doi:10.3969/j.issn.1672-1292.2019.04.017

基于卷积神经网络的仓储物体检测算法研究

王　飞[1],陈亮杰[2],王　梨[2],王　林[2]

(1.贵州民族大学人文科技学院,贵州 贵阳 550025)
(2.贵州民族大学数据科学与信息工程学院,贵州 贵阳 550025)

[**摘要**]　针对仓储环境中物体检测公开数据集匮乏的问题,通过摄像机采集真实仓储环境中包含货物、托盘和叉车的大量图像进行标注,创建了一个仓储物体数据集. 同时针对传统物体检测算法在仓储环境中检测准确率较低的问题,将基于卷积神经网络的 DSOD 应用于仓储环境中,通过在自己创建的仓储物体数据集上从零开始训练 DSOD 模型,实现了仓储物体的准确性检测. 该算法的 mAP 达到了 93.81%,比 Faster R-CNN、SSD 分别提高了 0.04%、1.44%;并且模型大小仅有 51.3 MB,比 Faster R-CNN、SSD 分别减小了 184.5 MB、43.4 MB. 实验结果表明,该算法获得了较为满意的仓储物体检测效果,其在仓储物体检测领域具有一定的实用价值.

[**关键词**]　卷积神经网络,仓储环境,物体检测,DSOD
[**中图分类号**]TP391.41　[**文献标志码**]A　[**文章编号**]1672-1292(2019)04-0099-07

Research on Warehouse Object Detection Algorithm Based on Convolutional Neural Network

Wang Fei[1],Chen Liangjie[2],Wang Li[2],Wang Lin[2]

(1.College of Humanities & Sciences of Guizhou Minzu University,Guiyang 550025,China)
(2.College of Data Science and Information Engineering,Guizhou Minzu University,Guiyang 550025,China)

Abstract:Considering the lack of public datasets for object detection based on the warehouse environment,a large number of images containing cargos,trays and forklifts in real warehouse environment are collected and labeled to build the warehouse object dataset. Meanwhile,aiming at the problem that the traditional object detection algorithm has lower detection accuracy in warehouse environment,the deeply supervised object detectors(DSOD) based on convolutional neural network is applied to the warehouse environment,and the DSOD model is trained from scratch on the self-built warehouse object dataset,and the accuracy detection of the warehouse object is realized. The mean Average Precision(mAP) of this algorithm reaches 93.81%,which is higher than that of Faster R-CNN and SSD by 0.04 and 1.44 points respectively,and the model size of this algorithm is only 51.3 MB,which is lower than that of Faster R-CNN and SSD by 184.5 MB and 43.4 MB respectively. The experimental results show that the algorithm has a relatively satisfying warehouse object detection effect,and it has certain practical values in the field of warehouse object detection.

Key words:convolutional neural network,warehouse environment,object detection,deeply supervised object detectors(DSOD)

物体检测是计算机视觉领域的热门研究课题之一,其在智能交通、视频监控、医疗诊断、工业检测和智能机器人等领域具有非常广泛的应用. 物体检测的主要任务就是定位出感兴趣物体在图像中的位置,并指出其所属类别名称.

物体检测算法主要包括传统物体检测算法和深度学习物体检测算法两大类. 传统物体检测算法采用“人工设计特征+分类器”的设计思路:首先提取物体的特征,然后将提取到的物体特征送入训练好的分类器进行分类. 例如,Haar-like 特征[1-2]+Adaboost 分类器[3]、方向梯度直方图(histogram of oriented gradient,

收稿日期:2019-07-05.
基金项目:贵州省教育厅创新群体重大研究项目(黔教合 KY 字[2018]018)、贵州省科技厅重点实验室(黔科合计 Z 字[2009]4002)、贵州民族大学人文科技学院基金科研项目(18rwjs016).
通讯联系人:王飞,助教,研究方向:图像处理、模式识别. E-mail:wangfei10248@163.com

HOG)[4]+支持向量机(support vector machine,SVM)分类器[5]、尺度不变特征变换(scale-invariant feature transform,SIFT)[6]+SVM 分类器以及可变形的组件模型(deformable part model,DPM)[7]等. 卷积神经网络(convolutional neural network,CNN)通过学习就能获得颜色、轮廓等底层的特征和更高级、抽象的特征. 与传统人工设计特征相比,基于 CNN 的物体检测算法对局部遮挡、光照条件和尺度变化等影响因素具有更好的鲁棒性和泛化能力,同时还能对多类物体进行检测.

近年来,基于 CNN 的仓储物体检测逐渐成为一个新兴的研究方向,其属于通用物体检测的范畴,而基于 CNN 的通用物体检测算法大体可以分为两大类:基于区域提议的算法和基于无提议的算法.

基于区域提议的物体检测算法采用两个阶段来解决物体检测问题. 该类算法中最具有代表性的是 R-CNN[8]、Fast R-CNN[9]以及 Faster R-CNN[10]等. R-CNN 首先采用选择性搜索(selective search,SS)[11]从图像中提取出物体区域,然后对提取出的物体区域进行分类. Fast R-CNN 和 Faster R-CNN 主要通过共享计算和采用神经网络生成区域提议来提高物体检测效率.

基于无提议的物体检测算法仅采用单个阶段来处理物体检测任务. 该类算法中最具有代表性的是 YOLO(you only look once)[12]和 SSD(single shot multibox detector)[13]. 此类算法去除区域提议阶段,采用单个前馈卷积神经网络来直接预测物体类别和具体位置.

当前,公开的物体检测数据集大多都是基于自然场景. 例如,常用的物体检测数据集 ImageNet、Pascal VOC(pascal visual object classes)均是在自然场景中采集的图像. 基于工业检测应用场景的公开数据集很少[14],而基于仓储环境中的物体检测公开数据集几乎没有. 因此,通过 CNN 来实现仓储环境中的物体检测,创建一个质量高且规模大的仓储物体数据集尤为重要.

本文通过摄像机采集真实仓储环境中包含货物、托盘和叉车的大量图像进行标注,创建了一个仓储物体数据集. 同时将基于 CNN 的 DSOD(deeply supervised object detectors)[15]应用于仓储环境中,通过在自己创建的仓储物体数据集上从零开始训练 DSOD 模型,实现了仓储环境中的货物、托盘、叉车检测.

1　卷积神经网络

CNN 是当前深度学习算法实现的主要途径和研究计算机视觉、图像处理以及人工智能等方向的重要工具之一,其主要由输入层、卷积层、池化层、全连接层和输出层构成,基本结构如图 1 所示.

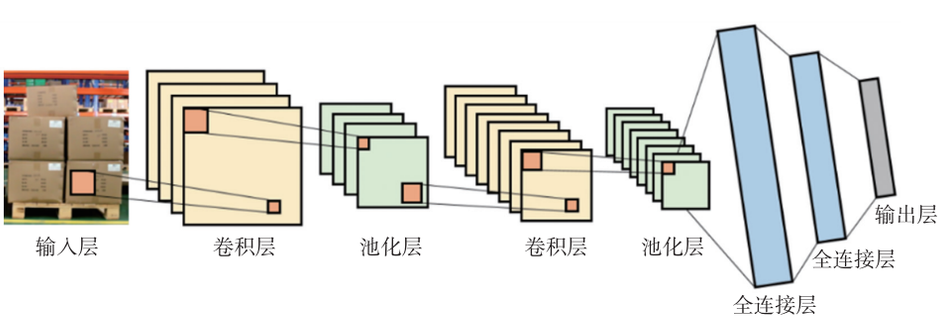

图 1　CNN 的基本结构

Fig. 1　Basic structure of CNN

1.1　卷积层

卷积层是对输入层输出的特征做卷积运算,再经过激活函数得到特征图,其作用是提取一个局部区域特征,每一个卷积核相当于一个特征提取器,第 l 层卷积层中第 j 个特征图的计算公式如下:

$$\begin{cases} u_j^l = \sum\limits_{i \in M_j} x_i^{l-1} * \boldsymbol{k}_{ij}^l + b_j^l, \\ x_j^l = f(u_j^l). \end{cases} \tag{1}$$

式中,$\boldsymbol{k}_{ij}^l$是卷积核矩阵,b_j^l 是卷积层的偏置,M_j 表示前一层输出的特征图集合,u_j^l 表示卷积层 l 层的第 j 个神经元,x_j^l 是 l 层卷积层的第 j 个通道的输出结果,$*$ 表示卷积运算,对于 l 层某一特征图 u_j^l,其对应的每个输入图的卷积核大小可以是不同的,卷积核的大小一般为 3×3 或 5×5. $f(\cdot)$代表非线性激活函数,常用的激活函数有:

$$\mathrm{Sigmoid}(x)=\frac{1}{1+\mathrm{e}^{-x}}. \tag{2}$$

$$\mathrm{Tanh}(x)=\frac{\mathrm{e}^{x}-\mathrm{e}^{-x}}{\mathrm{e}^{x}+\mathrm{e}^{-x}}. \tag{3}$$

$$\mathrm{ReLU}(x)=\max(0,x)=\begin{cases}0, & x\leqslant 0,\\ x, & x>0.\end{cases} \tag{4}$$

1.2　池化层

池化层是通过对不同位置的局部特征区域进行特征统计,实现特征合并和特征降维操作,以有效地缩小矩阵尺寸,从而减少参数数量,并且经过池化后的特征具有平移不变性,这对图像分类至关重要. 常用的池化操作有:最大池化、平均池化和随机池化,其原理如图 2 所示. 最大池化是选取图像区域的最大值作为该区域池化后的值;平均池化是计算图像区域的平均值作为该区域池化后的值;随机池化是对特征图中的元素按照其概率值大小随机选择,即元素值大的被选中的概率也大.

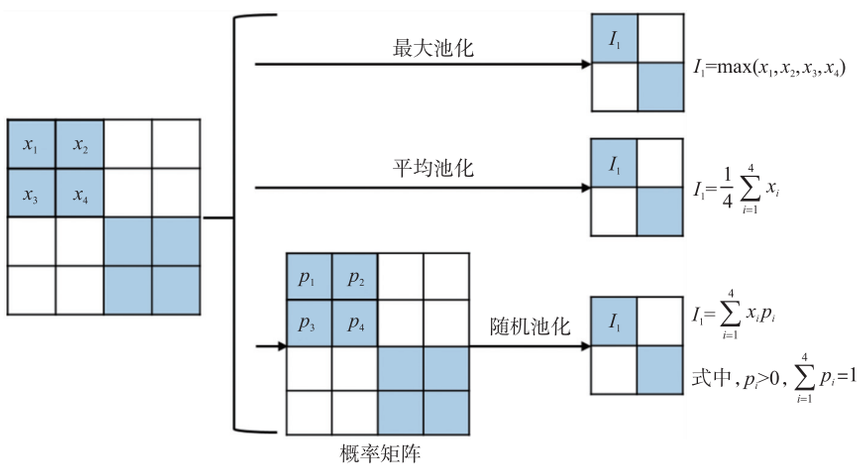

图 2　常用的池化操作原理

Fig. 2　Commonly used pooling operation principle

1.3　全连接层

在整个 CNN 中,全连接层类似于分类器. 卷积层和池化层的作用是将原始数据映射到隐含层特征空间中,而全连接层的作用是将 CNN 学习到的特征映射到样本的标记空间中,同时将卷积输出的 2 维特征图转化为一个 1 维向量. 最终,网络将这些特征输入到 Softmax 分类器中,并通过最小化损失函数来进行全局训练.

1.4　Softmax 分类器

分类是机器学习和人工智能领域的基本问题之一,例如,字符识别、图像识别、语音识别等都可以转化为分类问题. 逻辑回归(logistic regression,LR)是机器学习领域经典的二分类器,而 Softmax 回归是 LR 在多分类上的推广,其将分类问题转化成概率问题,即首先求解统计所有可能的概率,然后概率最大的就被认为是该类别.

假设有 m 个训练样本$\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\dots,(x^{(m)},y^{(m)})\}$,对于 Softmax 回归,它的输入特征为:$x^{(i)}\in R^{n+1}$,类标记为:$y^{(i)}\in\{0,1,\dots,k\}$. 假设函数是对于每一个样本估计其所属的类别的概率 $p(y=j|x)$,其具体表达式如下:

$$\boldsymbol{h}_{\theta}(x^{(i)})=\begin{bmatrix}p(y^{(i)}=1\mid x^{(i)};\theta)\\ p(y^{(i)}=2\mid x^{(i)};\theta)\\ \vdots\\ p(y^{(i)}=k\mid x^{(i)};\theta)\end{bmatrix}=\frac{1}{\sum_{j=1}^{k}\mathrm{e}^{\theta_j^{T}x^{(i)}}}\begin{bmatrix}\mathrm{e}^{\theta_1^{T}x^{(i)}}\\ \mathrm{e}^{\theta_2^{T}x^{(i)}}\\ \vdots\\ \mathrm{e}^{\theta_k^{T}x^{(i)}}\end{bmatrix}. \tag{5}$$

式中,$\theta_i(i=1,2,\dots,k)\in R^{n+1}$表示 Softmax 分类器的参数,$\frac{1}{\sum_{j=1}^{k}\mathrm{e}^{\theta_j^{T}x^{(i)}}}$表示归一化因子,其使所有类别概率和为 1.

对于每一个样本估计其所属类别的概率为:

$$p(y^{(i)}=j|x^{(i)};\theta)=\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}. \tag{6}$$

Softmax 分类器的训练过程就是不断地调整参数,优化损失函数,使损失函数达到最小值的过程.

2　基于 DSOD 的仓储物体检测算法

2.1　DSOD 的网络结构

基于 CNN 的 DSOD 框架类似于 SSD 多尺度无提议框架. 它的网络结构主要由用于特征提取的主干子网(backbone sub-network)和用于在多尺度响应图上进行预测的前端子网(front-end sub-network)两个部分构成. 主干子网是密集连接卷积网络(densely connected convolutional networks,DenseNet)[16]结构的变体,它由主干块(stem block),四个密集块(dense blocks),两个过渡层(transition layers)和两个无池化层的过渡层(transition w/o pooling layers)组成. 前端子网(即 DSOD 预测层)融合了多尺度预测密集连接结构.

2.2　DSOD 的设计准则

DSOD 最大的亮点在于其可以在相应的数据集上从零开始训练网络,而不需要 ImageNet 预训练模型. 这有效地避免了从预训练模型领域到物体检测领域可能存在巨大差异的困难,同时 DSOD 训练的模型比较小,在内存开销上也占有相对优势. DSOD 的设计原则如下:

(1)免提议. R-CNN 和 Fast R-CNN 需要额外的区域提议提取算法,如:选择性搜索(SS);Faster R-CNN 需要区域提议网络(region propose network,RPN)来生成相对较少的区域提议;YOLO 和 SSD 是不需要提议的算法,由于 DSOD 框架与 SSD 框架类似,故 DSOD 不需要区域提议.

(2)深度监督. 深度监督的核心思想是通过综合的目标函数来对浅层的隐含层进行直接监督,而不仅仅是输出层,附加在隐含层中的目标函数能有效缓解梯度消失问题. 免提议检测框架包含分类损失和位置损失,增加复杂的侧输出层是一个比较好的解决方案,以在每个隐藏层为检测任务引入附加目标函数. 根据文献[16]所述,使用 DenseNet 的密集层级连接,其中密集块是块中所有先前的层都连接到当前层. 这样,目标函数可以通过跳跃连接直接监督 DenseNet 中的浅层. 无池化层的过渡层,使得在增加密集块的数量时不会降低最终的特征图分辨率.

(3)主干块. 主干块由 3 个 3×3 的卷积层连接 1 个 2×2 的最大池化层组成,其目的是减少信息损失.

(4)密集预测结构. DSOD 与 SSD 的预测结构略有不同,在 SSD 的简单预测结构中,每个后续的尺度直接由相邻的前一尺度转换得到,而 DSOD 一方面会综合每个尺度获取的多尺度信息. 另一方面 DSOD 中的每个尺度,一半的特征图是通过之前尺度连接卷积层学习得到,剩余的一半由相邻的高分辨率的特征图直接下采样得到. 每个尺度只学习一半新的特征图,并复用前一个特征图剩余的一半. 因此,DSOD 的密集预测结构比 SSD 的简单预测结构产生的参数少.

2.3　DSOD 的损失函数

DSOD 的损失函数是位置损失(localization loss)和置信度损失(confidence loss)的加权和,其表达式为

$$L(x,c,l,g)=\frac{1}{N}(L_{\text{conf}}(x,c)+\alpha L_{\text{loc}}(x,l,g)). \tag{7}$$

式中,N 是匹配的默认框数目,l 和 g 分别是预测框和真值框,c 为多类置信度,α 为权重项. 位置损失如文献[9]所述,是 l 和 g 之间的 $smooth_{L_1}$ 损失,其表达式如式(8)所示. 类似于 Faster R-CNN,DSOD 回归得到边界框的中心及其宽度和高度的偏移.

位置损失:

$$L_{\text{loc}}(x,l,g)=\sum_{i\in \text{Pos}}^{N}\sum_{m\in\{cx,cy,w,h\}} x_{ij}^{k}\,smooth_{L_1}(l_i^m-\hat{g}_j^m). \tag{8}$$

式中,$smooth_{L_1}(x)$ 可以表示为:

$$smooth_{L_1}(x)=\begin{cases}0.5x^2 & \text{如果}|x|<1,\\ |x|-0.5 & \text{其他情况.}\end{cases} \tag{9}$$

置信度损失是多类置信度 c 下的 Softmax 损失,并且通过交叉验证将权重项 α 设置为 1,其表达式如式(10)所示.

置信度损失:

$$L_{\text{conf}}(x,c)=-\sum_{i\in\text{Pos}}^{N}x_{ij}^{p}\lg(\hat{c}_i^{p})-\sum_{i\in\text{Neg}}\lg(\hat{c}_i^{0}). \tag{10}$$

式中,$\hat{c}_i^{p}=\dfrac{\exp(c_i^{p})}{\sum_{p}\exp(c_i^{p})}$.

3　实验

3.1　实验数据采集和图像标注

本文实验所需的数据均是在真实仓储环境中通过摄像机采集获得的图像,其分辨率为 1 920×1 080. 为了减少模型的训练时间,以使网络更快地收敛,我们将图像的分辨率缩小了 3 倍,所以实验数据的最终分辨率为 640×360. 本文创建的仓储物体数据集共有 10 450 张图像,这些图像中包含货物、托盘和叉车 3 个物体类别,其中包含货物和托盘的图像共有 7 893 张,包含叉车的图像有 2 557 张.

图像标注采用图像标注工具 LabelImg,通过该工具可以直接将仓储物体用矩形框的形式标注出来,并且还能将标注的矩形框生成为可扩展标记语言(extensible markup language,XML)文本文件. XML 文件记录了标注矩形框的最小/最大坐标、宽、高以及物体类别等信息,以方便训练模型时直接读取图像信息. 仓储物体标注的实例图如图 3 所示.

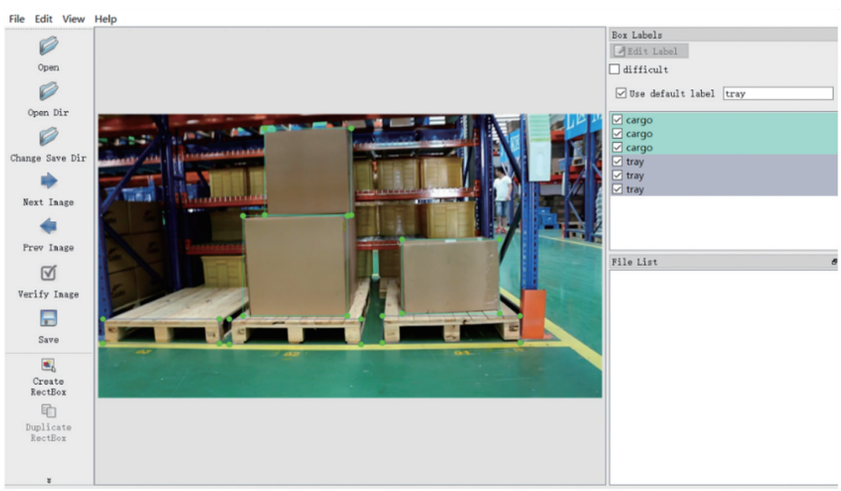

图 3　仓储物体标注的实例图

Fig. 3　Example of the warehouse object labeling

3.2　实验环境与模型训练

本文算法实验所需的软硬件配置如表 1 所示,并且使用了深度学习中的 Caffe 框架.

表 1　实验配置

Table 1　Experimental configuration

Operating System	CPU	Memory	GPU	CUDA
Ubuntu16.04	Intel i7-7700k	16 GB	NVIDIA TITAN X	CUDA8.0

在训练模型之前,我们首先按 4:1 的比例将自己创建的仓储物体数据集随机分为训练验证集和测试集,然后将训练验证集也按 4:1 的比例随机分为训练集和验证集. 其中,训练集有 6 688 张,测试集有 2 090 张,验证集有 1 672 张.

采用 CNN 训练模型时,设置学习率的大小尤为重要. 如果学习率设置过大,模型收敛速度太快,但是训练误差会出现震荡,无法收敛到全局最优值;反之,如果学习率设置得过小,网络收敛速度就会很慢,需要花费大量的时间才能达到最优.

本文中,我们将网络训练的迭代次数设置为 60 000 次,根据动态调整学习率的策略,我们将学习率具体设置为:(1)初始迭代阶段(0~20 000)设为 0.1;中间迭代阶段(20 001~40 000)设为 0.01;(3)最后优化阶段(40 001~60 000)设为 0.001. 同时,我们采用动量为 0.9,权重衰减为 0.000 5 以及批量尺度为 5 的随机梯度下降(stochastic gradient descent,SGD)在自己创建的仓储物体数据集上从零开始训练 DSOD 模型. 为了能更好地观察训练过程中损失(Loss)随迭代次数(Iterations)的变化情况,我们画出了 0~60 000 次的 Loss-iterations 曲线图,如图 4 所示.

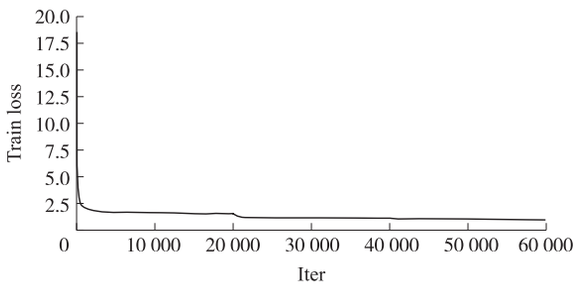

图 4　Loss-iterations 曲线图

Fig. 4　Loss-iterations curve

我们从图 4 可以观察到,Loss 曲线比较平滑,几乎没有出现震荡现象,说明我们设置的学习率比较适当,使模型达到了最优收敛的状态.

3.3　实验结果及分析

在物体检测领域中,平均准确率均值(mean Average Precision,mAP)是评价物体检测算法优劣的主要评价指标. 因此,我们同样采用 mAP 来评价本文仓储物体检测算法的优劣. 在相同训练集和测试集的情况下,本文算法与 Faster R-CNN、SSD 进行了比较,对比实验结果如表 2 所示.

表 2　不同算法对比实验结果

Table 2　Comparison of experimental results with different algorithms

算法	测试图像数量/张	模型大小/MB	mAP/%
Faster R-CNN	2 090	235.8	93.77
SSD	2 090	94.7	92.37
本文算法	2 090	51.3	93.81

由表 2 可知,本文算法在自己创建的仓储物体数据集上的 mAP 达到了 93.81%,比 Faster R-CNN 提高了 0.04%,比 SSD 提高了 1.44%. 并且模型大小仅有 51.3 MB,比 Faster R-CNN 减少了 184.5 MB,比 SSD 减少了 43.4 MB. 因此,本文算法能较好地满足仓储物体检测的准确性要求,并且模型占用内存小,适用于移动、嵌入式电子设备等低端设备.

我们给出部分仓储物体测试图像的检测效果图,如图 5 所示.

图 5　部分仓储物体测试图像的检测效果图

Fig. 5　Detection effect diagram of partial warehouse object test images

由图 5(a)、图 5(b)可知,对于不同颜色的货物,本文算法均能较好地检测出来. 由图 5(c)可知,当仓储环境中的光照条件发生变化时,本文算法也能较好地检测出货物、托盘. 由图 5(a)、图 5(d)可知,无论是尺寸大的货物、托盘,还是尺寸小的货物、托盘,本文算法同样能较好地检测出来. 由图 5(f)可知,对于局部遮挡的货物,本文算法依然能较好地检测出来. 因此,无论是光照条件、物体尺寸以及颜色等变化,还是存在局部遮挡,本文算法都能较好地检测出仓储物体,获得了较为满意的检测效果.

4　结语

基于仓储环境中的物体检测应用场景,我们创建了一个仓储物体数据集,将基于 CNN 的 DSOD 应用于仓储环境中,通过在自己创建的仓储物体数据集上从零开始训练 DSOD 模型,实现了仓储环境中的物体检测. 该算法的 mAP 比 Faster R-CNN、SSD 高,模型大小比 Faster R-CNN、SSD 小,并且对仓储物体颜色、尺寸大小以及光照条件等变化具有较好的鲁棒性,获得了较为满意的检测效果. 但本文算法对于尺寸较小的仓储物体仍然会出现漏检问题. 因此,下一步打算对 DSOD 的网络结构作相应的改进,并扩充仓储物体数据集,以进一步提升仓储物体检测准确率.

[参考文献](References)

[1] PAPAGEORGIOU C P,OREN M,POGGIO T. A general framework for object detection[C]//Sixth International Conference on Computer Vision. Bombay,India,1998.

[2] LIENHART R,MAYDT J. An extended set of Haar-like features for rapid object detection[C]//International Conference on Image Processing. Rochester,NY,USA,2002.

[3] FREUND Y,SCHAPIRE R E. A desicion-theoretic generalization of on-line learning and an application to boosting[J]. Journal of computer and system sciences,1997,55(1):119-139.

[4] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego,CA,USA,2005.

[5] SUYKENS J A K,VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural processing letters,1999,9(3):293-300.

[6] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision,2004,60(2):91-110.

[7] FELZENSZWALB P F,GIRSHICK R B,MCALLESTER D,et al. Object detection with discriminatively trained part based models[J]. IEEE transactions on pattern analysis and machine intelligence,2010,32(9):1627-1645.

[8] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus,OH,USA,2014.

[9] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision. Santiago,Chile,2015.

[10] 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 and machine intelligence,2017,39(6):1137-1149.

[11] UIJLINGS J R R,SANDE K E A V D,GEVERS T,et al. Selective search for object recognition[J]. International journal of computer vision,2013,104(2):154-171.

[12] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:unified,real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,NV,USA,2016.

[13] LIU W,ANGUELOV D,ERHAN D,et al. SSD:single shot MultiBox detector[C]//European Conference on Computer Vision. Amsterdam,Netherlands,2016:21-37.

[14] 李天剑,黄斌,刘江玉,等. 卷积神经网络物体检测算法在物流仓库中的应用[J]. 计算机工程,2018,44(6):176-181.
LI T J,HUANG B,LIU J Y,et al. Application of convolution neural network object detection algorithm in logistics warehouse[J]. Computer engineering,2018,44(6):176-181.(in Chinese)

[15] SHEN Z,LIU Z,LI J,et al. DSOD:learning deeply supervised object detectors from scratch[C]//IEEE International Conference on Computer Vision. Venice,Italy,2017.

[16] HUANG G,LIU Z,MAATEN L V D,et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Honolulu,Hawaii,USA,2017.

[责任编辑:陈　庆]