

深度多分支模型融合网络的胡萝卜缺陷识别与分割

谢为俊, 魏 硕, 郑招辉, 杨光照, 丁 鑫, 杨德勇^{*}

(中国农业大学工学院, 北京 100083)

摘 要: 缺陷检测是胡萝卜上市销售前的重要环节, 开裂缺陷区域的分割提取是开裂胡萝卜修整的必要条件。基于图像处理的传统的胡萝卜表面缺陷识别算法复杂, 通用性、鲁棒性较差。该研究提出一种集胡萝卜缺陷种类识别 (C-Net) 和开裂缺陷分割 (S-Net) 为一体的深度多分支模型融合网络 (CS-Net)。C-Net 将预训练的 ResNet-50 作为胡萝卜图像特征提取器, 分别输出 ResNet-50 不同卷积层特征, 将其作为支持向量机 (Support Vector Machine, SVM) 的输入训练不同的分类模型, 并利用不同策略将其融合以获取最终的分类模型; S-Net 将预训练的 ResNet-50 作为分割网络的编码器, 根据不同的分割网络构造思想设计解码器, 构造胡萝卜开裂区域分割提取网络。结果表明, C-Net 中, ResNet-50 第 49 层输出模型在测试集上的准确率为 94.71%, 利用 Stacking 融合方法得到的模型性能最好, 在测试集上的准确率为 98.40%; S-Net 中, 根据 U-net 构造思想构造的网络分割效果最好, 分割像素准确率 (Pixel Accuracy, PA)、类别平均像素准确率 (Mean Pixel Accuracy, MPA) 和平均交互比 (Mean Intersection over Union, MIOU) 分别为 98.31%、96.05% 和 92.81%。该研究构建的胡萝卜缺陷识别分割网络对胡萝卜外观品质的量化评价和表面缺陷的修整具有重要意义。

关键词: 图像分割; 深度学习; 分类; 胡萝卜; 外观品质; 模型融合

doi: 10.11975/j.issn.1002-6819.2021.2.021

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2021)-2-0177-10

谢为俊, 魏硕, 郑招辉, 等. 深度多分支模型融合网络的胡萝卜缺陷识别与分割[J]. 农业工程学报, 2021, 37(2): 177-186. doi: 10.11975/j.issn.1002-6819.2021.2.021 http://www.tcsae.org

Xie Weijun, Wei Shuo, Zheng Zhaohui, et al. Classification and segmentation of defect carrots using deep multi-branch models fusion network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(2): 177-186. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.2.021 http://www.tcsae.org

0 引 言

胡萝卜 (*Daucus carota* L.) 是世界上主要的块根块茎类蔬菜, 富含胡萝卜素、多酚及多种微量元素。中国胡萝卜产量最大且出口最多, 2018 年产量达 2 000 万 t, 接近世界总产量的一半^[1]。及时剔除缺陷胡萝卜有助于提高胡萝卜的市场竞争力, 因此对胡萝卜的外观品质评价十分必要。传统的胡萝卜缺陷识别分拣具有劳动力消耗大、成本高、主观性强等缺点, 而机器视觉识别技术具有快速无损、客观标准等优点, 越来越多地应用于农业生产中, 提高了农业生产现代化水平与生产效率^[2]。

20 世纪末, 有学者开始利用机器视觉技术对胡萝卜外观进行评价。但限于当时的技术条件, 机器视觉仅能识别胡萝卜的肩部与根部位置^[3]、表面颜色缺陷^[4]、部分形状缺陷^[5]、根尖形状特征^[6]等, 且识别准确率低, 鲁棒性差。随着机器视觉技术的飞速发展, 研究人员开始利用先进的软硬件条件对胡萝卜分级进行深入研究。Hahn 和 Sanchez^[7]开发了一种简单的算法, 仅使用相距 90° 的两幅图像, 就可以在不旋转相机的情况下准确预测胡萝卜

的体积。许多学者设计了特定的图像处理算法对胡萝卜不同外观缺陷进行检测。韩仲志等^[8]和 Deng 等^[9]针对胡萝卜青头、须根、开裂、弯曲等不同缺陷分别设计了图像处理算法, 平均检测准确率 90% 以上; Xie 等^[10]和谢为俊等^[11]根据胡萝卜青头、弯曲、断裂、分叉和开裂等不同缺陷特点设计不同的图像处理算法, 总体识别率达 90% 以上。上述图像处理算法可实现胡萝卜缺陷的检测, 但存在准确率较低、算法复杂、鲁棒性差等缺点。

机器学习和深度学习技术在图像分类领域获得巨大成功, 在农业生产领域也得到了大量应用^[12], 越来越多学者将其引入胡萝卜分级领域。Xie 等^[13]提取正常胡萝卜的 12 个特征作为反向传播神经网络 (Back Propagation Neural Network, BPNN)、SVM、极限学习机 (Extreme Learning Machine, ELM) 的输入, 将正常胡萝卜分为 4 个等级, 分级准确率达 96.67%, 但该方法未涉及缺陷胡萝卜的分级研究。Zhu 等^[14]和倪建功等^[15]利用深度学习对胡萝卜的黑斑、须根、弯曲缺陷进行检测, 平均识别准确率达 95% 以上, 但对胡萝卜外观影响更大的机械损伤、断裂、分叉、畸形和开裂等缺陷识别研究还未涉及。

缺陷胡萝卜通常用作动物饲料或直接丢弃, 降低了胡萝卜种植效益。而分叉、开裂等缺陷胡萝卜可作为深加工原料, 制作胡萝卜汁、胡萝卜粉等。但开裂胡萝卜开裂处会藏有泥土, 清洗不彻底会影响胡萝卜制品质量, 甚至导致食品安全问题, 因此在深加工前对胡萝卜开裂区域进行修整十分必要。胡萝卜开裂区域的精准分割是

收稿日期: 2020-11-20 修订日期: 2020-12-28

基金项目: 国家重点研发计划项目 (2018YFD0700102-02)

作者简介: 谢为俊, 博士生, 研究方向为农产品智能加工技术。

Email: BS20193070629@cau.edu.cn

*通信作者: 杨德勇, 博士, 副教授, 研究方向为农产品物理特性。

Email: ydy@cau.edu.cn

实现开裂胡萝卜自动化修整的必要条件。韩仲志等^[8]、Deng 等^[9]和谢为俊等^[11]利用胡萝卜开裂区域与正常区域的颜色、纹理差异提取开裂区域,但提取精准度较差。因此对胡萝卜开裂区域的精准提取研究十分必要。

为更加全面地评价胡萝卜外观品质和开裂胡萝卜的自动化修整,本研究提出一种集缺陷识别与分割为一体的深度多分支模型融合网络(CS-Net)。该网络由胡萝卜缺陷识别分类网络 C-Net 和开裂区域分割网络 S-Net 组成, C-Net 将 ImageNet 数据集上预训练的 ResNet-50 作为胡萝卜图像特征提取器,将提取的特征作为 SVM 的输入,得到不同的分类模型,再利用模型融合思想将不同分类模型融合得到最终的分类型模型; S-Net 将预训练的 ResNet-50 作为编码器,根据不同的分割网络构造思想设计解码器,构造胡萝卜开裂区域分割提取网络。该网络为胡萝卜外观品质评价和开裂缺陷修整奠定了基础。

1 材料与方法

1.1 试验材料

试验用胡萝卜购自绿龙食品加工有限公司(山东省潍坊市)。试验共包括 7 203 幅胡萝卜图像,包括正常胡萝卜 2 109 幅、断裂 1 424 幅、开裂 998 幅、畸形 665 幅、机械损伤 1 029 幅和分叉 978 幅,胡萝卜形态种类如图 1 所示。对于分类网络 C-Net,每种形态胡萝卜随机选取 80% 作为训练集,20% 作为测试集。训练集共包含 5 765 幅图像,其中断裂、机械损伤、开裂、分叉、畸形和正常胡萝卜图像分别为 1 140、824、799、783、532 和 1 687 幅。因不同形态胡萝卜的图片数量不平衡,本研究在训练时采用分层 5 折交叉验证,保证每次划分中包含类别比例与原数据相同。测试集共包含 1 438 幅胡萝卜图像,其中断裂、机械损伤、开裂、分叉、畸形和正常胡萝卜图像分别为 284、205、199、195、133 和 422 幅。对于分割网络 S-Net,随机选取 543 幅开裂胡萝卜图像,按照 6:2:2 划分训练集、验证集和测试集。

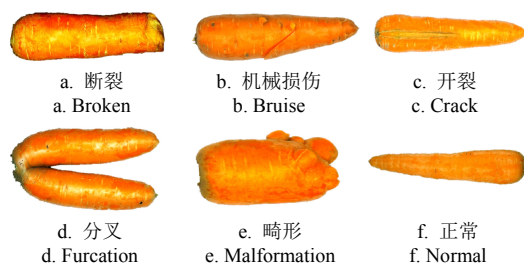


图 1 胡萝卜形态种类
Fig.1 Morphological types of carrots

1.2 图像获取及预处理

胡萝卜图像由彩色 CCD 相机(DFK-33UX265)采用 8 mm 镜头(The Imaging Source, Germany)获得。胡萝卜图像采集系统^[11]如图 2 所示,该系统在光照箱中配置了两块互成 120° 的平面镜,胡萝卜放在两块镜子中间,上方的相机可同时捕捉到 3 幅胡萝卜图像,以尽可能多地获得胡萝卜表面图像,图像尺寸为 1 920×1 080 像素。图像处理算法和卷积神经网络(CNN)搭建基于 Python 3.7、计算机视觉库 OpenCv4.0.0 和深度学习库 Keras 2.3.1、Tensorflow 2.0.0。电

脑配置为 Intel corn Xeon 3.70 GHz CPU, 128 GB 运行内存, NVIDIA Quadro P2000 (5 G) GPU, Windows 10.1 操作系统。

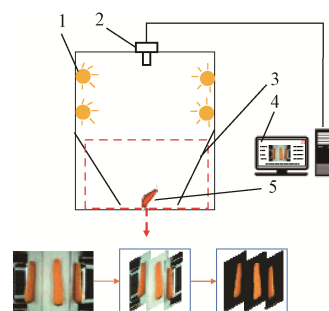
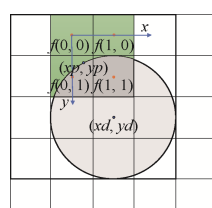


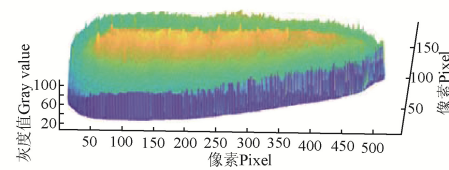
图 2 图像采集系统
1.光源 2.CCD 相机 3.平面镜 4.计算机 5.胡萝卜
1.Light source 2.CCD camera 3.Mirror 4.Computer 5.Carrot

图 2 图像采集系统
Fig.2 Image acquisition system

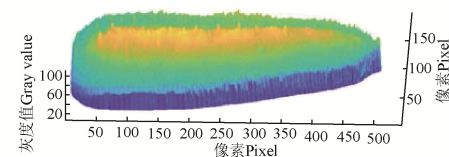
图像预处理中,利用灰度直方图双峰法提取胡萝卜区域图像并删除背景^[11]。为使胡萝卜图像不变形,将图像用 0 填充为尺寸 1 024×1 024×3 的图像,再利用线性插值算法将图像统一调整为适合 CNN 输入的尺寸(224×224×3)。为去除图像边缘高频噪声,提出一种基于圆形区域的均值滤波器,如图 2a 所示。



a. 圆形均值滤波器
a. Circular average filter



b. 滤波前
b. Image before filtering



c. 滤波后
c. Image after filtering

图 3 滤波前后胡萝卜效果图

Fig.3 Carrot image before and after filtering

对于给定中心点 (x_d, y_d) , 其邻域像素位置 (x_p, y_p) 由式(1)计算得到。

$$\begin{aligned} x_p &= x_d + R \cos\left(\frac{2\pi p}{P}\right) \\ y_p &= y_d - R \sin\left(\frac{2\pi p}{P}\right) \end{aligned} \quad (1)$$

式中 R 为圆形滤波器半径, P 为采样点总数, p 为采样点序号。通过此操作得到的采样点坐标可能均不是整数,因此通过插值获得其位置的像素值,本研究采用双线性插值法,数学表达式如式(2)

$$f(x_p, y_p) = \begin{bmatrix} 1-x_p & x_p \end{bmatrix} \begin{bmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{bmatrix} \begin{bmatrix} 1-y_p \\ y_p \end{bmatrix} \quad (2)$$

式中 $f(x_p, y_p)$ 为点 (x_p, y_p) 的像素值, $f(ip, jp)$ ($ip, jp=0,1$) 为点 (x_p, y_p) 相邻四点像素值。则中心点 (x_d, y_d) 像素

值由式 (3) 计算

$$f(x_d, y_d) = \frac{1}{P} \sum_{p=0}^P f(x_p, y_p) \quad (3)$$

图像三维空间显示是将图像灰度值作为图像 Z 轴进行三维显示。

由图 3b、3c 使用圆形滤波器滤波前后胡萝卜图像的三维空间显示图可知, 圆形滤波器对图像的高频噪声有很好的抑制效果。图像经预处理后, 既去除了背景又将胡萝卜图像变为适合 CNN 输入的正方形图像, 同时保证了胡萝卜无变形, 保留了原始图像的拓扑结构, 避免识别结果受胡萝卜变形的影响。

2 深度多分支模型融合网络

2.1 网络总体结构

本研究提出了一种集胡萝卜缺陷识别分类与开裂缺陷区域分割提取为一体的深度多分支模型融合网络 (CS-Net), 该网络由胡萝卜缺陷识别网络 (C-Net) 和胡萝卜开裂区域分割提取网络 (S-Net) 组成, 总体结构如图 4 所示。ResNet-50 包含的残差模块能够解决深度网络训练退化问题^[16], 大量研究表明残差网络是十分优秀的图像特征提取器^[17-18], 因此本网络以预训练的 ResNet-50 为基础, C-Net 将 ResNet-50 作为图像特征提取器, 输出不同层的卷积特征, 将其作为 SVM^[19] 的输入, 训练得到不同的分类模型, 再通过不同的模型融合策略 (硬投票 (Hard voting)、软投票 (Soft voting) 和 Stacking 法) 将各个分类器融合为最终的胡萝卜缺陷识别网络。S-Net 将 ResNet-50 作为分割网络的编码器, 使用三种经典分割网络 (U-net^[20]、FCN^[21] 和 SegNet^[22-23]) 的构造思想设计分割网络的解码器, 构造胡萝卜开裂区域分割提取网络。

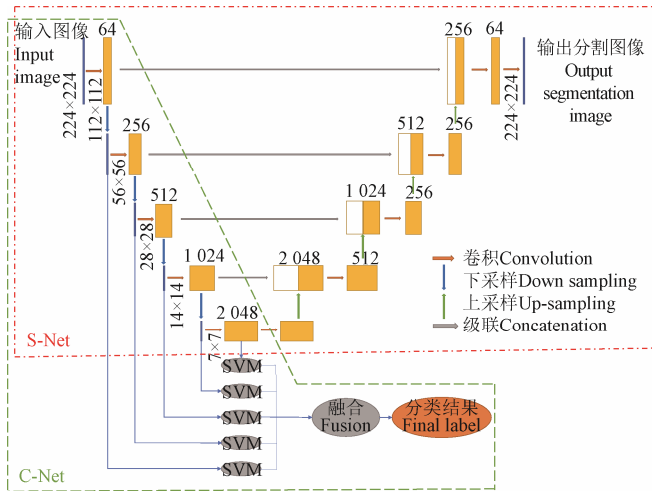


图 4 网络总体结构

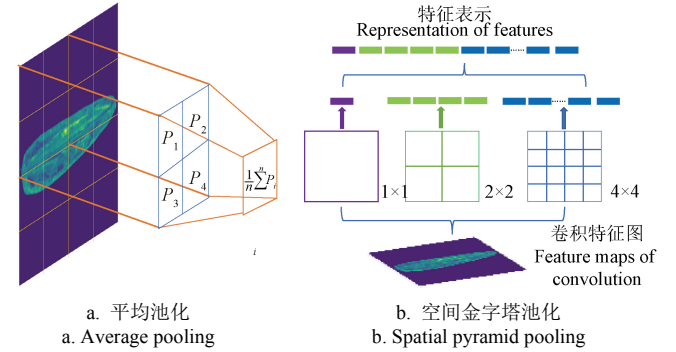
Fig.4 Overall structure of the network

2.2 缺陷识别网络 (C-Net)

2.2.1 池化方法

池化层是 CNN 中重要组成部分, 它能够实现图像特征下采样、去除图像冗余信息、简化网络复杂度从而减少模型的参数量。常见池化方法包括一般池化、重叠池

化和空间金字塔池化 (Spatial Pyramid Pooling, SPP)。在 C-Net 中采用的池化方法包括平均池化 (Average Pooling, AVP)、全局平均池化 (Global Average Pooling, GAP) 和 SPP。AVP 如图 5a 所示, 通过计算池化区域的平均值作为该区域池化后的值; GAP 是平均池化的一种特殊形式, 此时池化窗口尺寸与特征图尺寸相同; SPP 将一个池化操作分成若干个不同尺寸的池化操作, 用不同大小的池化核作用于上层的卷积特征, 如图 5b 所示。SPP 能使任意大小的特征图转换成固定大小的特征向量, 实现特征图的多尺度提取。图像卷积特征通过池化操作再经过全连接层后输出 4 096 维数据, 作为 SVM 的输入。



注: P_i 为该点像素值。

Note: P_i is the pixel value of point i .

图 5 池化方法

Fig.5 Methods of Pooling

2.2.2 数据降维方法

通过 CNN 提取的特征维度较高, 维度越高的数据在每个特征维度上的分布就越稀疏, 这不利于机器学习模型的训练, 因此对高维数据进行降维十分必要。本研究采用主成分分析 (Principal Component Analysis, PCA) 和 Relief 降维方法^[24]。PCA 通过某种线性映射将高维向量转为低维空间表示, 并使得在所投影的低维空间上数据方差尽可能大。PCA 具有使用较少的数据维度, 保留尽可能多的原数据信息。Relief 是一种基于特征权重的特征选择算法, 其设计了一个“相关统计量”来度量特征的重要性, 该统计量为向量, 它的每个分量是对其中一个初始特征的评价值。特征子集的重要性就是子集中每个特征所对应的相关统计量之和, 因此这个“相关统计量”也可以视为是每个特征的“权值”。指定阈值, 选择大于阈值的相关统计量分量所对应的特征。

本研究中, PCA 降维后数据维数分别为 256、500、1 000、2 000; Relief 方法中 K 近邻参数 k 的取值为 150, 阈值选择为 0.05。

2.2.3 SVM 训练

将 ResNet-50 输出的特征向量作为 SVM 的输入, 使用非线性映射算法, 通过核函数在高维空间中找到最优超平面 H , 使样本变得线性可分, 实现胡萝卜不同缺陷种类的正确识别。超平面 H 可用 $\omega^T x + b$ 表示, 样本空间中任一点 x 到超平面的距离 γ 可由式 (4) 计算

$$\gamma = \frac{|\omega^T x + b|}{\|\omega\|} \quad (4)$$

式中 ω 是超平面 H 的法向量, b 为超平面 H 的截距。

本研究中, 选择径向基函数 (Radial Basis Function, RBF) 作为 SVM 的核函数, 惩罚系数为 0.5, 停止训练时的误差值为 10^{-3} , 多分类方式采用一对一形式 (ovo)。

2.2.4 模型融合策略

模型集成学习策略已经在农业领域得到广泛的应用^[25-26]。传统 CNN 使用单一的分类模型^[27], 但不能适应胡萝卜的所有特征。集成学习思想则是利用训练数据首先得到一组基模型, 再利用模型融合策略将他们组合起来。集成学习综合各个基模型的优点, 往往可获得比单一模型更显著的性能^[26]。本研究由 ResNet-50 的第 1、10、22、40、49 层输出特征作为输入训练得到 5 个支持向量机 (SVM) 模型作为基模型。在此基础上, 利用硬投票、软投票和 Stacking 集成方法, 得到最终胡萝卜缺陷分类模型。本研究所采用的三种模型融合策略执行步骤如下:

1) 硬投票, 又叫相对多数投票法。首先获得 5 个基模型在测试集上的预测类别 C_i ($i=0, 1, 2, 3, 4$), 然后统计 5 个基模型输出结果中每个类别的票数, 最终输出结果为得票最多的类别。当票数相同时, 则随机选择一个类别作为最终输出结果。

2) 软投票, 又叫加权平均法。首先获得 5 个基模型在测试集上的每个类别的预测概率 pr_i ($i=0, 1, 2, 3, 4$), 再利用式 (5) 计算出每个类别融合后的预测概率。

$$PR = \sum_{i=0}^{N-1} \left(\frac{k_i}{\sum_{j=1}^N k_j} pr_i \right) \quad (5)$$

式中 N 为分类器个数, k_i 根据各分类器在测试集上的准

确率排序确定, 准确率最高的 $k_i=5$, 最低的 $k_i=1$ 。

3) Stacking 集成方法由两级分类器构成, 其中低级别的模型称为基模型, 高级别的称为元模型。本研究中低级别模型所用分类器为 SVM, 高级别模型所用分类器为梯度提升决策树 (Gradient Boosting Decision Tree, GBDT)^[25]。本研究中 GBDT 算法的学习率为 0.01, 最大迭代次数为 150, 树的深度为 10。具体训练步骤如下 (图 6): ①将胡萝卜图片数据集根据 4:1 划分为训练集与测试集, 数据集在经过 ResNe-50 网络特征提取, 分别输出第 1、10、22、40 和 49 层的训练集 D_{Train}^{it} ($it=1, 10, 22, 40, 49$) 和测试集数据 D_{Test}^{it} ($it=1, 10, 22, 40, 49$)。②将第 1 层输出的训练集数据 D_{Train}^1 利用 5 折交叉验证的方法随机均等划分为 5 个子集 D_1, D_2, D_3, D_4, D_5 , 依次选取其中一个子集 D_{iv} ($iv=1, 2, \dots, 5$) 作为验证集, 其他 4 个作为训练子集 $D_{-iv}=D_{Train}^1-D_{iv}$ 。③将 D_{-iv} 作为 SVM 的训练集, D_{iv} 作为验证集并输出验证集的预测结果 a_{iv} , 同时对原始测试集 D_{Test}^1 进行预测, 输出预测结果 b_{iv} 。④对 (3) 循环 5 次得到 $\{a_1, a_2, a_3, a_4, a_5\}$, 将这 5 次结果按列合并得到和 D_{Train}^1 标签 Y 相同长度的向量 A_1 , 对测试集预测值 $\{b_1, b_2, b_3, b_4, b_5\}$ 取平均值得到和 D_{Test}^1 标签 Y 相同长度的向量 B_1 。⑤对另外四个数据集 (第 10、22、40、49 层输出特征) 执行上面步骤得到由原始训练集 D_{Train}^{it} ($it=10, 22, 40, 49$) 产生的 $A_{10}, A_{22}, A_{40}, A_{49}$ 和原始测试集 D_{Test}^{it} ($it=10, 22, 40, 49$) 产生的 $B_{10}, B_{22}, B_{40}, B_{49}$ 。⑥将 $A_1, A_{10}, A_{22}, A_{40}, A_{49}$ 和原始训练集 D_{Train}^{it} 的标签 Y 合并得到新样本数据集 $\{(A_1, A_{10}, A_{22}, A_{40}, A_{49}), Y\}$ 作为二级分类器 GBDT 的输入特征, $\{(B_1, B_{10}, B_{22}, B_{40}, B_{49})\}$ 作为 GBDT 测试集生成最终结果。

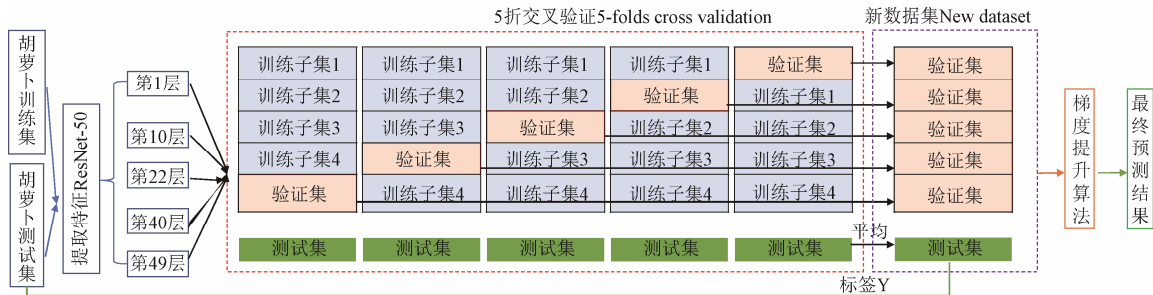


图 6 Stacking 集成学习训练过程
Fig.6 Training process of Stacking-ensembling

2.3 分割网络 (S-Net)

为了对胡萝卜开裂缺陷区域进行自动化修整, 本研究将 ResNet-50 作为编码器, 设计分割网络的解码器实现对胡萝卜开裂区域的精准提取。

2.3.1 构造方法

全卷积神经网络^[21] (Fully Convolutional Network, FCN) 是最早的分割网络, 它将 CNN 全连接结构用卷积结构代替, 将原本输出的一维向量改为二维特征图的形式, 保存了图像的二维信息。利用反卷积操作对特征图进行上采样, 并将编码器相应尺度的特征图复制过来与上采样后的特征图进行对应像素相加 (Add), 再通过卷积操作组合特征后输入下一个反卷积层, 如图 7a

所示。U-net 分割网络^[20]为了使网络在上采样的过程中获得更多信息, 恢复图像更多细节, 将编码器对应尺度特征图与反卷积后相同尺寸特征图进行级联 (Concat), 再通过卷积操作后作为下一个反卷积层的输入, 如图 7b 所示。SegNet^[22]为了解决网络编码器在池化操作时丢失位置信息的问题, 在编码器执行最大池化操作时, 将特征最大值的位置作为最大池化索引保存。网络解码器在上采样的时候根据最大池化索引直接恢复图像特征值, 如图 7c 所示。本研究将 ResNet-50 作为分割网络的编码器, 分别根据 FCN、U-net 和 Segnet 网络构造思想构造网络的解码器并将 ResNet-50 第 1、10、22、40、49 层输出特征尺寸作为解码器的反卷积后特征尺寸。

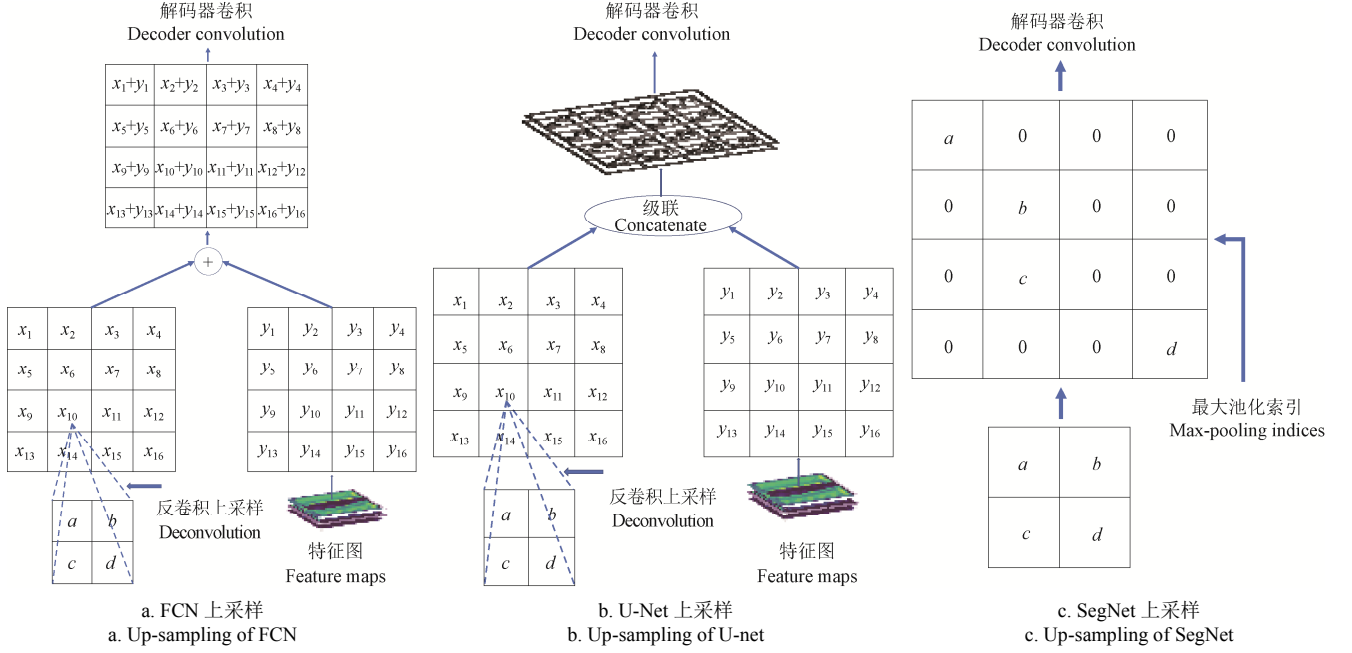


图 7 不同分割网络上采样方法
Fig.7 Up-sampling of different segmentation models

2. 3. 2 损失函数

损失函数的选择会直接影响分割网络的分割效果，交叉熵损失函数（Cross Entropy loss）^[28]、骰子系数差异函数（Dice loss）^[29]、焦点损失函数（Focal loss）^[30]是分割网络常用的损失函数。交叉熵损失函数主要是用来判定实际输出与标签的差异，网络即可利用这种差异经过反向传播去更新网络参数。该损失函数在分割网络中分别评估每个像素的类预测，然后对所有像素求平均，这对于类别数据不平衡时效果较差。骰子系数差异函数是一种集合相似度度量函数，用于计算两个样本之间的相似度，适用于样本不均的情况，但是会对反向传播造成不利的影响，造成训练波动大。Lin 等^[30]提出焦点损失函数用于解决上述两种激活函数存在的问题，焦点损失函数是动态缩放的交叉熵损失函数，通过增加动态缩放因子，实现在训练过程中动态调整简单样本的权重，让模型快速关注于困难样本的分类，由式（6）计算得到。由于本研究中胡萝卜开裂区域在图像中占比较小，因此本研究选择焦点损失函数作为分割网络损失函数。

$$FL(p_i) = -(1 - p_i)^\beta \log(p_i), \quad p_i = \begin{cases} p_r & y = 1 \\ 1 - p_r & y \neq 1 \end{cases} \quad (6)$$

式中 p_r 是类别预测概率， y 是标签值， β 是动态缩放因子。

2. 3. 3 分割网络性能评价指标

为了衡量分割网络的性能，需要经过标准、公认的指标进行评估。目前常用指标包括分割网络模型尺寸、分割时间、准确率^[31]，其中准确率包括 PA、MPA 和 MIoU。PA 是最简单的度量标准，只需计算正确分类的像素数量与像素总数之间的比率；MPA 是改进的 PA，按类计算正确像素的比率并取平均；MIoU 是实际区域和预测区域的交集和并集之间的比率，作为分割网络的标准度量。

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

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

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

式中 k 为分割目标类别数， p_{ii} 表示分类正确的像素数， p_{ij} 表示 i 类被误判断为 j 类的像素数， p_{ji} 表示 j 类被误判断为 i 类的像素数。

3 结果与分析

3.1 卷积层特征、池化方法与数据降维方法对模型性能的影响

表 1 为 ResNet-50 不同层（1、10、22、40、49）输出特征在不同池化方法和不同数据降维方法下训练的 SVM 模型的性能。准确率为模型在分层 5 折交叉训练下验证准确率的平均值。

由 ResNet-50 不同层输出卷积特征训练的模型性能各不相同。由表 1 可知，同样的池化和降维方法，高层网络（40、49）输出特征训练的模型准确率比低层（1、10、22）的高。为解释该现象，将 ResNet-50 模型的第 1、10、22、40 和 49 层输出特征进行可视化，如图 8 所示。由图可知，ResNet-50 网络第 1、10、22 层输出特征更多的是图像的一般特征（边缘、纹理、色彩等），而第 40 和 49 层输出特征则是图像高层语义特征，其将图像特征抽象为更加容易区分的特征，因此高层网络（40、49）输出特征训练的模型准确率要比低层（1、10、22）的高。

表 1 卷积特征、池化方法与数据降维方法对模型性能的影响
Table 1 Effects of convolution feature maps, pooling method and dimensionality reduction methods on model performance

输出特征 Output features	池化方法 Pooling method	降维方法 Dimensionality reduction methods	特征数量 The number of features	准确率 Accuracy/%
第 1 层卷积层 1 st convolution layer	AVP	/	4 096	74.48
		PCA	500	81.39
		ReliefF	1 950	75.85
	SPP	/	4 096	84.07
		PCA	1 000	86.06
		ReliefF	1 390	83.87
	GAP	/	64	55.76
		PCA	32	59.32
		ReliefF	60	56.23
第 10 层卷积层 10 th convolution layer	AVP	/	4 096	81.26
		PCA	500	88.83
		ReliefF	2 250	83.28
	SPP	/	4 096	89.10
		PCA	2 000	89.17
		ReliefF	2 269	89.88
	GAP	/	256	59.60
		PCA	128	76.48
		ReliefF	216	58.35
第 22 层卷积层 22 nd convolution layer	AVP	/	4 096	87.84
		PCA	256	93.16
		ReliefF	2 388	89.13
	SPP	/	4 096	93.70
		PCA	2 000	93.03
		ReliefF	2 452	93.40
	GAP	/	512	79.22
		PCA	256	86.38
		ReliefF	481	80.15
第 40 层卷积层 40 th convolution layer	AVP	/	4 096	91.47
		PCA	256	95.11
		ReliefF	2 010	91.79
	SPP	/	4 096	94.53
		PCA	2 000	94.17
		ReliefF	2 232	94.21
	GAP	/	1 024	89.37
		PCA	500	92.34
		ReliefF	880	89.82
第 49 层卷积层 49 th convolution layer	AVP	/	4 096	93.55
		PCA	1 000	94.68
		ReliefF	3 185	93.70
	SPP	/	4 096	92.24
		PCA	2 000	86.60
		ReliefF	2 977	90.71
	GAP	/	2 048	93.55
		PCA	1 000	95.31
		ReliefF	1 566	93.72

注: AVP、SSP、GAP 分别为平均池化、空间金字塔池化、全局平均池化; PCA 和 ReliefF 分别为主成分分析和 ReliefF 降维方法。下同。

Note: AVP, SSP and GAP represent the average pooling, spatial pyramid pooling and global average pooling, respectively. PCA and ReliefF are the dimension reduction methods of principal component analysis and ReliefF. Same as below.

不同池化方法对模型性能影响不同。由表 1 可知, 在低层特征下不同池化层的模型性能从大到小排序为 SPP、AVP、GAP, 这可能是因为 SPP 有 3 个不同尺寸的池化核卷积特征进行提取, 因而可以在不同尺度上提

取卷积特征, 在某种程度上使用 SPP 相当于额外增加了 1 层卷积层, 所以在低层特征的基础上可以抽象出更高层次的特征, 因此采用 SPP 的模型性能最好, 在第 1 层卷积特征上准确率达到 84.07%, 比采用 AVP 的准确率高出近 10 个百分点; AVP 使用的池化核尺寸小于低层卷积特征尺寸, 使用窗口滑动法在卷积特征上依次滑动提取特征, 而 GAP 的池化核尺寸与低层卷积特征尺寸相同, 所以在相同的低层卷积特征上 AVP 提取的特征要远远多于 GAP, 这就导致了使用 AVP 的模型性能要高于采用 GAP 的模型。在高层语义特征下不同池化层对模型性能影响较小, 因为高层卷积特征图尺寸较小, 包含的信息数据量少, 用不同的池化方法提取得到的特征差别小。在第 49 层卷积特征下, 采用 SPP 的模型性能略差于采用其他两种池化方法的, 因为第 49 层输出的卷积特征尺寸只有 7×7 , 使用 SPP 会额外提取到冗余的信息, 从而降低模型性能。此时的 AVP 与 GAP 的池化核尺寸相同, 因此采用 GAP 与 AVP 的模型性能表现一致。

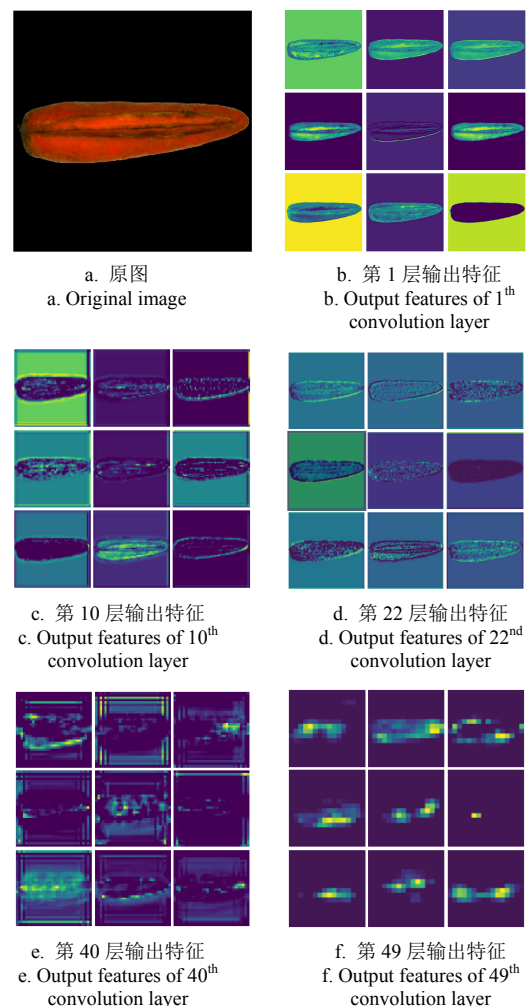


图 8 ResNet-50 不同卷积层输出特征图

Fig.8 Output features of different convolution layers of ResNet-50

在机器学习中, 对数据进行降维处理是必要的。对数据进行降维处理不仅能减少特征数量还会提高模型的性能。由表 1 可知, PCA 在卷积特征上对模型性能影响比 ReliefF 法更明显, 可能是因为该方法能更有效地剔除

经过池化层提取后特征包含的冗余信息。从表 1 中也可以看出, 一些模型在降维后性能下降, 原因在于降维过程中不可避免地去除了一些对分类结果有积极意义的数据, 从而影响了模型性能。

3.2 不同分类模型的性能比较

为进一步提升模型的分类精度, 本研究采用模型融合的思想对模型进行集成学习。根据表 1, 在保证融合后模型包含图像多尺度特征, 本研究最终选择的基模型 L1、L2、L3、L4、L5 分别为 ResNet-50 第 1、10、22、40、49 层输出卷积特征训练的 SVM 模型, 池化方法、降维方法如表 2 所示。采用硬投票、软投票、Stacking 方法分别得到 3 个不同的元模型, 各模型在测试集上的准确率排序为硬投票<软投票<Stacking 方法。硬投票以简单的少数服从多数原则, 对模型的性能提升不明显。软投票根据各个分类器在验证集上的准确率排序给予不同的权重, 使得表现更好的模型在最终的预测输出中占有更大的比重, 因此软投票方法对模型提升作用大于硬投票方法。Stacking 方法则是利用各模型在验证集和测试集上的预测结果作为新的训练集和测试集训练二级分类器, 它可以任意函数形式逼近真实值, 因此 Stacking 法所获得的模型在测试集上的精度最高, 达 98.40%, 相比单一分类器性能提升了 3 个百分点以上。传统的 HOG+SVM 和 LBP+SVM 模型性能仅相当于第 1 层卷积层提取特征模型 L1 的性能, 说明 CNN 自动提取特征优于人工手动提取特征。

表 2 模型性能比较

Table 2 Performances comparison of different models

项目 Item	池化方法 Pooling method	降维方法 Dimensionality reduction methods	特征数量 The number of features	验证集 准确率 Validation accuracy/%	测试集 准确率 Test accuracy/%
L1	SPP	PCA	1 000	86.06	84.35
L2	SPP	ReliefF	2 263	89.88	86.93
L3	SPP	ReliefF	2 452	93.40	93.39
L4	AVP	PCA	256	95.11	93.60
L5	GAP	PCA	1 000	95.31	94.71
硬投票 Hard voting					95.48
软投票 Soft voting					96.45
Stacking 方法					98.40
LBP			2 560	86.30	84.14
HOG			6 084	80.83	79.90

注: L1、L2、L3、L4、L5 分别为 ResNet-50 第 1、10、22、40、49 层输出卷积特征训练的 SVM 模型; LBP、HOG 分别为局部二值模式、方向梯度直方图提取特征训练的 SVM 模型。

Note: L1, L2, L3, L4 and L5 are the output convolution feature training SVM models with 1st, 10th, 22nd, 40th and 49th layer of ResNet-50, respectively. LBP and HOG represent SVM models with features extracted by local binary pattern and histogram of oriented gradients, respectively.

计算由 Stacking 方法获得的模型在测试集上的混淆矩阵, 由图 9 可知, 模型共正确识别出断裂、机械损伤、开裂、分叉、畸形和正常胡萝卜分别为 278、203、195、194、127 和 418 幅, 识别率分别为 97.89%、99.02%、97.99%、99.49%、95.49%和 99.05%, 总体识别率为

98.40%, 表明融合后的模型可以正确区分各缺陷的类间差异与类内差异, 但误识别情况依然存在, 因为所有的图片并不都是完美的, 存在一些不可避免的瑕疵。

分类混淆矩阵

Classification confusion matrix

真实值 True label	断裂 Broken	278	1				5
	机械损伤 Bruise	1	203	1			
	开裂 Crack			195	2		2
	分叉 Furcate	1			194		
	畸形 Malformation					127	6
	正常 Normal		3			1	418
	断裂 Broken	机械损伤 Bruise	开裂 Crack	分叉 Furcate	畸形 Malformation	正常 Normal	
	预测值 Predicted label						

图 9 Stacking 模型在测试集上的混淆矩阵

Fig.9 Confusion matrix of Stacking model on test dataset

3.3 不同方法构造的分割网络性能比较

本研究以 ResNet-50 为编码器, 分别以 FCN、U-net、Segnet 思想构造出 3 种不同的解码器, 得到 3 种分割模型 Res-FCN、Res-U-net、Res-Segnet。由表 3 可知, Res-U-net 综合表现最好, 在测试集上的 PA、MPA、MIoU 分别为 98.31%、96.05%、92.81%, 且模型最小, 为 0.24G, 单幅图像分割时间为 0.72 s。Res-FCN 的表现略低于 Res-U-net, 但其模型大小达到 1.5 G, 不适合配置较低的设备使用。Res-Segnet 的表现最差, MIoU 仅为 88.68%, 但其模型尺寸相比 Res-FCN 减少了 78%, 原因在于 Res-Segnet 复制了编码器的最大池化索引, 而 Res-FCN 则是复制了编码器特征, 因此在内存占用上 Res-Segnet 比 Res-FCN 更高效。为更好地说明本研究设计的网络模型的先进性, 将其与目前先进的分割模型 DeeplabV3⁺^[32] 进行比较, Res-U-Net 的 PA、MIoU 与 DeeplabV3⁺ 的相当, 而 MPA 优于 DeeplabV3⁺, 模型大小仅为 DeeplabV3⁺ 的一半, 单幅图像分割速度快于 DeeplabV3⁺, 说明本研究设计的网络结构可以快速分割出胡萝卜开裂区域。

表 3 不同分割网络的性能

Table 3 Performance of different segmentation models

分割模型 Segmentation models	模型 大小 Model size/G	检测 时间 Detection time/s	像素准确率 Pixel accuracy/%	类别平均像素 准确率 Mean pixel accuracy/%	平均交互比 Mean intersection over union/%
Res-FCN	1.50	0.67	98.20	94.54	92.48
Res-Segnet	0.33	1.32	97.28	91.40	88.68
Res-U-net	0.24	0.72	98.31	96.05	92.81
DeeplabV3 ⁺	0.46	0.92	98.38	95.77	93.17

3.4 Res-U-net 训练过程

Res-U-net 模型训练时损失函数为焦点损失函数, α, β 分别为 0.25 和 2, 优化器选择为 Adadelata, 不预设学习率。Res-U-net 模型在训练集和验证集上的准确率、损失如

图 10 所示。由图可知, Res-U-net 模型在前 50 轮训练的时候, 准确率快速上升、损失值迅速下降, 250 轮训练以后模型准确率和损失趋于稳定, 经过 300 轮训练后, 在验证集上的准确率和损失分别为 98.40% 和 0.072, 说明设计的网络结构、选择的损失函数和优化算法可以使模型快速收敛, 达到较高的准确率和较低的损失。

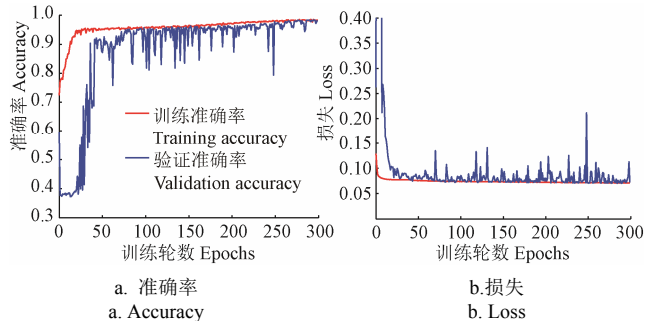
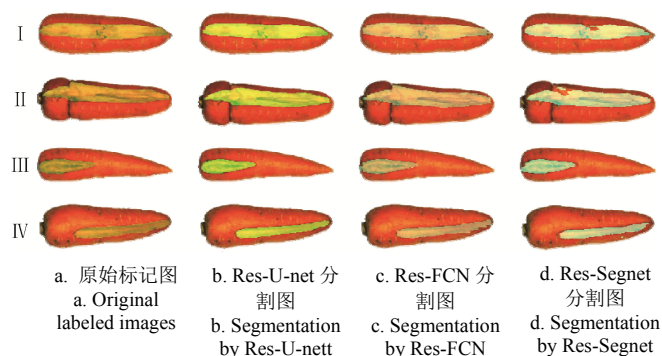


图 10 Res-U-net 训练过程中的准确率和损失
Fig.10 Accuracy and loss of Res-U-net during the training process

3.5 模型分割结果可视化比较

为更直观比较 Res-U-net、Res-FCN、Res-Segnet 模型的分割效果, 在测试集中随机选取 4 幅面积、位置不同的开裂区域图像, 将不同模型的分割效果可视化, 如图 11 所示。由图可知, Res-U-net 的分割效果最接近于原始标记结果, Res-FCN 分割区域边缘存在明显的锯齿状, Res-Segnet 分割区域比原始标记区域小, 且存在空洞现象。原因在于 Res-FCN 通过上采样卷积层生成比较粗糙的分割图, 而 Res-U-net 和 Res-Segnet 在网络中引入跳跃连接生成边缘平滑的分割图。由 I、III 可知, Res-U-net 分割效果均接近于原始标记图, 表明 Res-U-net 分割效果不受开裂区域面积大小的影响。由 III、IV 可知, Res-U-net 对胡萝卜不同位置的开裂缺陷均有较好的分割效果。由 II 可知, Res-FCN、Res-Segnet、Res-U-net 均未将胡萝卜的周向凹陷分割为开裂, 说明这些模型均很好地学习到了胡萝卜开裂区域特征。综上, Res-U-net 模型分割精度高、鲁棒性强, 可作为胡萝卜开裂区域自动分割模型。



注: I 纵贯浅型开裂, II 纵贯深型开裂; III 头部开裂, IV 中下部开裂。
Note: I represents longitudinal shallow crack, II represents longitudinal deep crack, III represents the crack in the head, IV represents the crack in the middle and lower part.

图 11 不同模型分割效果可视化

Fig.11 Visualization of different model segmentation results

4 结 论

1) 本文提出了一种集胡萝卜表面缺陷识别与开裂区域分割为一体的深度多分支模型融合网络, 在自制的胡萝卜表面缺陷数据集上, 实现了胡萝卜不同表面缺陷识别和开裂区域分割提取。

2) 以预训练的 ResNet-50 为胡萝卜图像特征提取器, 分别输出 ResNet-50 的第 1、10、22、40、49 层特征作为支持向量机 (SVM) 的输入, 得到 5 个不同的分类模型。随着网络层数的增加, 模型准确率不断提高, 第 49 层输出特征训练的模型准确率最高, 测试准确率为 94.71%。

3) 为提高胡萝卜缺陷识别分类模型的准确率, 采用硬投票、软投票和 Stacking 方法对 5 个分类模型进行融合得到的 3 个融合模型, 其准确率从小到大排序为硬投票、软投票、Stacking 方法, Stacking 融合方法得到的模型准确率最高为 98.40%。

4) 以预训练的 ResNet-50 为编码器, 根据 FCN、U-net、Segnet 网络的思想构建了多个胡萝卜开裂区域分割模型。结果表明 Res-U-net 分割模型综合性能最好, 其像素准确率、类别平均像素准确率、平均交互比为 98.31%、96.05%、92.81%, 模型尺寸为 0.24 G, 单幅图像分割时间为 0.72 s。

本文提出的深度多分支模型融合网络能够准确识别出胡萝卜表面缺陷类型, 并精确地分割提取出胡萝卜开裂区域, 为后续更加全面地评价胡萝卜表面质量和开裂胡萝卜的修整奠定了基础。

[参 考 文 献]

- [1] FAO, 联合国粮食及农业组织, 2020. [2020-11-20] <http://www.fao.org/>.
- [2] Hameed K, Chai D, Rassau A. A comprehensive review of fruit and vegetable classification techniques[J]. Image and Vision Computing, 2018, 80: 24-44.
- [3] Batchelor M M, Searcy S W. Computer vision determination of the stem/root joint on processing carrots[J]. Journal of Agricultural Engineering Research, 1989, 43: 259-269.
- [4] Howarth M S. Reflectance characteristics of fresh-market carrots[J]. Transactions of the ASAE, 1990, 33(3): 961-964.
- [5] Howarth M S, Searcy S W. Fresh market carrot inspection by machine vision[C]//Optics in Agriculture International Society for Optics and Photonics, 1991, 1379: 141-150.
- [6] Howarth M S, Brandon J R, Searcy S W, et al. Estimation of tip shape for carrot classification by machine vision[J]. Journal of Agricultural Engineering Research, 1992, 53: 123-139.
- [7] Hahn F, Sanchez S. Carrot volume evaluation using imaging algorithms[J]. Journal of Agricultural Engineering Research, 2000, 75(3): 243-249.
- [8] 韩仲志, 邓立苗, 徐艳, 等. 基于图像处理的胡萝卜青头、须根与开裂的检测方法[J]. 农业工程学报, 2013, 29(9): 156-161.
Han Zhongzhi, Deng Limiao, Xu Yan, et al. Image processing method for detection of carrot green-shoulder, fibrous roots and surface cracks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(9): 156-161. (in Chinese with English

- abstract)
- [9] Deng L, Du H, Han Z. A carrot sorting system using machine vision technique[J]. *Applied Engineering in Agriculture*, 2017, 33(2): 149-156.
 - [10] Xie W, Wang F, Yang D. Research on carrot surface defect detection methods based on machine vision[J]. *IFAC-PapersOnLine*, 2019, 52(30): 24-29.
 - [11] 谢为俊, 魏硕, 王凤贺, 等. 基于机器视觉的胡萝卜表面缺陷识别方法研究[J]. *农业机械学报*, 2020, 51(S1): 458-464.
Xie Weijun, Wei Shuo, Wang Fenghe, et al. Machine vision based detection method of carrot external defects[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(S1): 458-464.
 - [12] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey[J]. *Computers and Electronics in Agriculture*, 2018, 147: 70-90.
 - [13] Xie W, Wang F, Yang D. Research on carrot grading based on machine vision feature parameters[J]. *IFAC-PapersOnLine*, 2019, 52(30): 30-35.
 - [14] Zhu H, Deng L, Wang D, et al. Identifying carrot appearance quality by transfer learning[J]. *Journal of Food Process Engineering*, 2019, 42(6): 1-10.
 - [15] 倪建功, 李娟, 邓立苗, 等. 基于知识蒸馏的胡萝卜外观品质等级智能检测[J]. *农业工程学报*, 2020, 36(18): 181-187.
Ni Jiangong, Li Juan, Deng Limiao, et al. Intelligent detection of appearance quality of carrot grade using knowledge distillation[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(18): 181-187. (in Chinese with English abstract)
 - [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: 770-778.
 - [17] 王春山, 周冀, 吴华瑞, 等. 改进 Multi-scale ResNet 的蔬菜叶部病害识别[J]. *农业工程学报*, 2020, 36(20): 209-217.
Wang Chunshan, Zhou Ji, Wu Huarui, et al. Identification of vegetable leaf diseases based on improved Multi-scale ResNet[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(20): 209-217. (in Chinese with English abstract)
 - [18] 李善军, 胡定一, 高淑敏, 等. 基于改进 SSD 的柑橘实时分类检测[J]. *农业工程学报*, 2019, 35(24): 307-313.
Li Shanjun, Hu Dingyi, Gao Shumin, et al. Real-time classification and detection of citrus based on improved single short multibox detector[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(24): 307-313. (in Chinese with English abstract)
 - [19] 梁习卉子, 陈兵旗, 李民赞, 等. 基于 HOG 特征和 SVM 的棉花行数动态计数方法[J]. *农业工程学报*, 2020, 36(15): 173-181.
Liang Xihuizi, Chen Bingqi, Li Minzan, et al. Method for dynamic counting of cotton rows based on HOG feature and SVM[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(15): 173-181. (in Chinese with English abstract)
 - [20] 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*. Cham, Springer, 2015: 234-241.
 - [21] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
 - [22] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
 - [23] 杨建宇, 周振旭, 杜贞容, 等. 基于 SegNet 语义模型的高分辨率遥感影像农村建设用地提取[J]. *农业工程学报*, 2019, 35(5): 251-258.
Yang Jianyu, Zhou Zhenxu, Du Zhenrong, et al. Rural construction land extraction from high spatial resolution remote sensing image based on SegNet semantic segmentation model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(5): 251-258. (in Chinese with English abstract)
 - [24] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
 - [25] 谢文涌, 柴琴琴, 甘永辉, 等. 基于多特征提取和 Stacking 集成学习的金线莲品系分类[J]. *农业工程学报*, 2020, 36(14): 203-210.
Xie Wenyong, Chai Qinqin, Gan Yonghui, et al. Strains classification of *Anoectochilus roxburghii* using multi-feature extraction and Stacking ensemble learning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(14): 203-210. (in Chinese with English abstract)
 - [26] 冯海林, 胡明越, 杨垠晖, 等. 基于树木整体图像和集成迁移学习的树种识别[J]. *农业机械学报*, 2019, 50(8): 235-242.
Feng Hailin, Hu Mingyue, Yang Yinhui, et al. Tree species recognition based on overall tree image and ensemble of transfer learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(8): 235-242. (in Chinese with English abstract)
 - [27] 谢为俊, 丁冶春, 王凤贺, 等. 基于卷积神经网络的油茶籽完整性识别方法[J]. *农业机械学报*, 2020, 51(7): 13-21.
Xie Weijun, Ding Yechun, Wang Fenghe, et al. Integrity recognition of *Camellia oleifera* seeds based on convolutional neural network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(7): 13-21. (in Chinese with English abstract)
 - [28] Shore J, Johnson R. Properties of cross-entropy minimization[J]. *IEEE Transactions on Information Theory*, 1981, 27(4): 472-482.
 - [29] Milletari F, Navab N, Ahmadi S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation[C]// *2016 Fourth International Conference on 3D vision (3DV)*, Stanford, USA, 2016: 565-571.
 - [30] Lin T, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]// *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 2980-2988.

- [31] Alberto G, Sergio O, Sergiu O, et al. A survey on deep learning techniques for image and video semantic segmentation[J]. *Applied Soft Computing*, 2018, 70: 41-65.
- [32] 易诗, 李俊杰, 贾勇. 基于红外热成像的夜间农田实时语义分割[J]. *农业工程学报*, 2020, 36(18): 174-180.
- Yi Shi, Li Junjie, Jia Yong. Real-time semantic segmentation of farmland at night using infrared thermal imaging[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(18): 174-180. (in Chinese with English abstract)

Classification and segmentation of defect carrots using deep multi-branch models fusion network

Xie Weijun, Wei Shuo, Zheng Zhaohui, Yang Guangzhao, Ding Xin, Yang Deyong^{*}

(College of Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Detection of carrot defects plays an important role in the sale of carrots. The segmentation and extraction of carrot crack regions have become necessary to automatically evaluate the crack degree of carrots, and further trim the area of the crack. In the traditional detection of carrot external quality, different image processing was designed using the features of different defects, showing high complexity while low robustness. In this study, a deep multi-branch models fusion network (CS-net) was proposed to integrate the recognition of carrot defects and segmentation of crack regions. The network contained two parts: the classification of carrot defects (C-Net), and segmentation extraction of carrot crack regions (S-Net). In C-Net, the ResNet-50 pre-trained on the ImageNet dataset was taken as an image feature extractor of carrot. The output features in the 1st, 10th, 22nd, 40th and 49th layers of ResNet-50 were processed by different pooling methods, including Average Pooling (AVP), Global Average Pooling (GAP), and Spatial Pyramid Pooling (SPP), as well as dimension reduction (principal component analysis, ReliefF). The extracted features were then used as input of Support Vector Machines (SVM) to obtain five classification models. Besides, the five classification models were ensemble with different fusion strategies (hard voting, soft voting and stacking) to obtain the final classification model. In S-Net, the pre-trained ResNet-50 was served as the encoder of segmentation network, and then the network decoder was designed to build the segmentation network of carrot crack regions. The results showed that the output features in the 49th layer of the ResNet-50 with SVM model performed best with the test accuracy of 94.71% among the single model. The fusion model with the stacking ensemble performed best with the accuracy of 98.40%, indicating a better performance in the fusion model than the single model. Different pooling methods had different effects on the performance of the model. In the low-level feature maps, the order of performance for different pooling methods was SPP > AVP > GAP. However, the pooling methods had little impact on the model performance with the high-level semantic features. It was found that dimensionality reduction reduced the number of features and then improved the performance of the model. In the segmentation part, the constructed segmentation network with the U-net construction ideas (Res-U-net) performed best with the Pixel Accuracy (PA), Mean Pixel Accuracy (MPA) and mean intersection over union (MIoU) of 98.31%, 96.05% and 92.81%, respectively. The performance of Res-U-net was not affected by the cracking area and different positions of crack. Comparing with Deeplabv3⁺, the PA and the MIoU in the Res-U-net were similar to those of Deeplabv3⁺, while the MPA was better than that of Deeplabv3⁺, and the model size was only half of that of Deeplabv3⁺. In addition, the segmentation speed of single image was faster than that of Deeplabv3⁺. The Res-U-net reached an advanced level in the segmentation task of carrot crack defects. The defect recognition and segmentation network have a positive significance on the quantitative evaluation of carrot external quality and the automatic trim of carrot crack.

Keywords: image segmentation; deep learning; classification; carrot; external quality; fusion model