): 68—74.
- [12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems, 2012: 1097—1105.
- [13] 孙俊, 谭文军, 毛罕平, 等. 基于改进卷积神经网络的多种植物叶片病害识别[J]. 农业工程学报, 2017, 33(19): 209—215.
Sun Jun, Tan Wenjun, Mao Hanping, et al. Recognition of multiple plant leaf diseases based on improved convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(19): 209—215. (in Chinese with English abstract)
- [14] Lecun Y, Bengio Y. Convolutional networks for images, speech, and time series[M]//The Handbook of Brain Theory and Neural Networks. MIT Press, 1998.
- [15] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//IEEE Computer Society. IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [17] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Computer Vision and Pattern Recognition. IEEE, 2015.
- [18] 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.
- [19] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013.
- [20] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//International Conference on Neural Information Processing Systems. MIT Press, 2015.
- [21] 王璨, 武新慧, 李志伟. 基于卷积神经网络提取多尺度分层特征识别玉米杂草[J]. 农业工程学报, 2018, 34(5): 144—151.
Wang Can, Wu Xinhui, Li Zhiwei. Recognition of maize and weed based on multi-scale hierarchical features extracted by convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(5): 144—151. (in Chinese with English abstract)
- [22] 孙俊, 何小飞, 谭文军, 等. 空洞卷积结合全局池化的卷积神经网络识别作物幼苗与杂草[J]. 农业工程学报, 2018, 34(11): 159—165.
Sun Jun, He Xiaofei, Tan Wenjun, et al. Recognition of crop seedling and weed recognition based on dilated convolution and global pooling in CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(11): 159—165. (in Chinese with English abstract)
- [23] Tang J L, Wang D, Zhang Z G, et al. Weed identification based on K-means feature learning combined with convolutional neural network[J]. Computers and electronics in agriculture, 2017, 135: 63—70.
- [24] Milioto A, Lottes P, Stachniss C. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.
- [25] Chebrolu N, Lottes P, Schaefer A, et al. Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields[J]. The International Journal of Robotics Research, 2017, 36(10): 1045—1052.
- [26] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv: 1704.04861, 2017.
- [27] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. arXiv preprint arXiv: 1511.00561, 2015.
- [28] Paszke A, Chaurasia A, Kim S, et al. Enet: A deep neural network architecture for real-time semantic segmentation[J].

arXiv preprint arXiv: 1606. 02147, 2016.

using CNNs[J]. arXiv preprint arXiv: 1802. 08960, 2018.

[29] Milioto A, Stachniss C. Bonnet: An open-source training and deployment framework for semantic segmentation in robotics

[30] Redmon J, Farhadi A. Yolo3: An incremental improvement[J]. arXiv preprint arXiv: 1804. 02767, 2018.

Real-time recognition of sugar beet and weeds in complex backgrounds using multi-channel depth-wise separable convolution model

Sun Jun, Tan Wenjun, Wu Xiaohong, Shen Jifeng, Lu Bing, Dai Chunxia

(School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: Mechanical weeding can reduce the use of pesticides and is of great significance to ensure high yield of crops. Real-time and accurate identification of crops is a key technical problem needs to be solved in mechanical weeding equipment. Because of the subjectivity of feature extraction process in weed recognition, the accuracy of traditional methods in actual field environment is low. In recent years, the method of weed identification based on convolution neural network has been widely studied. Although the accuracy is obviously improved, there are still problems such as large parameters and poor real-time performance. In order to solve the above problems, a four-channel input image is constructed by collecting near infrared and visible images of sugar beet in the field, and a lightweight convolution neural network based on codec structure is proposed. In this paper, Sugarbeet and weed images collected from a farm in Bonn, Germany, in 2016 were used as data sets, which covered images of different growth stages of sugar beet, and 226 pictures of which were randomly selected as training sets, and the remaining 57 pictures were used as test sets. The experimental data set was composed of three channels of visible light image and one channel of near infrared image, which are merged into a four-channel image by pixel level superposition, and the depth-wise separable convolution was used in the deep model. Firstly, the input feature image was convoluted in 2 dimensions convolution kernel and the number of channels was expanded. Then, the 1×1 convolution kernel was used to make the 3 dimensions convolution which combined channel feature and compressed the channels to enhance the nonlinear mapping ability of the model. In order to avoid the problem of the gradient disappearing, the residual block was used to connect the input and output of the depth-wise separate convolution. Finally, the coding and decoder structure was designed and the shallow features were combined with deep features to refine the segmentation effect. Due to the imbalance of pixel proportions of soil, crops and weeds, the weighted loss function was used to optimize the model. The segmentation accuracy, parameters and operating efficiency of models at different input resolutions and different width factor were introduced to evaluate the model. When the width factor was 1, the segmentation accuracy of the model increased with the increase of the input image resolution, the model accuracy of four channel input was higher than that of the model based on original visible image input, which showed that the near-infrared image features can compensate the defects of ordinary RGB images to some extent, and make the model more suitable for the dark environment. Under the same input image resolution, the model with a width factor of 2 or 4 performs better than the model with a width factor of 1. With the increases of width factor, the parameters of the model increase greatly. The amount of calculation is related to the size of the input image, so the frame rate gradually decreases with the increase of the size of input image. The experimental results show that the optimal model in this paper is a four channel input model with a width coefficient of 2, and the average intersection union ration is 87.58%, the average pixel accuracy is 99.19%, the parameters are 525 763 and the frame rate is 42.064 frames/s. The model has high segmentation and recognition accuracy and good real-time performance, and can provide theoretical basis for the development of intelligent mechanization weeding equipment.

Keywords: image segmentation; crops; convolutional neural network; deep learning; sugar beet; weed; real-time