计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(11):128−135 [doi: 10.15888/j.cnki.csa.006631]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于改进 Faster RCNN 与 Grabcut 的商品图像检测①

胡正委, 朱　明

(中国科学技术大学 信息科学与技术学院, 合肥 230031)
通讯作者: 胡正委, E-mail: zhengwei_hu@163.com

摘　要: 近年来, 图像检测方法已经被应用于很多领域. 然而, 这些方法都需要在目标任务上进行大量边框标注数据的重新训练. 本文基于 Faster RCNN 方法, 并对其进行改进, 解决了在小数据且无需边框标注的情况下的商品图像检测问题. 首先对 Faster RCNN 的边框回归层进行改进, 提出了一种非类别特异性的边框回归层, 仅使用公开数据集训练, 无需在目标数据集上进行再训练, 并将其用于数据预标定与商品检测. 然后结合 Grabcut 与非类别特异性 Faster RCNN 提出了一种样本增强方法, 用来生成包含多个商品的训练图像; 并为 Faster RCNN 添加了重识别层, 提高了检测精度.

关键词: 商品检测; Faster RCNN; Grabcut; 重识别层; 边框标注

引用格式: 胡正委,朱明.基于改进 Faster RCNN 与 Grabcut 的商品图像检测.计算机系统应用,2018,27(11):128–135. http://www.c-s-a.org.cn/1003-3254/6631.html

Product Image Detection Method Based on Improved Faster RCNN and Grabcut

HU Zheng-Wei, ZHU Ming

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230031, China)

Abstract: In recent years, object detection has been applied to many fields. However, retraining using large amount of bounding-box labeled data is needed. This study improves the Faster RCNN method and solves the problem of detecting multi-object in images using few-shot single object training data without bounding-box annotation. We propose a non-classwise bounding-box regression layer, which is only trained by public dataset and used for product training image labeling and testing image detection. Combined with Grabcut method, a data augmentation method is proposed to generate multi-object product training image. The improved faster RCNN model is re-trained by these images. In addition, a re-identification layer is added to the Faster RCNN architecture and improves the detection performance.

Key words: product detection; Faster RCNN; Grabcut; re-identification layer; bounding-box label

近年来, 深度学习被用于很多领域. 如人脸识别, 对象检测等[1–4]. 深度学习可以从大量的数据捕捉有用的信息, 同时也由于大数据时代的来临与计算设备性能的提高, 使深度学习的应用成为现实. VGG[1,5]模型起初被用于物体识别, 然后被扩展到人脸识别等任务中. GoogleNet[6]提出的 Inception 结构, 使用多个小尺寸的卷积核来代替大尺寸卷积核, 减少了网络参数, 也提高了模型性能. Resnet[7]的提出很大程度上改进了传统的网络结构, 并在 Imagenet 上获得最好的效果. 通过迁移学习[8], 这些网络模型也已经被广泛地应用到了各种识别任务中. 同时, 对象识别任务逐渐扩展到对象检测任务, 即从单个对象的分类扩展到了多个对象的分

① 基金项目: 中科院先导专项课题 (XDA06011203)
Foundation item: Strategic Priority Project of Chinese Academy of Sciences (XDA06011203)
收稿时间: 2018-03-27; 修改时间: 2018-04-23, 2018-05-04; 采用时间: 2018-05-08; csa 在线出版时间: 2018-10-24

类和定位. Faster RCNN[3]及其扩展版本成为近年来最有效的方法之一.

然而若将上述方法迁移到新的任务中,都需要在新的任务上使用大量标定数据重新调整模型. 但是, 在实际场景中, 数据的标定是非常难的工作, 需要消耗很大的财力物力和人力[9]. Vinyals, Koch[10,11]等人开始致力于研究如何在仅有少量训练样本甚至没有训练样本时进行分类工作. 然而相比较于分类任务的数据标定, 检测任务中边框定位数据的标定更难获得. 如何在没有边框标定数据的情况下, 将分类任务迁移到检测任务也是目前研究的难点.

本文提出一种方法, 解决了数据瓶颈问题, 可以在无需边框标定的情况下进行商品的检测定位. 本文构建的数据集中, 训练图像仅包含单个商品, 且没有边框标定, 而测试图像中包含多个商品. 本文首先对 Faster RCNN 进行改进, 提出非类别特异性 Faster RCNN, 并结合迁移学习, 对训练数据进行预标定; 然后结合 Grabcut[12]无监督方法, 对训练数据进行样本增强, 生成逼真的多个物体的训练图像; 然后再对非类别特异性 Faster RCNN 进行训练, 使其可以检测多个物体; 最后提出基于 Faster RCNN 的重识别方法, 在 Faster RCNN 中添加重识别层, 来提高多个物体检测精度.

1 非类别特异性 Faster RCNN

传统的 Faster RCNN 包括两部分, 如图 1 所示: 区域候选网络 (RPN) 和头网络 (Network Head). 其中区域候选网络为头网络提供 Feature Map(特征图) 和 ROI(感兴趣区域). 头网络利用 ROIAlign/ROIPool[13]从特征图中提取特定 ROI 的特征, 并利用分类层和边框回归进行物体分类和边框回归.

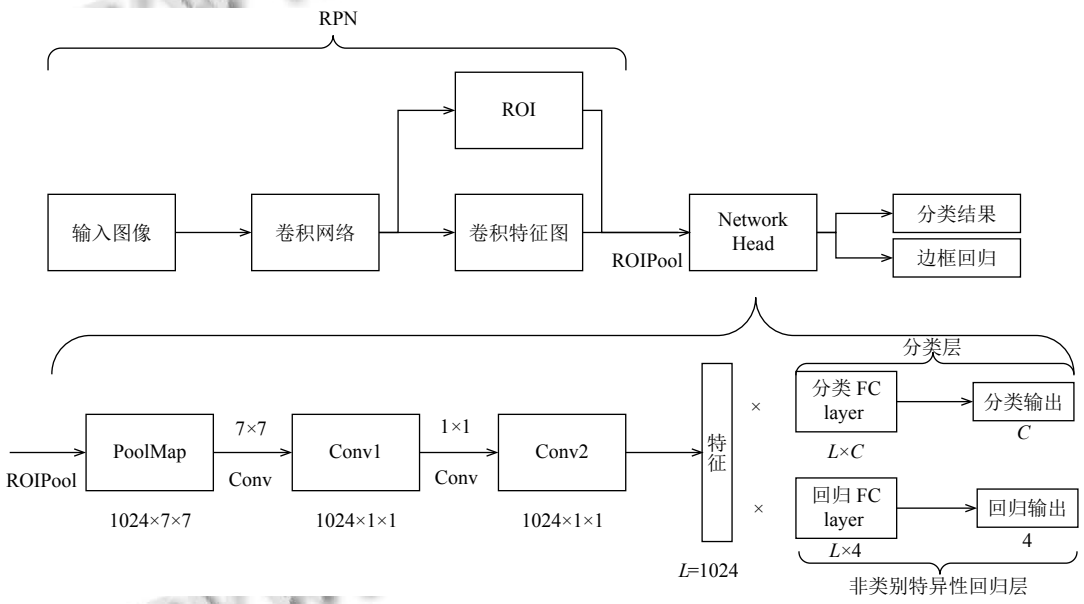


图 1 Faster RCNN 框架

其中分类层的输出维度为分类类别数目 C 与背景类别, 即 C+1. 回归层输出维度为类别数目的 4 倍, 即 $4C$, 为预测的每个类别的中心坐标与长宽. 本文将其中的回归层称为类别特异性的回归, 因为其对每个类别都会产生对应的边框预测. 然而在预测时, 只有 1 个类别的边框会被选中. 因此其它类别的边框回归结果可以认为是一定程度的冗余. 而且不同的数据集具有不同的类别数目, 则相应的边框回归层也需要采集大量的边框标定数据进行训练.

本文对回归层进行改进, 提出非类别特异性回归层. 其中, 回归层输出维度与类别无关, 输出维度由 $4C$ 改为 4. 并在公开的 COCO[14]数据集上进行训练, 然后迁移到商品数据集上, 且不需要再训练. 一方面减少了模型的冗余, 另一方面解决了数据的瓶颈问题.

本文利用提出的非类别特异性 Faster RCNN, 对构建的商品数据集中的训练图像进行预标定. 其中训练图像仅包含单个商品. 因为实际应用中, 单个商品的图像非常容易采集, 而边框的标定工作则需要大量的成

本. 如图 2 显示了本文提出的非类别特异性 Faster RCNN 对训练图片的预标定效果.

图 2 非类别特异性 Faster RCNN 边框预标定

2 结合 Grabcut 的样本增强方法

非类别特异性的 Faster RCNN 解决了边框回归的问题, 但是在训练数据中只有单个商品, 而测试图片中有多个商品时, 分类问题是目前很难解决的. 即使可以用单个商品来训练分类模型, 但测试图像中的多个商品会存在边框重叠甚至遮挡的情况. 在训练图像没有出现商品重叠的情况下进行模型的训练, 会使模型的分类能力不够鲁棒, 不足以识别边框重叠甚至遮挡的商品, 增大了分类的难度.因此本文提出一种样本增强方法, 通过对单个商品训练图片的处理, 来生成具有多个商品的训练图像.

通过类别非特异性的 Faster RCNN, 可以得到单个商品图像中商品的边框. 一个直接的想法是将图像中的商品边框部分提取出来, 经过旋转或者平移之后, 与其它商品进行组合, 如图 3 所示.

图 3 仅基于边框预标定的样本增强

然而, 这种方法会导致商品边框中的背景区域会覆盖其它商品区域, 与实际图片相差很大. 因此仅利用商品图像的边框不足以完成逼真的样本生成. 若能获得商品的精确区域信息, 例如商品对象掩码, 则可以分离出背景区域, 解决生成的样本中商品被背景遮挡的问题. 因此本文利用 Gabcut 方法对训练图像中商品进行分割.

Grabcut 在用户交互的基础上, 分别为背景和前景构建了 GMM(高斯混合模型). 背景和前景模型都分别包含K个 GMM 函数. RGB 图像中的像素集定义为z_n, 图像中每个像素都有一个高斯函数标记, 表示一个 GMM 函数. 最终组成了向量$\mathbf{k}=k_1,\cdots,k_n, k_n\in 1,\cdots,K$. 每个像素还有一个表示是否为背景的标记$a_n$, 值为 0 表示背景, 为 1 表示前景. 并定义能量函数:

$$E(a,\mathbf{k},\theta,z)=U(a,\mathbf{k},\theta,z)+V(a,z) \tag{1}$$

其中, U 定义为:

$$U(a,\mathbf{k},\theta,z)=\sum_n D(a_n,k_n,\theta,z) \tag{2}$$

其中, $D(a_n,k_n,\theta,z)=-\log p(z_n|a_n,k_n,\theta)-\log(a_n,k_n)$. $p()$表示高斯函数$\pi()$是高斯混合模型中的混合权重系数. 因D表示为:

$$\begin{aligned}D(a_n,k_n,\theta,z)=&\log\pi(a_n,k_n)+\frac{1}{2}\det\sigma(a_n,k_n)\\&+\frac{1}{2}[z_n-\mu(a_n,k_n)]^{\mathrm{T}}\sigma(a_n,k_n)^{-1}[z_n-\mu(a_n,k_n)]\end{aligned} \tag{3}$$

参数$\theta=\{\pi(a,k),\mu(a,k),\sigma(a,k)\}$, 因为对于前景和背景分别有$K$个高斯函数, 所以$\theta$对应$2K$个高斯函数的权重$\pi$, 均值$\mu$和协方差矩阵$\sigma$.

Grabcut 与 Graphcut[15]方法都是交互式图像分割方法. 其中 Graphcut 需要在交互时提供精确前景和背景像素种子区域, 并计算其它像素与前景和背景的相似度, 利用图论算法计算最佳分割. 而 Grabcut 算法的用户交互较少, 仅需要提供一个包含前景的矩形边框, 分割步骤如下:

(1) 通过用户交互提供前景对象的边框, 将边框外的区域初始化为T_B, 边框内的区域初始化为不确定区域$T_U=\overline{T_B}$, 前景区域初始化为空$T_F=\emptyset$. 对于区域T_B总的像素n, 置$a_n=0$, 区域T_U中像素n, 置$a_n=1$. 然后分别对$a_n=1$和$a_n=0$的像素分别初始化K个高斯函数.

(2) 对区域T_U内的像素n重新分配高斯成分$k_n=\arg\min_{k_n} D(a_n,k_n,\theta,z)$.

(3) 从图像数据z中学习高斯函数的参数$\theta=\arg\min_{\theta} U(a,\mathbf{k},\theta,z)$.

(4) 使用最小割算法最小化能量函数E, 重新计算区域分配T_B,T_U.

(5) 重复步骤 (2), 直至E收敛.

步骤 (2) 直接列计算每个像素对应的k_n. 步骤 (3) 进行高斯参数的估计. 对于一个确定的 GMM 组成

部分k, 前景像素的集合$F(k)=\{z_n : k_n=k \& a_n=1\}$也是确定的. 然后根据这些参数估计参数. 步骤 (3) 使用最小割进行全局优化. Grabcut 算法可以保证收敛, 因为步骤 (2) 和 (4) 均使能量函数E趋向减小的方向. 因此E是单调递减的, 所以算法至少会收敛到局部最小值. 当能量函数E趋于不变时, 算法停止.

在利用本文提出的非类别特异性 Faster RCNN 之后, 可以获得训练图像中单个商品的矩形边框, 因此只需要再结合 Grabcut 算法, 对商品的精确区域进行分割. 然后再将训练集中的单个商品区域进行随机旋转和平移, 并进行随机组合, 即可生成多个商品的训练图像, 效果如图 4 所示. 值得注意的是, 考虑到数据的准确性, 商品之间不能完全覆盖, 因为如果商品过度覆盖, 会导致区域内的真实商品几乎被覆盖, 而占大面积区域的商品与实际标签不符, 这样会误导识别模型的训练. 因此随机组合时需要对商品的重叠区域面积进行约束, 假设重叠面积的上限为s_{up}. 考虑三种融合策略: (1) 进行随机旋转和平移, 仅约束重叠面积的上限, 即$s_c \leqslant s_{up}$,即融合时, 商品可能会距离较远, 这是$s_c=0$. (2) 在限制重叠面积上限的同时, 对重叠面积的下限做约束, 即$s_c \geqslant 0$. 这种方案使商品之间必须有重叠, 保证了商品之间的距离较近, 但又没有大面积覆盖. (3) 增大对重叠面积下限的约束, 即$s_c \geqslant s_m$, 这样做是为了商品之间重叠的可能性更大, 并通过模型的训练来区分重叠的情况.

图 4 结合 Grabcut 的样本增强

3 重识别层

Faster RCNN 是一种两级 (tow stage) 方法. 第一级由 RPN(候选区域网络) 先筛选出候选区域, 过滤掉一部分背景区域. 第二级由头网络对候选区域进行细分类, 同时对每个候选区域的边框进行矫正, 即边框回归. 显然 RPN 提取的候选区域是不精确的, 这会影响头网络识别准确度. 因此本文提出重识别层, 来提高 Faster RCNN 识别的准确度.

因为经过头网络的边框回归层之后的边框位置会更精确, 这里的边框回归层为本文提出的非类别特异性回归方法. 而且头网络的分类层又过滤了一大部分背景区域. 本文将利用头网络回归之后的精确区域, 并结合 ROIAlign 方法, 对这些区域作为输入, 再一次经过头网络的分类层. 如图 5 所示.

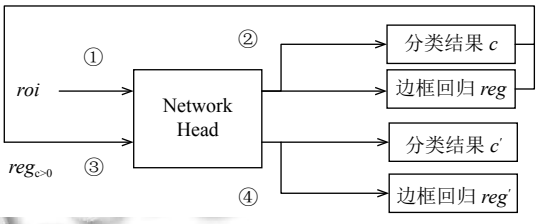


图 5 重识别层模型

传统 Faster RCNN 可以定义为:

$$c = f_{cls}(roi) \tag{4}$$

$$reg = f_{reg}(roi) \tag{5}$$

其中, roi表示 RPN 所产生的候选区域, f_{cls}表示分类层, f_{reg}表示回归层, c表示候选区域所对应的分类结果, reg表示候选区域的边框回归结果.

所添加的重识别层, 选出候选区域中被分类为非背景的区域, 背景类别用 0 表示, 然后将其回归边框作为新的候选区域再进行分类和回归, 表示为:

$$c' = f_{cls}(reg_{c>0}) \tag{6}$$

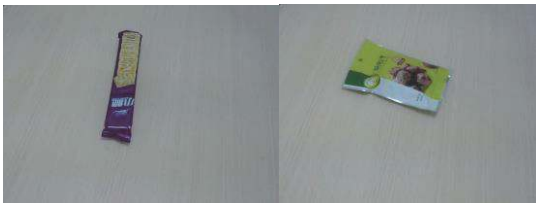
$$reg' = f_{reg}(reg_{c>0}) \tag{7}$$

其中, c'表示重识别后的最终分类结果, reg'表示最终的边框回归结果.

4 实验分析

4.1 数据集

本文在构建的商品数据集上验证了提出了方法. 本文提出的数据集如图 6 所示. 利用本文的方法. 我们不需要在商品数据集上训练边框回归. 所以本文构建的商品数据集训练图像仅包含类别信息. 训练集中共包含 3200 张训练图像, 400 张测试图像, 共计 40 个商品类别. 训练图像通过 2 个摄像头在 4 个不同的视角下拍摄的, 每张图像只有一个商品对象. 而测试图像包含多个商品, 使用另外一个摄像头拍摄, 且图像中的商品位置, 角度多样, 且包含跨背景的测试图像, 如购物车背景下采集的测试图像.



(a) 商品图像训练集示例

(b) 商品图像测试集示例

图 6 本文构建的商品数据集图片示例

所提出的非类别特异性 Faster RCNN 是在 COCO 数据集训练完成的, 并直接应用于商品训练图像数据的预标注. COCO 数据集共 80 个类别, 并包括非常多的图片以及边框和类别标注.

本文构建的商品数据集和 COCO 数据集主要的区别在于, 本文商品数据集中的物体时可旋转的, 且训练数据远远少于 COCO 数据集. 且本文构建的数据集中, 训练图片仅包含单个商品, 且不需边框标定.

4.2 网络模型训练

本文提出的非类别特异性回归层来改进原始 Faster RCNN 的类别特异性回归层, 组成非特异性 Faster RCNN 模型. 希望能够从公开数据集中学习边框回归知识, 并直接应用于单个商品训练图片的预标注. (1) 首先利用原始的 FasterRCNN 在 COCO 上进行训练, 其主干网络中 Resnet 模型使用 ImageNet 预训练模型, 然后进行分类层和回归层的训练, 最后进行整体网络模型的联合训练. 这样做是为了使的模型从 COCO 数据集中学习到有效的特征泛化能力. (2) 将训练好的 Faster RCNN 模型中的边框回归层改为所提出的非类别特异性回归层, 其他部分的参数保持不变, 仅在 COCO 数据集训练非类别特异性回归层. (3) 对于新的非类别特异性 Faster RCNN, 再使用 COCO 数据集调优整个网络. 这样是为了使 ROIAlign 得到的特征可以兼顾分类和边框回归的能力. 通过以上步骤训练完成的模型可以直接用于商品训练图像的标注. 训练非类别特异性 Faster RCNN 时, 其基本超参数设置如表 1 所示.

表 1 非类别特异性 Faster RCNN 基本参数设置

参数	参数值
初始学习率	0.001
Momentum	0.9
权重衰减	0.0001
每张图片 ROI 数量	2000
ROI 正负样本比例	1:3
bach size	1
每次循环训练次数	1000
最大训练循环次数	300

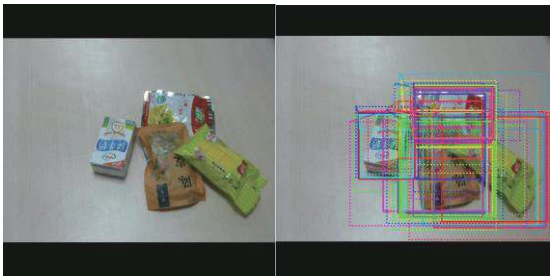
通过非类别特异性 Faster RCNN 与 Grabcut 的结合, 可以生成大量的多个商品图像样本. 并用于整体模型的训练. 训练步骤如下: (1) 首先, 在用于训练样本预标定的非类别特异性 Faster RCNN 的参数基础上进行训练, 保持主干网络和非类别特异性回归层参数不变, 仅训练分类层模型. (2) 然后保持非类别特异性回归层参数不变, 同时训练 RPN 网络的分类层和回归层, 以及步骤 (1) 中的分类层. (3) 训练整个网络, 包括主干网络中的 Resnet 参数, 仅固定非类别特异性回归层. 这是因为主干网络中的特征由 COCO 训练完成, 为了使其更好地提取商品数据中的特征, 需要对其主干网络参数进行训练. 通过以上步骤, 利用生成的样本训练完成的模型可以用于真实商品图像的检测任务. 而且其中的非类别特异性回归层无需在目标数据集进行再训练, 更加印证了其知识迁移的能力.

在实验中, 我们使用 Faster RCNN 的扩展版 Mask RCNN, 其除 Faster RCNN 方法外利用了特征金字塔网络 (FPN) 和兴趣区域对齐 (ROIAlign) 方法[6]. 在训练过程中不需要进行分割预测, 因此我们移除了 Mask RCNN 中的分割分支, 只使用其分类和回归分支. 实验在 2 个 NVIDIA TITANX GPU 上进行. 初始的学习率为 0.001, 并在训练时手动调整. 动量参数 Momentum 为 0.9.

4.3 实验结果

首先利用 COCO 数据集训练 Faster RCNN 中的分类分支与本文提出的非类别特异性回归分支. 如图 7 所示, 图中的黑色边界是由于 Mask RCNN[13]方法中的零填充 (Zero Padding) 导致的, Mask RCNN 方法是 Faster RCNN 的扩展版本. 其中虚线代表候选区域, 实线代表候选区域对应的边框回归结果, 可以看出本文提出的非类别特异性回归分支可以预测候选区域的真实边框, 而且商品周围的候选区域对应的回归边框趋向于同一位置. 同时相较于传统 Faster RCNN, 回归层

参数量减少很多, 且不需再训练即可迁移到其它数据集.



(a) 输入图像 (b) 候选区域与位置修正

图 7 候选区域与其对应的回归边框

提取训练图像的商品边框之后, 结合 Grabcut 算法对商品区域进行分割. 因为训练图像包含大面积背景, 若直接利用 Grabcut 算法对原始训练图像进行分割, 其分割效果非常不理想. 因为没有边框来标定图像的背景区域, 一般取图像的最外围的像素作为背景. 然而其所占面积非常小, 很难对整个背景进行建模. 在结合本文提出的非类别特异性 Faster RCNN 预标注算法与 Grabcut 算法进行训练集的商品图像分割. 然后使用简单的图像处理方法, 来生成多个商品的训练图像, 用于 Faster RNN 模型的训练, 生成的图像数据如图 4 所示. 在使用 Grabcut 算法进行图像生成时, 对象重叠面积上限s_{up}设置为 10 000. 对于不同的重叠面积s_c, 本文对其效果进行了对比. 如图 8, 当重叠面积为 0 时, 即商品距离较远时, 效果不好, 因为商品距离较远, 很难出现折叠等情况, 使得网络得不到折叠情况的训练, 所以效果相对较差. 当重叠面积为 6000 时, 模型的召回率 (Recall) 和精度 (Precision) 分别达到 93.8% 和 96.3%, 效果最好. 重叠面积过大时, 会使商品之间大面积覆盖, 会倾向于误导网络误识别.

模型在进行识别和定位时, 对于每个区域都会输出其对应类别的概率, 在进行模型的布置时通常需要对概率进行阈值化, 过滤概率低的预测, 保留概率高的结果. 因此, 我们分析了不同的阈值对于模型的召回率和精度的影响, 如图 9. 一般情况下概率阈值越高, 精度越高, 召回率越低. 概率阈值越低, 精度越低, 召回率越高. 图 9 中, 概率阈值为 0.3 时, 我们的模型能同时达到较高的精度和召回率, 这是因为模型对类别的预测概率较高, 低阈值对其影响不大, 模型预测能力强. 本文为了权衡准确率和召回率, 确定概率阈值为 0.7, 这是召回率为 93.8%, 精度为 96.3%.

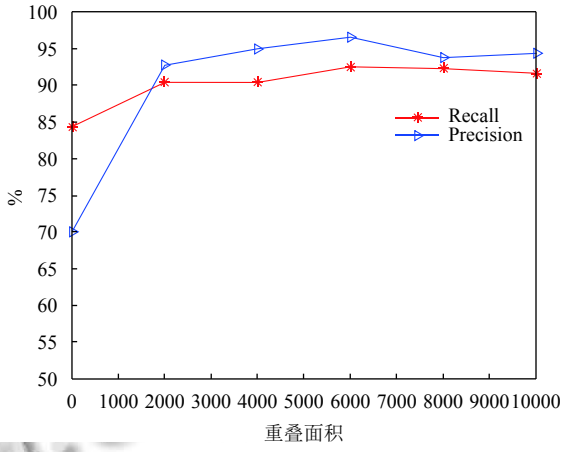


图 8 不同组合策略的检测结果

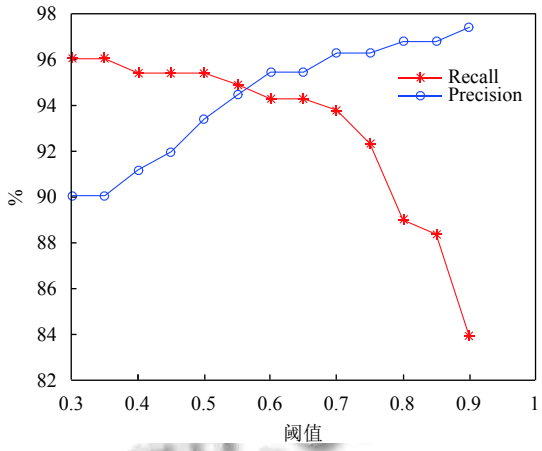


图 9 不同概率阈值的检测效果

如表 2, 我们通过对模型各部分进行分析, 所提出的结合 Grabcut 的样本增强方法, 使检测召回率提升超过 40%, 精确度提升了 30%.为了提高多个商品检测的精度, 本文提出了重识别层, 将分类与回归之后的候选区域, 经过边框回归层矫正之后, 再次输入分类层. 经过非特异性边框回归层的矫正, 可以有效避免候选区域不精确带来的分类误差. 在使用重识别层时, 比不使用重识别层时召回率提高了 3%, 精率提高了 4%.

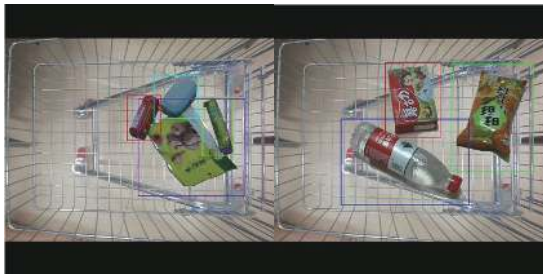
表 2 本文所提方法各部分效果分析

方法	Recall(%)	Precision(%)
非类别特异性 Faster RCNN	58.13	50.95
非类别特异性 Faster RCNN+Grabcut	90.42	92.75
非类别特异性 Faster RCNN+Grabcut+重识别	93.80	96.29

因为所提出的非类别特异性 Faster RCNN 可以检

测到单个商品边框, 当应用到多个商品的检测时, 其主要问题是当商品边框内存在其他商品的区域时会对识别造成干扰. 而非类别特异性 Faster RCNN 的边框回归不受多个商品的影响. 因此, 在使用本文生成的多个商品的训练数据进行训练时, 仅训练分类层的参数, 同时保持非类别特异性回归层参数不变. 检测结果如图 10 所示, 通过所提出的图像增强技术, 实现了多个商品的检测, 且其中的非类别特异性回归层仅使用公开数据集训练, 并学习到了回归知识, 且迁移到商品图像检测时并不需要再训练.

(a) 购物平台数据检测结果



(b) 购物车数据检测结果

图 10 商品检测结果

本文在构建的商品数据集中量化验证了所提出的方法. 由于本文旨在解决数据瓶颈问题. 所构建的训练数据集中仅有类别标签, 没有边框标定. 这种情况下, 传统的图像检测方法一般使用无监督的 SIFT[16]特征, 计算被检索图像的局部特征并与训练集中图像的特征做相似度匹配. 目前效果最好的深度学习方法, 如 VGG16[1]、VGG19[1]、Xception[17]、Resnet[7], 一般将其视为多标签分类任务进行识别. 本章对这些方法进行了比较. 如表 3 所示, SIFT 和其它目前最优的深度学习方法的性能明显低于本文所提出的方法. 一方面 SIFT 没有区别背景特征, 从而导致背景特征影响了匹配效果; 另一方面其为无监督人工特征, 在识别效果上不及有监督方法, 而且商品包装会有严重的反光, 也使其特征性能较低. 其它深度学习方法由于从单个商品训练图像推广到多个商品训练图像时, 没有学习到多个商品相近时的区分信息, 同时也没有区别背景特征, 因此识别率低. 而有些深度学习模型如 VGG16 和 VGG19, 其效果和 SIFT 相差不大, 这是因为从单个商品训练图像到多个商品识别与定位这种跨任务识别任务使得深度学习模型性能很低. 而本文方法通过提出一种无需目标数据集训练的样本标注以及样本增强方法, 可以利用单个商品的训练图像来学习到多个商品的区分信息, 起到了跨任务的桥梁作用, 对性能有很大提升.

表 3 不同的方法对比

方法	Recall(%)	Precision(%)
SIFT	33.47	20.30
VGG16	41.21	29.38
VGG19	36.50	26.25
Xception	58.50	42.50
Resnet	58.92	43.75
本文方法	93.80	96.29

5 结论与展望

本文基于 Faster RCNN 提出了一种非类别特异性的边框回归层, 仅使用公开数据集训练, 无需在目标数据集上进行再训练, 并将其用于数据预标定与商品检测. 同时结合 Grabcut 与非类别特异性 Faster RCNN 提出了一种样本增强方法, 来生成包含多个商品的训练图像, 用于模型的训练; 并为 Faster RCNN 添加了重识别层, 提高了检测精度. 未来, 我们将致力于研究如何在没有数据标定的情况下, 将本文方法拓展到图像分割领域.

参考文献

1 Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. Proceedings of the British Machine Vision Conference. Swansea, UK. 2015.

2 Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 815–823.

3 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91–99.

4 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 21–37. [doi: 10.1007/978-3-319-46448-0_2]

5 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.

6 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015.

7 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.

8 Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359. [doi: 10.1109/TKDE.2009.191.]

9 Dai JF, He KM, Sun J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1635–1643.

10 Vinyals O, Blundell C, Lillicrap T, *et al*. Matching networks for one shot learning. Proceedings of the 30th Conference on Neural Information Processing Systems. Barcelona, Spain. 2016. 3630–3638.

11 Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. ICML 2015 Deep Learning Workshop. Lille, France. 2015. 2.

12 Rother C, Kolmogorov V, Blake A. “Grabcut”: Interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics, 2004, 23(3): 309–314. [doi: 10.1145/1186562.1015720]

13 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. 2017. 2980–2988. [doi: 10.1109/ICCV.2017.322].

14 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 740–755. [doi: 10.1007/978-3-319-10602-1_48].

15 Yi FL, Moon I. Image segmentation: A survey of graph-cut methods. Proceedings of 2012 International Conference on Systems and Informatics (ICSAI2012). Yantai, China. 2012. 1936–1941. [doi: 10.1109/ICSAI.2012.6223428].

16 Lowe DG. Object recognition from local scale-invariant features. Computer vision, 1999. Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece. 1999. 1150–1157. [doi: 10.1109/ICCV.1999.790410].

17 Carreira J, Madeira H, Silva JG. Xception: A technique for the experimental evaluation of dependability in modern computers. IEEE Transactions on Software Engineering, 1998, 24(2): 125–136. [doi: 10.1109/32.666826]