

基于 Faster RCNN 的生活垃圾智能识别

文灿华, 李佳*, 董雪

上海交通大学中英国际低碳学院, 上海 201306

摘要 利用 Faster RCNN 算法实现生活垃圾的高精度识别。选取典型的 6 种生活垃圾建立数据集, 采用数据增强方法提升了数据集目标数量及目标类别、尺度均衡性, 分析对比三种具有显著差异的主干网络 VGG-16、Res101、MobileNet_v1 的精度、速度及泛化性能。采用结合特异层微调的端到端训练策略, 对低识别率样本开展增强训练, 由此获得了最低为 92.85% 的均值平均精度 (mAP), 随后对误识别样本中提取的三种典型错误进行优化, 将最高 mAP 提高到 99.23%。此外, 设计含 816 张图片的背景数据集测试算法在多变背景下的泛化性能, 发现复杂垃圾背景对检测精度的影响最大, 且泛化性能与网络收敛性能趋势一致, 即从优到劣排序依次为 Res101、VGG-16、MobileNet_v1。最后, 基于可回收垃圾倾向高精度指标及有害垃圾倾向高召回率指标的原则, 分析并得到算法最优检测概率阈值的设置方法。

关键词 图像处理; 目标检测; 垃圾识别; 数据增强; 神经网络训练; 背景干扰

中图分类号 TP391.4; X705

文献标志码 A

doi: 10.3788/LOP57.201014

Intelligent Domestic Garbage Recognition Based on Faster RCNN

Wen Canhua, Li Jia*, Dong Xue

China-UK Low Carbon College, Shanghai Jiao Tong University, Shanghai, 201306, China

Abstract In this paper, we presented the Intelligent Domestic Garbage Recognition using Faster RCNN to realize high-precision identification of domestic garbage. Specifically, 6 kinds of domestic garbage were selected to build the dataset. The data augmentation technique was adopted to expand the quantity and category of the targets, and improve balance on the size of the targets. Moreover, we used three different types of backbone networks including VGG-16, Res101, and MobileNet_v1 to analyze and compare the accuracy, speed, and generalization performance. The research used end-to-end training network finely tuned by the special layer, and carried out enhanced training on low recognition rate samples to obtain a minimum mean average precision (mAP) of 92.85%. Subsequently, we captured three typical errors and optimized from the misidentified samples, and thus the highest recognition mAP increased to 99.23%. To analyze the generalization performance of different backbone networks embedded in the algorithm, we built a dataset with 816 pictures derivatized from the different backgrounds and used it to test the impact of changing the background on garbage detection. As a result, we found that the complex backgrounds from surrounding garbage put the greatest impact on detection accuracy. Thus, the generalization performance takes the same trend as convergence performance, which changes Res101, VGG-16, MobileNet_v1 from good to bad. Therefore, the setting method of the optimal probability threshold for algorithm detection was analyzed and obtained based on the principles of the high-precision requirement for recyclable garbage and high recall requirements for hazardous garbage.

Key words image processing; object detection; garbage recognition; data augmentation; neural network training; background interference

OCIS codes 100.3008; 150.1135; 100.4996

1 引言

近年来,深度卷积神经网络(CNN)在图像处理领域取得长足进步。2013年,RCNN^[1]算法将CNN引入目标检测领域,极大地提高了目标检测性能。目前,目标检测主流方法分为一阶段与二阶段两类,前者以YOLO^[2-4]、SSD^[5]算法为代表,后者以Faster RCNN^[6]及其变种为主,后者比前者多了一个感兴趣区域提取网络(RPN)^[6],以生成初选目标框,供最终分类及定位使用。基于CNN的目标检测技术在垃圾检测领域仅有少量研究,Chen等^[7]利用Faster RCNN算法检测流水线上的199个垃圾目标(仅bottle),得到系统漏识率为3%,误识率为9%。Mikami等^[8]利用SSD算法实时检测并跟踪视频中的垃圾袋目标,在NVIDIA TX2上实现了6.2 frame/s的检测速度,垃圾袋识别平均精度(AP)为0.62。Zhang等^[9]利用Faster RCNN算法检测681张街道图片中共9种类别的垃圾目标,检测的均值平均精度(mAP)为0.82,但存在类别不平衡问题。总体而言,基于CNN的目标检测技术在垃圾处理领域的研究不够充分,存在数据集规模较小、算法泛化性能及垃圾实际特性研究不足等问题,亟需进一步研究,以满足实际垃圾检测的需要。

本文采用二阶段的Faster RCNN算法实现对多类垃圾目标的高精度检测。由于目前目标检测领域没有针对生活垃圾的公开数据集,故研究的重点之一是制作数据集。考虑到不同的主干网络对识别精度、速度及泛化性能影响较大,实验对比测试了三种典型网络结构;此外,采用数据增强、困难样本增强训练、典型错误优化等多种手段提升算法性能;最后,对于背景干扰及算法实际检测中最佳概率阈值的设定问题,分析并提出解决方法。

2 数据获取方法与算法

2.1 实验装置及数据采集

实验装置如图1所示,主要包括计算机及拍摄系统:计算机安装了一块NVIDIA 1080Ti显卡,用于加速算法的训练及检测;拍摄系统包括4个摄像头及一个可调光照摄影棚。计算机通过程序控制4个摄像头从顶部、前部、后部及侧部四个角度不同距离拍摄图片并保存以供算法训练,这样获取的目标受光照影响较小,且角度、目标大小多样,依此训练的算法对不同角度出现的目标均有较强的检测能力。数据集选取可回收垃圾和有害垃圾中常见的

6种类别,其中可回收垃圾包含金属、塑料、纸盒三类,有害垃圾包含电池、灯管、药片三类。每一类目标,均来源于网络爬取的公开图片及拍摄系统实拍的图片。最终采集的图片总数6076张,其中实拍2566张,采集的典型图片样本见图2。

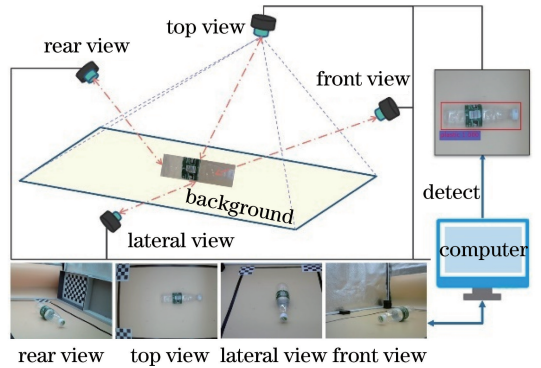


图1 实验装置

Fig. 1 Experimental equipment



图2 采集的各类别图片典型样本

Fig. 2 Typical image samples from each class

2.2 数据集建立与数据增强

图片标注格式采用开源目标检测数据集PASCAL VOC2007^[10]的格式,训练集与测试集的比例设置为8:2,各类别的图片按比例随机分配。为提升算法性能,此处采用一种新的数据增强手段,利用标注信息截取目标并随机旋转缩放后贴入图片,达到了平衡各类别目标数量并扩充目标尺度多样性的目的。具体实现如下:

- 1) 通过标注框将图片中标注的目标全部截取出来,组成单元目标库;
- 2) 对每一张图片,从备选单元目标图片库中随机选取其未涵盖的类别目标,将其随机缩放旋转后贴入图片以平衡各类别数据量,更新标注文件。
- 3) 单元目标贴入参数:交并比(IoU)取0,旋转

角度随机参数取 $0^\circ, 90^\circ, 180^\circ, 270^\circ$, 像素大小规格取 256, 160, 96, 48 pixel, 像素大小规格对应的像素大小随机参数取 64, 32, 32, 16 pixel, 如此贴入的目标大小多样, 覆盖 32~320 pixel 的尺度范围。

数据增强操作在测试集及训练集上分别进行, 最终数据集的图片数量为 6076, 各类别的目标

数量见表 1, 对比数据集目标数量, 可知数据增强后目标数量提升很大, 且各类别更为均衡。此外, Faster RCNN 算法在训练阶段提取感兴趣区域 (RoI) 时背景样本数量往往远大于前景样本数量, 数据增强后有利于缓解正负样本不均衡的问题, 提升训练效果。

表 1 数据集各类别目标数量

Table 1 Object quantity on garbage dataset

Dataset	Metal	Plastic	Carton	Battery	Bulb	Pill	Total
Original dataset	1321	1392	807	1058	1129	1402	7109
Augmented train dataset	2429	2435	1938	2389	2398	2528	14117
Augmented test dataset	630	585	483	597	617	630	3542
Augmented dataset	3059	3020	2421	2986	3015	3158	17659

2.3 Faster RCNN 算法

Faster RCNN 算法整体结构如图 3 所示。主干网络用于提取图片的高维特征图, RPN 用于提取

图片的 RoI。通过 RoIAlign^[11], 即用 RoI 截取特征图, 并采样到固定尺寸, 送入 Fast RCNN 完成最终 7 种目标 (6 种目标加一种背景) 的分类及定位。

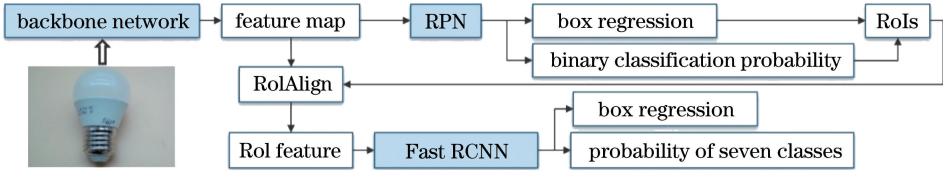


图 3 Faster RCNN 算法结构

Fig. 3 Network structure of Faster RCNN

训练时, Faster RCNN 通过 RPN 层损失 L_{RPN} 及 Fast RCNN 层损失 L_{Fast_RCNN} 反向传播优化网络参数, 其中 L_{RPN} 及 L_{Fast_RCNN} 均由分类损失及回归损失组成。算法主干网络采用经典结构 VGG-16^[12]、残差结构网络 Res101^[13]、轻量化结构网络 MobileNet_v1^[14], 以观察这三种典型的网络结构对算法识别精度、速度及泛化性能的影响。三种网络的参数量、计算量 (每秒浮点运算次数, FLOPs) 及网络层数见表 2, 计算时输入图片大小 600 pixel × 800 pixel ($H \times W$), RPN 通道数设为 512。

表 2 不同网络参数量、计算量及网络层数

Table 2 Number of parameters, FLOPs and layers for different networks

Network	Number of parameters / 10^7	Number of FLOPs / 10^{10}	Layer numbers
VGG-16	136.79	166.37	20
Res101	47.21	167.25	105
MobileNet_v1	5.61	19.02	32

2.4 Faster RCNN 算法训练策略

采用结合特异层微调的端到端训练策略, 在训练过程中提取低识别率样本, 构建困难样本数据集进行增强训练, 提高了算法对困难样本的识别精度。

端到端训练速度较文献[6]中采用的 RPN 及 Fast RCNN 网络交替训练的方式更快, 但精度略低, 结合特异层微调的端到端训练策略可较好地兼顾训练速度与训练精度的要求。具体实现如下。

1) 端到端训练: 反向传播网络总损失 L_{total} , L_{total} 定义为

$$L_{total} = L_{RPN} + L_{Fast_RCNN} \quad (1)$$

2) 微调 RPN: 固定除 RPN 外的全部网络参数, 反向传播 L_{RPN} 。

3) 微调 Fast RCNN: 固定除 Fast RCNN 外的全部网络参数, 反向传播 L_{Fast_RCNN} 。

最终训练分为 4 个步骤, 见图 4, 困难样本的筛选条件满足以下任一条件即可:

1) 用最新权重测试图片, 分析所有概率不小于 0.3 的检测目标, 结果中含有 FP (False Positive) 或 FN (False Negative);

2) 提取最近两个 epoch 的各图片损失数据, 满足

$$l \geq E_{loss} + k\sigma_{loss} \quad (2)$$

式中: E_{loss} 为 epoch 的损失均值; σ_{loss} 为 epoch 的损失标准差; k 为系数, 此处取 0.5。

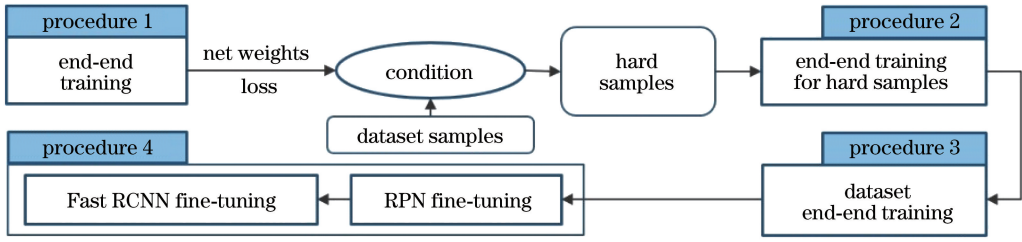


图4 结合困难样本增强及特异层微调的 Faster RCNN 算法训练步骤

Fig. 4 Faster RCNN train process combined with hard samples enhancement and special layer fine-tuning

学习率采用自动下降策略,初始学习率设为 0.0001,最小学习率设为 0.00001,学习率衰减比例为 0.35,学习率衰减条件为每 epoch 的 E_{loss} 下降率低于 0.05。为提高网络精度及收敛性,采用迁移学习策略,利用 ImageNet LSVRC^[15]数据集上的预训练权重初始化主干网络。

3 结果与讨论

3.1 Faster RCNN 算法训练及检测结果

网络训练过程的 epoch 平均损失曲线及各训练步骤在测试集上的 mAP 见图 5,从中可以看出:收敛趋势正确,其中按照收敛性从大到小排序依次为 Res101、VGG-16、MobileNet_v1,且 Res101 优势明显,这主要得益于残差结构更易于学习,网络深度更深;MobileNet_v1 网络深度与 VGG-16 相当,但可分离卷积会损失图像通道信息,导致收敛性较差。网络收敛值越小,一般精度越高,此处因为数据集规模偏小,Res101 与 VGG-16 的差异未反映出来。综合图5各训练步骤在测试集上的 mAP,可以看出困

难样本增强训练(训练步骤 2)、特异层微调(训练步骤 4)有效提升了网络识别精度。各网络检测结果见表 3,对比测试集与训练集数据,可认为不存在明显的过拟合现象;测试集及训练集中各类别 AP 较为均衡,可认为数据集在经过数据增强后实现了较好的类别均衡性。此外,从检测速度来看,MobileNet_v1 检测速度优势明显。

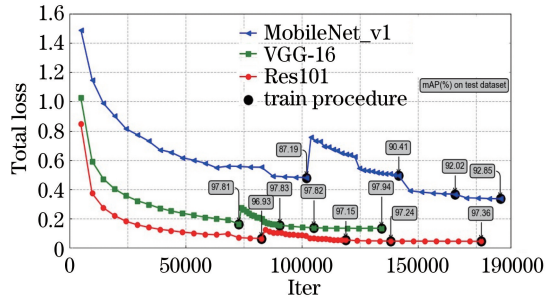


图5 各网络总损失收敛情况及各训练步骤在测试集上的 mAP

Fig. 5 Total loss convergence and mAP of test dataset during training procedure

表3 各网络训练集(TR)及测试集(TE)测试结果

Table 3 Network results on train dataset (TR) and test dataset (TE)

Backbone network		AP						mAP	Optimized mAP	Detection speed / (frame · s ⁻¹)
		Metal	Plastic	Carton	Battery	Pill	Bulb			
Res101	TR	1.0	0.9996	0.9985	0.9996	0.9973	1.0	0.9992	0.9993	~7
	TE	0.9770	0.9597	0.9817	0.9695	0.9728	0.9811	0.9736	0.9857	
VGG-16	TR	1.0	0.9997	0.9997	0.9996	0.9970	1.0	0.9993	0.9992	~9
	TE	0.9758	0.9639	0.9866	0.9835	0.9813	0.9853	0.9794	0.9923	
MobileNet_v1	TR	0.9817	0.9715	0.9732	0.9851	0.9831	0.9879	0.9804	0.9833	~20
	TE	0.9139	0.8737	0.9204	0.9408	0.9671	0.9554	0.9285	0.9490	

3.2 结果分析及优化

根据目标检测通用的评价手段,将检测目标划分为 TP(True Positive)、FP(False Positive)、TN(True Negative)、FN(False Negative)。检测完成后,对某一类别,将 TP、FP 按概率从大到小排序,取不同的概率阈值,可得相应的精度(P)及召回率

(R)。由此可得到各类别 PR 曲线,且 PR 曲线与概率曲线完全对应,涵盖目标检测评估的完备信息,优化的目的是消除 FP 并消除 FN,其中 MobileNet_v1 网络的金属类别如图 6 所示。

提取出各网络所有的 FP、FN,分析可得三类典型错误及相应的优化方法,如下所示:

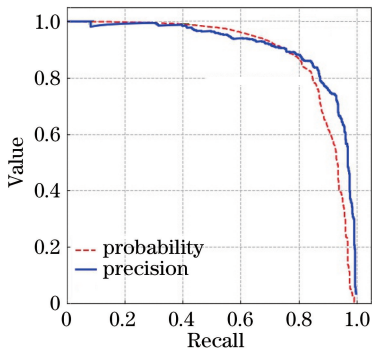


图6 MobileNet_v1网络的金属类别精度及概率曲线

Fig. 6 Precision curve and probability curve of metal on MobileNet_v1

1) 易混淆类别误检及背景物体误识导致FP。

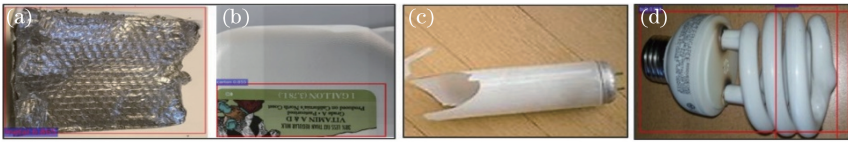


图7 误检实例。(a)易混淆特征误检；(b)背景物体误识别；(c)漏识别；(d)重复识别

Fig. 7 Error detection instances. (a) Misdetection of confusing features; (b) background misidentification; (c) missing detection; (d) duplicate detection

各网络优化后的 mAP 见表 3,对比优化前后数据可见,错误优化后性能得到有效提升。

3.3 泛化性能评估

实际垃圾检测时存在复杂的背景影响,故需引入背景干扰来测试算法的泛化性能。背景影响主要来自两方面:拍摄大背景变化,如传送带颜色、纹理不同;其他垃圾的干扰,这个主要由于目标检测的对象为图像框,若框内存在其他物体就会干扰原目标识别,同时可能误识背景为目标。据此,人为导入多种测试背景:216种纯色背景,15种纹理不同的传送带背景,15种密集散垃圾背景。为可靠评估小数据集下不同网络结构对不同背景的泛化能力,将已正确检测的目标引入不同背景中建立背景数据集,以测试网络泛化性能,方法如下:首先从训练集中取出300张检测无误且背景较为干净的源图片,由此保证算法检测的是图中目标而非背景;而后,将图片内目标轮廓标注并截取后贴入上述背景图中,由此制作出含原目标的新背景测试数据集,其中纯色背景216张,传送带背景300张,密集散垃圾背景300张。背景数据集样本示例见图8。

背景数据集的测试结果见表4,对比表中源图片集及背景测试集的 mAP,可得到背景对精度的影响大小:纯色背景影响较小,纹理次之,密集散垃圾背景的影响最大。原因在于:纯色背景灰度值固定,

如图7(a)和(b)所示,网络将金属锡箔纸误检为塑料,将塑料盒上的贴纸误检为纸盒,原因在于对象特征较为接近误检类别特征,且两类目标在训练集中数量差异大。优化方法为数据增强,即扩充误检对象的数量。

2) 漏识别导致 FN。如图7(c)所示,某灯管未被检测出来,原因在于在训练集中此对象数量较少。优化方法为数据增强。

3) 局部特征重复识别导致 FP。如图7(d)所示,灯管存在局部重复检测,原因在于训练集其他图片中含有不完整灯管样本,算法学习其他图片后会对局部特征进行识别。解决办法为进行后处理抑制,即若检测目标被同类其他目标覆盖则只取大目标。



图8 背景数据集样本示例

Fig. 8 Samples of background dataset

其影响极易在网络浅层中消除;纹理特征相对复杂,但对纹理特征响应较大的层级仍处于中低层级^[16],对深层网络特征影响有限,另外由于数据集中各类别目标并不具备明显的纹理特征,神经网络在学习目标特征的过程中对纹理的响应也相对较小,因此纹理特征对精度的影响不大;密集分布的散垃圾背景丰富复杂,对检测精度的影响最大。

背景干扰对各网络检测精度影响的大小反映了各网络泛化性能的优劣,此性能对算法的实际应用较为关键。对比表4中各网络的测试 mAP 数据,可知泛化性能与收敛性能趋势一致,Res101 泛化性能大大优于其余两者,这一点得益于残差网络结构对特征的学习更为充分,且网络深度远大于 VGG-16 及 MobileNet_v1(表2),而背景影响容易逐层递减。之后,将此数据集送回网络重新训练,由表4中

表4 背景数据集测试结果
Table 4 Test results on background dataset

Backbone network	Original mAP	Status	mAP under different background types		
			Pure color	Texture	Garbage
Res101	1.0	Before re-training	0.9913	0.9222	0.9050
		After re-training	1.0	1.0	1.0
VGG-16	1.0	Before re-training	0.9901	0.8835	0.6494
		After re-training	1.0	1.0	1.0
MobileNet_v1	0.9917	Before re-training	0.9691	0.7433	0.4204
		After re-training	0.9999	0.9933	0.9793

重新训练后的 mAP 数据可知,训练后背景的干扰已被基本消除。因此,拍摄大背景的变化对检测精度的影响较小,原因在于此类背景较为固定,可提前在网络中训练以消除影响;而其他垃圾目标密集进入检测框,对检测精度的影响更大,原因在于此类背景多样且未知,但事先将多种背景引入网络训练有利于增强其抗背景干扰的能力。

3.4 最优阈值设置及最优检测性能

目标检测算法根据设定概率阈值来检测目标,不同阈值下各类别的召回率、精度不同。高召回率表征低的漏识率,高精度表征低的误识率,因此各类别召回率及精度可以直接反映检测的优劣程度,为此需要设置最佳阈值以使检测最优。垃圾检测更为特殊,可回收垃圾应倾向较高的精度指标,有害垃圾应倾向较高的召回率指标。因此,本研究综合权衡召回率及精度指标,并区别对待可回收垃圾与有害垃圾,其最佳阈值 P_1, P_2 按以下条件计算:

1) 为避免单个目标被同时识别为可回收垃圾、有害垃圾,设置: $P_1 + P_2 = 1$;

2) 最大化以下判别指标: $\sum_i (k_1 \times P_i + k_2 \times R_i)$, $i = 1, 2, \dots, 6$, 分别对应 6 种类别。对于可回收垃圾, $k_1 = 1.1, k_2 = 1$; 对于有害垃圾, $k_1 = 1, k_2 = 1.1$ 。

根据上述条件可生成阈值判定曲线, MobileNet_v1 下的阈值判定曲线如图 9 所示,随着

概率阈值增加,各类别精度递增,召回率递减,判别指标两端最小,最大时相应最佳阈值是 P_1 为 0.56, P_2 为 0.44。据此,可得到各网络的最优阈值及类别精度、召回率,见表 5。对于可回收垃圾, Res101 检测精度最高,为 97.64%; 对于有害垃圾, VGG-16 召回率最高,为 99.08%, 即检测遗漏率最低,为 0.92%。此外,对比三种网络可回收垃圾的最佳阈值,网络按照最佳阈值从大到小排序依次为 Res101、VGG-16、MobileNet_v1, 此趋势与网络收敛性能一致,说明在网络收敛性提高时,对应的精度指标的阈值设置可更大,原因在于收敛性越高,反映目标被正确检测的概率的平均值越大,设置更高的概率阈值对检测精度的提升效应高于对召回率的下降效应。

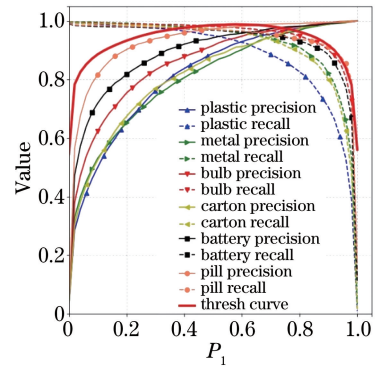


图9 MobileNet_v1 概率阈值判定曲线

Fig. 9 Probability threshold decision curve on MobileNet_v1

表5 各网络最优阈值下测试集精度及召回率

Table 5 Precision and recall on test dataset under optimal threshold of each network

Backbone network (P_1, P_2)	Parameter	Recyclable garbage				Hazardous garbage			
		Metal	Plastic	Carton	Mean	Battery	Pill	Bulb	Mean
Res101 (0.76, 0.24)	Precision	0.9796	0.9662	0.9834	0.9764	0.9497	0.9792	0.9698	0.9662
	Recall	0.9889	0.9778	0.9834	0.9834	0.9799	0.9714	0.9887	0.9800
VGG-16 (0.62, 0.38)	Precision	0.9583	0.9487	0.9775	0.9615	0.9657	0.9658	0.9534	0.9491
	Recall	0.9857	0.9795	0.9876	0.9843	0.9899	0.9873	0.9951	0.9908
MobileNet_v1 (0.56, 0.44)	Precision	0.8943	0.9235	0.9109	0.9096	0.9266	0.9636	0.8867	0.9256
	Recall	0.9609	0.9322	0.9654	0.9528	0.9728	0.9739	0.9854	0.9774

4 结 论

基于 Faster RCNN 算法实现了生活垃圾的高精度识别,涵盖数据集建立到检测结果分析与优化的全过程,可得以下结论:

1) 建立了涵盖可回收垃圾及有害垃圾中典型 6 种类别的有效数据集,图片总数 6076 张,数据增强后包含目标总数量 17659,且类别均衡、目标大小多样。

2) 困难样本增强训练、特异层微调网络及典型错误优化可有效提升算法精度,对于 MobileNet_v1 网络, mAP 从 87.19% 依次提升到 92.02%、92.85% 及 94.90%。

3) 不同网络的收敛性、精度、速度及泛化性能指标差异较大,收敛性从优到劣排序为 Res101、VGG-16、MobileNet_v1,收敛性直接反映泛化性能,并基本反映精度;速度从大到小排序为 MobileNet_v1、VGG-16、Res101。综合来看,Res101 性能最优,MobileNet_v1 速度优势明显。

4) 给出了最佳概率阈值的设定方法,且网络收敛性越好,最佳概率阈值越大。

参 考 文 献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [2] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [3] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [4] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-01-03]. <https://arxiv.org/abs/1804.02767>.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [6] Ren S Q, He K M, Girshick R, et al. Faster R-

- CNN: towards real-time object detection with region proposal networks[EB/OL]. (2015-06-04)[2020-01-03]. <https://arxiv.org/abs/1506.01497>.
- [7] Chen Z H, Zou H B, Wang Y B, et al. A vision-based robotic grasping system using deep learning for garbage sorting [C] // 2017 36th Chinese Control Conference (CCC), July 26-28, 2017, Dalian, China. New York: IEEE Press, 2017: 11223-11226.
- [8] Mikami K, Chen Y, Nakazawa J, et al. DeepCounter: using deep learning to count garbage bags[C]//2018 IEEE 24th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), August 28-31, 2018, Hakodate, Japan. New York: IEEE Press, 2018: 1-10.
- [9] Zhang P C, Zhao Q, Gao J, et al. Urban street cleanliness assessment using mobile edge computing and deep learning[J]. IEEE Access, 2019, 7: 63550-63563.
- [10] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88 (2): 303-338.
- [11] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2020-01-03]. <https://arxiv.org/abs/1409.1556>.
- [13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [14] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2020-01-03]. <https://arxiv.org/abs/1704.04861>.
- [15] 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.
- [16] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision. Cham: Springer, 2014: 818-833.