

# 基于卷积神经网络的田间多簇猕猴桃图像识别方法

傅隆生<sup>1,2</sup>, 冯亚利<sup>1</sup>, Elkamil Tola<sup>3</sup>, 刘智豪<sup>1</sup>, 李瑞<sup>1</sup>, 崔永杰<sup>1,2</sup>

(1. 西北农林科技大学机械与电子工程学院, 杨凌 712100; 2. 农业部农业物联网重点实验室, 杨凌 712100;  
3. Precision Agriculture Research Chair, King Saud University, Riyadh 11451, Saudi Arabia)

**摘要:**为实现田间条件下快速、准确地识别多簇猕猴桃果实, 该文根据猕猴桃的棚架式栽培模式, 采用竖直向上获取果实图像的拍摄方式, 提出一种基于 LeNet 卷积神经网络的深度学习模型进行多簇猕猴桃果实图像的识别方法。该文构建的卷积神经网络通过批量归一化方法, 以 ReLU 为激活函数, Max-pooling 为下采样方法, 并采用 Softmax 回归分类器, 对卷积神经网络结构进行优化。通过对 100 幅田间多簇猕猴桃图像的识别, 试验结果表明: 该识别方法对遮挡果实、重叠果实、相邻果实和独立果实的识别率分别为 78.97%、83.11%、91.01%和 94.78%。通过与 5 种现有算法进行对比试验, 该文算法相对相同环境下的识别方法提高了 5.73 个百分点, 且识别速度达到了 0.27 s/个, 识别速度较其他算法速度最快。证明了该文算法对田间猕猴桃图像具有较高的识别率和实时性, 表明卷积神经网络在田间果实识别方面具有良好的应用前景。

**关键词:** 图像处理; 图像识别; 算法; 深度学习; 卷积神经网络; 猕猴桃

doi: 10.11975/j.issn.1002-6819.2018.02.028

中图分类号: TP391.41

文献标志码: A

文章编号: 1002-6819(2018)-02-0205-07

傅隆生, 冯亚利, Elkamil Tola, 刘智豪, 李瑞, 崔永杰. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205-211. doi: 10.11975/j.issn.1002-6819.2018.02.028 http://www.tcsae.org  
Fu Longsheng, Feng Yali, Elkamil Tola, Liu Zhihao, Li Rui, Cui Yongjie. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(2): 205-211. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2018.02.028 http://www.tcsae.org

## 0 引言

中国是猕猴桃栽培面积最大的国家<sup>[1]</sup>, 大都是手工收获果实<sup>[2]</sup>。在乡村劳动力向城镇转移的大背景下, 发展猕猴桃自动化采摘技术, 特别是研发猕猴桃采摘机器人, 具有重要的意义<sup>[3-7]</sup>。猕猴桃采摘机器人的首要及关键技术之一是果实的快速有效识别<sup>[7]</sup>。在自然场景下, 由于猕猴桃果实颜色与枯叶、枝干、果柄等复杂背景的颜色相近<sup>[8]</sup>, 果实成簇并存在大量重叠与遮挡果实。因此, 对田间环境下猕猴桃果实的特征学习从而进行识别是猕猴桃采摘机器人急需解决的一个关键性问题<sup>[9]</sup>。

近年来, 科研工作者对自然环境下猕猴桃果实的识别进行了深入研究。总体分为 2 部分, 从果实斜侧面获取图像进行识别和从果实底部竖直向上采集图像进行识别。对于从斜侧面获取的图像, 丁亚兰等<sup>[10]</sup>利用  $R-B$  颜色因子, 采用固定阈值进行猕猴桃图像分割, 但无法有效识别强反光及暗影区的果实; 崔永杰等<sup>[11]</sup>利用  $L^*a^*b^*$  颜色空间  $a^*$  通道进行猕猴桃图像分割, 采用椭圆形 Hough 变换拟合单个果实轮廓, 但对田间背景下的猕猴

桃果实分割效果不理想; 崔永杰等<sup>[8]</sup>通过对比不同颜色空间, 提出利用  $0.9R-G$  颜色特征, 结合椭圆形 Hough 变换进行果实的识别, 但识别的果实针对特定类型, 在实际应用中其适用性受到抑制; 詹文田等<sup>[5]</sup>基于 Adaboost 算法, 通过引入 RGB、HIS、 $L^*a^*b^*$  颜色模型构建识别猕猴桃果实的分类器, 但识别速度有待提高; 慕军营等<sup>[12]</sup>利用 Otsu 算法在  $a^*$  通道进行图像分割, 基于正椭圆 Hough 变换提取 Canny 算子获取的猕猴桃果实边缘图像进行识别, 但不能很好识别远处的果实。对于竖直向上获取的图像, Scarfe 等<sup>[13]</sup>采用固定阈值法去除背景, 提取 Sobel 边缘后, 利用模板匹配的方法识别猕猴桃, 但未利用果实的形状信息; Fu 等<sup>[14]</sup>提出  $1.1R-G$  颜色特性进行夜间猕猴桃图像分割, 并结合最小外接矩形法和椭圆形 Hough 变换识别每个果实, 但只能识别单簇果实; 傅隆生等<sup>[9]</sup>利用竖直向上成像时果萼都显现且与果实有区别的特点, 进行基于果萼的夜间猕猴桃识别, 但未涉及遮挡与重叠果实的识别, 且对多果簇识别效果不佳。田间环境下的猕猴桃果实图像特征多样、背景复杂且形态特征差异大。已有识别方法主要根据经验, 受样本和人为主观性的影响, 很难具有普适性, 鲁棒性差, 难以用一种方法同时识别所有类型的猕猴桃果实, 且不能同时识别多簇果实, 不能满足复杂田间环境下的应用需求。

相比常规方法, 近年兴起的卷积神经网络<sup>[15]</sup> (convolutional neural network, CNN) 方法直接由数据本身驱动特征及表达关系的自我学习, 对图像具有极强的数

收稿日期: 2017-08-28 修订日期: 2017-12-26

基金项目: 陕西省重点研发计划一般项目 (2017NY-164); 陕西省科技统筹创新工程计划项目 (2015KTCQ02-12); 国家自然科学基金资助项目 (61175099); 西北农林科技大学国际合作种子基金 (A213021505)

作者简介: 傅隆生, 江西吉安人, 副教授, 博士, 主要从事农业智能化技术与装备研究。Email: fulsh@nwfufu.edu.cn

中国农业工程学会会员: 傅隆生 (E042600025M)

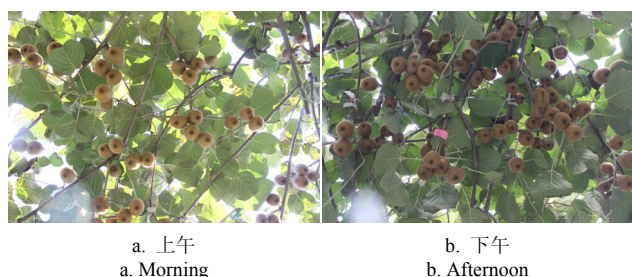
据表征能力。CNN 已在手写字符识别<sup>[16-18]</sup>、人脸识别<sup>[19-21]</sup>、行为识别<sup>[22-23]</sup>以及农作物识别<sup>[24-25]</sup>等方面, 获得了较好的效果。学者们也开始 CNN 在果实识别方面的研究, 王前程<sup>[26]</sup>将 CNN 应用于处理后的 6 种水果图像数据集进行识别, 证明了 CNN 在水果图像识别上的有效性; Sa 等<sup>[27]</sup>基于 CNN 模型建立果实的深度网络识别模型, 对不同的果实图像进行测试, 取得不错的效果。以上研究的开展为 CNN 应用于果实识别提供了参考和可行性依据, 同时也表明 CNN 在图像识别中可以克服传统方法的不足。

本文在采集大量田间样本图像的基础上, 通过 CNN 对复杂背景下的猕猴桃果实进行识别, 避免人为主观因素影响识别结果。依据田间环境下猕猴桃图像的特点, 优化 LeNet 卷积神经网络的结构与参数, 从而建立一种基于卷积神经网络的田间猕猴桃果实图像的识别模型, 以实现田间复杂环境下多簇猕猴桃果实的快速有效识别。

## 1 材料与方法

### 1.1 图像采集

试验供试图像于 2016 年 10 月—11 月采集自陕西省眉县猕猴桃试验站 (34°07'39"N, 107°59'50"E, 海拔 648 m), 将数码相机 (Canon EOS 40D) 通过三脚架置于猕猴桃果实下方 100 cm 左右对“海沃德”品种进行拍照。共采集原始图像 700 幅, 晴天上午、下午 2 个不同时间段各 350 幅, 图像格式为 JPEG, 分辨率为 2 352×1 568 像素, 如图 1 所示。



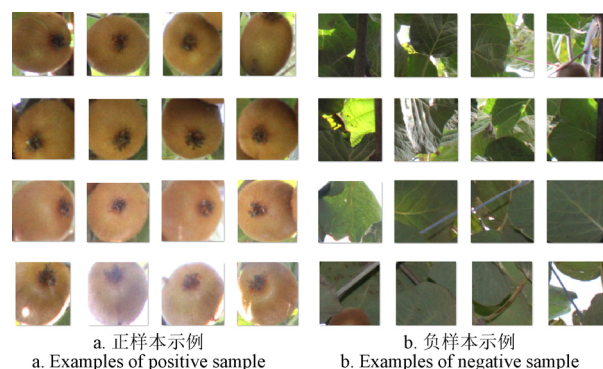
a. 上午  
a. Morning  
b. 下午  
b. Afternoon

图 1 田间自然环境下的猕猴桃图像

Fig.1 Kiwifruit image under natural conditions in field

### 1.2 样本数据集与试验方法

猕猴桃采用棚架式栽培方式形成果实自然下垂且位于枝叶下方的特点, 底部竖直向上成像后, 每个果实的果萼部分都显现。该文随机选取 600 幅 (上午和下午各 300 幅) 图像, 截取具有萼的单果作为目标区域, 并剔除无效的图像区域, 所截取的样本图像最小尺寸为 74×76 像素。再由人工对原始采集的图片进行筛选, 从而避免数据样本的错误选定和单一性。最终试验所用数据集由正样本 (6 000 幅) 和负样本 (4 020 幅) 组成, 为 2 个不同时间段均匀分布 (上午和下午各 5 010 幅)。数据集均用于卷积神经网络的训练和参数优化验证, 分别从正、负样本中随机选择 80% 样本构建训练集, 20% 作为验证集。部分正、负样本图样如图 2 所示。



a. 正样本示例  
a. Examples of positive sample  
b. 负样本示例  
b. Examples of negative sample

图 2 试验部分数据集样本示例

Fig.2 Example images of partial dataset

模型训练完成后, 将剩余的 100 幅猕猴桃原始图像 (上午和下午各 50 幅) 作为模型效果验证的测试集, 为减少计算量及运行时间, 将原图像缩放为 600×400 像素进行测试, 训练数据集与测试图像间不重叠。最后本文将与已有的猕猴桃识别方法进行对比分析。由于测试数据集中两个不同时间段样本数量呈均衡分布, 因此可将测试结果的平均准确率作为本文模型的识别效果评价指标<sup>[28]</sup>。

### 1.3 卷积神经网络的构建

本文使用 Matlab 的 MatConvnet 工具箱<sup>[29]</sup>建立卷积神经网络。LeNet<sup>[30]</sup>是典型的卷积神经网络, 最初成功用于手写数字识别。由于猕猴桃果实的识别亦是对某一未知猕猴桃果实图像进行识别和匹配, 该过程与 LeNet 手写数字的识别相似。因此可以将卷积神经网络 LeNet 作为基础网络架构, 并对其重要的结构参数和训练策略进行优化, 以获取适合猕猴桃果实图像识别的模型架构。LeNet 卷积神经网络具体算法描述如下:

#### 1) 卷积层

卷积核的尺寸与数量对于 CNN 的性能至关重要。输入图像通过  $j$  个不同的卷积核卷积, 生成  $j$  个不同的特征图, 卷积层如式 (1) 所示。

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} W_{ij}^l + b_j^l) \quad (1)$$

式中  $x_j^l$  表示在  $l$  层 (卷积层) 的第  $j$  个特征图,  $f()$  表示激活函数,  $M_j$  代表输入图像的特征量,  $W_{ij}^l$  为卷积核,  $b_j^l$  为偏置项。

#### 2) 下采样层

下采样层对输入进行抽样, 如式 (2) 所示。

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l) \quad (2)$$

式中  $\text{down}()$  为下采样函数,  $\beta_j^l$  输出特征的偏置。

## 2 基于卷积神经网络的猕猴桃果实识别

### 2.1 试验软件及硬件

本文处理平台为笔记本电脑, 处理器为 Inter (R) Core (TM) i3, 主频为 2.40 GHz, 4 GB 内存, 500 GB 硬盘, 运行环境为: Windows 7 64 位, Matlab R2016a, Microsoft Visual Studio 12.0。

## 2.2 网络模型的试验方法

若将 LeNet 结构直接引入猕猴桃果实图像特征提取与分类, 考虑与原始网络所用样本(手写字符)的差异以及猕猴桃果实图像的成像通道, 本文将所用正、负样本图像通过插值缩放变化为  $3 \times 32 \times 32$  的矩阵, 并将正、负样本分别标记为“2”和“1”, 作为网络训练的输入。由于猕猴桃图像受扭转、变形等因素影响较小, 因此可以约减原始 LeNet 网络中各卷积层中局部感受野的数量, 以提高网络的训练速度。该文对不同结构的卷积神经网络进行训练, 然后通过验证对比不同模型识别的准确率及耗时试验, 本研究最终采用的局部感受野的尺寸均为  $5 \times 5$ , 3 个卷积层 C1、C3、C5 局部感受野个数分别是 6、16 和 120 个。

针对各层分布不均和精度弥散的问题, 该文引入批次规则化(batch normalization, BN)法减小影响, 加快网络收敛, 防止过拟合。在原网络第 1、3、5 卷积层后添加 BN 层, 将输出按照同一批次的特征数值规范化至同一分布, 具体如下所示

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (3)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (4)$$

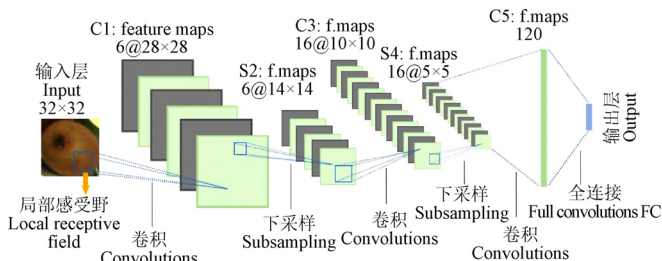
$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (5)$$

$$y_i = \gamma \hat{x}_i + \beta = \text{BN}_{\gamma, \beta}(x_i) \quad (6)$$

式中  $x_1, x_2, \dots, x_m$  为输入值,  $m$  是批量化的数目,  $\gamma$  和  $\beta$  是学习参数,  $\mu_B$  是均值,  $\sigma_B^2$  是方差,  $\hat{x}_i$  表示规范化,  $\varepsilon$  是常量, 保证数值的稳定性,  $y_i = \text{BN}_{\gamma, \beta}(x_i)$  是输出。

激活函数采用非饱和线性修正单元(rectified linear units, ReLU)。由于 Max-pooling 作为一种非线性的下采样方法, 可以在一定的程度上降低卷积层参数误差造成的估计均值偏移所引起的特征提取的误差, 试验选用 Max-pooling 作为下采样方法。网络的训练阶段采用批量随机梯度下降法(mini-batch stochastic gradient descend)。

本文选用损失函数 Softmax loss (对应 Softmax 回归分类器) 进行网络性能的对比分析。最终确定的卷积神经网络结构可表示为  $32 \times 32\text{-}6\text{C}\text{-}2\text{S}\text{-}16\text{C}\text{-}2\text{S}\text{-}120\text{C}\text{-}2$ , 如图 3 所示。



注: C1、S2、C3、S4、C5 和 FC 分别为第 1 卷积层、第 2 下采样层、第 3 卷积层、第 4 下采样层、第 5 卷积层和全连接层。  
Note: C1, S2, C3, S4, C5 and FC are the 1<sup>st</sup> convolutional layer, 2<sup>nd</sup> subsampling layer, 3<sup>rd</sup> convolutional layer, 4<sup>th</sup> subsampling layer, 5<sup>th</sup> convolutional layer and full convolutional layer, respectively.

图 3 基于 LeNet 的猕猴桃果实图像训练网络结构

Fig.3 Training architecture for kiwifruit image recognition based on LeNet

基于 LeNet 的猕猴桃果实识别步骤<sup>[31]</sup>如下所示:

1) 对裁剪后的猕猴桃图像进行分类并作相应预处理, 使图像符合网络训练的要求;

2) 对 1) 中图像进行随机采样, 获得适量的数据集, 初始化 LeNet 结构得到初始化滤波器的权值;

3) 将 2) 的滤波器与 1) 的训练集图像卷积, 获得预定数量的特征图, 用 BN 法对数据进行处理;

4) 将 3) 中获得的特征图通过式(2)进行最大化采样, 得到泛化后的图像;

5) 分别利用上边 3) 和 4) 的方法对 4) 中输出的特征图进行二次卷积, 二次批量归一化处理, 二次下采样, 获得所需的特征图;

6) 用同样的方法对 5) 中输出的特征图进行三次卷积, 三次批量归一化处理;

7) 将 6) 中所有特征图转化为一个列向量, 作为全连接层的输入, 计算识别结果和标记的差异, 通过反向传播算法自顶向下调节更新网络参数;

8) 输入处理后的测试图像, 利用训练得到的网络模型对测试图像进行分类, 通过 Softmax 分类器, 并结合多尺度滑动窗算法显示识别结果。

## 2.3 分类定义

由于田间拍摄的猕猴桃图像中果实并非全部相互独立, 因此本文按照图像中果实轮廓的完整程度将果实分为 4 种类型: 第 1 类是指果实的部分区域被遮挡而导致轮廓不完整的果实, 称为遮挡果实, 如图 4a 所示; 第 2 类是指两个及以上果实区域互相遮挡不易于区分开的果实, 称为重叠果实, 如图 4b 所示(为矩形框所标记的果实); 第 3 类是 2 个及以上果实轮廓相接, 称为相邻果实, 如图 4c 所示; 第 4 类是指果实轮廓独立完整且相互分离的果实, 称为独立果实, 如图 4d 所示。



a. 遮挡果实  
a. Occluded fruit  
b. 重叠果实  
b. Overlapped fruit  
c. 相邻果实  
c. Adjacent fruit  
d. 独立果实  
d. Separated fruit

图 4 猕猴桃图像类别

Fig.4 Categories of kiwifruit images

## 3 结果与分析

### 3.1 训练误差率

采用上文描述的 CNN 结构, 使用训练集样本来训练 CNN, 网络初始权重的初始化采用标准差为 0.01, 均值为 0 的高斯分布。样本迭代次数均设置为 45 次, 批处理 BatchSize 为 100, 并设置权重参数的初始学习速率为 0.001, 动量因子设置为 0.9。对上述训练集进行 45 次迭代的训练, 其变化曲线如图 5 所示。

结果表明, 随着迭代次数不断增加, 训练集和验证集的分类误差逐渐降低, 当训练迭代到第 28 次时, 网络可以实现对训练集和验证集的误识别率都降至 0, 之后分



类准确率趋于稳定, 且从第 3 次迭代以后训练集和验证集两者的误差差值相差不大, 说明模型状况良好。模型在经过 28 次迭代后, 训练损失基本收敛到稳定值, 表明卷积神经网络达到了预期的训练效果。

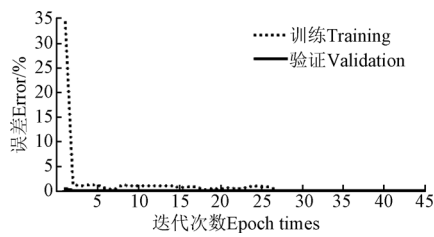


图 5 训练和验证误差曲线  
Fig.5 Errors curves of training and validation

3.2 特征图可视化分析

按照图 3 所示的网络结构, 使用训练好的模型对猕猴桃果实样本进行识别。图 6 为输入的猕猴桃果实图像经过 3 个卷积层所对应的各层特征图的输出结果, 输出层输出 1 和 2 分别指代背景和果实。由图 6 所示的各层显示结果可知, 卷积操作能够有效提取猕猴桃果实特征, 说明本试验的网络结构通过局部感受野和权值共享, 能够降低背景干扰、增强目标特征。

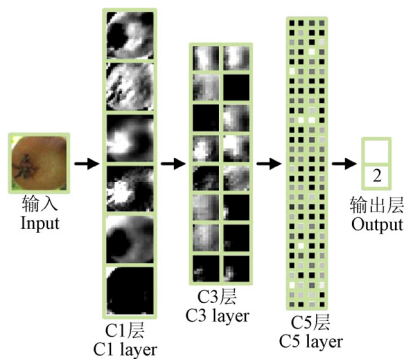


图 6 卷积神经网络各卷积层的处理结果示例  
Fig.6 Results example of each convolutions layer of CNNs

3.3 模型的效果验证

为了验证模型的可靠性与稳定性, 对测试集的 100

幅田间猕猴桃果实图像（上午和下午各 50 幅, 共包含目标猕猴桃果实 5 918 个）进行识别。该文选用重叠系数<sup>[32]</sup>作为试验结果有效性的评价指标, 重叠系数是指识别后的目标与真实目标重合的比率。该课题组设计的末端执行器<sup>[7]</sup>根据猕猴桃生长特点, 从果实底部旋转上升伸入毗邻间隙, 采用逐渐包络的方式分离毗邻果实并抓持, 试验结果表明允许的误差半径为 10 mm, 因此只需知道果实的大部分区域（80%）即可进行果实的采摘, 避免果实实际区域难以精确定位的问题。因此当重叠系数大于等于 80%, 即为正确识别。果实识别成功率=成功识别的果实数与实际目标果实数的比值。果实识别时间为一幅图片的运行时间除以该图片中成功识别的果实数。识别结果如表 1 所示, 识别效果如图 7 所示。

表 1 猕猴桃果实识别结果					
Table 1 Recognition results of kiwifruit					
参数 Parameter	遮挡果实 Occluded fruit	重叠果实 Overlapped fruit	相邻果实 Adjacent fruit	独立果实 Separated fruit	总计 Total
正确识别果实数 Recognized fruits number	642	438	3259	945	5284
果实数量 Fruits number	813	527	3581	997	5918
正确识别率 Recognition rates/%	78.97	83.11	91.01	94.78	89.29

由表 1 知, 独立果实识别率最高（94.78%）, 其次是相邻果实（91.01%）, 再次是重叠果实（83.11%）, 识别效果最差的是遮挡果实（78.97%）。当猕猴桃果实距离图像中心较远从而发生变形或被枝叶遮挡面积较大时, 果实易被误识别; 当图像中多个果实连续重叠时, 易将后边果实与前方果实判断为一个果实或者无法识别后方重叠严重的果实, 出现漏识别现象, 如图 7a 所示; 当图像中果实的部分或整体区域被阳光直射形成强烈反光, 该区域不易识别或无法识别, 影响识别精度; 图像中的多个相邻果实两侧相邻部分轮廓, 易被识别为一个果实, 出现误识别, 是造成识别率低的主要原因, 如图 7b 所示。图 7b 所示的误识别情况, 可能原因是制作训练数据集时, 单个猕猴桃果实裁剪效果不理想（裁剪重叠区域时没有处理好边缘问题）。

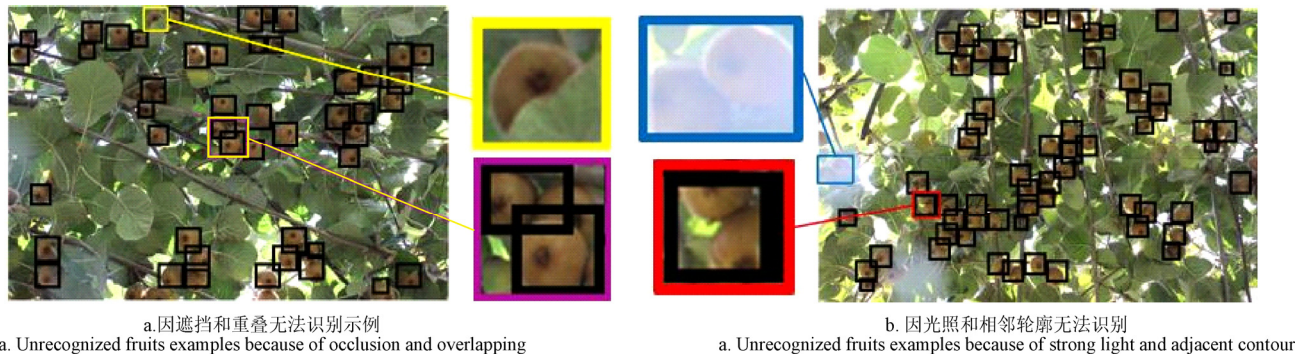


图 7 猕猴桃果实识别结果以及误识别示例  
Fig.7 Examples for recognition results and wrongly recognized kiwifruit images

在图 7 中有一些猕猴桃果实所占的区域不能被精确的识别, 识别区域（图 7 中黑色的框）稍有偏离果实实

际区域。但整体上而言, 果实的主要区域已被识别, 采用本课题组开发的猕猴桃采摘机器人末端执行器<sup>[7]</sup>能够

实现果实的采摘。

### 3.4 与常规方法的比较

相关文献提出了基于田间环境下猕猴桃图片的识别

方法，为了验证本文提出的算法性能，与 Scarfe<sup>[13]</sup>、詹文田等<sup>[5]</sup>、崔永杰等<sup>[8]</sup>、Fu 等<sup>[14]</sup>、傅隆生等<sup>[9]</sup> 5 种常规方法进行比较，结果如表 2 所示。

表 2 不同猕猴桃果实识别方法的性能比较  
Table 2 Comparing of CNNs and other methods of kiwifruit recognition

算法 Algorithm	图像拍摄方式 Shooting mode	图像特征 Image features	果实特征 Fruit categories	成功率 Recognition rates/%	单个果实识别时间 Single fruit recognition time/s
Scarfe <sup>[13]</sup> (2012)	远距离底部成像 Long distance and bottom imaging	多簇果实 Multiple clusters of fruits	遮挡、重叠、相邻、独立 Occlusion, Overlap, Adjacency, Separated	83.56	0.28
詹文田等 <sup>[5]</sup> (2013)	近距离侧面成像 Long distance and side imaging	单簇少果 Single cluster of less fruits	遮挡、重叠、相邻、独立 Occlusion, Overlap, Adjacency, Separated	96.70	—
崔永杰等 <sup>[8]</sup> (2013)	近距离侧面成像 Long distance and side imaging	单簇少果 Single cluster of less fruits	遮挡、重叠、相邻、独立 Occlusion, Overlap, Adjacency, Separated	89.10	—
Fu et al <sup>[14]</sup> (2015)	近距离底部成像 Close distance and bottom imaging	单簇少果 Single cluster of less fruits	相邻、独立 Adjacency, Separated	88.30	1.64
傅隆生等 <sup>[9]</sup> (2017)	近距离底部成像 Close distance and bottom imaging	单簇少果 Single cluster of less fruits	相邻、独立 Adjacency, Separated	94.30	0.5
本文算法	远距离底部成像 Long distance and bottom imaging	多簇果实 Multiple clusters of fruits	遮挡、重叠、相邻、独立 Occlusion, Overlap, Adjacency, Separated	89.29	0.27

从表 2 中可以看到，崔永杰等<sup>[8]</sup>、Fu 等<sup>[14]</sup>提出的算法的识别率与本文算法的识别率相近，詹文田等<sup>[5]</sup>、傅隆生等<sup>[9]</sup>提出的算法的识别率比本文算法的识别率分别高 7.41 个百分点和 5.01 个百分点。但是 Fu 等<sup>[14]</sup>和傅隆生等<sup>[9]</sup>识别的猕猴桃果实图像是近距离底部拍摄的图像，并且只针对果实相互独立和相邻的单簇少果类型，识别 5 果及以上效果不好；詹文田等<sup>[5]</sup>和崔永杰等<sup>[8]</sup>识别的猕猴桃果实图像是近距离侧面拍摄的图像，图像的获取在很大程度上是人为拍摄特定角度的图片，并且用于试验的测试图片每一张只针对一种果实特征，在实际应用中具有一定局限性。此外，Scarfe<sup>[13]</sup>、詹文田等<sup>[5]</sup>、崔永杰等<sup>[8]</sup>、Fu 等<sup>[14]</sup>、傅隆生等<sup>[9]</sup>的识别方法均需对猕猴桃图像提取人工选取的底层特征，对图像进行大量的预处理，操作复杂。另外已有常规算法缺乏高层次表达，难以体现所选底层特征间的空间关系，因此对于识别多果相对困难。

本文提出的算法只需对图像进行简单的预处理，用于试验的测试图片每一张都包含果实的 4 种特征，一幅图片至少包含 30 个果实以上，并且在单个猕猴桃果实识别速度上本文算法比其他 3 种算法有了明显提升。在相同的识别环境下，本文算法的识别率 89.29% 高于 Scarfe<sup>[13]</sup> 的 83.56%，提高了 5.73 个百分点。总体而言，本文提出的基于卷积神经网络的识别方法具有较强的抗干扰能力，可以同时识别田间复杂环境下的多簇猕猴桃果实，且识别过程耗时短，对光线变化、枝叶遮挡均具有相对较好的鲁棒性，更加满足猕猴桃采摘机器人实际应用中的采摘要求。

## 4 结 论

1) 针对猕猴桃采摘的需求，提出了一种基于卷积神经网络的田间猕猴桃果实识别方法，本文对 LeNet 模型进行参数优化和结构约简，并通过试验验证，表明识别模型可以自动从复杂数据中有效学习到猕猴桃的特征，

从而避免了常规方法中由研究者主观选取特征的不足。同时，简约后的模型在很大程度上满足了在实际中的应用，加强了该模型在常规性能计算平台上的适应性。

2) 本文构建的 32×32-6C-2S-16C-2S-120C-2 结构卷积神经网络，经过训练后对 100 幅图像中共包含 5 918 个猕猴桃果实的识别率达到 89.29%，相对其它远距离底部成像识别多簇猕猴桃果实的识别方法提高了 5.73 个百分点。在果实识别速度上，本算法达到平均 0.27 s 识别一个猕猴桃果实，基本上满足猕猴桃采摘机器人的工作需求。

3) 本文所用模型可以应用于田间环境下多果猕猴桃识别，突破了大多数常规识别算法不能同时识别多簇猕猴桃果实的不足，为猕猴桃采摘机器人多机械臂作业的研究提供有力支撑。

目前，该文所用模型可以准确地识别出猕猴桃果实是否存在，但对于一些遮挡和重叠果实没有达到很好的效果，尤其是两个或两个以上相邻或重叠果实两侧部分轮廓，易被识别为一个果实，从而出现误识别现象，这种现象有待进一步研究。同时，为了达到推广应用的成果，下一步将深化网络结构，增加学习样本的种类与数量，提高分类器的识别能力。

### [参 考 文 献]

- [1] 张计育, 莫正海, 黄胜男, 等. 21 世纪以来世界猕猴桃产业发展以及中国猕猴桃贸易与国际竞争力分析[J]. 中国农学通报, 2014, 30(23): 48—55.  
Zhang Jiyu, Mo Zhenghai, Huang Shengnan, et al. Development of kiwifruit industry in the world and analysis of trade and international competitiveness in china entering 21st century[J]. China Agricultural Science Bulletin, 2014, 30(23): 48—55. (in Chinese with English abstract)
- [2] 陈军, 王虎, 蒋浩然, 等. 猕猴桃采摘机器人末端执行器设计[J]. 农业机械学报, 2012, 43(10): 151—154.  
Chen Jun, Wang Hu, Jiang Haoran, et al. Design of end-effector for kiwifruit harvesting robot[J]. Transactions of

- the Chinese Society for Agricultural Machinery, 2012, 43(10): 151—154. (in Chinese with English abstract)
- [3] Zhang L, Wang Y, Yang Q, et al. Kinematics and trajectory planning of a cucumber harvesting robot manipulator[J]. International Journal of Agricultural & Biological Engineering, 2009, 2(1): 1—7.
- [4] Rakun J, Stajanko D, Zazula D. Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry[J]. Computers & Electronics in Agriculture, 2011, 76(1): 80—88.
- [5] 詹文田, 何东健, 史世莲. 基于 Adaboost 算法的田间猕猴桃识别方法[J]. 农业工程学报, 2013, 29(23): 140—146. Zhan Wentian, He Dongjian, Shi Shilian. Recognition of kiwifruit in field based on Adaboost algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(23): 140—146. (in Chinese with English abstract)
- [6] Bechar A, Vigneault C. Agricultural robots for field operations: Concepts and components[J]. Biosystems Engineering, 2016, 149: 94—111.
- [7] 傅隆生, 张发年, 槐岛芳德, 等. 猕猴桃采摘机器人末端执行器设计与试验[J]. 农业机械学报, 2015, 46(3): 1—8. Fu Longsheng, Zhang Fanian, Gejima Yoshinori, et al. Development and experiment of end-effector for kiwifruit harvesting robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(3): 1—8. (in Chinese with English abstract)
- [8] 崔永杰, 苏帅, 王霞霞, 等. 基于机器视觉的自然环境中猕猴桃识别与特征提取[J]. 农业机械学报, 2013, 44(5): 247—252. Cui Yongjie, Su Shuai, Wang Xiaxia, et al. Recognition and feature extraction of kiwifruit in natural environment based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(5): 247—252. (in Chinese with English abstract)
- [9] 傅隆生, 孙世鹏, Vázquez-Arellano Manuel, 等. 基于果萼图像的猕猴桃果实夜间识别方法[J]. 农业工程学报, 2017, 33(2): 199—204. Fu Longsheng, Sun Shipeng, Vázquez-Arellano Manuel, et al. Kiwifruit recognition method at night based on fruit calyx[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(2): 199—204. (in Chinese with English abstract)
- [10] 丁亚兰, 耿楠, 周全程. 基于图像的猕猴桃果实目标提取研究[J]. 微计算机信息, 2009, 18(4): 294—295. Ding Yalan, Geng Nan, Zhou Quancheng. Research on the object extraction of kiwifruit based on images[J]. Microcomputer Information, 2009, 18(4): 294—295. (in Chinese with English abstract)
- [11] 崔永杰, 苏帅, 吕志海, 等. 基于 Hough 变换的猕猴桃毗邻果实的分离方法[J]. 农机化研究, 2012, 34(12): 166—169. Cui Yongjie, Su Shuai, Lü Zhihai, et al. A method for separation of kiwifruit adjacent fruits based on Hough transformation[J]. Journal of Agricultural Mechanization Research, 2012, 34(12): 166—169. (in Chinese with English abstract)
- [12] 慕军营, 陈军, 孙高杰, 等. 基于机器视觉的猕猴桃特征参数提取[J]. 农机化研究, 2014, 36(6): 138—142. Mu Junying, Chen Jun, Sun Gaojie, et al. Characteristic parameters extraction of kiwifruit based on machine vision[J]. Journal of Agricultural Mechanization Research, 2014, 36(6): 138—142. (in Chinese with English abstract)
- [13] Scarfe A J. Development of an Autonomous Kiwifruit Harvester[D]. New Zealand, Manawatu: Massey University, 2012.
- [14] Fu L, Wang B, Cui Y, et al. Kiwifruit recognition at nighttime using artificial lighting based on machine vision[J]. International Journal of Agricultural and Biological Engineering, 2015, 8(4): 52—59.
- [15] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// Curran Associates Inc. International Conference on Neural Information Processing Systems. 2012: 1097—1105.
- [16] Alwzawy H, Albehadili H, Alwan Y, et al. Handwritten digit recognition using convolutional neural networks[J]. International Journal of Innovative Research in Computer & Communication Engineering, 2016, 4(2): 1101—1106.
- [17] Yang W, Jin L, Tao D, et al. Dropsample: A new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition[J]. Pattern Recognition, 2016, 58(4): 190—203.
- [18] Albu R D. Human face recognition using convolutional neural networks[J]. Journal of Electrical & Electronics Engineering, 2009, 2(2): 110—113.
- [19] Ramaiah N P, Ijjina E P, Mohan C K. Illumination invariant face recognition using convolutional neural networks[C]// IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems, 2015: 1—4.
- [20] Singh R, Om H. Newborn face recognition using deep convolutional neural network[J]. Multimedia Tools & Applications, 2017, 76(18): 19005—19015.
- [21] Dobhal T, Shitole V, Thomas G, et al. Human activity recognition using binary motion image and deep learning [J]. Procedia Computer Science, 2015, 58: 178—185.
- [22] Ronao C A, Cho S B. Human activity recognition with smartphone sensors using deep learning neural networks[J]. Expert Systems with Applications, 2016, 59: 235—244.
- [23] 王忠民, 曹洪江, 范琳. 一种基于卷积神经网络深度学习的人体行为识别方法[J]. 计算机科学, 2016, 43(s2): 56—58. Wang Zhongmin, Cao Hongjiang, Fan Lin. Method on human activity recognition based on convolutional neural networks[J]. Computer Science, 2016, 43(s2): 56—58. (in Chinese with English abstract)
- [24] 高震宇, 王安, 刘勇, 等. 基于卷积神经网络的鲜茶叶智能分选系统研究[J]. 农业机械学报, 2017, 48(7): 53—58. Gao Zhenyu, Wang An, Liu Yong, et al. Intelligent fresh-tea-leaves sorting system research based on convolution neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(7): 53—58. (in Chinese with English abstract)
- [25] 周云成, 许童羽, 郑伟, 等. 基于深度卷积神经网络的番茄主要器官分类识别方法[J]. 农业工程学报, 2017, 33(15): 219—226. Zhou Yuncheng, Xu Tongyu, Zheng Wei, et al. Classification and recognition approaches of tomato main organs based on DCNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(15): 219—226. (in Chinese with English abstract)

- [26] 王前程. 基于深度学习的水果图像识别算法研究[D]. 保定: 河北大学, 2016.  
Wang Qiancheng, The Algorithm Research of Fruit Image Recognition Based on Deep Learning[D]. Baoding: Hebei University, 2016. (in Chinese with English abstract)
- [27] Sa I, Ge Z, Dayoub F, et al. Deepfruits: A fruit detection system using deep neural networks[J]. *Sensors*, 2016, 16(8): 1–23.
- [28] 杨国国, 鲍一丹, 刘子毅. 基于图像显著性分析与卷积神经网络的茶园害虫定位与识别[J]. *农业工程学报*, 2017, 33(6): 156–162.  
Yang Guoguo, Bao Yidan, Liu Ziyi. Localization and recognition of pests in tea plantation based on image saliency analysis and convolutional neural network[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2017, 33(6): 156–162. (in Chinese with English abstract)
- [29] Vedaldi A, Lenc K. MatConvNet: Convolutional neural networks for MATLAB[C]// 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015: 689–692.
- [30] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324.
- [31] 李辉, 石波. 基于卷积神经网络的人脸识别算法[J]. *软件导刊*, 2017, 16(3): 26–29.
- [32] 宋怀波, 张卫园, 张欣欣, 等. 基于模糊集理论的苹果表面阴影去除方法[J]. *农业工程学报*, 2014, 30(3): 135–141.  
Song Huaibo, Zhang Weiyuan, Zhang Xinxin, et al. Shadow removal method of apples based on fuzzy set theory[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2014, 30(3): 135–141. (in Chinese with English abstract)

## Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks

Fu Longsheng<sup>1,2</sup>, Feng Yali<sup>1</sup>, Elkamil Tola<sup>3</sup>, Liu Zhihao<sup>1</sup>, Li Rui<sup>1</sup>, Cui Yongjie<sup>1,2</sup>

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China;

2. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture, Yangling 712100, China;

3. Precision Agriculture Research Chair, King Saud University, Riyadh 11451, Saudi Arabia)

**Abstract:** China is the largest country for cultivating kiwifruit, and Shaanxi Province provides the largest production, which accounts for approximately 70% of the production in China and 33% of the global production. Harvesting kiwifruit in this region relies mainly on manual picking which is labor-intensive. Therefore, the introduction of robotic harvesting is highly desirable and suitable. The fast and effective recognition of kiwifruit in the field under natural scenes is one of the key technologies for robotic harvesting. Recently, the study on kiwifruit recognition has been limited to a single cluster and multi clusters in the field have seldom been considered. In this paper, according to growth characteristics of kiwifruit grown on sturdy support structures, an RGB (red, green, blue) camera was placed around 100 cm underneath the canopy so that kiwifruit clusters could be included in the images. We proposed a kiwifruit image recognition system based on the convolutional neural network (CNN), which has a good robustness avoiding the subjectivity and limitation of the features selection by artificial means. The CNN could be trained end to end, from raw pixels to ultimate categories, and we optimized the critical structure parameters and the training strategy. Ultimately, the network was made up of 1 input layer, 3 convolutional layers, 2 sub-sampling layers, 1 full convolutional layer, and 1 output layer. The CNN architecture was optimized by using batch normalization (BN) method, which normalized the data distribution of the middle layer and the output data, accelerating the training convergence and reducing the training time. Therefore, the BN layers were added after the 1, 3 and 5th convolutional layer (Conv1, Conv3, and Conv5 layer) of the original LeNet network. The size of all convolutional kernels was  $5 \times 5$ , and that of all the sub-sampling layers was  $2 \times 2$ . The feature map numbers of Conv1, Conv3, and Conv5 were 6, 16 and 120, respectively. After manual selection and normalizing, the RGB image of kiwifruit was transferred into a matrix with the size of  $32 \times 32$  as the input of the network, stochastic gradient descent was used to train our models with mini-batch size of 100 examples, and momentum was set as 0.9. In addition, the CNN took advantages of the part connections, the weight sharing and Max pooling techniques to lower complexity and improve the training performance of the model simultaneously. The network used rectified linear units (ReLU) as activation function, which could greatly accelerate network convergence. The proposed model for training kiwifruit was represented as  $32 \times 32$ -6C-2S-16C-2S-120C-2. Finally, 100 images of kiwifruit in the field (including 5918 fruits) were used to test the model, and the results showed that the recognition ratios of occluded fruit, overlapped fruit, adjacent fruit and separated fruit were 78.97%, 83.11%, 91.01% and 94.78%, respectively. The overall recognition rate of the model reached 89.29%, and it only took 0.27 s in average to recognize a fruit. There was no overlap between the testing samples and the training samples, which indicated that the network had a high generalization performance, and the testing images were captured from 9 a.m. to 5 p.m., which indicated the network had a good robustness to lightness variations. However, some fruits were wrongly detected and undetected, which included the fruits occluded by branches or leaves, overlapped to each other and the ones under extremely strong sunlight. Particularly, 2 or more fruits overlapped were recognized as one fruit, which was the main reason to the success rate not very high. This phenomenon demands a further research. By comparing with the conventional methods, it suggested that the method proposed obtained a higher recognition rate and better speed, and especially it could simultaneously identify multi-cluster kiwifruit in the field, which provided significant support for multi-arm operation of harvesting robotic. It proves that the CNN has a great potential for recognition of fruits in the field.

**Keywords:** image processing; image recognition; algorithms; deep learning; convolutional neural network; kiwifruit