

基于 Faster R-CNN 的垃圾分类检测

张智先 李丹

四川大学锦城学院 四川 成都 610065

【摘要】自 2019 年上海全面开展垃圾分类，出台相关管理条例以来，垃圾分类不仅多次作为公众焦点，而且成为了一种“新时尚”。然而不足 50% 的垃圾分类正确率，更是加速催化了相关市场的成长。针对垃圾产出量持续攀升且垃圾投放正确率低的情况，这里提出应用垃圾分类检测的 faster rcnn 算法。通过目标检测获取垃圾所属类别，为后续的垃圾投放提供技术支持，为环境治理以及垃圾回收处理等方面减轻工作量。

【关键词】垃圾分类；Faster R-CNN；目标检测

引言

近年来，我国经济增长和城市化建设发展十分迅猛，随之出现的就是各类生产生活垃圾量的增大，也就意味着处理难度也会比较大^[1]。垃圾分类看似很容易，然而实际推广和实施起来却不轻松。早在 2016 年 12 月，中央召开的会议就在强调垃圾分类，并且将其看作后续工作的一大重点，同时在会议中也指出了要大力推广和普及垃圾分类知识，推行垃圾分类制度，在最近 5 年之内，大部分城市会落实垃圾分类。

在国家政策和垃圾分类管理条例的大环境下，我国在垃圾分类上的问题仍然很多。在北京市的 14 区 2 县，居民垃圾分类的正确率仅为 16.92%~34.56%^[2]，说明居民在这方面的意识仍然不足，并且宣传教育存在走过场的情况，没有落实到位；还存在基层社区经费不足的情况，在上海徐汇区三村，街道社区会进行补贴，但是时间一长，高昂补贴的压力不利于垃圾分类的长期开展^[3]；还有城市遇到政策不够完善的问题，垃圾分类的罚款不能过高，然而过低又起不到想要的效果，甚至可能会出现的一系列的问题^[4]。

陶威远等^[5]使用基于 SMC 滤波器的 Faster R-CNN 实现简易垃圾分类识别系统，在塑料袋和衣物上的检测平均准确率能达到近 80%。周滢懋^[6]结合 Faster R-CNN 与 Hyper-Column 算法，应用于生活垃圾智能分拣系统中。彭昕昀^[7]等选择 SSD 算法，通过数据增强以及算法的改进，达到对垃圾精准又快速识别的标准。

1 数据集

通过爬取、下载以及自定义采集等方式，构建了由 540 张垃圾图片组成的垃圾数据集，其中包含香蕉皮、饮料瓶、西瓜皮、蔬菜叶和鞋子这 5 种垃圾。为了便于

后续使用，这里使用的工具 x2coco 制作 COCO 数据集。在图片的选取中，对于水果蔬菜种类，收集了水果蔬菜成熟和腐烂阶段、水果蔬菜的经过压缩和撕扯之后的多个角度图片；对于饮料瓶，则是收集了饮料瓶经过挤压的不同形态图片。这样能够消除一些客观因素对结果带来的影响，从而更好地对目标垃圾进行检测。

2 Faster R-CNN 目标检测算法

Faster R-CNN 算法^[8]是 Fast R-CNN 算法^[9]的升级版，顾名思义，“Faster”就意味着本节介绍的算法比 Fast R-CNN 算法的检测速度更快，更能接近“实时”的效果。使用训练后的模型能够识别出垃圾在图片中的位置以及垃圾所属的类别，为各类工业部署提供支持。

2.1 Faster R-CNN 相比 Fast R-CNN 检测速度更快

Fast rcnn 使用 SS 算法进行候选框的选择，但只能在 CPU 上实现的，而 Faster rcnn 算法可以使用 GPU，大大提升了计算速度。虽然两个算法其他方面也有些许不同，但计算速度是差别最大的。

2.2 Faster R-CNN 算法流程

首先是要被检测的图象输入到 CONV 层，提取后续共享的特征映射。RPN 有两个分支，一个分支产生 anchor，然后剪枝和使用阈值进行筛选，最后通过激活函数判断 anchor 是否为物体；另一条路线是使用框回归来调整 anchor，使其更接近物体真实的位置。经过 RPN 之后，就得出候选框。将其与特征映射输入到 RoI 池化层，它的作用就是将不同大小的 RoI 统一规格，最后进行第二次分类和回归的训练。

接下来的小节将会对 RPN、anchor、损失函数进行详细介绍。

2.3 RPN

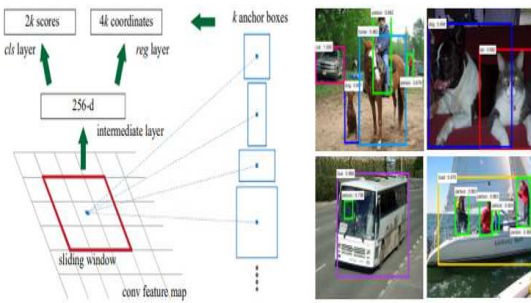


图 1 区域建议网络 RPN

RPN 的输入端为特征映射，而输出包含了表示真实框的偏移量的 4 个坐标和 $2k$ 个分数（图 1）。RPN 是这个算法独有的，看起来很复杂，实质上还是卷积神经网络。简单来说，就是由一个 $n*n$ 的卷积层（图 1 中展示的是 $n=3$ ），以及用于实现分类和回归的两个并列的全连接层组成，最后结合 NMS 算法筛选出 proposals，完成定位需要识别的物体的任务。

接下来解释一下图 2 中展示的 $3*3$ 窗口滑动。在滑动的过程中，这个 $3*3$ 的中心与原图的中心一直是重合的，也就是意味着在窗口的各个位置可以通过比例推断到原图片的区域和位置。而 $3*3$ 的窗口就是原图片通过 $SSP^{[10]}$ 池化得到的，池化的部分就是前文所提及的 anchors。

2.4 anchors

以窗口中心为基准，使用 3 种尺寸、3 组比例，得到 9 个 anchor。那么宽 W 长 H 的图片，最终 anchor 的数量是 $W*H*K^{[11]}$ 个。这种设置提高了在不同大小和长宽比的图片上的泛化能力，节省了使用图像金字塔和滤波器金字塔在多个维度上运行的计算时间，同时也比基础选择的感受更大。

算法需要保持函数（anchor 和候选框之间）平移不变的这个特点，当目标物体发生平移或者位置发生移动时，相应地候选框也应该移动，在不同的位置上仍然能够使用相同的函数生成预测候选框。正是因为这个特性也使得模型更轻量级了，所涉及的参数也比之前的方法少了很多^[8]，也降低了过度拟合在较小的数据集的风险。

针对多尺度的预测，在这种 anchor 出现之前有两种比较常用的方法。如图 3 (a) 所示，一种基于“图片金字塔”的方法。将图片进行缩放，在不同尺度下进行特征映射计算。这种方法非常有用，不过每进行一次缩放和计算，都会消耗大量的时间，就会导致算法耗时比较大。另一种方法就是“滤波器金字塔”^[8]，如图 3 (b) 所示。可以看到在原图片处理得到特征映射后，在图中使用了 3 种尺寸不同的滤波器构成的“滤波器金字塔”，达到了多尺度预测的目标。不过这两种方法，通常都是结合起来一起使用。这里的 anchor 相比来看，既快速又准确。它使用一种尺寸图片生成的特征映射，在此基础上使用多种 anchor，进行边界框的调整。这样不仅能够多尺度上进行预测，而且特征也是共享的，后续不需要重复计算，也就提升了算法的速度。

在 RPN 训练时，anchor 会得到两种标签。在两种情况的 anchor 下，会得到“是物体”的标签（正标签）。第一种是与真实框交并比值最大的情况，第二种是与真实边界框交并比高于设定的阈值（这里设置为 $0.7IoU$ ）的 anchor。在使用的时候，一般都会使用第一种分配方式，因为第二种分配方式有可能会交并比均小于设定的阈值，也就是没有正标签分配，这不是愿意出现的情况。

2.5 损失函数

损失函数由分类和回归两个部分的损失组成，前半部分是分类损失，后半部分是回归损失。这两种损失下分别还有两个主要网络的损失，也就是细分的话，有四部分损失。

在 RPN 中把 anchor 分成了两类，也就是前文说的物体和背景的二分类。因此这里的 L_{cls} 选用是二分类的交叉熵， N_{cls} 表示的是 anchor 的总数。对每个 anchor 的交叉熵损失进行求和运算，再求平均。后面的 Fast rcnn 类别数通常大于 2，使用多分类的交叉熵损失。

公式后半部分的回归损失中的 t_i 是预测框坐标的 4 个值构成的向量， t_i^* 与 t_i 结构相同，只不过是真实框坐标的向量，这 4 个量分别就是中心点的坐标 x 和 y ，以及框的宽 w 和高 h 。回归损失 $L_{reg}^{[11]}$ 中的 R 是 smooth L1 损失函数。在回归损失前乘了一个 pi^* ，当值为 1，就是正类时，进行损失的回归；为 0 时，就不考虑。

对于框回归，预测出的边界框一般来说与实际物体的框是有偏差的，需要对其进行平移和缩放来进行调整。这里沿用 R-CNN 中学习位置和大小的变化参数。然后将这些参数应用在 Faster R-CNN 中的框回归的公式，得到新的预测边界框。因为 anchor 的设计，所以可以对多种尺寸的边界框预测进行调整。

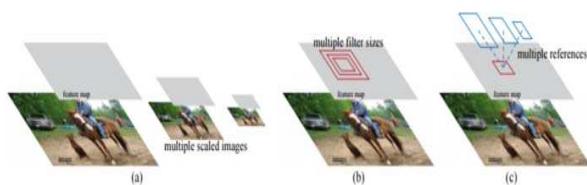


图 2 多尺度预测的方法

3 在垃圾分类上的实验结果与分析

本文使用的是飞桨发布的目标检测库，这是一种端到端的目标检测套件。模块化的设计以及模型压缩能力，对后续的算法部署提供了便利。实验在 Tesla V100 环境下进行 GPU 计算。

实验使用 Resnet101 作为主干提取网络来搭建 Faster R-CNN 模型，并且采用迁移学习的方式，也就是使用模型预训练权重初始化网络加快模型的收敛速度。这里对模型进行了 50000 次训练迭代，结果图 5 所示。从图中明显看出，在 40000 次迭代之后时，训练的 mAP 基本稳定在 0.88 左右。

检测单垃圾图片时，可以从检测结果上看出成功率很高。图 6 为其中两种垃圾(西瓜皮、饮料瓶)的检测效果。

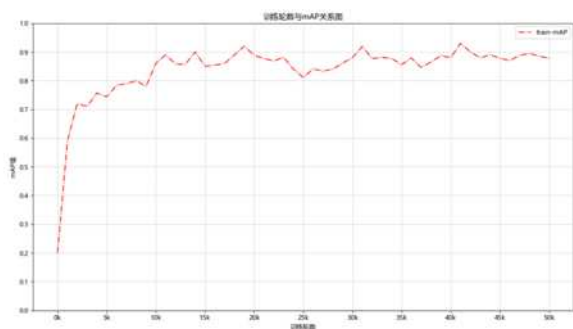


图 3 训练轮数与 mAP

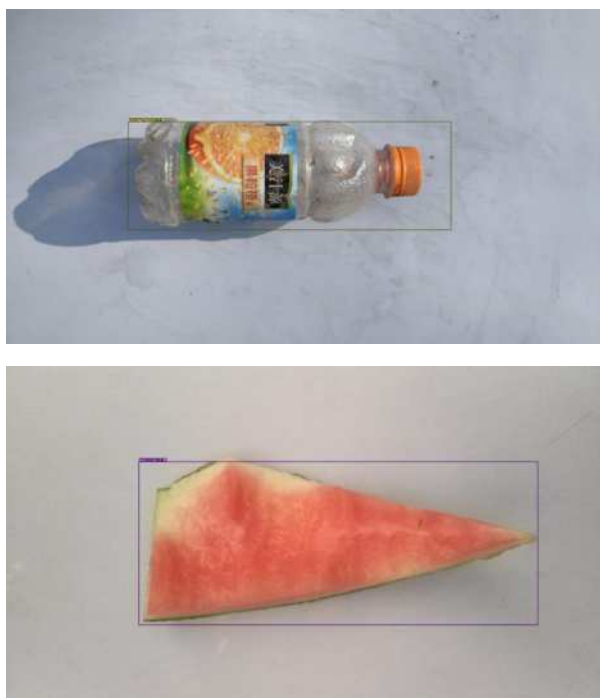


图 4 其中两种垃圾的检测效果

4 结束语

本文提出了一种基于 Faster R-CNN 算法实现了对几类垃圾的识别和分类。此方法虽然能达到一定高度的 mAP 值，也展现出了一定的效果。然而在分析和实现上仍然有一些地方可以提升，后续将从多个方面深入研究：（1）受设备和时间限制，这里只选取了少量的数据集和垃圾种类，也影响了最终的结果。（2）这里对于该算法原论文上提出的结构没有作大的改动，但是在后续的深入研究中，可以着手对特征提取网络的选取以及损失函数的改进等方面展开，进一步提高模型的检测能力，达到更精准地识别垃圾的效果，为各类工业部署的使用提供支撑。

【参考文献】

- [1] 徐林, 凌卯亮, 卢昱杰. 城市居民垃圾分类的影响因素研究 [J]. 公共管理学报, 2017, 14(01): 142-153+160.
- [2] 邓俊, 徐琬莹, 周传斌. 北京市社区生活垃圾分类收集实效调查及其长效管理机制研究 [J]. 环境科学, 2013, 34(01): 395-400.
- [3] 严宇珺, 严运楼. 城市生活垃圾分类试点政策执行问题与对策探讨——以上海市徐汇区梅陇三村为例 [J]. 生态经济, 2020, 36(03): 197-200.
- [4] 陆柯兵. 杭州市生活垃圾分类治理的多元共治模式：现状、问题与对策 [J]. 环境保护与循环经济, 2020, 40(03): 16-20.
- [5] 陶威远, 姜太平, 郑一然. 基于 Faster R-CNN 算法的垃圾分类识别系统 [J]. 网络安全技术与应用, 2020(03): 45-47.
- [6] 周滢懋. 基于机器视觉的生活垃圾智能分拣系统的设计与实现 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
- [7] 彭昕昀, 李嘉乐, 李婉, 刘杏洲, 张程发, 林显新, 欧嘉诚. 基于 SSD 算法的垃圾识别分类研究 [J]. 韶关学院学报, 2019, 40(06): 15-20.
- [8] Ren Shaoqing, He Kaiming, Girshick Ross, 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.
- [9] Girshick R. Fast R-CNN [J]. Computer Science, 2015
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in European Conference on Computer Vision (ECCV), 2014.
- [11] 张亚辉. 基于 Faster R-CNN 目标检测的机器人抓取系统研究 [D]. 北京: 中国科学院大学 (中国科学院深圳先进技术研究院), 2019.
- [12] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.