

# 基于深度神经网络的货架商品识别方法

刘照邦, 袁明辉

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** **目的** 为快速统计货架商品信息, 提出一种基于深度神经网络的货架商品自动识别方法。**方法** 摄像头采集的货架商品图像经过深度神经网络算法处理, 得到了图像中商品的 SKU 和位置。针对货架商品识别这种密集检测场景, 文中方法改进了通用深度神经网络目标检测算法, 将算法分为检测和分类 2 个阶段且重新设计了部分网络结构。最后, 将文中方法和传统货架商品识别方法以及通用深度神经网络目标检测方法进行了比较。**结果** 实验证明该方法的检测阶段的模型平均正确率达到 96.5%, 分类阶段的分类准确率达到 99.9%。整图测试的查准率为 97.56%, 查全率为 99.26%。**结论** 相较于以往使用传统的目标检测模型进行货架商品识别以及使用 SIFT 等人工算子提取特征并分类识别商品具体 SKU, 文中方法的商品检出率和分类准确率都有了大幅度的提升, 具有很好的应用潜力。

**关键词:** 货架商品识别; 深度神经网络; 目标检测; 图像分类; 存货单位

中图分类号: TP312 文献标识码: A 文章编号: 1001-3563(2020)01-0149-07

DOI: 10.19554/j.cnki.1001-3563.2020.01.023

## Product Recognition on Shelves Based on Deep Neural Network

LIU Zhao-bang, YUAN Ming-hui

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**ABSTRACT:** The work aims to fast count the product information on shelves, and propose an automatic recognition method of shelf products based on deep neural network. The image of the shelf products collected by the camera was processed by the deep neural network algorithm to obtain the SKU and position of the products in the image. Aiming at the dense detection scenario of shelf product recognition, this method improved the general deep neural network object detection algorithm: the algorithm was divided into two stages of detection and classification and a part of the network structure was redesigned. At the same time, this method was compared with the traditional shelf product recognition methods and the general deep neural network object detection methods. From the experiment results, the average precision of the model reached 96.5% in the detection stage and 99% in the classification stage. In whole image test, the precision was 98.17% and the recall was 97.05%. Compared with prior works by traditional object detection methods for product recognition on shelves or SIFT artificial operators to extract features and classify product SKU, this method greatly improves the detection rate and classification precision rate, which has good application potential.

**KEY WORDS:** product recognition on shelves; deep neural network; object detection; classification; Stock Keeping Unit

收稿日期: 2019-04-27

作者简介: 刘照邦 (1992—), 男, 上海理工大学硕士生, 主攻深度学习图像处理。

货架商品信息统计对于超市具有重要的作用,它能够帮助超市得到铺货率、货架占有率、排面数、促销执行度等关键指标。长期以来,货架商品信息统计都是由人工来完成,效率低下,成本高昂,并且容易出错。以上海某超市为例,共有货架 126 个,需要一名超市员工 3 d 才能盘点完毕。

近年来,随着图像处理技术的飞速发展。利用图像采集和算法处理自动化地获取货架商品信息成为图像处理技术与零售业结合的一个热点研究方向。并且,基于视觉的处理方法可以帮助超市进行货架合规优化,据文献[1],当货架能够保持 100% 合规时,两星期内,超市销售额和利润分别增长了 7.8% 和 8.1%。

基于传统图像处理方法的货架商品识别主要分为 2 个阶段:第 1 阶段使用目标检测算法或者图像分割算法检测出货架上的商品;第 2 阶段使用特征匹配算法或图像识别算法识别出商品的具体种类。在以往的文献中,第 1 阶段大都为传统图像处理方法,如基于简单特征瀑布流的快速目标检测、DPM、滑窗法<sup>[2]</sup>等,或者是利用 Hough 直线检测、Canny 边缘检测<sup>[3]</sup>、BRISK 特征<sup>[4]</sup>等将货架图像分割提取单个商品图像。这些方法都有着检测准确率不高,难以适应不同场景、不同形态货架的问题。第 2 阶段中的单个商品图像识别算法也采用了传统的 ORB 特征、SIFT 特征、SURF 特征<sup>[5]</sup>以及 HOG<sup>[6]</sup>等工设计的特征方法进行匹配,这些特征方法有的对商品局部图案敏感,有的需要较高的图像分辨率,还有的无法识别多尺度商品图像。自从 Krizhevsky 等<sup>[7]</sup>提出了 AlexNet 深度神经网络以及 Girshick 等<sup>[8]</sup>提出了 RCNN 目标检测模型,深度神经网络相较于传统算法大大提升了在图像检测与识别任务上的性能,因此,将深度神经网络应用于货架商品识别成为了可能。

文中基于 RetinaNet 通用目标检测框架<sup>[9]</sup>和 ResNet 特征提取网络<sup>[10]</sup>,并针对货架商品识别任务改进 RetinaNet 框架,提出使用深度神经网络识别货架商品的方法。文中方法提升了货架商品识别的查准率和查全率,对不同场景和不同形态货架更具适应性,并且采用特征提取与聚类的方法降低数据标注成

本,使得快速深度人工智能图像处理方法在实际使用中成为可能。

## 1 深度神经网络货架识别方法原理

### 1.1 网络整体设计

文献[4]尝试使用单个目标检测模型完成货架商品识别,但有 2 个问题:当商品 SKU 类别比较多时,使用单个模型需要精确标注,即标注每个商品的具体 SKU,这大大增加了标注难度和成本以及降低标注质量,同时数据质量下降引入了较多噪声;使用单个模型在检测商品目标的同时进行多 SKU 分类,较多的 SKU 类别会使模型性能下降。为了减少数据集标注量,增加商品识别准确率,该方法级联了 2 个深度神经网络,见图 1。第 1 个深度神经网络是通用目标检测网络,基于目标检测网络 RetinaNet 并进行了改进,作用是检测出图像中一些特定的商品且无需预测类别。第 2 个深度神经网络是图像分类网络 ResNet-50,它的作用是分类第 1 个网络中检测出的商品,得到商品的具体 SKU。

### 1.2 检测阶段设计

检测阶段基于 RetinaNet 通用目标检测框架,结构如图 1 检测阶段所示。针对货架商品检测,此框架作了如下改进:将 FPN 特征融合网络由 P3-P7 改为 P2-P5;改变锚框大小;舍弃掉分类子网络;在回归子网络中添加是否是感兴趣商品的置信度预测。

考虑到货架商品图像中大多为密集小目标以及计算速度的折中,输入图像的分辨率都统一按比例缩放为 800 pixel×800 pixel,不足部分填充为 0。原始图像中单个商品的目标尺度约在 25 pixel×25 pixel 到 120 pixel×120 pixel 之间,处于一个适当的分辨率大小。卷积网络使用了 ResNet-50,具体卷积层参数见表 1,其中 conv3\_1, conv4\_1, conv5\_1 的 3×3 卷积核的步长为 2。在 conv2\_4 后面新加上 2 个卷积核大小分别为 1 和 3 的 2 个卷积层得到特征图 P2。同样的,

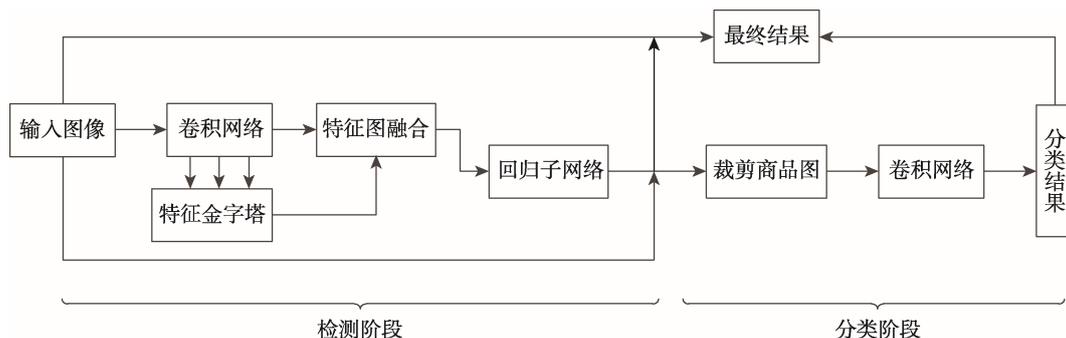


图 1 网络整体设计

Fig.1 Design of whole network

由 conv3\_4, conv4\_6, conv5\_3 得到特征图 P3, P4, P5。RetinaNet 中的特征图为 P3-P7, 但由于商品目标最小尺寸为 25 pixel×25 pixel, 而 P3 的感受野为 105, 远远大于最小商品尺寸, 所以将起始特征图改为感受野为 41 的 P2 特征图。又由于 P6 和 P7 的感受野已远远超过输入图像尺寸且商品尺寸最大为 120 pixel×120 pixel, 为了减少计算量和训练干扰, 舍弃了 P6 和 P7 特征图。不同于 RetinaNet 中的锚框设置, 该方法在 P2-P5 特征图上分别设置了面积为 {16<sup>2</sup>, 32<sup>2</sup>, 64<sup>2</sup>, 128<sup>2</sup>} 的基本锚框, 每个基本锚框有 {2<sup>0</sup>, 2<sup>1/3</sup>, 2<sup>2/3</sup>} 3 种大小尺寸和 {1 2, 1 1, 2 1} 3 种长宽比。特征图上每个位置有 9 个不同的锚框, 覆盖了边长 16~203 pixel 尺寸的商品目标。为了降低标注难度和提升模型性能, 检测阶段只进行单类检测, 舍弃掉 RetinaNet 中的分类子网络。原本的回归子网络中只有对锚框进行回归长度为 4 的向量, 为了预测回归框是否是目标商品, 预测向量长度被扩展为 6, 最后 2 位连接 softmax 激活判定是否为目标。在特征图融合中, 各特征图自顶向下进行上采样融合进下层特征图(上层特征图使用最邻近插值上采样后与下层特征数值相加), 融合后的 4 层特征图分别连接一个回归子网络, 得到坐标预测和是否为目标的置信度。

表 1 网络结构参数  
Tab.1 Network structure parameters

Layer name	ResNet-50	Layer name	Feature map
conv1	conv, 7×7, 64, stride2 max pool, 3×3, stride2		
conv2_	$\begin{bmatrix} \text{conv}, 1 \times 1, 64 \\ \text{conv}, 3 \times 3, 64 \\ \text{conv}, 1 \times 1, 256 \end{bmatrix} \times 3$	P2	
conv3_	$\begin{bmatrix} \text{conv}, 1 \times 1, 128 \\ \text{conv}, 3 \times 3, 128 \\ \text{conv}, 1 \times 1, 512 \end{bmatrix} \times 4$	P3	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \end{bmatrix}$
conv4_	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \\ \text{conv}, 1 \times 1, 1024 \end{bmatrix} \times 6$	P4	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \end{bmatrix}$
conv5_	$\begin{bmatrix} \text{conv}, 1 \times 1, 512 \\ \text{conv}, 3 \times 3, 512 \\ \text{conv}, 1 \times 1, 2048 \end{bmatrix} \times 3$	P5	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \end{bmatrix}$

检测模型区分目标和背景置信度的损失函数由式 (1) 定义:

$$L_{\text{conf}}(p, y) = \begin{cases} -(1-p)^\gamma \log(p) & \text{if } y = 1 \\ -p^\gamma \log(1-p) & \text{otherwise} \end{cases} \quad (1)$$

式中:  $p$  为检测网络预测是目标商品的概率;  $y \in \{1, 0\}$  分别表示真实值是否是目标商品;  $\gamma$  是平衡系数, 它有效解决了 one-stage 模型的类别不平衡问

题<sup>[9]</sup>, 提升了模型性能。

检测模型目标框回归损失函数由式 (2) 定义:

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos} m \in \{cx, cy, w, h\}} \sum_{j=1}^N x_{ij} \text{smooth}_{L1}(l_i^m - \hat{g}_j^m)$$

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^w \hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^h$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right) \hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right) \quad (2)$$

式中:  $x_{ij} \in \{0, 1\}$  代表第  $j$  个真实目标框是否属于第  $i$  个默认锚框;  $N$  代表有  $N$  个默认锚框被分配了真实目标;  $d_i^{cx}, d_i^{cy}, d_i^w, d_i^h$  分别代表第  $i$  个默认锚框的中心点  $x$  坐标、中心点  $y$  坐标以及锚框的宽度和高度;  $g_j^{cx}, g_j^{cy}, g_j^w, g_j^h$  分别代表第  $j$  个真实目标框的中心点  $x$  坐标、中心点  $y$  坐标以及真实目标框的宽度和高度;  $\hat{g}_j^m$  是变换后的真实目标框,  $l_i^m$  是预测的目标框,  $m \in \{cx, cy, w, h\}$ 。

检测模型的总损失函数如式 (3) 所示, 式中采用系数  $\alpha$  来平衡置信度损失和回归损失。

$$L_{\text{tol}}(p, y, x, l, g) = L_{\text{conf}}(p, y) + \alpha L_{\text{loc}}(x, l, g) \quad (3)$$

### 1.3 分类阶段设计

检测阶段得到的是货架商品图像上的商品坐标, 分类模型根据得到的坐标在原始图像上裁剪出单个商品图像并编号。由于裁剪出的单个商品图像较多但图像尺寸不大, 分类网络会将所有裁剪得到的 SKU 图片都调整到 128 pixel×128 pixel 的尺寸组合成一个 Batch 输入分类神经网络, 以进行并行计算, 减少运行时间。分类神经网络预测好每张单个商品图像的 SKU 后, 对应每张图像的编号和它们的坐标, 在原始货架商品图像上映射最终结果。如上节所说分类阶段的网络同样使用了 ResNet-50, 这样做的好处是可以利用检测阶段训练好的模型参数“微调”分类模型的权重参数, 分类阶段训练好的模型参数“微调”检测模型的权重参数, 交替训练使得网络很容易就收敛, 且一定程度上可以提升模型性能。ResNet-50 特征提取网络之后连接均值池化层和全连接层以及 softmax 作为分类器。

在分类阶段, 使用了 A-Softmax 损失函数<sup>[11]</sup>训练网络, 如式 (4) 所示:

$$L_{\text{ang}} = \frac{1}{N} \sum_i -\log \left( \frac{\exp(\|x_i\| \cos(m\theta_{y_i}))}{\exp(\|x_i\| \cos(m\theta_{y_i})) + \sum_{j \neq y_i} \exp(\|x_i\| \cos(\theta_{j,i}))} \right) \quad (4)$$

式中:  $x_i$  是第  $i$  个训练样本在 softmax 激活后的输出;  $\theta_{y_i}$  是  $x_i$  与对应真实分类权重参数的夹角;  $\theta_{j,i}$  是  $x_i$  与其他分类权重参数夹角;  $m$  是调整分类间距的

系数。A-Softmax 损失函数使得不同类别的数据在超球面上更为聚集,相互之间间隔更大,从而在更多类别更相似数据集上取得相较普通 softmax 损失函数更好的分类结果。

## 2 实验

### 2.1 数据集制作

该方法的数据集包含了用手机拍摄的 500 张某品牌饮料在超市货架上的摆放照片,标注并统计出其中有 22 种 SKU 和 20 000 个以上的目标商品。在这 500 张图片中,挑选出 100 张图片作为验证集,50 张图片作为测试集,剩下 350 张图片作为训练集。

在检测阶段中,使用 labeling 软件标注图片。由于标注时间复杂度为  $O(n \cdot m)$ ,式中  $n$  为需要标注的类别(当  $n > 10$  时,由于同类商品的相似性,就难以进行标注了), $m$  为商品数量,并且分类阶段网络能够区分具体 SKU,因此在标注时只标注是否为需要识别的商品。

在分类阶段中,为了获得足够多的带标签的分类数据和降低标注难度,采用了一种独特的数据集制作策略。首先,在原始图片中根据标记好的目标框裁剪出只包含每件商品自身的图片,500 张原始图片共提取了 22 761 张商品小图。然后使用在 IMAGENET 数据集上训练好的 ResNet-50 网络提取这些图片的最后一层 2048 维特征。提取出的特征进行  $k$ -means 聚类,聚类完成后的每一类图片实际上包含了太多同 SKU 的图像。将这些图片筛选一遍,直接删除其他少数 SKU 图片,只保留最多的 SKU 图片。这个策略让数万张 22 种 SKU 且商品外包装非常相似的图片的分类标记变得简单。对于每一类 SKU,随机抽取 500 张作为训练集,100 张作为验证集,50 张作为测试集。分类数据集共计 22 类,14 350 个正样例,随机裁剪 11 200 张背景图片作为负样例,其中 1000 张为验证集负样例,200 张为测试集负样例。22 种 SKU 数据集见图 2。



图 2 数据集中的 SKU  
Fig.2 SKUs in dataset

### 2.2 数据增强

传统的一些方法<sup>[2-6]</sup>无法使用数据增强,但数据

增强能够有效提升深度神经网络性能,它解决了如下几个问题:由于货架商品图片中商品目标数量很多,标注成本高昂,因此训练数据较少,数据增强可以增加数据集;原始的标注数据容易产生过拟合,数据增强可以使模型得到更好的泛化效果;本方法使用的神经网络较大,网络容易学习到无意义的高频特征,数据增强可以防止网络学习到不相关的模式,根本上提升网络性能。有 5 种图像数据增强方法被采用,见图 3。



图 3 标注样例和数据增强样例  
Fig.3 Labeled example and augmentation examples

### 2.3 模型训练

神经网络模型实现及训练基于 TensorFlow 框架。学习率策略:初始值为 0.01,1000 steps 处 warmup 为 0.02,采用 cosin 衰减直至训练结束。Batchsize 设为 8,总训练步数为 30 000,动量设为 0.9, IOU 阈值为 0.5,检测阈值为 0.8。由于数据集不够庞大,为了防止模型过拟合以及训练更容易收敛,使用在 IMAGENET 数据集上训练好的 ResNet-50 参数进行“微调”训练。训练结果见图 4。

训练时,每个锚框只能匹配一个前景目标,当锚框和目标框的 IOU 大于等于阈值 0.5 时,此锚框被分配给对应的目标框,且回归向量倒数第 2 位被设为 1,最后一位设为 0;当 IOU 处于[0, 0.4)时,锚框被认为是背景,回归向量最后一位被设为 1,倒数第 2 位被设为 0;当 IOU 处于[0.4, 0.5)时,锚框训练时被忽略掉。

同样使用 TensorFlow 构建分类阶段的 ResNet-50 模型。使用在 IMAGENET 上训练过的 ResNet-50 模型进行“微调”,学习率为 0.01, batchsize 为 256。经

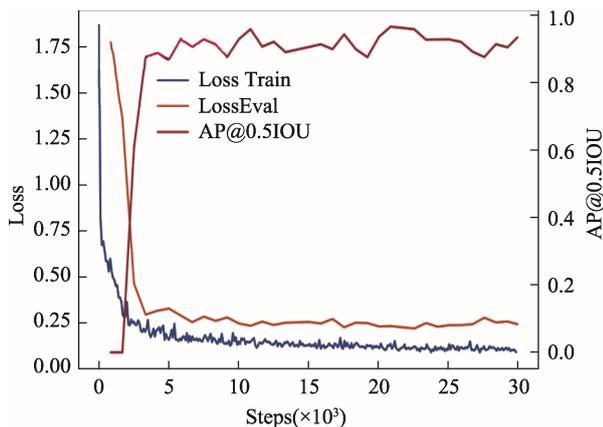


图 4 训练结果  
Fig.4 Training results

过 5 个 epoch 训练后验证集准确率就已经达到了 99.9%，并且训练好的模型在测试集上同样达到了 99.9%。

### 3 结果与讨论

#### 3.1 实验结果与分析

改进前后的检测网络的 AP (平均正确率) 分别

为 85.3%和 92.1% 加上数据增强的 AP 分别为 90.2%和 96.5%。

整体测试的结果见表 2。在测试集中随机选取了 10 张货架商品图片进行 2 个阶段的串联测试，检测阶段阈值设为 0.8。统计出共有 418 个 SKU，测试结果中真正例 (TP) 400 个，假正例 (FP) 10 个，假负例 3 个，最终查准率为 97.56%，查全率为 99.26%。

该方法和其他方法在检测阶段和分类阶段的结果分别见表 3 和 4。无论是在检测阶段还是分类阶段，本方法都取得了最好的效果。

该方法优于上述其他方法的原因主要有：特征提取效果优异，使用的 ResNet-50 特征提取网络能够提取 2048 维度的特征信息，相较于方法<sup>[3-4,13-15]</sup>中的 Hough, HOG, SIFT, SURF 等特征提取器，提取的特征信息更加丰富，语义信息层次更多；用 Anchor, FPN 和数据增强，能够在多尺度、多角度以及不同光线效果下密集地检测图片中的商品，这优于其他方法<sup>[3-4,13]</sup>使用的滑窗匹配；更丰富的特征下，A-softmax 分类器可比其他方法<sup>[13-15]</sup>中的 SVM 更好地分类相似商品。

表 2 实际整图测试结果 (阈值为 0.8)  
Tab.2 Test result of real images with threshold of 0.8

测试图片序号	SKU 数量	TP	FP	FN	Precision/%	Recall/%
1	29	26	3	0	89.66	100.00
2	35	34	0	0	100.00	100.00
3	26	25	0	0	100.00	100.00
4	65	62	3	1	95.38	98.41
5	19	18	0	0	100.00	100.00
6	65	61	0	0	100.00	100.00
7	78	76	2	1	97.44	98.70
8	46	43	1	1	97.73	97.73
9	18	18	1	0	94.74	100.00
10	37	37	0	0	100.00	100.00
合计	418	400	10	3	97.56	99.26

表 3 商品检测方法对比  
Tab.3 Comparison of product detection methods

文献	数据类型	训练集	测试集	方法	识别率/%
[3]	饮料 2SKU	未知	未知	单层货架分割；重复区域提取	85.6
[4]	洗发水	未知	400 张自建货架图片	Hough；BRISK	90.0
[13]	香烟	未知	375 正样例 195 负样例	HOG；Cascade Object Recognition Algorithm	Precision：88.8 Recall：87.9
[16]	香烟	375 张货架图片	354 张货架图片	DPM；HOG	Recall：89 Precision：69
文中	饮料	600 张货架图片	50 张货架图片	Deep Neural Network；FPN； ResNet-50	AP：96.5

表4 商品SKU分类方法对比  
Tab.4 Comparison of product SKU classification methods

文献	数据类型	训练集	测试集	方法	准确率/%
[13]	香烟	47 正样例, 87 负样例	31 正样例, 24 负样例	HOG; Color Histogram; SVM	96
[14]	饮料 6SKU	195 样例	194 样例	SURF; Bag of Words; SVM	89.0
[15]	饮料 794SKU	20000 样例	70000 样例	SIFT; Bag of Words; SVM	68.45
[15]	饮料 794SKU	70000 样例	10000 样例	Chain Structured Graphical Model; SVM	78.40
文中	饮料 21SKU	13200 正样例, 11000 负样例	1100 正样例, 200 负样例	ResNet-50; A-softmax	99.9

### 3.2 补充实验

为了检验算法的效果,补充了实际检测效果实验和扩展到其他种类商品上的实验。实际检测效果见图5。将算法应用于调味品类商品的检测效果见图6。本方法在调味品数据集上以及在饮料和调味品混合数据上的结果分别为95.7%和90.3%(AP@0.5IOU)。因为检测种类的增加和正样本之间的数据不平衡,混合数据集的算法性能会有所下降。算法在 GTX 1080Ti 显卡上的运行速度:检测阶段、分类阶段、整体的时间分别为476,352,978 ms,考虑到算法应用场景,其能够满足实时性要求。



图5 实际检测效果  
Fig.5 Actual test results



图6 调味品检测效果  
Fig.6 Test results of condiments

## 4 结语

目前超市都是人工进行货架商品信息统计,而传统的计算机视觉处理方法在检出率和准确率上都较低,并且对商品摆放和货架图像质量都有要求,离实际应用要求还有一定距离。研究了采用改进型深度神经网络进行货架商品识别的方法。该方法仅仅利用商品外包装的视觉特征即可准确定位与检测出具体商品的SKU,相较于其他方法在检出率和准确率上都有了很大的提高,同时对货架场景形态及识别数量和商品相似度的容忍度更高。该方法虽然在检测某些简单类型(图片识别度较高)的货架商品(如酒水、调味品、饮料等),具有很好的识别效果,但检测更多复杂包装形态(如柔性包装等)以及密集堆放商品时,算法性能有一定的下降,还需要在下一步的工作中进一步优化。

### 参考文献:

- [1] SHAPIRO M. Executing the Best Planogram[J]. Professional Candy Buyer, 2009, 3(1): 24—28.
- [2] MELEK C G, SONMEZ E B, ALBAYRAK S. A Survey of Product Recognition in Shelf Images[C]// Computer Science and Engineering (UBMK), 2017 International Conference on IEEE, 2017: 145—150.
- [3] 梁青青. 超市货架区域分割与商品识别研究[D]. 南京: 南京理工大学, 2013.  
LIANG Qing-qing. Research on Supermarket Shelf Area Segmentation and Product Recognition[D]. Nanjing: Nanjing University of Science and Technology, 2013.
- [4] 刘永豪. 基于深度学习的货架商品检测技术研究[D]. 杭州: 浙江大学, 2017.  
LIU Yong-hao. Product Detection on Shelves Based on Deep Learning[D]. Hangzhou: Zhejiang University, 2017.
- [5] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB:

- An Efficient Alternative to SIFT or SURF[C]// Computer Vision (ICCV), 2011 IEEE International Conference on IEEE, 2011: 2564—2571.
- [6] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on IEEE, 2005: 886—893.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolution Neural Networks[C]// Advances in Neural Information Processing Systems, 2012: 1097—1105.
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580—587.
- [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980—2988.
- [10] 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.
- [11] LIU W, WEN Y, YU Z, et al. Sphere Face: Deep Hyper Sphere Embedding for Face Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 212—220.
- [12] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211—252.
- [13] BURUŞ M E, KUZU R S, BEYAZIT S, et al. Market Raflarında Dönüşüm Ve Ölçeklendirmeye Dayanıklı Nesne Tanıma[J]. Akademik Bilişim, 2014, 25(3): 5—7.
- [14] HAFIZ R, ISLAM S, KHANOM R, et al. Image Based Drinks Identification for Dietary Assessment[C]// 2016 International Workshop on Computational Intelligence (IWCI). IEEE, 2016: 192—197.
- [15] BAZ İ, YORUK E, ÇETİN M. Retail Product Recognition with a Graphical Shelf Model[C]// 2017 25th Signal Processing and Communications Applications Conference (SIU). IEEE, 2017: 1—4.
- [16] VAROL G, KUZU R S. Toward Retail Product Recognition on Grocery Shelves[C]// Sixth International Conference on Graphic and Image Processing (ICGIP 2014). International Society for Optics and Photonics, 2015: 9443—9449.