

基于无标签半监督学习的商品识别方法

刘文豪 姜胜明

(上海海事大学信息工程学院 上海 201306)

摘要 深度学习商品识别技术存在有标签数据获取成本高、无标签数据获取简单高效的问题,给技术的实际应用造成了一定障碍。为解决在少量有标签数据下用大量无标签数据提高模型准确率的问题,提出一种半监督学习方法。使用基于 self-training 的方式训练两个目标检测模型;在检测模型的损失函数中加入基于熵的正则化项,提高无标签预测的准确性;为避免错误伪标签数据不断被使用,采取 co-training 的训练方式。通过构建包含 16 000 幅图片的商品识别数据集,最终实验结果表明,相比没有利用无标签数据之前,SSD 模型的精度从 90.6% 提高到了 97.2%,YOLOv3 模型精度从 91.7% 提高到了 96.9%。

关键词 半监督学习 商品识别 目标检测 无标签数据 伪标签

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2022.07.026

PRODUCT IDENTIFICATION METHOD BASED ON UNLABELED SEMI-SUPERVISED LEARNING

Liu Wenhao Jiang Shengming

(College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract Commodity recognition technology based on deep learning has the problems of high cost of acquiring labeled data and simple and efficient acquisition of unlabeled data, which poses certain obstacles to the practical application of the technology. In order to solve the problem of using a large amount of unlabeled data to improve the accuracy of the model under a small amount of labeled data, we propose a semi-supervised learning method. We used self-training-based method to train two target detection models. The entropy-based regularization term was added to the loss function of the detection model to improve the accuracy of label free prediction. To avoid the continuous use of false pseudo-label data, we adopted the co-training method. We constructed the product recognition data set containing 16 000 images, and carried out the experiment on it. The final experimental results show that the accuracy of SSD model has increased from 90.6% to 97.2%, and the accuracy of YOLOv3 model has been improved from 91.7% to 96.9% compared to that without unlabeled data.

Keywords Semi-supervised learning Product recognition Object detection Unlabeled data Pseudo-label

0 引言

新零售^[1]是最近几年的一个热门话题,一般采用条形码来进行商品识别,然后通过人工结算,在一定程度上提高了结算效率。然而随着电子射频技术的发展,RFID 技术开始在商品识别中应用^[2],该技术是在商品上粘贴一个电子标签,然后通过 RFID 读取器来

读取商品的详细信息,最后在后台自动进行结算,由于零售行业对成本比较敏感,现实生活中,该技术只能在部分场景下应用。

自深度学习技术在 2012 年的 imageNet^[3] 图像识别大赛上取得了突破性的进展之后,深度学习不管是在理论研究,还是在实践应用中都取得了长足的发展。随着新零售概念日益流行,文献[4-5]提出了利用深度学习技术来解决商品识别的问题,不但可以降低成本,还能

做到全自动化结算,为用户带来了全新的购物体验^[6]。

现有基于深度学习的目标检测^[7]技术主要是识别图片中物体的种类以及确定物体位置信息,常见的有 one-stage^[8-11] 和 two-stage^[12,13] 两大类检测算法。但这两类算法本身需要大量训练样本,只有训练样本数量足够多、多样性足够广,才能使模型准确率足够高。然而在商品识别中,图像数据往往是复杂多变的,例如数据采集困难、标注需要大量的人工、无标签数据大量存在等问题,都在一定程度上影响了目标检测技术在商品识别中的应用。Tonioni 等^[5,14] 提出的基于深度学习的货架商品识别技术是目前最优秀的解决方案,利用深度学习技术识别物体,使用全局描述符搜索匹配物体,但是没有考虑到实际情况下数据集的问题。通常货架商品数据的获取是通过部署在货架上某个固定位置的摄像头,使用网络将数据传输到后台服务器,该途径是一种比较便捷常规的方式,会产生海量的样本数据。然而只是简单的图像数据并不能直接被神经网络利用,需要额外的人工标注,来确定图像中商品的种类和位置信息,这是一种强度高、劳动密集的工作,往往需要多人耗时数周乃至数月的时间来完成。现有的深度学习商品识别方案中就是使用已经标注好的数据来训练模型,最后将模型部署进行预测,严重依赖已标注好的数据。

此外,在实际应用中,也要考虑到被识别物品本身的变化,例如在该套系统下添加一个新的产品、更换产品的包装等一系列变化。因此,要让已有的模型快速地利用无标签的海量数据来适应各种变化,降低训练的成本,构建一套适用的商品识别数据集^[15] 也成为了要解决的问题。

针对商品识别中的问题,即有标签数据集获取成本较高,无标签数据大量存在的问题。在特征分布上,无论是有标签数据还是无标签数据都具有相同的特征分布,为充分利用无标签数据来提高模型的特征学习能力,本文提出了一种半监督学习^[16] 框架,具体是:

1) 使用少量的有标签数据分别训练 SSD^[8] 和 YOLOv3^[9] 两个目标检测模型,在达到稳定之后,通过 self-training 的方式来利用无标签数据;

2) 在 self-training 阶段中,对目标检测网络损失函数添加基于熵的正则化损失项,使预测的结果更加准确。

3) 解决 self-training 阶段中由于错误伪标签数据重复利用导致模型精度变糟糕问题。采用 co-training 训练方法,将无标签数据分别在两个不同的模型上进行预测,产生一些伪标签数据,再将伪标签结果分别在另一个模型上训练,不断迭代,使其能够从数据中学习更多的特征信息。

1 数据集

将基于 RFID 的无人售货柜进行改装,在每层加装两个摄像头,我们选用了市面上销售量比较大的 20 种饮料,模拟实际使用场景,20 种类别的饮料被放置在同一售货柜的不同层,这里选用其中的一层进行说明,其他层具有相同的场景设置。

对于每一层,经过测试发现,在该硬件条件下能够放置 8 列饮料,这一特点受到摄像头的镜头、层高、箱体宽度等因素影响。考虑到实际应用过程,每行饮料的类别是完全相同的,其中每个摄像头能够看到 4 种类型的饮料,每个摄像头看到的排列情况会因为实际情况不同会有所差别,但仍然可以看成是在同一个摄像头下拍摄的图片。图 1 展示了同一时刻左右两边两个摄像头看到的商品。



(a) 左侧摄像头图像

(b) 右侧摄像头图像

图 1 左右两侧同一时刻拍摄的画面

针对 20 种饮料,每个摄像头能够看到的是 4 种类型,根据数学中排列组合理论,我们可以得出全排列的结果是 116 280,即该数据集的所有解为 116 280 种结果。为了在较少标注工作量以及满足实际应用场景,并没有列举所有的结果,而是采集了 20 种饮料抽取 4 种组合的方式,即 4 845 种组合方式,作为基础数据集,并进行了标注,用于基础模型的训练。为了模拟实际应用中产生的其他的组合方式,我们随机采集了无重复的 10 000 多幅无标签数据,作为无标签数据集,图 2 展示了有标签数据和无标签数据不同。



(a) 使用 labelImg 标注的有标签数据



(b) 无标注的无标签数据

图 2 数据集展示

此外为了验证模型的训练效果,从 4 845 种组合中选取 480 种不同的组合方式,每种组合按照与训练集不同的排列方式进行采集,最后进行标注,形成标准训练集,表 1 展示了构建的数据集详细信息。

表 1 数据集信息

数据集名称	图片数量	标注方式
标准训练集 L	4 845	Labelimg
标准测试集 T	480	Labelimg
无标签数据集 U	11 236	无标注

2 SSD 和 YOLOv3 模型及初始化

SSD (Single Shot MultiBox Detector) 模型^[8]和 YOLOv3^[9]模型分别是刘伟在 2016 年提出和 Joseph Redmon 在 2018 年提出的 one-stage 深度学习目标检测模型,该类模型将目标检测中分类和位置回归整合到一个网络中,使模型的训练更加简单,相比于 two-stage 目标检测模型来说,在运行效率上快了许多,十分适合于商品检测对场景的要求。SSD 模型和 YOLOv3 模型使用了卷积神经网络中的多层特征,每层特征代表着不同的语义信息,总体上提高了模型的表达能力。

在商品识别中,除了对准确率等技术指标有要求之外,也应该充分考虑实际的使用场景中,对实时性的要求,SSD 和 YOLOv3 相对于 R-CNN 系列模型 FPS 普遍要高。这两个模型采用完全不同特征提取网络,SSD 使用 VGG-16,如表 2 所示,YOLOv3 使用 darknet-53,如表 3 所示。两个神经网络通过特征提取之后,将位置预测和类别预测整合到一个网络中,从而达到 one-stage 目标检测的目的。从深度学习抽象层理解的话,两个模型分别代表着不同的视角来学习数据集中

存在的特征,避免不同的特征在后期的训练中出现无效的重复学习。

表 2 SSD 提取网络特征参数

层数	类型	卷积信息
2 ×	Conv1	64 3 × 3
	Maxpool_1	
2 ×	Conv2	128 3 × 3
	Maxpool_2	
2 ×	Conv3	256 3 × 3
1 ×	Conv3_3	256 1 × 1
	Maxpool_3	
2 ×	Conv4	512 3 × 3
1 ×	Conv4_3	512 1 × 1
	Maxpool_4	
2 ×	Conv5	512 3 × 3
1 ×	Conv5_3	512 1 × 1
	Maxpool_5	
	Conv6(FC6)	1 024 3 × 3
	Conv7(FC7)	1 024 1 × 1
	Conv8_1	256 1 × 1
	Conv8_2	512 3 × 3/2
	Conv9_1	128 1 × 1
	Conv9_2	256 3 × 3/2
	Conv10_1	128 1 × 1
	Conv10_2	256 3 × 3/1
	Conv11_1	128 1 × 1
	Conv11_2	256 3 × 3/1

表 3 YOLOv3 特征提取网络模型参数

层数	类型	卷积信息	特征图大小
	Conv	32 3 × 3	256 × 256
	Conv	64 3 × 3/2	128 × 128
1 ×	Conv	32 1 × 1	
	Conv	64 3 × 3	
	Residual		128 × 128
	Conv	256 3 × 3/2	64 × 64
2 ×	Conv	128 1 × 1	
	Conv	256 3 × 3	
	Residual		64 × 64
	Conv	128 3 × 3/2	32 × 32
8 ×	Conv	64 1 × 1	
	Conv	128 3 × 3	
	Residual		32 × 32

续表 3

层数	类型	卷积信息	特征图大小
8 ×	Conv	512 3 × 3/2	16 × 16
	Conv	256 1 × 1	
	Conv	512 3 × 3	
	Residual		16 × 16
4 ×	Conv	1 024 3 × 3/2	8 × 8
	Conv	512 1 × 1	
	Conv	1 024 3 × 3	
	Residual		8 × 8

SSD 主要是使用 VGG 作为基础的网络,将全连接层(FC6 和 FC7)替换为卷积层,并在后面加上额外的卷积层,用于多层特征的预测。

dark-net 网络主要是由一系列的 1×1 和 3×3 的卷积层组成,每个卷积层后都会跟一个 BN 层和一个 LeakyReLU 层,一共有 53 层卷积(包括全连接层),所以称为 darknet-53。

分别利用 VOC 数据集训练好的权重参数,初始化 SSD 和 YOLOv3 模型,这种 pre-training 训练方式在目前的深度学习参数初始化中方法十分常见。将同样的训练集分别转换为 SSD 和 YOLOv3 可以识别的格式,按照作者原文中的训练方法对模型继续训练,直到训练结束,这种被称作有标签的全监督训练。

3 半监督学习框架

3.1 自训练

模型的初始化训练是一个利用有标签数据的全监督学习过程。为了提高大量无标签数据的利用率,从中提取到有用的特征信息,可以采用自训练的方式。先使用训练好的模型在无标签数据上进行预测,选择置信度比较高的伪标签数据加入到训练数据集中,直到无标签数据集中的数据不再变化为止,参考算法 1。

算法 1 Self-training 算法

假定有标签数据集为 $L = \{(x^r, y^r)\}_{r=1}^R$, 无标签数据为 $U = \{x^u\}_{u=1}^{R+U}$, 通常情况下 $U \gg L$ 。

Input: 训练集 U 和 L 、学习率 η

repeat:

1: 使用有标签数据集分别训练 SSD 和 YOLOv3 模型,得到

模型参数为 f_{SSD}^* 和 f_{YOLOv3}^* 。

2: 分别使用 f_{SSD}^* 和 f_{YOLOv3}^* 在无标签数据集上进行前向预测,获得伪标签数据 $\{(x^u, y^u)\}_{u=l}^{R+U}$ 。

3: 根据预测出来的结果,将 (x^u, y^u) 加入到 L 中,并将 x^u 从 U 中移除。

until: $U = \{x_u\}_{u=l}^{R+U}$ 中没有无标签数据或者无标签数据不再变化。

自训练的方式虽然能够在一定程度上利用无标签数据,但伪标签的选择一直是自训练的核心,本文中采用基于熵正则化的方法来选择伪标签数据。假设模型预测的结果呈现某种分布,如果在结果中比较集中于某个类别,那么就有大概率认为是属于这个类别,通常用熵表示分布的集中程度,分布越集中,则熵越小。用 y^u 代表模型预测中分类结果,则:

$$E(y^u) = - \sum_{m=1}^n y_u^m \ln(y_u^m)$$

在 SSD 和 YOLOv3 中对类别预测添加熵的正则化,并使得类比的熵越小越好,使用 y^r 是标准训练数据集中的真实标签, \hat{y}^r 是模型预测的结果, y^u 是模型在无标签数据集上预测的类别结果,模型的损失函数为:

$$L = \sum_{x^r} C(y^r, \hat{y}^r) + \lambda \sum_{x^u} E(y^u) \quad (1)$$

式中:等式右边第一项是针对有标签的损失函数,和全监督训练时使用的损失函数一样,第二项是针对无标签预测的类别熵, λ 为协调有标签数据和无标签数据对损失函数的影响系数,最后通过最小化损失函数求解模型的参数。

3.2 协同交叉训练

SSD 和 YOLOv3 模型产生了不同的标签数据,这些数据结果从一定意义上来说可信的,但即使可信度再高,依然有一部分是错误数据。为了避免模型重复使用错误的的数据,致使模型朝着糟糕的方向训练,引入了一种类似于对抗机制。如图 3 所示,使用自训练之后的模型在无标签数据集 U 下进行预测,SSD 和 YOLOv3 模型分别获得伪标签数据集 PLa 和 PLb ,将标准训练集分别与 PLa 和 PLb 合并,形成新的有标签数据集 NLa 和 NLb ,最后使用 SSD 模型在 NLb 下训练,使用 YOLOv3 模型在 NLa 数据集下训练,通过交叉协同训练直到满足迭代停止条件。

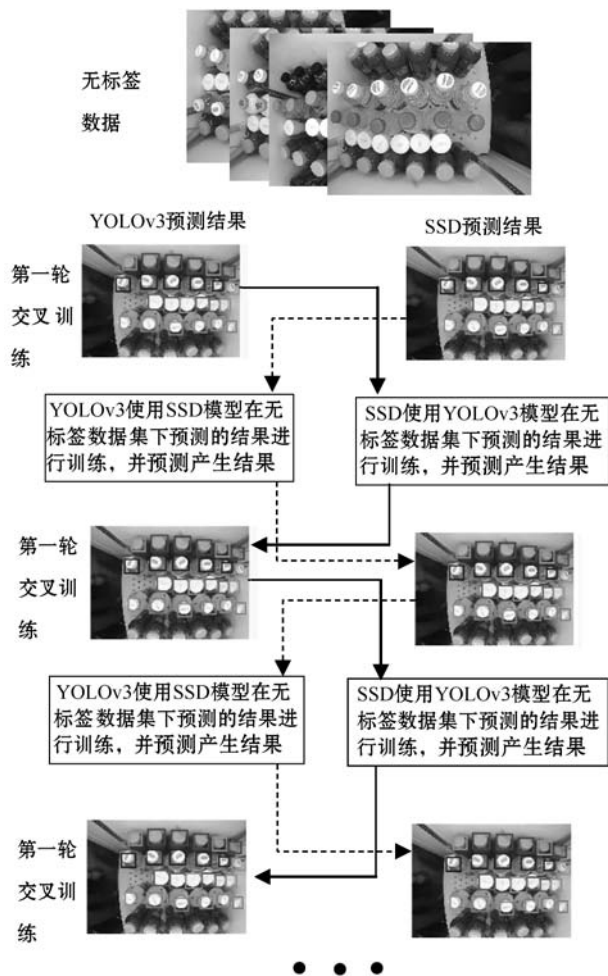


图 3 协同交叉训练流程

4 实验

4.1 模型的训练

为了快速验证半监督学习框架的有效性,在深度学习平台 Pytorch 框架下,使用 Python 代码编写 SSD 算法和 YOLOv3 算法,实验软件环境为 ubuntu 16.04 系统,硬件平台使用的是一块 GTX 2080Ti 11G 显卡, intel E5 处理器,模型的训练和测试都是在显卡 GPU 上进行,CPU 处理器主要完成简单的数据处理工作。

4.1.1 有标签数据的训练

为了使模型的初始训练能够获得更好的效果,两个模型在训练之前都采用了图像增强技术,如随机裁剪、图像归一化、色彩扭曲、随机水平翻转等。其中 SSD 模型的初始化参数使用基于 VOC 数据集权重结果,图片的输入大小为 300×300 ,采用 SGD 权重更新策略,学习率为 0.001,在整个数据集上迭代了 200 个 epoch 之后停止训练,YOLOv3 模型的初始化采用了官方^[9]权重结果,图片的输入大小为 416×416 ,同样经过了 200 个 epoch 的训练,采用了 adam 权重更新方

法,学习率为 0.001,在数据集迭代训练 200 次。

4.1.2 无标签数据的训练

无标签数据的训练主要分为两个部分,自训练和交叉协同训练。在自训练中,我们采用 4.1.1 节训练好的模型在无标签上进行预测,将结果和标准训练集合并成新的训练数据集,之后在新数据集下训练 50 个 epoch,学习率调整为更小的 0.0005,整个过程经过多轮迭代,完成 SSD 模型和 YOLOv3 模型训练。

由于自训练固有的缺陷,接下来采用交叉协同训练,其中主要是学习率的设置与前面有所不同,因为在训练的后期,主要是模型的微调,所以我们进一步降低学习率,将两个模型的学习率都调整为 0.0001,在达到 10 次交叉训练之后终止训练,每次交叉训练,模型迭代 15 个 epoch,最终经过 1 天左右的时间完成交叉协同训练。

4.2 实验结果及分析

采用原始的 SSD 模型和 YOLOv3 算法模型,对有标签的数据集进行了训练,随着迭代次数的增加,模型的损失函数不断下降,直到趋于稳定;对于无标签数据的训练分为自训练阶段和协同交叉训练阶段,加入自训练和协同交叉训练之后,模型的损失值进一步下降,图 4 展示了在数据集下不同阶段训练的结果。

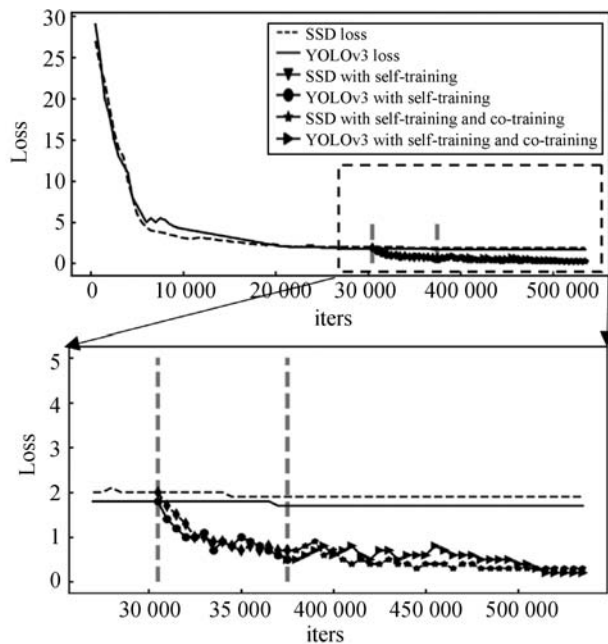


图 4 不同阶段训练损失值曲线

为了防止训练的模型存在过拟合的问题,需要在测试集上进行评估,首先需要评估的是模型在有标签数据集下的预测效果,我们采用在目标检测中常用的 mAP(mean average precision) 指标来确定网络模型分类和定位性能,半监督学习的检测也采取相同的评价指标,确保半监督学习前后评估方法的一致性,更加

科学合理地评估半监督学习算法。

和传统的分类任务类似,首先需要计算 precision 和 recall,这些指标依赖于真正例(TP)、假正例(FP)、真负例(TN)和假负例(FN)。 IOU (交并比)的设置决定检测结果是正确还是错误,最常用的阈值是 0.5。使用 P 代表 precision,使用 R 代表 recall,则:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

式中: TP 是根据 IOU 的值检测正确的数量; FP 为检测错误的数量; FN 表示漏检的数量。

根据对预测出来每种类别的所有物体概率进行排序,并取不同的阈值,得到每个类别的 PR 曲线,在每个类别的 R 上取 $[\frac{1}{N}, \dots, \frac{N}{N}]$ 时的最高 P 均值,就可得出该类别的 AP 值,表 4 展示了全监督学习下的 20 种饮料类别 AP 值。最终在测试集上,SSD 和 YOLOv3 算法分别达到了 90% mAP 和 91% mAP 的训练结果,两个模型具有基本相同的性能表现,能够在接下来的自训练和交叉训练中使用。

表 4 有标签数据集下模型 AP 值结果

饮料品种	SSD	YOLOv3
阿萨姆	0.949 1	0.950 4
贝纳颂	0.900 9	0.896 8
冰红茶	0.934 2	0.914 9
果粒橙	0.761 6	0.948 7
康师傅	0.761 2	0.930 6
美汁源	0.993 4	0.919 7
每日 C	0.953 4	0.919 7
果粒奶优	0.986 2	0.892 3
拿铁	0.768 6	0.905 8
能量王	0.961 5	0.910 3
农夫果园	0.909 3	0.899 1
农夫山泉	0.951 5	0.898 3
噗嗤	0.906 1	0.898 6
雀巢	0.936 8	0.890 7
营养快线	0.990 8	0.927 0
王老吉	0.660 8	0.787 6
小茗同学	0.937 1	0.959 8
怡宝	0.926 7	0.947 6
怡泉	0.955 6	0.861 7
可口可乐	0.985 9	0.982 9

在完成有标签的学习之后,采用 self-training 和 co-training 对伪标签数据进行学习,这些伪标签数据的学习效果直接影响着最终模型的性能。最后的验证部分,以是否使用伪标签为基准,分别对比使用伪标签前后模型的性能是否有所提高。伪标签性能验证时,每经过 10 个 epoch,在验证集 T 上进行一轮验证实验,图 5 和图 6 分别展示了 SSD 和 YOLOv3 模型使用伪标签前后的实验结果。实验发现,无论是 SSD 模型还是 YOLOv3 模型在使用伪标签数据之后,相比没有单纯使用有标签数据,模型的 mAP 值都有所提高。

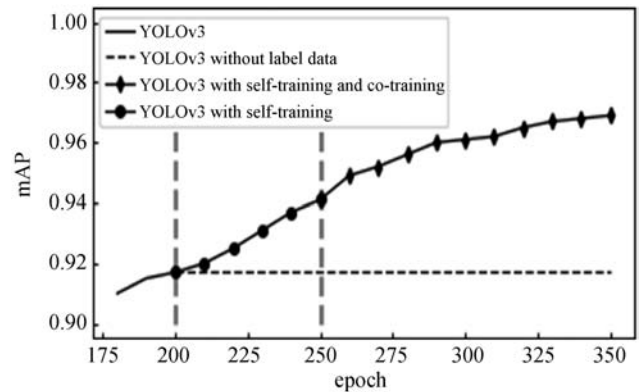


图 5 YOLOv3 使用伪标签前后性能结果

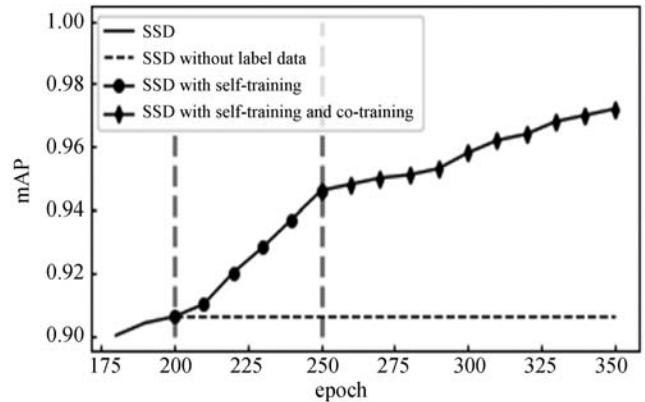


图 6 SSD 使用伪标签前后性能结果

在半监督学习框架下完成自训练和交叉协同训练,并在每个训练阶段的结束后,在测试集上进行一次性能测试,结果如表 5 所示。

表 5 半监督框架实验结果对比

模型名称	mAP/%	训练时间
SSD	90.6	15 h
YOLOv3	91.7	16 h
SSD + self-training	94.4	1 天
YOLOv3 + self-training	93.5	1 天
SSD + self-training + co-training	97.2	2 天
YOLOv3 + self-training + co-training	96.9	2 天

实验结果表明,在经过无标签数据集的训练后,

SSD模型的精度从90.6%提高到了97.2%,YOLOv3模型精度从91.7%提高到了96.9%,半监督学习框架相比全监督学习方法有明显提高,说明了半监督学习框架的有效性,在模型的预测中有很好的效果,能够准确地预测图片中商品的位置和类别,如图7所示。在实验环境中达到了60帧/s的实时处理效果,完全能够满足实际需要。

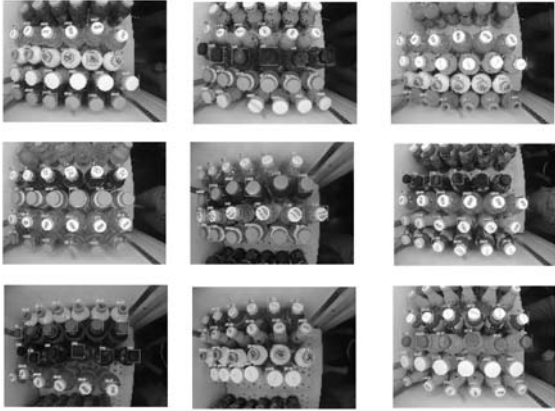


图7 半监督学习框架预测结果展示

5 结 语

基于半监督框架的商品识别算法在一定程度上解决了有标签数据不足、无标签数据无法利用等实际问题,为基于图像的商品识别应用提供了一种可行的解决思路。利用有标签数据进行预训练,再用无标签数据进行交叉训练,最终提高了模型的准确性,但依然存在一些问题,例如如何减少模型训练时长等,这些问题也将成为下一阶段的研究重点。未来无人售货这种新型的零售方式也将给人们带来全新的购物体验。

参 考 文 献

[1] 赵树梅,徐晓红.“新零售”的含义、模式及发展路径[J]. 中国流通经济,2017,31(5):12-20.

[2] Yi Q, Li P. Design and implementation on supermarket shopping guide system based on RFID and internet of things [J]. Journal of Computer Research and Development,2010, 47(2):350-354.

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems,2012:1097-1105.

[4] 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: 944309.

[5] Tonioni A, Serra E, Di Stefano L. A deep learning pipeline

for product recognition on store shelves[C]//2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS). IEEE, 2018: 25-31.

- [6] Alibaba Tech. New Meaning to Grab and Go; Alibaba Unveils its Unmanned Store[EB/OL]. 2018-11-07. <https://hackernoon.com/new-meaning-to-grab-and-go-alibaba-unveils-its-unmanned-store-48b9fce72a89>.
- [7] 张慧,王坤峰,王飞跃.深度学习在目标视觉检测中的应用进展与展望[J].自动化学报,2017,43(8):1289-1305.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European conference on computer vision. Springer, 2016: 21-37.
- [9] Redmon J, Farhadi A. YoloV3: An incremental improvement [EB]. arXiv:1804.02767, 2018.
- [10] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [11] 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.
- [12] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149.
- [13] Girshick R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [14] Tonioni A, Di Stefano L. Domain invariant hierarchical embedding for grocery products recognition [J]. Computer Vision and Image Understanding, 2019, 182:81-92.
- [15] Wei X S, Cui Q, Yang L, et al. RPC: A large-scale retail product checkout dataset [EB]. arXiv:1901.07249, 2019.
- [16] Zhu X, Goldberg A B. Introduction to semi-supervised learning [J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, 3(1): 1-130.
- ~~~~~
- (上接第43页)
- [22] Yu F, Koltun V, Funkhouser T. Dilated residual networks [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017.
- [23] Chu X, Yang W, Ouyang W, et al. Multi-context attention for human pose estimation [C]//IEEE Conference on Computer vision and Pattern Recognition, 2017.
- [24] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB]. arXiv:1502.03167, 2015.