

集成的卷积神经网络在智能冰箱果蔬识别中的应用

李思雯 吕建成 倪胜巧

(四川大学计算机学院, 成都, 610065)

摘要: 冰箱内物体繁多且摆放随意, 给基于图像分析的冰箱物体识别带来了很大挑战。本文提出一种用集成的卷积神经网络方法来解决冰箱食物种类识别问题。其基本思想是首先分别训练两个卷积神经网络, 一个用于果蔬种类识别, 一个用于果蔬的颜色识别, 然后用一个多层感知器将两个独立的网络集成进行分类训练。集成训练之后的模型能将两个网络的信息进行补偿和强化。本文方法能有效提升颜色在物体识别中的主导作用, 改善了由于遮挡、视角变化导致识别准确性不高的问题。最后通过对从冰箱获取大量真实的图片数据进行实验, 验证了本文方法在解决智能冰箱物体识别问题的有效性。

关键词: 卷积神经网络; 智能冰箱; 物体识别; 深度学习; 集成网络

中图分类号: TP18 文献标志码: A

Integrated Convolutional Neural Network and Its Application in Fruits and Vegetables Recognition of Intelligent Refrigerator

Li Siwen, Lü Jiancheng, Ni Shengqiao

(College of Computer Science, Sichuan University, Chengdu, 610065, China)

Abstract: As an important part of the household appliances, the refrigerator becomes more intelligent. Object recognition of the food in a refrigerator is a key technology of a smart refrigerator. However, the foods in the refrigerator are diverse and disordered, which brings a lot of challenges to identify the varieties of foods. A method using an integrated convolutional neural network is proposed to solve this problem. The basic idea is that two convolutional neural networks are firstly trained separately. One is used to recognize the kinds of fruits and vegetables, the other is to recognize the color of them. Then, a multi-layer perceptron is used to integrate the two independent networks to carry out classification. The two separate convolutional neural networks can complement and improve each other in the integrated network. In the method, color information, an important feature in the recognition, can be enhanced. The proposed structure also improves the recognition rate which is influenced by object occlusion and view variations. Finally, the effectiveness of the proposed method is validated on a dataset which contains a large amount of images obtained from a real situation of a refrigerator.

Key words: convolutional neural network; intelligent refrigerator; object recognition; deep learning; integrated network

引言

智能冰箱的许多关键功能都依赖于对冰箱的食物种类进行有效识别。目前智能冰箱采用的物体识别方法主要有^[1]: 用户手动或语音输入食物种类方法、无线射频识别(Radio frequency identification, RFID)扫描方法^[2]、条形码或二维码扫描方法^[3-4]以及图像识别技术方法^[5]。前3种方法均需用户人为配合,过程繁琐,实用性较差。图像识别技术无需用户参与,极大方便了用户,但是识别难度较大。冰箱内存储的食物主要是以果蔬为主。为解决基于图像分析的果蔬识别问题,国内外学者已经提出了许多方法^[6-9]。Bolle等^[6]在1996年首次提出Veggie vision的概念,并开发了一个超市智能果蔬识别系统。该系统通过提取图片的颜色、纹理和密度特征可以识别任意数量、任意摆放的果蔬,但该方法对光照变化较为敏感。Arivazhagan等^[7]提出将色调、饱和度和亮度(Hue, Saturation, Value, 即HSV)颜色模型中的HS直方图作为颜色统计特征,将V成分用离散小波变换分解,并用其子波段构造自相关矩阵作为纹理特征进行果蔬识别。Tao^[8]提出一种颜色完全局部二值模式纹理特征提取算法,采用HSV颜色直方图和外点/内点直方图提取图像颜色特征,通过最近邻分类器实现果蔬农产品分类,但是其实验采用的图片果蔬仅考虑了光照强度变化因素。Paul等^[9]提出一种3D颜色直方图结合卷积神经网络的方法,对视角和光照变化的果蔬进行分类,但是其实验图片物体数目多数仅一个,而且多数视角变化较不明显。Fasel^[10]提出一种对卷积神经网络进行集成训练的方法,并在人脸识别上取得了一定的效果。

图1给出了一张从冰箱获取的典型真实图片。由图可见,冰箱内果蔬种类繁多,摆放随意。果蔬视角多变,物体间容易相互遮挡,这给智能冰箱果蔬种类识别带来了很大挑战。近几年,卷积神经网络在图像识别等领域备受关注。它可以直接输入二维图像并隐式地从训练数据中学习数据从低层到高层的本质特征,具有旋转、放缩、平移和扭曲不变性。深层卷积神经网络具有十分强大的学习能力,在大型图像数据集等复杂图像识别上取得了显著的成果^[11-13]。本文提出一种用集成训练的卷积神经网络的方法来解决冰箱果蔬种类识别问题。



图1 一张典型的从冰箱获取的真实图片
Fig. 1 A typical real image from a refrigerator

1 卷积神经网络与空间金字塔池化

1.1 卷积神经网络

卷积神经网络是由Lecun在1989年提出的一种网络结构^[14],它最初被应用于手写数字识别并取得了非常好的效果。卷积神经网络来源于猫的视觉皮层,它的卷积和池化的思想是对视觉皮层中S细胞和C细胞的模拟^[15]。卷积神经网络将权值共享、空间下采样和局部连接结合起来,不仅降低了网络的复杂性,而且保持了对图像的边缘模式信息和空间位置信息强大的检测能力。此外,卷积神经网络可以把2维图片直接作为网络的输入,解决了传统方法中需要人工提取图像特征进行训练等问题。一个典型的卷积神经网络通常由输入层、卷积层、池化层、全连接层和输出层组成,其中卷积层和池化层一般交替出现多次。

卷积层是卷积神经网络从输入图像进行特征抽取时非常重要的一部分。卷积层的输出定义如下

$$\mathbf{A}_j = f\left(\sum_{i=1}^N \mathbf{I}_i \times \mathbf{K}_{i,j} + \mathbf{B}_j\right) \quad (1)$$

式中: I_i 代表输入矩阵; $K_{i,j}$ 代表与 I_i 相应的卷积核矩阵; B_j 代表与 I_i 相应的偏置矩阵; f 代表一个非线性函数; A_j 表示卷积层的输出矩阵。

每个卷积层都拥有一系列卷积核,这些卷积核对前一层的输入进行卷积运算得到相应的特征图。通常,第 1 个卷积层会从输入图像中学习水平或垂直的边、环和颜色信息等低级特征,后面的层将逐渐学习到关于物体的角点、形状等高级特征。

1.2 空间金字塔池化

池化层是卷积神经网络中非常重要的一部分,它主要用于扩大感受野,保持特征的空间位置关系。池化的主要方法有:最大池化、平均池化和空间金字塔池化。空间金字塔池化引入空间金字塔尺度划分思想,按等级对输入的特征图进行池化^[16]。假设从卷积层输入空间金字塔池化层的特征图尺寸为 $a \times a$,金字塔划分的 bins 为 $n \times n$,那么空间金字塔池化层的池化窗口大小(win)和池化步长(str)的计算关系式分别为: $\text{win} = \left\lceil \frac{a}{n} \right\rceil$

$\text{str} = \left\lfloor \frac{a}{n} \right\rfloor$ 。图 2 给出了一个池化尺寸为 $\{3 \times 3, 2 \times 2, 1 \times 1\}$ 的空间金字塔池化方法。图中的每个特征图经过该金字塔池化后仅向第一个全连接层输入一个 14 维向量,该 14 维向量由 $3 \times 3 + 2 \times 2 + 1 \times 1$ 计算得到。空间金字塔池化方法可以保证池化层向全连接层输入固定尺寸的特征向量,这不仅有效减少输入第一个全连接层的参数个数,而且使得卷积神经网络将任意尺寸的图片作为输入,有利于保持图像尺寸比例,提升整个网络的准确率。在采用金字塔池化时,特征图的划分尺度与图像的尺寸保持一定的比例,使得整个网络对物体形变和空间分布的鲁棒性增强,能够更好地保持特征的空间结构关系。

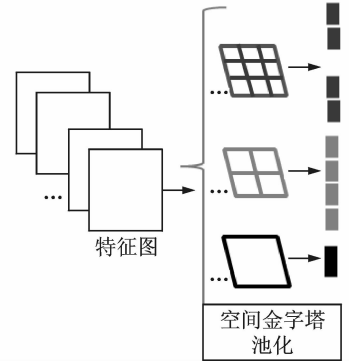


图 2 池化尺寸为 $\{3 \times 3, 2 \times 2, 1 \times 1\}$ 的空间金字塔池化层

Fig. 2 A spatial pyramid pooling layer with pooling size $\{3 \times 3, 2 \times 2, 1 \times 1\}$

2 基于卷积神经网络的识别模型

2.1 集成的网络模型

冰箱内果蔬识别容易受到视角变化、物体之间相互遮挡和图片背景等因素的影响。先前的实验发现在用训练好的单个卷积神经网络对测试集进行测试时,分类结果中的果蔬种类和目标果蔬的颜色不相关的情形很常见。但当减少有遮挡的图片和视角变化的图片后,整个网络的识别准确率有所提升。颜色作为果蔬的重要特征,它对视角变化和物体间遮挡具有较强的鲁棒性。为此,考虑强化水果颜色对水果种类识别的作用以减少视角和遮挡对果蔬识别的影响。本文采取了一种两个卷积神经网络集成学习的方法来强化颜色信息对果蔬识别的影响。首先,分别构建两个单独的卷积神经网络,一个用于识别水果的种类,另一个用于识别水果的颜色,然后用一个多层感知器将两个网络集成训练。

图 3 给出了本文所采用的单独的网络结构。该结构将 Zeiler 网络^[17]的第 5 个卷积层后的池化层替换为一个空间金字塔池化层。表 1 给出了该网络每层的具体参数。其中, 96×11^2 表示卷积核个数是 96,卷积核尺寸是 11^2 ,卷积核移动的步长为 4,最大池化的池化窗口为 3×3 ,池化窗口的步长为 2,网络采用局部响应归一化方法,即对一个局部的输入区域进行的归一化。空间金字塔池化 $\{4 \times 4, 3 \times 3, 2 \times 2, 1 \times 1\}$ 表示池化尺寸为 $\{4 \times 4, 3 \times 3, 2 \times 2, 1 \times 1\}$ 的空间金字塔池化。此外,网络模型中使用 pad 参数,通过在输入的每一边加上相应个数的像素以得到预期大小的特征图。网络采用 ReLU 激活函数,其计算公式为: $f(x) = \max(0, x)$ 。网络的权值更新基于带动量因子的梯度下降的反向传播算法,权重更

新规则如下

$$v_{i+1} = \alpha v_i - \lambda \eta w_i - \eta \left(\frac{\partial L}{\partial w} \Big|_{w_i} \right)_{D_i} \tag{2}$$

$$w_{i+1} = w_i + v_{i+1} \tag{3}$$

式中： i 为迭代次数； η 为学习率； α 为动量系数； λ 为权重衰减率； w 表示权重向量； $\left(\frac{\partial L}{\partial w} \Big|_{w_i} \right)_{D_i}$ 为第*i*批训练样本 D_i 的损失函数在权值为 w_i 时关于 w 的导数的平均值。

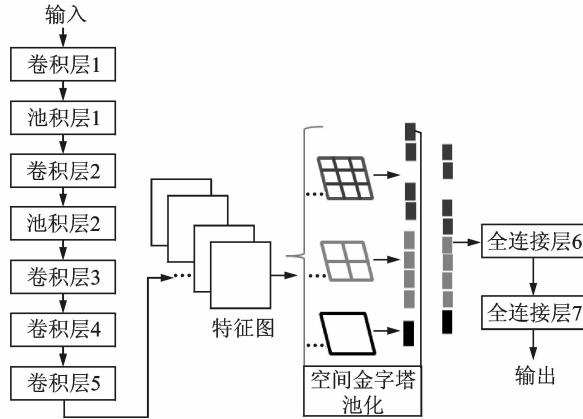


图 3 单个卷积神经网络结构

Fig. 3 Single convolutional neural network structure

表 1 单个网络结构的各层参数

Tab. 1 Parameters of layers in the single network model

卷积层 1	卷积层 2	卷积层 3	卷积层 4	卷积层 5	全连接层 6	全连接层 7	输出
96×11^2 , 步长 4 池化层 1: 3^2 , 步长 2; 局部响应 归一化	256×5^2 池化层 2: 3^2 , 步长 2; 局部响应 归一化	384×3^2	384×3^2	256×3^2 空间金字塔池化 $\{4 \times 4, 3 \times 3, 2 \times 2, 1 \times 1\}$	4 096 单元	4 096 单元	14 或者 5 Softmax 分类

本文采用的集成网络中的多层感知器含有 50 个隐层单元, 该多层感知器的输入为两个独立的卷积神经网络的输出层神经元拼接后的向量。该集成后的网络结构如图 4 所示。

2.2 基于参数迁移的网络初始化

深度卷积神经网络在检测和识别物体方面显现出超出手工特征提取方法的巨大性能, 但它的强大性能与可用的大规模数据集、强大的计算性能以及有效率的算法等密不可分。参数迁移^[18]是利用现有的预先训练好的具有强大学习能力的网络模型参数初始化某个小型训练集模型参数。这种参数初始化方法可以将已有的强大的学习能力迁移到另一个网络中, 使得新网络具有快速学习能力。更重要的是, 当训练的数据集规模不够大时, 该方法能也保证网络具有较强的泛化能力。

AlexNet 模型^[11]是目前较好的深度卷积神经网络模型之一。bvlc_reference_caffent 与 AlexNet 模型除了池化层和归一化层顺序倒置、没有通过亮度变化对数据集进行扩大外没有其他区别。现有的在大规模数据集 ImageNet^[11]训练好的 bvlc_reference_caffent 模型具有强大的学习能力。本文用该模型的前 4 个卷积层分别初始化水果种类识别网络 and 水果颜色识别网络的对应层。图 5 给出了参数迁移过程。待水果种类识别网络和水果颜色识别网络训练结果趋于稳定, 再抽取这两个模型的参数去初始化集成的网络模型。

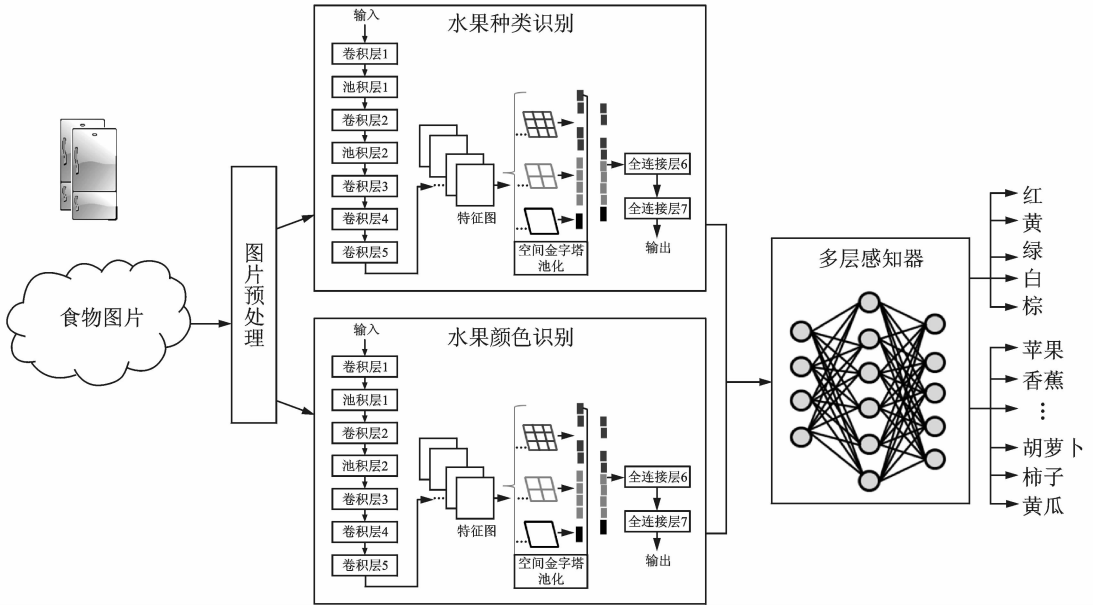


图 4 集成的网络模型
Fig. 4 Integrated network model

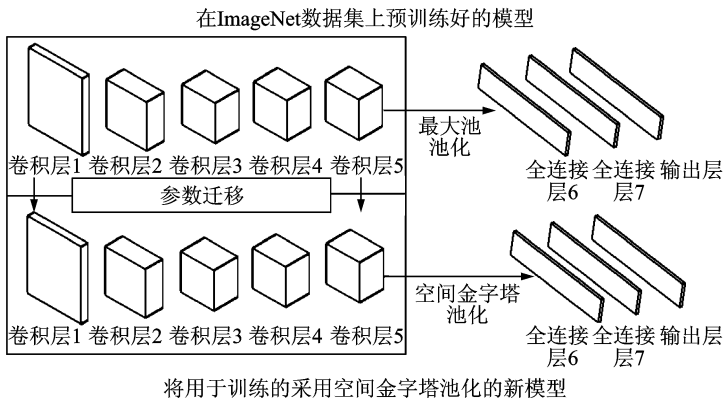


图 5 参数迁移过程
Fig. 5 Process of parameters transferring

3 实验及结果分析

3.1 数据集

从冰箱获取的图片通常同时包含多种不同的果蔬,先用 Selective search 方法^[19]检测并定位具体果蔬,然后将定位后的图片作为数据集的一个重要组成部分。此外,为了避免因采集冰箱图片时同一果蔬种类样本单一导致难以学到某种果蔬通用特征的情况,从互联网上采集了一些相应果蔬种类的图片,与冰箱获取的图像数据混合在一起作为训练集,以增加数据集的通用性和鲁棒性。

本文采用的数据集包含 14 种冰箱中常见果蔬,分别为苹果、香蕉、皇冠梨、橙子、菠萝、火龙果、猕猴桃、花菜、黄瓜、尖椒、柿子、甜椒、玉米和胡萝卜。由于图片采集于真实的冰箱环境,所以物体视角多变

和遮挡情况很常见。图 6 给出了数据集中部分图片样本,其中每一行的前 3 张图片是从冰箱拍摄的图片分割得来的,其余是从网上爬取搜集得到。数据集中每种食物照片约为 230~270 张,其中冰箱实际拍摄的图片与网上搜集的图片的比例约为 2:1,所有食物图片总共 3 685 张。

3.2 实验结果比较与分析

实验中采用的单个网络结构如表 1 所示,采用的集成网络结构如图 4 所示。采用参数迁移的集成网络中第一层学到的特征如图 7 所示。分割后的图片按照长宽比例进行重新调整,使得图片长或宽的最小值为 256,然后输入网络进行训练。在进行果蔬颜色识别训练时,先将数据集中果蔬依照自身主颜色分为 5 类。将苹果、火龙果、胡萝卜、柿子和红甜椒依据颜色划分为红类;将香蕉、皇冠梨、橙子、菠萝和玉米划分为黄类;将黄瓜、尖椒划分为绿类;将猕猴桃划分为棕类;将花菜划分为白类。此外,本文采用上文提到的参数迁移方法对网络进行初始化。实验中,本文采用二重交叉验证评估本文提出的方法。本文将每个类别的图片分别进行排序,奇数序列作为训练集,其余作为验证集。本文分别对未采用迁移学习的单个网络、采用参数迁移的单个网络和采用参数迁移的集成网络 3 种情形进行训练,其识别率如图 8 所示。单个卷积神经网络和集成网络采用分段学习率训练方式,初始学习率为 0.001,待网络准确率没有显著上升,再将网络的学习率改变为原先学习率的 1/10,以此类推,待网络准确率趋于稳定为止。所有网络的动量系数均为 0.9,权重衰减率均为 0.005。



图 6 数据集中的一些图片实例

Fig. 6 Some sample images in the dataset

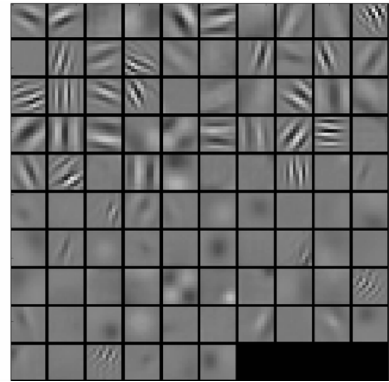


图 7 参数迁移的集成网络的第一层特征

Fig. 7 Features for the first layer of the integrated network with transferred parameters

对比表 2 中实验(1)与(3)、(2)与(4)以及表 3 中实验(1)与(3)、(2)与(4)可知,采用参数迁移进行模型初始化能够有效提升果蔬识别的准确率。从图 8 曲线走势也可以看出,采用参数迁移进行学习可以加快网络训练,使得网络有更加泛化的识别能力,而且在学习复杂特征时更为显著。

表 2 单一网络情形下各种算法准确率比较

Tab. 2 Comparison of the accuracy of different algorithms in the single network circumstance

网络结构	平均准确率/%
(1) 单一网络(种类)+未迁移训练	73.7
(2) 单一网络(颜色)+未迁移训练	88.2
(3) 单一网络(种类)+迁移训练	90.7
(4) 单一网络(颜色)+迁移训练	92.25

表 3 集成网络情形下各种算法准确率比较

Tab. 3 Comparison of the accuracy of different algorithms in the integrated network circumstance

网络结构	平均准确率/%
(1) 集成网络(种类)+未迁移训练	78.9
(2) 集成网络(颜色)+未迁移训练	88.01
(3) 集成网络(种类)+迁移训练	91.67
(4) 集成网络(颜色)+迁移训练	92.27

从表 2 中实验(3)和表 3 中实验(3)可知,集成的卷积神经网络比单一网络在冰箱果蔬种类的识别率上有一定幅度的提升。对比图 9,10 可知,图 10 中集成的卷积神经网络对果蔬种类识别的准确率比图 9 中的准确率大部分均有提高,尤其是苹果、黄瓜和尖椒被正确分类的概率明显提升。此外,苹果、花菜、黄瓜、火龙果、胡萝卜、尖椒和猕猴桃被误分为皇冠梨的概率下降幅度都在 6%及以上。由此可见,集成的两个卷积神经网络可以进行信息补偿和强化。集成网络可以通过强化颜色信息对果蔬种类识别的结果产生影响,从而提升了卷积神经网络在果蔬种类识别的准确率。

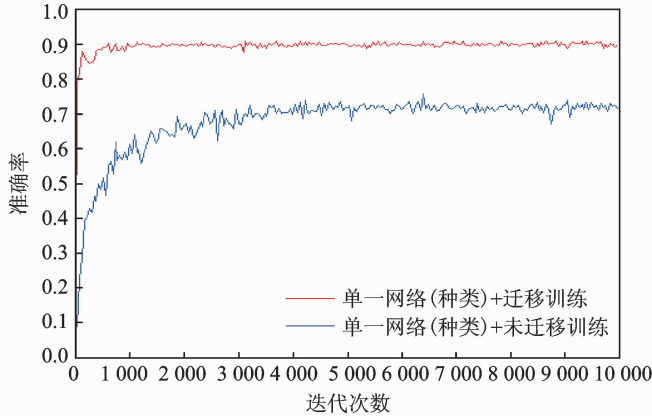


图 8 两种训练方法下的果蔬种类识别率

Fig. 8 Recognition rate of fruits and vegetables under two training situations

Confusion matrix for the single network (species) without transferred parameters. The table shows classification results for 15 categories: 菜果, 香蕉, 菠萝, 橙子, 花菜, 黄瓜, 火龙果, 胡萝卜, 尖椒, 猕猴桃, 皇冠梨, 柿子, 玉米, 甜椒. The diagonal elements represent correct classifications, with values ranging from 0.32 to 1.00.

图 9 单一网络(种类)在未迁移训练时的混淆矩阵

Fig. 9 Confusion matrix for the single network (species) without transferred parameters

Confusion matrix for the integrated network (species) without transferred parameters. The table shows classification results for the same 15 categories as Figure 9. The diagonal elements represent correct classifications, with values ranging from 0.65 to 1.00, indicating improved performance compared to the single network.

图 10 集成网络(种类)在未迁移训练时的混淆矩阵

Fig. 10 Confusion matrix for the integrated network (species) without transferred parameters

4 结束语

本文将两个独立的卷积神经网络进行集成训练,集成的模型能对两个网络各自拥有的信息进行补偿和强化,可以使得冰箱果蔬种类的识别率有一定提升。此外,采用参数迁移进行网络初始化,不仅加快了网络训练的速度,而且使得网络有更加泛化的识别能力,避免因为样本集较小过拟合。未来工作将专注于研究果蔬的成熟程度对果蔬识别的影响以及如何有效对现有网络模型进行加速和扩展训练以适应更多的果蔬分类任务等。

参考文献:

- [1] Sandholm T, Lee D, Tegelund B, et al. CloudFridge: A testbed for smart fridge interactions[J]. arXiv:1401.0585, 2014.
- [2] Gu Hanshen, Wang Dong. A content-aware fridge based on RFID in smart home for home-healthcare [C]//Proc of the 11th International Conference on Advanced Communications Technology. Beijing: IEEE, 2009: 987-990.
- [3] 戴扬, 于盛林. 二维条形码编码与译码的计算机实现[J]. 数据采集与处理, 2003, 18(3): 356-360.
- [4] Dai Yang, Yu Shenglin. Computer encoding and decoding of two-dimensional bar codes [J]. Journal of Data Acquisition and Processing, 2003, 18(3): 356-360.
- [5] 祁文娟. 基于物联网技术的智能家电管理模型设计与验证[D]. 青岛: 中国海洋大学, 2013.
- [6] Qi Wenjuan. A design and validation on management model of smart household appliances based on the Internet of things [D]. Qingdao: Ocean University of China, 2012.
- [7] 李英. 基于图像处理技术的冰箱食物管理系统研究[D]. 大连: 大连理工大学, 2012.
- [8] Li Ying. A research on a management system of goods in the refrigerator based on image technology [D]. Dalian: Dalian University of Technology, 2012.
- [9] Bolle R M, Connell J H, Haas N, et al. Veggievision: A produce recognition system [C]// WACV96. Sarasota, USA: IEEE, 1996: 1-8.
- [10] Arivazhagan S, Shebiah R N, Nidhyandhan S S, et al. Fruit recognition using color and texture features[J]. Journal of Emerging Trends in Computing and Information Sciences, 2010, 1(2): 90-94.
- [11] Tao Huawei, Zhao Li, Xi Ji, et al. Fruits and vegetables recognition based on color and texture features[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(16): 305-311.
- [12] Earnest Paul Ijjina, Krishna Mohan C. View and illumination invariant object classification based on 3D color histogram using convolutional neural networks [C] //Proc of the 12th Asian Conference on Computer Vision. New York: Springer International Publishing, 2015: 316-327.
- [13] Fasel B. Robust face analysis using convolutional neural networks [C]//Proc of the 16th International Conference on Pattern Recognition. Canada: IEEE, 2002: 11-15.
- [14] Krizhevsky A, Sutskever I, Hinton G E. Image net classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. US: MIT Press, 2012: 1097-1105.
- [15] Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks [J]. ArXiv Preprint arXiv:1312.6229, 2013.
- [16] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [J]. arXiv Preprint arXiv:1409.4842, 2014.
- [17] Lecun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
- [18] Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex [J]. The Journal of Physiology, 1959, 148(3): 574-591.
- [19] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [C]// Computer Vision-ECCV 2014. New York: Springer International Publishing, 2014: 346-361.
- [20] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [C]// Computer Vision-ECCV 2014. New York: Springer International Publishing, 2014: 818-833.
- [21] Wei Y, Xia W, Huang J, et al. CNN: Single-label to multi-label [J]. ArXiv Preprint arXiv:1406.5726, 2014.
- [22] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

作者简介:



李思雯(1990-),女,硕士研究生,研究方向:机器学习,深度学习, E-mail: lswinscu@163.com.



吕建成(1973-),男,教授,博士生导师,研究方向:机器学习,神经网络。



倪胜巧(1982-),男,博士研究生,研究方向:机器学习,数据挖掘,数据降维。

