

引文格式:

陈欢欢, 汪建晓, 王高杰, 等. 基于改进 Faster R-CNN 的瓶装饮料商品目标检测方法 [J]. 集成技术, 2021, 10(3): 1-11.

Chen HH, Wang JX, Wang GJ, et al. Target detection method of bottled drinks based on improved faster R-CNN [J]. Journal of Integration Technology, 2021, 10(3): 1-11.

基于改进 Faster R-CNN 的瓶装饮料商品 目标检测方法

陈欢欢^{1,2} 汪建晓^{1*} 王高杰² 陈 勇¹

¹(佛山科学技术学院 佛山 528000)

²(广东顺德创新设计研究院 佛山 528000)

摘 要 该文以无人售货机售卖瓶装饮料商品为研究场景, 提出一种基于改进 Faster R-CNN 算法的瓶装饮料商品目标检测方法。首先, 采用残差网络 ResNet-50 进行特征提取, 加深网络对目标特征的提取和学习的深度; 然后, 根据瓶装饮料商品形态学特征, 增加区域建议网络 (Regional Proposal Network) 的锚框数量和样式; 最后, 基于所提出的多维特征图融合网络, 增强了网络对小目标的检测性能。实验结果表明, 模型训练迭代 10 000 次后的损失值趋于收敛, 10 个类别的瓶装饮料商品平均精度值均在 90% 以上, 综合检测识别率平均精度均值 (MAP) 为 93.26%, 较改进前的模型测试精度提升 20%, 取得良好检测效果。

关键词 Faster R-CNN; 目标检测; 残差网络; 区域提议网络; 多维特征融合

中图分类号 TP 391 文献标志码 A doi: 10.12146/j.issn.2095-3135.20201020001

Target Detection Method of Bottled Drinks Based on Improved Faster R-CNN

CHEN Huanhuan^{1,2} WANG Jianxiao^{1*} WANG Gaojie² CHEN Yong¹

¹(Foshan University of Science and Technology, Foshan 528000, China)

²(Guangdong Shunde Innovation and Design Institute, Foshan 528000, China)

*Corresponding Author: jwang90@163.com

Abstract This paper presents an improved faster R-CNN algorithm based on the application of unmanned vending machine selling bottled drinks. Firstly, the residual network ResNet-50 is used as the feature

收稿日期: 2020-10-20 修回日期: 2020-11-29

基金项目: 广东省科技计划项目 (2017A010102018)

作者简介: 陈欢欢, 硕士研究生, 研究方向为工业机器人、机器视觉; 汪建晓 (通讯作者), 硕士研究生导师, 研究方向为机械设计理论与方法、机电系统控制工程等, E-mail: jwang90@163.com; 王高杰, 硕士研究生导师, 研究方向为嵌入式系统; 陈勇, 副教授, 硕士研究生导师, 研究方向为复杂系统控制与优化、人工智能和图像处理等。

extraction network to deepen the depth of target feature extraction and learning. Then, the number and style of anchor frame in regional proposal network (RPN) is improved according to the morphological characteristics of bottled beverage products. Finally, a multi-dimensional feature map fusion network is proposed to enhance the detection performance of small targets. The experimental results showed that, the loss value tends to converge after 10 000 iterations of model training. Average precision values of 10 categories of bottled beverage products are all larger than 90%. And the comprehensive detection recognition rate mean average precision value is 93.26%, which is improved 20% compared with the original model.

Keywords faster R-CNN; target detection; residual network; regional proposal network; multi-dimensional feature fusion

Funding This work is supported by Guangdong Province Science and Technology Plan Project (2017A010102018)

1 引 言

随着“新零售”概念的兴起,无人超市、无人售货机甚至无人货架如雨后春笋般涌现^[1]。其中,对商品目标“快”且“准”的检测是未来“新零售”最重要的技术需求,也是未来人工智能领域一个重要的研究和应用方向^[2]。目前有很多应用于商品检测和识别的技术,但传统的手工提取图像特征与浅层机器学习相结合的商品检测方法存在一定的局限性^[3]:如方向梯度直方图(Histogram of Oriented Gradient, HOG)^[4]、局部二值模式算子(Local Binary Pattern, LBP)^[5]和尺度不变特征转换算子(Scale Invariant Feature Transform, SIFT)^[6]等多用于人工选定的颜色特征、纹理特征、形状特征和空间关系特征等浅层次特征方面,而在实际商用中遇到复杂场景或具有复杂特征的目标检测效果较差。相较于传统方法,利用深度学习(Deep Learning)^[7]算法能对商品图像进行深层次的特征提取、处理和分析,有较强应对复杂场景的目标检测识别优势。因此,未来无人售货的发展趋势更多将采用深度学习来实现商品的视觉检测与识别。

目前基于深度学习的目标检测任务主要有两大技术路线:一是以YOLO(You Only Look Once)^[8]系列算法和SSD(Single Shot Detection)^[9]系列算法为主的One-Stage目标检测算法;二是以Faster R-CNN(Faster Region with Convolutional Neural Network)^[10]系列算法为主的Two-Stages目标检测算法。其中,One-Stage目标检测算法是将目标检测问题转化成直接回归问题,不产生候选框;Two-Stages目标检测算法是通过卷积神经网络(Convolutional Neural Network, CNN)^[11]与区域建议网络(Region Proposal Network, RPN)^[10]相结合,以不同尺寸的滑动窗口自左上角向右下角滑动并扫描特征图(Feature Map)来生成若干候选区域,再提取候选区域相关的视觉特征。两种方法各有其特点和优势:One-Stage目标检测算法检测速度快,因此实时性更有优势。如方仁渊和王敏^[12]在对商品包装类型的检测中,先精简原始YOLO算法的骨干网络部分,使得训练时间成本大幅减低,再提出反向连接结构得到与原始网络在平均精度上持平的检测结果;廖珈艺等^[13]在对商品陈列货架的目标检测中,采用对比性实验在相同深度学习框架和数据集情况下以不同算法

为单一变量, 比较 SSD 和 Faster R-CNN 在检测速度和精确度两方面的优劣, 得出 SSD 检测速度更快、Faster R-CNN 检测精确度稍高的结果。

Two-Stages 目标检测算法则是精确度更高, 小目标检测方面更稳定。如周立旺等^[14]针对无人冰柜中小目标商品的检测, 提出一种由粗到精的两阶段检测结构 FocusNet, 使该场景下的小目标检测效果提升 3.5%; 胡正委和朱明^[15]针对商品小样本数据的情况下, 结合 Grabcut 与非类别特异性 Faster R-CNN 提出一种样本增强方法, 并添加重识别层, 检测精度提高 4%; 梅舒欢等^[16]基于 Faster R-CNN 网络和视觉基因库的食品菜品图像检索与分类, 对 20 类菜品获得 75.4% 精确率的结果; 王振^[17]提出一种密集连接的残差网络与 Faster R-CNN 基础网络相结合的方法, 以减轻梯度消失来增强深度学习能力, 但网络复杂度增加, 参数量变得更加庞大, 不利于实时性检测。

本文面向无人零售终端商品视觉检测的实际商用, 采用改进的 Faster R-CNN 算法实现对瓶装饮料商品检测。为解决原始 Faster R-CNN 算法检测精度和实时性的不足, 采用 ImageNet^[18]预训练模型进行迁移学习, 同时针对小目标检测和局部遮挡问题对原有特征提取网络进行多维特征融合, 并对 RPN 网络中锚框样式进行丰富。实验结果显示, 该模型实时检测性能稳定, 对商品平均检测精确率达到 93.26%, 具有良好的检测效果。

2 改进 Faster R-CNN 深度学习模型的建立

2.1 Faster R-CNN 简述

Faster R-CNN 由 Ross B. Girshick 在 2016 年提出^[10], 是目前比较流行的 Two-Stages 目标检测算法。网络结构上, Faster R-CNN 算法由主干网络、区域建议网络 (RPN) 和全连接层网络 3 部分构成, 具体如图 1 所示。

相比 R-CNN 和 Fast R-CNN, Girshick 在 Faster R-CNN 中提出采用 RPN 网络代替传统的选择性搜索 (Selective_Search) 网络, 解决了候选框生成的问题。RPN 网络原理如图 2 所示, 将特征提取网络中得到的特征图作为输入, 一般用 3×3 大小的卷积核跟随滑动窗口自左上角到右下角卷积操作, 得到 256 维长度的特征向量再

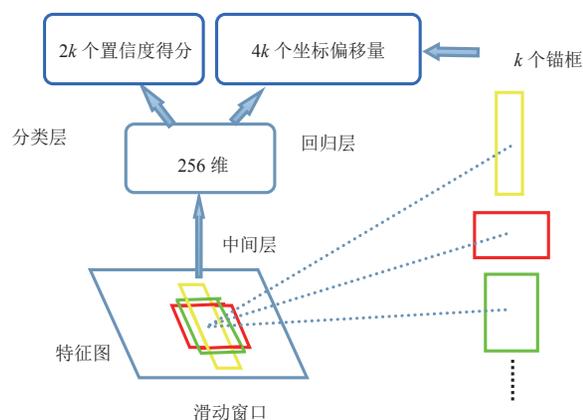


图 2 RPN 网络原理图

Fig. 2 The schematic diagram of RPN network

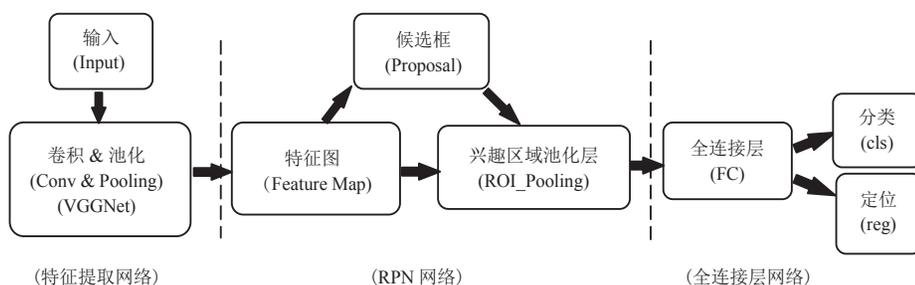


图 1 Faster R-CNN 网络结构图

Fig. 1 The network structure of Faster R-CNN

参与 2 个连续的全连接层。其中，第 1 个全连接层实际是一个二分类操作，完成前景与背景的区别；第 2 个全连接层实现候选框的初定位，输出的是锚框 (Anchor) 与真值框 (Ground True Boxes) 的 4 个坐标点偏移量，通过分类和回归方法生成若干候选框，再经过兴趣区域池化层池化操作去除部分候选框，并将剩余候选框固定尺寸进行输出。整个 RPN 网络完成前景与背景区分以及候选目标的初定位，打破了 R-CNN 和 Fast R-CNN 等采用传统选择性搜索非常耗时的瓶颈。

2.2 改进的 Faster R-CNN 算法及优化

本文改进的 Faster R-CNN 网络设计方案依然沿用迁移学习的方法：首先，对原始网络中的特征提取网络采用深度更深的 ResNet-50 来获得更多不同深度层次的特征图；其次，增加多维特征融合网络使不同层次特征图的丰富语义信息得以保留；最后，对 RPN 网络的锚框样式进行优化。图 3 为改进的 Faster R-CNN 网络结构设计思路方案。

2.2.1 ResNet-50 特征提取网络

与 VGGNet 系列的特征提取网络相比，ResNet 残差特征提取网络具有 3 方面的优势：(1) 提取深度更深，VGG16 由 13 个卷积层和 3 个全连接层构成，而 ResNet-50 由 49 个卷积层和 1 个全连接层构成，故 ResNet-50 算法精度更高；(2) 计算量更小，ResNet-50 具有 2 500 万

参数量，而 VGG16 参数量则达到 1.38 亿，因此 ResNet-50 内存功耗更低；(3) 增加网络深度的同时，ResNet 残差网络模块可以缓解梯度消失的问题。其中，梯度消失问题的出现，是因为随着深度的增加，求解神经网络梯度过程中，当梯度值小于 1 时，由于链式法则的连乘形式， N 层求解之后很容易衰减到 0。通常情况下，神经网络正向传播公式 $X_{i+1}=W_iX_i+b$ ，损失函数 $L=F(X,W)$ ，其反向传播的梯度公式为： $\frac{\partial L}{\partial X}=\frac{\partial F(X,W)}{\partial X}$ 。扩展到 n 层神经网络的损失函数公式为：

$$\begin{cases} L_2=F_1(X_1,W_1) \\ L_3=F_2(X_2,W_2) \\ \dots \\ L_n=F_{n-1}(X_{n-1},W_{n-1}) \end{cases} \quad (1)$$

根据链式法则可以推导出第 i 层的梯度值计算公式为：

$$\frac{\partial L_n}{\partial X_i}=\frac{\partial F_{n-1}(X_{n-1},W_{n-1})}{\partial X_{n-1}} \times \dots \times \frac{\partial F(X_i,W_i)}{\partial X_i} \quad (2)$$

由此可知，随着网络深度的增加，反向传播使误差回传，导致前层网络的梯度变得越来越小。为解决深度卷积神经网络梯度消失的问题，ResNet 残差卷积网络在结构上增加了输入层与输出层的直接通道。即将输出层 $H(X)=F(X)$ 改为 $H(X)=F(X)+X$ ，因此第 i 层的梯度值计算公式

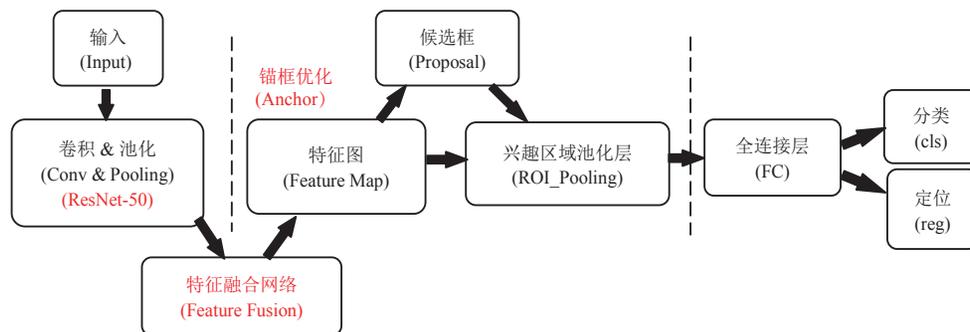


图 3 改进的 Faster R-CNN 网络结构设计图

Fig. 3 The network structure design diagram of improved Faster R-CNN

变为:

$$\frac{\partial X_{i+1}}{\partial X_i} = \frac{\partial X_i + \partial F(X_i, W_i)}{\partial X_i} = 1 + \frac{\partial F(X_i, W_i)}{\partial X_i} \quad (3)$$

因此, 采用 ResNet 残差卷积网络提取特征, 即使网络结构变得更深, 梯度也不会消失。

2.2.2 多维特征融合网络

Faster R-CNN 特征提取网络多次下采样过程中, 输入图像经过不同残差卷积块会产生大小不同的特征图, 但这些包含丰富语义信息的特征图都仅作为下一个残差卷积块的输入而没有直接参与到后面 RPN 网络的定位和分类任务中去, 造成部分有用语义信息的丢失, 导致模型对小目标的检测能力减弱。本文借鉴 SSD 网络的金字塔特征层级 (Pyramidal Feature Hierarchy) 对小目标检测的优势, 提出多维特征融合的训练方法。即在特征提取网络和 RPN 网络之间增加多维特征融合网络, 目的是将特征提取过程中高低维度特征图的语义信息尽可能少损失地保留合并到一个特征图中。特征融合网络就是把 Conv1~Conv5 残差卷积块得到的不同大小尺寸的特征图逐步上采样还原成原图尺寸大小, 再采用元素级相加法 (Element-Wise Add) 来进行特征融合, 从而弥补 Faster R-CNN 单特征图网络语义信息不足的缺陷。

高低层维度特征融合的经典方法主要有两种: 一种是元素级相加 (Element-Wise Add), 另一种是堆叠特征融合方法 Concat。两者的区别在于 Element-Wise Add 是描述图像特征的信息量增多, 但是描述图像的维度本身并没有增加。

而 Concat 是通道数的合并, 即描述图像本身的特征数 (通道数) 增加, 而每一特征的信息没有增加。由于 Element-Wise Add 的计算量比 Concat 的小, 也更适合 ResNet-50 残差网络, 因此本文设计的多维特征融合网络在策略方法上选择 Element-Wise Add, 具体如图 4 所示。

2.2.3 区域建议网络优化

滑窗 (Sliding Window) 是目标检测任务中常用的方法, 但也存在一些问题: 滑窗的大小尺寸、宽高比例以及位置不同将产生很大的计算量。Faster R-CNN 的 RPN 网络部分对输入特征图的滑窗大小设置为 3×3 矩形框, 但统一用 3×3 的滑窗难以很好地适应不同任务中对检测目标的拟合。因此为了适应多种形状的物体, RPN 定义了 k 种不同规格的锚框。原始网络中, 锚框是以 16 个像素作为基准、 $[0.5 : 1 : 2]$ 的比例作为一组, 在此基础上按一定比例增加到 3 组, 共 9 个大小尺寸和宽高比不同的锚框。锚框机制就是沿着滑窗自左向右、从上到下滑动直至走完整个特征图, 然后计算出每一个锚框与真值框的交并比 (Intersection Over Union, IOU), 记为 R_{IOU} 。当 R_{IOU} 大于设定阈值 A 时, 判定为该锚框含有检测目标, 为正样本; 而当 R_{IOU} 小于阈值 B 时, 则认为该锚框不包含目标物体, 为负样本。交并比的计算公式为:

$$R_{IOU} = \frac{A_{gt} \cap A_{an}}{A_{gt} \cup A_{an}} \quad (4)$$

其中, A_{gt} 为真值框面积; A_{an} 为锚框面积。

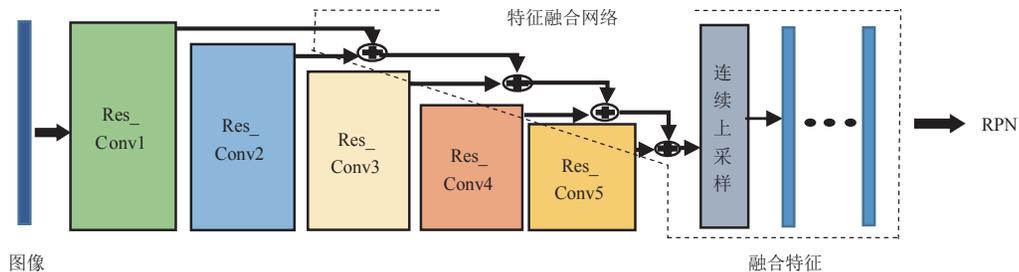


图 4 本文设计的多维特征融合网络结构图

Fig. 4 The network structure of multi-dimensional feature fusion designed in this paper

由于根据瓶装饮料商品物体的形态学特点,继续沿用原始网络中设置的锚框宽高比会造成正负样本比例严重失衡,不能满足实际任务需要。因此需要丰富锚框样式,增加 0.25 : 1 : 4 的宽高比,即由原来的锚框宽高比 $R = [0.5, 1, 2]$ 改为 $R = [0.25, 0.5, 1, 2, 4]$ 。此时每一组锚框由原来的 3 个变成了 5 个,但是单方面增加锚框数量会造成计算量增大。因此可以减少设置组数为 2,即 $5 \times 2 = 10$ 个锚框,比原来的 9 个锚框在数量上仅略微增加,但锚框样式上更满足本次任务的需要。不同宽高比例的锚框生成机制:在同一组中,不同宽高比的锚框的覆盖面积基本相同。记锚框面积为 A_{an} ,基础锚框边长为 H 、变形后锚框宽 w 、高 h 、宽高比 $R = w/h$,根据面积不变,推导过程为:

$$A_{an} = H^2 = w \times h = (h \times R) \times h = R \times h^2 \quad (5)$$

$$\text{变形后, 高度 } h = \sqrt{\frac{A_{an}}{R}} = \frac{H}{\sqrt{R}}; \quad (6)$$

$$\text{同理, } A_{an} = H^2 = w \times h = w \times \frac{w}{R} = \frac{w^2}{R} \quad (7)$$

$$\text{变形后, 宽度 } w = \sqrt{A_{an} \times R} = H \times \sqrt{R} \quad (8)$$

根据推导公式(6)、(8),当 $R=0.25$ 时, $h=2H$, $w=0.5H$; 当 $R=0.5$ 时, $h=1.414H$, $w=0.707H$; 当 $R=1$ 时,此时锚框为正方形框,与基础锚框长宽一致, $H=w=h$; 当 $R=2$ 时, $h=0.707H$, $w=1.414H$; 当 $R=4$ 时, $h=$

$0.5H$, $w=2H$; 保持原始基础锚框边长 $H=16$ 不变,则有 8×32 、 11.314×22.627 、 16×16 、 22.627×11.314 和 32×8 五种锚框样式,具体如图 5 所示。

3 实验

3.1 图像数据采集和预处理

本文实验所用图像数据由广东顺德创新设计研究院“新型无人售货机的研发”项目组提供。实验总共对 10 种类别常见品牌的饮料商品采集了 10 000 张图片以及若干背景图片。数据预处理阶段进行了图片筛选清洗、归一化以及适量的数据扩增,并采用 LabelImg_Master 标注软件进行数据标注。然后按照 [7 : 2 : 1] 的比例随机生成训练集、验证集和测试集,最终把数据集制作打包成 Pascal_Voc 固定数据格式。具体结果如图 6 所示。

3.2 实验软硬件环境配置及参数设置

实验硬件采用 Intel Z490 芯片组的服务器,处理器为 Intel 酷睿 i9-10900K,配置两块 NVIDIA 系列的 GPU 显卡、DDR4_32G 内存以及 250 G 固态硬盘。实验软件方面,操作系统为 Ubuntu18.04,安装 Anaconda 和 Caffe 深度学习框架,编辑语言为 Python,配置 Faster R-CNN 运行环境。实验参数设置方面,基础学习率为 0.001,学习策略选用 step(步长设为 2 500),

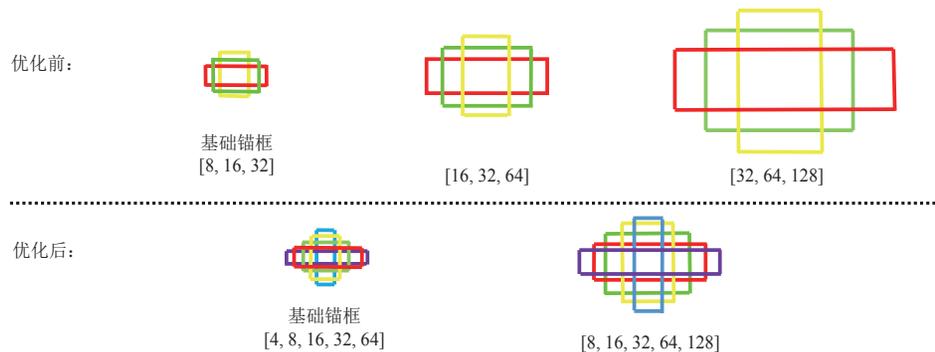


图 5 优化前后锚框样式变化

Fig. 5 Change of anchor frame style before and after optimization



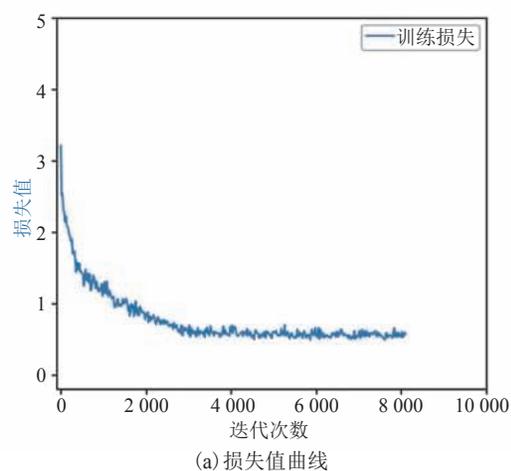
图 6 实验数据集原始图片及其 LabelImg_Master 软件标注图

Fig. 6 The original image of the experimental data set and its LabelImg_Master software annotation drawing

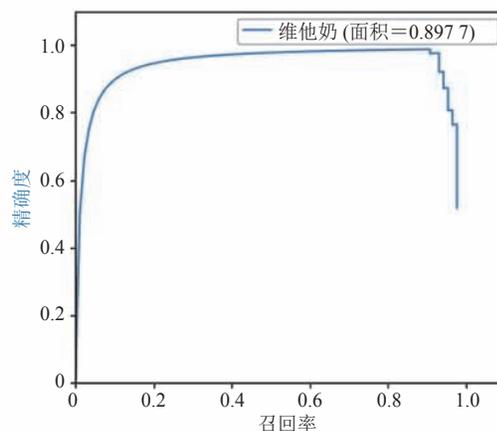
总共迭代次数设为 10 000 次, 权重衰减率设为 0.000 5, 动量设为 0.9, 界定正样本的 IOU 阈值设为 0.7。

3.3 网络模型训练与测试

网络模型阶段, 采用 ImageNet 的预训练模型和端到端的训练方法对改进前后的 Faster R-CNN 进行微调 (Fine-Turn)。为增强实验结果的对比性, 采用单一变量法共进行 8 组实验。实验步骤为: (1) 导入数据集训练, 比较 VGG16 和 ResNet-50 两种特征提取网络下, 原始 Faster R-CNN 训练消耗时间和模型测试平均精度均值; (2) 重新建立模型, 增加多维特征融合网络后再进行训练, 将所得模型与上述 (1) 中模型进行测试消耗时间和模型测试平均精度均值的比较; (3) 调整锚框参数, 将锚框宽高比 $R=[0.5, 1, 2]$ 改为 $R=[0.25, 0.5, 1, 2, 4]$, 比较改进锚框样式前后模型测试平均精度均值; (4) 根据上述步骤得到的实验结果, 采用 ResNet-50 特征提取网络与多维特征融合网络相结合的 Faster R-CNN 网络模型进行 10 000 次迭代的模型训练; (5) 测试训练后的模型, 获取损失值 (Loss)、召回率 (Recall)、各类别的平均精确度和平均精度均值以及每次迭代的耗时等关键数据, 并进行实验数据分析, 结果如图 7~8 所示; (6) 对平均精度均



(a) 损失值曲线



(b) 精确度-召回率曲线 (以维他奶为例)

图 7 VGG16+多维特征提取网络+RPN 的损失值曲线及精确率-召回率曲线

Fig. 7 Loss curve and precision-recall curve of VGG16 + multi-dimensional feature extraction network + RPN

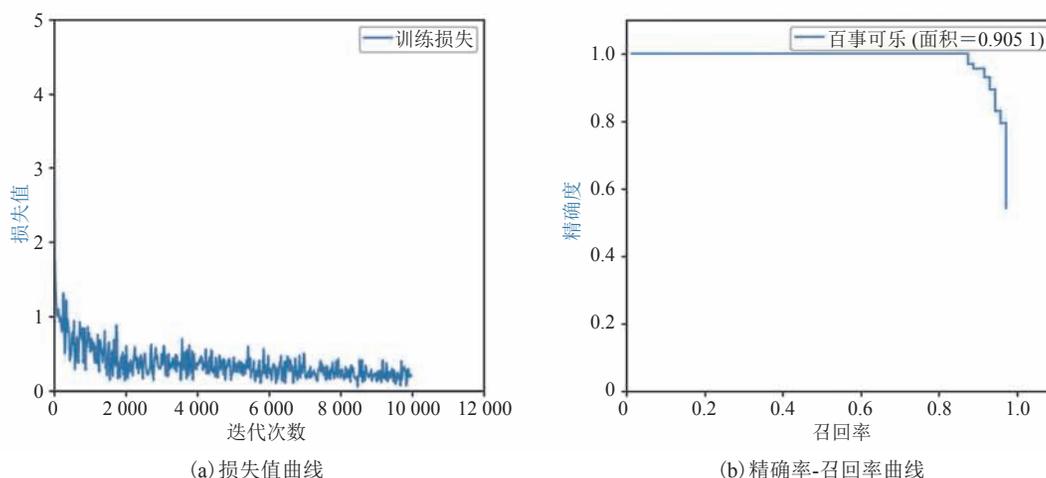


图 8 ResNet-50+多维特征提取网络+RPN 的损失值曲线及精确率-召回率曲线

Fig. 8 Loss curve and precision-recall curve of ResNet-50 + multi-dimensional feature extraction network + RPN

值最高的模型，进行实物检测。

4 实验结果与分析

4.1 实验结果

本文在原有 Faster R-CNN 网络基础上采用不同的特征提取网络 (VGG16、ResNet-50)，再通过增加多维特征融合网络对原有 Faster R-CNN 网络结构进行改进，同时对 RPN 网络中锚框样式进行改进。综合上述 3 个方面

的改进，共进行 8 组对比实验。为了定量评估改进后的 Faster R-CNN 算法网络模型对瓶装饮料商品检测识别的有效性，本次实验采用平均精度 (Average Precision, AP)、平均精度均值 (Mean Average Precision, MAP)、训练与测试耗时作为评价指标，实验结果见表 1。

本文在改进 Faster R-CNN 网络后，采用同样的数据集和硬件配置进行 YOLO 和 SSD 经典目标检测模型的对比实验。实验结果如表 2

表 1 实验结果汇总

Table 1 Summary of experimental results

网络结构	R	平均精确度 (%)										MAP (%)	T_{train} (s)	T_{test} (s)
		加多宝	雪碧	百事可乐	怡宝	农夫山泉	百岁山	脉动	维他奶	哈尔滨啤酒	芬达			
VGG16+RPN	R_1	61.54	95.89	87.80	78.13	67.24	69.73	78.37	52.55	65.71	73.64	73.06	0.689	0.239
ResNet-50+RPN		73.85	87.39	86.10	90.56	86.44	71.75	74.83	82.53	74.94	70.37	79.88	0.527	0.268
VGG16+FFN+RPN		75.72	83.69	88.47	93.21	78.23	74.39	70.64	69.36	66.43	71.54	77.17	0.725	0.571
ResNet-50+FFN+RPN		88.34	90.96	85.73	93.83	86.20	89.26	81.67	85.03	78.64	84.63	86.40	0.583	0.489
VGG16+RPN*	R_2	57.37	63.43	69.20	58.17	74.53	73.90	61.03	84.78	85.51	76.20	70.41	0.695	0.239
ResNet-50+RPN*		71.24	86.68	96.74	83.16	77.73	85.71	91.04	92.93	82.37	92.75	86.04	0.598	0.269
VGG16+FFN+RPN*		62.31	74.90	77.67	89.42	76.33	89.13	60.40	74.37	78.26	84.12	76.69	0.841	0.612
ResNet-50+FFN+RPN*		90.91	99.84	90.63	90.91	90.91	90.37	90.91	97.78	90.73	99.62	93.26	0.706	0.561

注：RPN 为区域建议网络 (Region Proposal Network)；FFN 为特征融合网络 (Feature Fusion Network)；“*”为优化后网络结构； R 为锚框宽高比，

$R_1=[0.5,1,2]$ ， $R_2=[0.25,0.5,1,2,4]$ ；MAP 为平均精度均值； T_{train} 为训练集平均单次迭代耗时； T_{test} 测试集平均单次迭代耗时

所示。

表 2 不同经典目标检测算法实验结果对比

Table 2 Comparison of experimental results of classical target detection algorithms

算法模型	平均精度均值 (%)	平均单次迭代耗时(s)
YOLO_v2	68.42	0.194
YOLO_v3	76.47	0.303
SSD300	75.15	0.215
SSD512	86.31	0.350
Faster R-CNN	73.06	0.239
本文改进的 Faster R-CNN	93.26	0.561

4.2 讨论与分析

表 1 结果表明, 改进后的 Faster R-CNN 网络模型对各类别目标的检测精确度均在 90% 以上, 整体平均精度均值为 93.26%, 较原始的 VGG16+RPN 网络结构的 73.06% 提升了近 20%, 虽然同时也增加了训练和检测耗时, 但仍然达到实时性的要求。表 2 结果表明, 与其他经典目标检测模型(如 YOLO 和 SSD)的检测效果相比, 本文改进后的 Faster R-CNN 模型综合性有一定提升。该结果与张瑞^[19]对商品 Logo 的检测识别结果相比, 在 $R_{IOU} \geq 0.7$ 的相同测试条件下, 本文方法在检测精度上提升 1.66%, 在实时性方面的测试耗时也更少。与罗辉^[20]采用的 SSD 经典目标检测算法相比, 其检测精度虽比本文方法略高, 但该文献采用先分 5 大类再分 10 小类的策略显然导致其最终实时性较差的结果, 而且固定检测瓶盖而忽略其他特征信息容易出现漏检。与袁豪^[21]同样采用 Faster R-CNN 的检测效果相比, 本文方法在特征提取网络采用更有学习深度的残差网络 ResNet-50, 而袁豪研究中依然沿用原始网络的 VGG16 特征提取网络, 故本文方法的检测精度比该文献提升 2.06%。由此可见, 本文改进的 Faster R-CNN 方法在对商品目标检测方面的整体性能都有所提升。

本文从目标检测精度和实时性能两方面

对改进前后的 Faster R-CNN 性能结果进行分析。(1)目标检测精度方面: ①残差特征提取网络 ResNet-50 明显优于 VGG16, 检测精度提升 6%~16%, 这是检测精度大幅提升的主要原因; ②多维特征融合网络的运用提升了模型对小目标的检测能力, 检测精度提升 4%~9%; ③锚框样式的改进对检测精度也有一定程度的提升。其中, 当特征提取网络采用 ResNet-50 时, 改进丰富锚框样式使得检测精度提升近 7%, 而当特征提取网络为 VGG16 时, 改进锚框样式对检测精度影响不大, 基本持平。(2)实时性能检测方面: 随着 8 组实验中深度学习网络复杂程度的加大, 训练过程中完成每次迭代的耗时从原始 Faster R-CNN 网络的 0.689 s 增加到 0.841 s, 测试耗时从 0.239 s 增加到 0.612 s。虽然整体上表现出训练与测试耗时逐步增加, 但增加的时间在实际任务中影响不大。

在实物测试中, 改进的 Faster R-CNN 网络展现出良好的检测效果, 其检测识别结果如图 9 所示。其中, 图 9(a)为实物图片检测结果, 图中显示 10 个类别的瓶装饮料商品全部正确检测识别且均达到 90% 以上的置信度得分, 达到预期的改进效果。图 9(b)为调用外部摄像头进行的视频实时检测, 同样获得较好的置信度得分, 同时每秒帧率(fps)为 44.8, 实时性检测效果良好。不足之处是对一些倾斜目标, 检测框有轻微漂移的迹象, 但仍然可以较好地满足对这 10 类饮料商品的检测识别要求。

5 结论

本研究以无人售货机售卖瓶装饮料商品为应用场景, 提出基于改进 Faster R-CNN 算法的瓶装饮料商品目标检测方法。首先, 通过在原有 Faster R-CNN 算法基础上试验不同特征提取网络(VGG16、ResNet-50)的性能; 其次, 在网络结

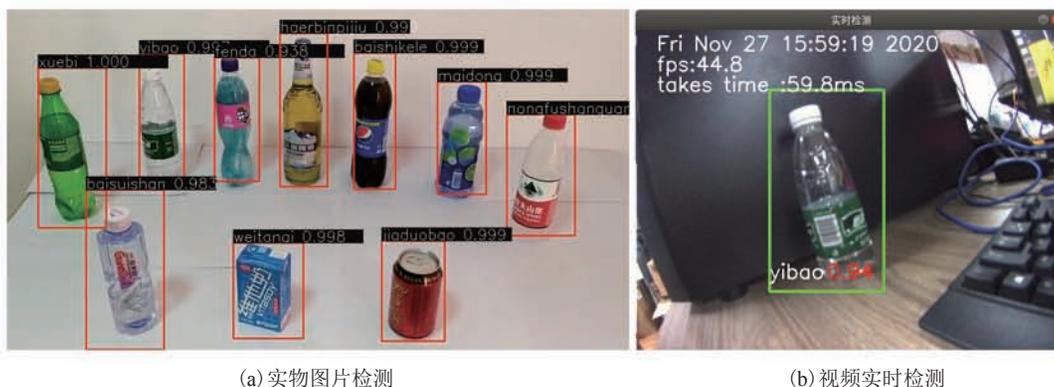


图 9 实物图片检测与视频实时检测效果图

Fig. 9 Effect picture of physical image detection and video real-time detection

构上增加多维特征融合网络模块以及对 RPN 网络中锚框样式进行改进，同时采用单一变量法对不同改进部分进行实验；最后，训练出改进 Faster R-CNN 算法的瓶装饮料商品目标检测模型，并利用该模型进行实时性检测获得良好效果。需要说明的是，该模型对局部遮挡和目标跟踪问题表现欠佳，在今后工作中将继续探究和改进。

参 考 文 献

- [1] 杨坚争, 齐鹏程, 王婷婷. “新零售”背景下我国传统零售企业转型升级研究 [J]. 当代经济管理, 2018, 40(9): 24-31.
Yang JZ, Qi PC, Wang TT. Research on the transformation and upgrading of China's traditional retail enterprises under the background of "new retail" [J]. Contemporary Economic Management, 2018, 40(9): 24-31.
- [2] 余京蕾. 浅谈计算机视觉技术进展及其新兴应用 [J]. 北京联合大学学报, 2020, 34(1): 63-69.
Yu JL. A study on the progress of computer vision and its newly developing applications [J]. Journal of Beijing Union University, 2020, 34(1): 63-69.
- [3] 方路平, 何杭江, 周国民. 目标检测算法研究综述 [J]. 计算机工程与应用, 2018, 54(13): 11-18, 33.
Fang LP, He HJ, Zhou GM. Research overview of object detection methods [J]. Computer Engineering and Application, 2018, 54(13): 11-18, 33.
- [4] Zhang Y, Zhao Z, Xu HJ, et al. Vibration pattern recognition using a compressed histogram of oriented gradients for snoring source analysis [J]. Bio-Medical Materials and Engineering, 2020, 31(3): 143-155.
- [5] Ojala T, Näppi J, Nevalainen O. Accurate segmentation of the breast region from digitized mammograms [J]. Computerized Medical Imaging and Graphics, 2001, 25(1): 47-59.
- [6] 蔺海峰, 马宇峰, 宋涛. 基于 SIFT 特征目标跟踪算法研究 [J]. 自动化学报, 2010, 36(8): 1204-1208.
Lin HF, Ma YF, Song T. Research on object tracking algorithm based on SIFT [J]. Acta Automatica Sinica, 2010, 36(8): 1204-1208.
- [7] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C] // European Conference on Computer Vision, 2016: 21-37.
- [10] Ren SQ, He KM, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

- [11] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [12] 方仁渊, 王敏. 基于改进型 YOLO 网络的商品包装类型检测 [J]. *电子测量技术*, 2020, 43(7): 108-112.
Fang RY, Wang M. Retail product packaging type detection based on improved YOLO network [J]. *Electronic Measurement Technology*, 2020, 43(7): 108-112.
- [13] 廖珈艺, 张新阳, 代娱乐. 基于 TensorFlow 的商品陈列货架目标检测算法研究与实现 [J]. *现代计算机*, 2019(35): 41-47.
Liao JY, Zhang XY, Dai YL. Research and implementation of object detection algorithm for commodity display shelves based on TensorFlow [J]. *Modern Computer*, 2019(35): 41-47.
- [14] 周立旺, 潘天翔, 杨泽曦, 等. 多阶段优化的小目标聚焦检测 [J]. *图学学报*, 2020, 41(1): 93-99.
Zhou LW, Pan TX, Yang ZX, et al. FocusNet: coarse-to-fine small object detection network [J]. *Journal of Graphics*, 2020, 41(1): 93-99.
- [15] 胡正委, 朱明. 基于改进 Faster RCNN 与 Grabcut 的商品图像检测 [J]. *计算机系统应用*, 2018, 27(11): 128-135.
Hu ZW, Zhu M. Product image detection method based on improved faster RCNN and Grabcut [J]. *Computer Systems & Applications*, 2018, 27(11): 128-135.
- [16] 梅舒欢, 闵巍庆, 刘林虎, 等. 基于 Faster R-CNN 的食品图像检索和分类 [J]. *南京信息工程大学学报(自然科学版)*, 2017, 9(6): 635-641.
Mei SH, Min WQ, Liu LH, et al. Faster R-CNN based food image retrieval and classification [J]. *Journal of Nanjing University of Information Science and Technology (Natural Science Edition)*, 2017, 9(6): 635-641.
- [17] 王振. 基于目标检测的快消品识别研究与应用 [D]. 长沙: 湖南大学, 2018.
Wang Z. Research and application of FMCG detection on object detection [D]. Changsha: Hunan University, 2018.
- [18] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks [C] // *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012: 1097-1105.
- [19] 张瑞. 基于卷积神经网络的 Logo 检测与识别研究 [D]. 杭州: 浙江大学, 2015.
Zhang R. Research on Logo detection and recognition based on convolutional neural networks [D]. Hangzhou: Zhejiang University, 2015.
- [20] 罗辉. 基于深度学习的智能饮料售卖机商品识别 [D]. 成都: 电子科技大学, 2020.
Luo H. Product identification of intelligent beverage vending machine based on deep learning [D]. Chengdu: University of Electronic Science and Technology, 2020.
- [21] 袁豪. 基于深度神经网络的便利店商品检测与识别算法研究 [D]. 西安: 西安电子科技大学, 2019.
Yuan H. Research on convenience store commodity detection and recognition algorithm based on deep neural network [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2019.