

基于迁移学习的蔬菜图像识别方法

赖佩霞, 王晓东*, 章联军

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要:为解决蔬菜识别领域缺少带标签样本的问题, 提出了一种基于迁移学习的图像识别方法. 首先, 将原始数据集利用数据增强扩大样本数据量后引入到大规模数据集上的预训练模型. 针对迁移过程中高层特征的领域特定性导致的网络泛化性能差, 通过加入两层自适应层参数初始化后重新训练得到基本模型; 对该基本模型再利用参数冻结的迁移方式进一步调优参数, 得到用于蔬菜图像识别的最终网络模型. 实验表明, 基于 CaffeNet 和 ResNet10 两个小型网络的迁移策略可以较好地处理小样本的蔬菜图像识别, 训练得到的模型准确率分别为 94.97%、96.69%. 与其他迁移算法及传统的神经网络方法相比, 该算法具有更高的识别性以及更强的鲁棒性.

关键词: 蔬菜图像识别; 卷积神经网络; 迁移学习; 小样本

中图分类号: TP391.4

文献标志码: A

文章编号: 1001-5132 (2019) 05-0036-06

随着经济的发展, 人们日益关注膳食均衡以及食品安全, 智能化管理蔬菜产品成为现代生活的发展趋势. 以往采用人工记忆输入、二维码或条形码扫描、无线射频识别^[1]等方式进行物体识别, 极大地耗费了人力和物力, 而智能化管理首先要解决高效识别技术的实施问题.

由于蔬菜识别领域缺乏公开标注数据集以供研究, 目前应用在小数据样本的识别主要有 2 种技术: (1)传统的机器学习算法^[2], 该方法依赖手工特征设计, 典型的有 IBM 公司根据提取统计直方图特征研发的 Veggie Version 果蔬分类系统^[3], 该系统识别性能虽有提升, 但易受条件(摆放位置或光照等)限制. (2)基于卷积神经网络的迁移学习方法. 迁移学习是利用已有的知识对不同但相关领域问题进行求解的一种新的机器学习方法^[4], 而卷积神经网络独特的仿生结构非常适用于处理图像视觉相关任务, 将两者相结合, 可以进一步发挥网络的语义概括能力. 如文献[5]在研究分析网络各层次表达性能后, 利用低层特征的通用性提出了层冻结方式进行网络微调, 获得了比源模型更好的准确率. 在此基础上, 文献[6]在修改源模型分类器

的同时进一步训练高层卷积块的参数, 但该方式在较深的网络结构训练时收敛速度较慢. 文献[7]将预训练后的最后一层全连层保留的权重和参数作为第二阶段微调的特征提取源, 有效地解决了小样本数据问题, 但是算法较为繁琐.

在现有算法基础上, 本文基于数据集 vegFru Dataset^[8]构建一个小型蔬菜库, 结合参数迁移方式提出一种识别方法. 利用卷积神经网络作为特征提取器, 并加入自适应层对迁移源模型的高层进行全局参数调整, 然后结合层冻结方式进一步微调, 得到最佳的识别模型. 实验结果表明, 本文提出的方法能在缺乏数据量的情况下学习到蔬菜图像的丰富特征, 取得比文献[6-7]更高的准确率.

1 基于小样本的蔬菜图像识别算法

1.1 卷积神经网络

卷积神经网络本质上可以看作是一种有监督的层次化特征学习方式, 输入原始图像, 通过连续的卷积层、池化层的交替堆叠, 自适应地学习图像的低层特征, 并在高层将这些特征抽象成具体的语义信息, 从而进行识别分类. 在该学习过程中,

收稿日期: 2018-12-11.

宁波大学学报(理工版)网址: <http://journal.lg.nbu.edu.cn/>

基金项目: 国家科技支撑计划项目(2012BAH67F01); 国家自然科学基金(U1301257); 浙江省自然科学基金(LY17F010005).

第一作者: 赖佩霞(1993-), 女, 浙江宁波人, 在读硕士研究生, 主要研究方向: 图像模式识别与深度学习. E-mail: jula199@163.com

*通信作者: 王晓东(1970-), 男, 浙江上虞人, 教授, 主要研究方向: 网络通信及计算机软件. E-mail: wxd@nbu.edu.cn

式(1)是关键的运算式:

$$X_j^l = f \left(\sum_{i \in M_j} X_i^{l-1} * K_{ij}^l + b_j^l \right), \quad (1)$$

其中, l 表示当前层; X_j^l 表示该层第 j 个卷积核的响应输出; f 为激活函数; K 表示不同的卷积核函数; M_j 表示前一层输入特征图的组合选择的集合, 每个输出的特征都有一个唯一的偏置 b_j^l .

权值共享的特点使得训练参数量只与卷积核种类有关, 因此响应输出为卷积核每次对局部区域操作后获取的与位置无关的小范围统计特征. 随着网络层次的加深, 卷积核的局部感受野也随之逐渐增大, 继而网络传递学习得到的特征也趋向全局化和抽象化.

由于卷积神经网络的训练需要海量数据的支撑, 在数据规模较小时, 直接应用网络模型会导致网络产生严重的过拟合现象. 如曾维亮等^[9]设计仅由 2 个卷积层组成的网络结构用于识别, 本质上网络规模和数据集都较小, 因此识别准确率不理想. 针对蔬菜样本稀少问题, 本文采用迁移学习方式, 将预训练模型及其特征分布迁移到本文网络, 以此提升识别性能.

1.2 迁移学习

迁移学习的初衷是节省样本标注时间, 随着卷积神经网络的兴起逐渐引起研究者的关注, 并得到广泛应用. 在医学图像、卫星图像等领域带标签的训练样本稀少的难题已取得了突破性的成果, 如 Shie 等^[10]基于 ImageNet 图像集以无监督学习方式获取图像中的关键元素, 并以该方式对中耳炎图像进行加权编码, 取得了比手工提取方式更显著的识别准确率. Xie 等^[11]通过卷积神经网络的训练, 获取遥感卫星拍摄的自然地形图像的灯光强度特征, 通过知识迁移成功应用于社会经济指标的预测中.

根据源域和目标域样本是否标注, 以及任务是否相同, 迁移学习主要分为: 归纳式迁移学习、直推式迁移学习以及无监督迁移学习等. 其中, 应用较广的为前 2 种. 归纳式迁移学习适用于 2 个领域相似, 并且任务相关但不同的情况. 如 Dai 等^[12]提出的 TradaBoost 算法, 利用 Boosting 技术滤除与目标域差异较大的源域样本后进行样本迁移来完成

新的学习任务; 而直推式迁移学习要求 2 个领域不同但任务相同, 如 Ando 等^[13]将用于结构学习的 SVD-ASO 算法的特征提取能力应用到包含大量无标签的汉字块数据域中, 通过三重训练数据选择方式提取目标数据最具预测性的低维特征.

尽管本文的蔬菜集与 ImageNet 的图像内容具有较大的差异, 然而边缘、纹理等这些抽象的低层特征属于不变的普适性特征, 是网络能够学习到的 2 个数据域存在的共性特征. 因此, 利用从源域中学习到的特征表示或模型参数迁移来进行蔬菜识别的研究.

1.3 算法设计

本文提出的算法主要由 2 个阶段构成, 训练流程如图 1 所示. 以数据增强后的蔬菜集作为输入, 常用的 CaffeNet 和 ResNet10 这 2 个小型网络作为迁移策略的源模型. 第一阶段去除源模型的最后分类层, 并扩展若干自适应层对源模型进行预训练, 引入小规模的网络参数随网络的连续反向传播算法(Back Propagation, BP)^[10]进行参数更新, 并保留扩展后的全连接模型作为第二阶段的特征提取源. 对训练好的基本模型, 第二阶段通过设置网络部分参数冻结进一步对模型微调, 使用 softmax 进行识别分类.

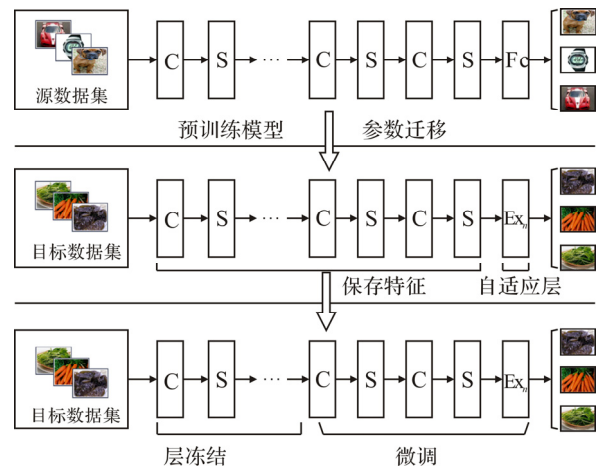


图1 算法训练流程

第一阶段中, 考虑到源域与蔬菜域的差异, 以及预训练模型的规模, 过多的扩展层反而会出现过拟合, 造成实验性能下降. 因此, 实验设计采取 2 层扩展层, 分别为 Ex_1 和 Ex_2 .

图2为基于 CaffeNet 的迁移模型, 微调时设置

较小的学习率或保持不变,将模型的 C1 至 Fc7 层特征迁移至蔬菜域中.针对迁移的高层特征的全局性以及领域特定性可能导致训练模型的泛化性能差,通过扩展层的自适应调整以充分获取蔬菜域图像信息.扩展层 Ex_1 为包含 512 个神经元的全连层, Ex_2 为 softmax 分类层,输出目标分类任务的类别数 15,这 2 层在重新训练时将参数设置为随机初始化,以此实现对最后一层全连层 Fc7 特征的微调.即 Fc7 输出向量 X_{Fc7} 输入到 Ex_1 层,得到的特征向量 X_{Ex_1} 作为 Ex_2 层的输入,最终获得分类输出结果.2 个自适应扩展层的计算公式为:

$$X_{Ex_1} = f(X_{Fc7} * K_{Ex_1} \Big|_p + B_{Ex_1}), \quad (2)$$

$$K_{Ex_1} \Big|_p = K_{Ex_1} \cdot p, \quad (3)$$

$$X_{Ex_2} = \psi(X_{Ex_1} * K_{Ex_2} + B_{Ex_2}), \quad (4)$$

式中:激活函数 f 是 ReLU 函数; ψ 是 softmax 预测函数; K_{Ex_1} 、 K_{Ex_2} 、 B_{Ex_1} 、 B_{Ex_2} 是网络训练中 2 个扩展层相应的卷积核参数以及偏置.

式(3)使用了 Dropout 技术,表示以概率 p 随机激活 Ex_1 层的神经元,可以在增加网络可训练参数的情况下保持网络结构的稀疏性,从而减少神经元之间的共同作用.

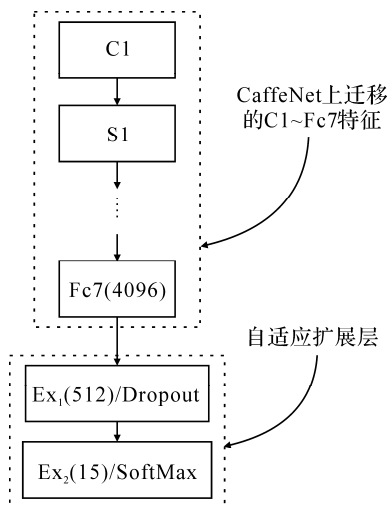


图2 基于 CaffeNet 的第一阶段模型

第二阶段主要是对第一阶段得到的基本模型的参数进行优化,使其更适用于本文的蔬菜集.在全连层中加载第一阶段的模型参数(包括矩阵权重和相应正则项系数、偏置项参数等),通过对部分高层设置较小的学习率 0.000 1 进行参数调整,提高

模型对蔬菜样本的全局概括能力.在训练过程中,当网络的损失率趋于稳定时将学习率再次调小,直至不再变化时获得本文算法的最优识别模型.

1.4 数据选取及预处理

实验选取的蔬菜图像均来自 vegFru Dataset,该数据集包含了食用菌类、根菜类、绿叶菜类等 15 个大类,符合日常饮食习惯.本文从中选取 15 个子类进行分类研究.选取的这 15 种类别图像均包含部分噪声干扰,如图 3 所示.如各式容器、人体、背景物体遮挡等非目标物体,属于复杂背景下的蔬菜识别.在训练前需要将数据统一归一化到 224×224 大小.



(a) 胡萝卜干扰实例 1

(b) 胡萝卜干扰实例 2



(c) 茄子干扰实例 1

(d) 茄子干扰实例 2

图3 部分蔬菜干扰样本

为减弱在小样本数据集上训练产生的过拟合,将数据集随机按 4:1 划分为训练集和验证集,然后采用数据增强方式进行样本扩增.在数据预处理中,对图像进行仿射变换、随机剪裁、随机旋转以保留图像的语义信息,并使用 HSV 扰动操作引入随机噪声来增强训练模型的泛化性能.将训练集扩大至每类 1 600 张,每个种类 400 张作为验证集,2 个数据集没有交叉部分.

由于蔬菜域没有权威数据集可供参考,为了验证算法的可行性,将国际上常用的 MNIST^[14]数据集作为对照数据集. MNIST 数据集是 Yann LeCun 等构建的包含 250 个人员手写数字的字体库,共 0~9 这 10 个数字,其中训练集和测试集分别有 60 000、10 000 个样本.实验中将数据还原成 png

格式图像, 每个类别分别取 1 600 和 400 张作为训练集和验证集.

注意到上述数据集的尺寸较小(MNIST 数据集 28×28), 采用训练数据集的像素平均值填充图像周围, 使其大小为 224×224 , 从而保证待识别原始图像位于识别中心, 避免尺寸过小造成图像信息丢失的问题.

1.5 实验设置

实验的开发环境是深度学习开源框架 Caffe^[15]. 迁移学习性能评价的标准为网络趋于稳定时验证集上的准确率. 网络的基本参数设置见表 1, 其中 2 个网络均使用随机梯度下降法(SGD)更新反向传播时的梯度和超参数. 模型的基础学习率设置为 0.001.

表 1 网络的基本参数设置

训练参数	参数设置
激活函数	ReLU
池化方式	AVE pooling, MAX pooling
优化算法	SGD, Momentum=0.9
Dropout	0.50, 0.39
最大迭代次数	50

2 实验结果与分析

2.1 特征可视化分析

图 4 为 CaffeNet 源模型迁移到目标蔬菜集学习

到的低层特征图和高层特征图. 对于输入的胡萝卜和空心菜蔬菜图像的第一层卷积(Conv1)后特征图和全连层前的池化层(Pool5)的特征图, 虽然 2 幅原始图像内容存在显著差异, 但是在 Conv1 的输出特征图中均能看出原蔬菜的大致轮廓, 而在高层, Pool5 的输出特征图已经十分抽象, 无法分辨特征图的内容. 低层卷积核和高层卷积核的局部细节信息表明, 卷积神经网络的低层特征抽象能力是可以共享的, 因此将 ImageNet 上训练好的成熟模型迁移到蔬菜识别上可以有效地提取图像的抽象特征, 从而提高模型的泛化能力.

2.2 MNIST 数据集上的实验结果与分析

在 MNIST 数据集上加载网络 ResNet10 以及使用本文算法, 得到的实验曲线如图 5 所示. 运行相同的迭代次数时, 仅加载 ResNet10 进行重新训练准确率提升速度缓慢, 且最高仅能达到 87.54%左右; 而使用本文算法的迁移技术引入预训练模型后, 由于模型通过第一阶段自适应快速训练, 已充分获取训练图像的语义信息, 网络的准确率在一个较高的平台上逐渐上升至 96%以上, 并且趋于稳定, 最终达到 97.53%. 虽然识别的准确率略低于文献[14]中给出的结果(98.8%), 但是本文极大地节省了时间开支, 且无需全部样本数据进行训练, 因此可以成功用于小数据样本上的识别. 通过在蔬菜库上与其他算法的对比, 可进一步验证本文算法的有效性.

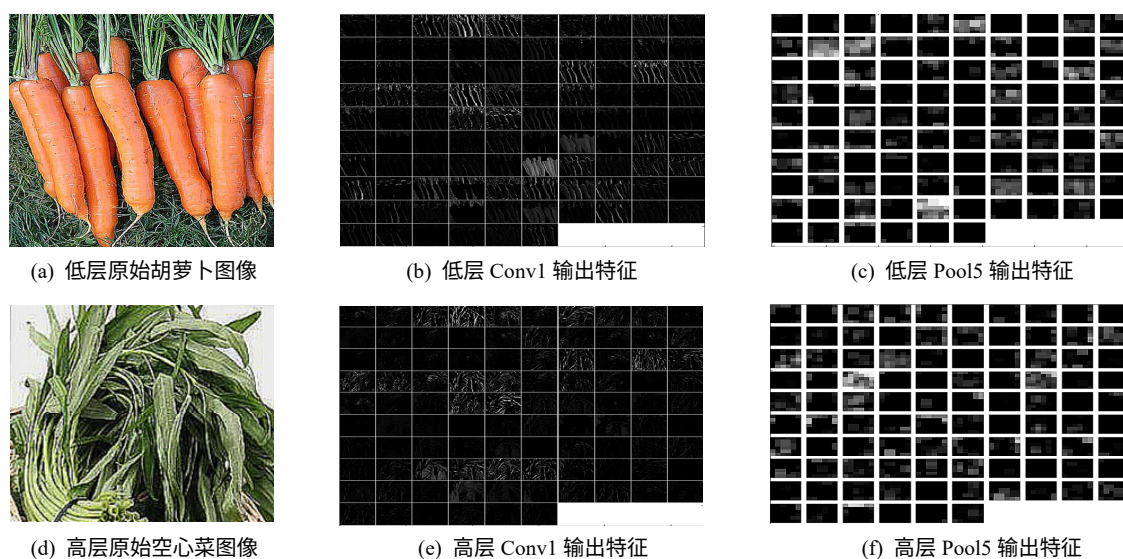


图 4 不同蔬菜图像以及相应低层和高层特征

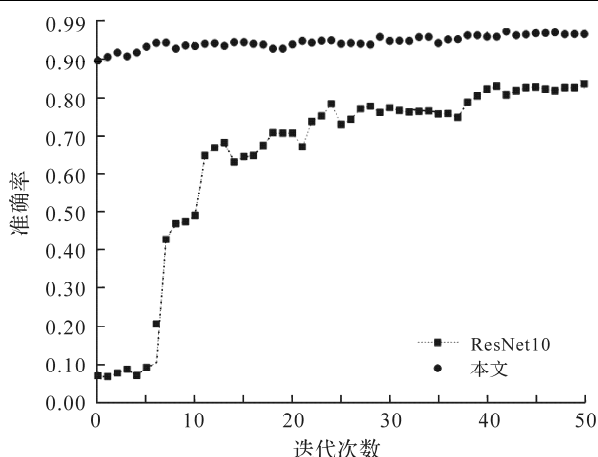


图5 ResNet10与本文算法结果对比

2.3 算法的对比分析

表2为各种算法的识别准确率对比结果,其中Raw-reTrain、Aug-reTrain为传统神经网络方法加载本文迁移策略采用的源网络结构从头开始训练。

表2 不同方法的性能对比 %

方法	CaffeNet	ResNet10
Raw-reTrain	54.21	56.93
Aug-reTrain	75.47	82.07
FT	93.93	96.20
	93.62	96.38
	94.15	95.72
	93.72	95.82
文献[6]	93.32	92.38
	92.97	-
	94.17	96.40
文献[7]	94.17	96.40
本文方法	94.97	96.69

注: Raw表示原始数据集; Aug表示数据增强后的样本; FT为仅更新分类器的迁移方式,为最简单的微调。

表2数据表明,神经网络方法应用在小型蔬菜库上,即使扩大数据样本后也不能取得很好的效果,说明深度网络对数据有极强的依赖性。而使用迁移策略后,即使是最简单的微调,也能取得远远优于神经网络方法的准确率,说明特征迁移能较好地处理小数据样本。ResNet10在验证集上的准确率达到96.69%,比CaffeNet提高了1.72个百分点,这是由于CaffeNet的网络训练参数少于ResNet10,网络的修正能力较弱。

表2中文献[6]数据的两列分别为依次增加CaffeNet卷积层的冻结层以及ResNet10残差块的

冻结层的实验性能。对于CaffeNet,冻结2个卷积模块时达到最佳识别率(94.15%),而对于ResNet10,低层残差块具有较强的泛化能力,冻结第一个残差块时有最佳准确率(96.38%)。之后冻结参数再逐渐增加时,网络的性能会有所下降,冻结全部卷积层参数时性能最差,说明高层特征能高度适应ImageNet源数据集,这与文献[5]的观点一致,可以直接迁移到目标域的是网络的通用性低层特征,而高层特征的特定性使得增加自适应调整层十分必要。同时将本文方法与文献[7]的方法进行对比,也证明本文方法由于自适应层的训练学习减弱了源模型高层的领域特定性,在引入较小训练参数的基础上进一步提升了识别准确率,且有效地避免了过拟合。

卷积神经网络虽然能够自主获取图像像素上的本质特征表示,然而重新训练一个网络由于缺乏海量数据的支撑,以及各网络层参数的精心设计,导致网络的识别准确率较差。本文的方法相比其他几种迁移学习方法,其识别准确率最高,鲁棒性最强,能够用于基于小数据集的蔬菜识别。

3 结语

本文将大规模数据集上充分训练得到的小型网络模型引入数据扩充后的蔬菜库中,在修改源模型结构的基础上,提出了一种基于小样本的蔬菜识别方法。该方法比简单的微调,以及参数冻结并训练分类器方式的网络学习特征更丰富。而与迁移全部高层卷积特征方式相比,自适应层的训练使得网络特征提取能力有了较大程度的提升。实验表明,即使是网络规模较小的浅层模型,其预训练模型也有强大的泛化能力和移植能力,可以完成蔬菜领域数据稀少的识别任务,并且具有较好的鲁棒性。

参考文献:

- [1] Gu H, Wang D. A content-aware fridge based on RFID in smart home for home-healthcare[EB/OL]. [2018-08-28]. <https://ieeexplore.ieee.org/document/4809580>.
- [2] Sharma A, Paliwal K K. Linear discriminant analysis for the small sample size problem: An overview[J]. International Journal of Machine Learning & Cybernetics, 2014, 6(3):443-454.

- [3] Bolle R M, Connell J H, Haas N, et al. Veggievision: A produce recognition system[EB/OL]. [2018-08-28]. <https://ieeexplore.ieee.org/document/572062?arnumber=572062&tag=1>.
- [4] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1):26-39.
- [5] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks?[EB/OL]. [2018-08-28]. https://www.researchgate.net/publication/268079628_How_transferable_are_features_in_deep_neural_networks.
- [6] 王柯力, 袁红春. 基于迁移学习的水产动物图像识别方法[J]. 计算机应用, 2018, 38(5):1304-1308.
- [7] 段萌, 王功鹏, 牛常勇. 基于卷积神经网络的小样本图像识别方法[J]. 计算机工程与设计, 2018, 39(1):224-229.
- [8] Hou S, Feng Y, Wang Z. VegFru: A domain-specific dataset for fine-grained visual categorization[EB/OL]. [2018-08-28]. <https://ieeexplore.ieee.org/document/8237328>.
- [9] 曾维亮, 林志贤, 陈永洒. 基于卷积神经网络的智能冰箱果蔬图像识别的研究[J]. 微型机与应用, 2017, 36(8): 56-59.
- [10] Shie C K, Chuang C H, Chou C N, et al. Transfer representation learning for medical image analysis[EB/OL]. [2018-08-28]. https://www.researchgate.net/publication/308298066_Transfer_representation_learning_for_medical_image_analysis.
- [11] Xie M, Jean N, Burke M, et al. Transfer learning from deep features for remote sensing and poverty mapping[EB/OL]. [2018-08-28]. <https://dl.acm.org/citation.cfm?id=3016457>.
- [12] Dai W, Yang Q, Xue G R, et al. Boosting for transfer learning[EB/OL]. [2018-08-28]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.76.7577>.
- [13] Ando R K, Zhang T. A high-performance semi-supervised learning method for text chunking[EB/OL]. [2018-08-28]. <https://dl.acm.org/citation.cfm?id=1219841>.
- [14] Mohapatra R K. Majhirning in feedforward neural network[J]. International Journal of Multimedia & Ubiquitous Engineering, 2014, 9(8):149-156.
- [15] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding [EB/OL]. [2018-08-28]. <https://arxiv.org/abs/1408.5093>.

Vegetable image recognition based on transfer learning

LAI Peixia, WANG Xiaodong*, ZHANG Lianjun

(Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: To solve the problem of lacking the labeled samples in vegetable recognition domain, a method for image recognition based transfer learning is proposed. Firstly, raw data are expanded by the data augmentation technique, then pre-trained models on the large-scale data sets are introduced to the target data set to train for the base model, where initializing two additional adaptive layers is performed. This step is aimed to solve the poor generalization performance caused by the transferred domain specific high-level features. Next, the final model is obtained for recognizing vegetable images by adopting the fine-tuning method with several layers frozen. Experimental results show that, based on CaffeNet and ResNet10, the proposed transfer approach reaches the accuracy of 94.97%, 96.69%, respectively, and can effectively process small samples in vegetable image recognition with higher accuracy and better robustness comparing to other transferring algorithms and the conventional convolution neural networks.

Key words: vegetable image recognition; convolution neural network; transfer learning; small samples

(责任编辑 史小丽)