****

**计算机视觉期末报告**



学 院 智能与计算学部

专  业 软件工程

姓 名 周玮康

学 号 3017207087

姓 名 赵为

学 号 3017210149

姓 名 杨欣捷

学 号 3017218209

目录

[基于改进Cascade R-CNN的垃圾目标检测系统 3](#_Toc39483387)

[1. 引 言 3](#_Toc39483388)

[2. 相关工作 4](#_Toc39483389)

[3. 方法 5](#_Toc39483390)

[3.1 模型结构 5](#_Toc39483391)

[3.1.1 R-CNN 5](#_Toc39483392)

[3.1.2 FASTER R-CNN 5](#_Toc39483393)

[3.1.3 Cascade R-CNN 7](#_Toc39483394)

[3.1.4 创新点 8](#_Toc39483395)

[3.2 数据增强 11](#_Toc39483396)

[4. 实验仿真及结果分析 13](#_Toc39483397)

[4.1 数据集 13](#_Toc39483398)

[4.2 模型构建 14](#_Toc39483399)

[4.3 模型评估指标 14](#_Toc39483400)

[4.4 结果 14](#_Toc39483401)

[5. 结 论 15](#_Toc39483402)

[参考文献 16](#_Toc39483403)

基于改进Cascade R-CNN的垃圾目标检测系统

赵为，周玮康，杨欣捷

天津大学

摘 要：近年来，随着人们生活品质的提高和物质生活的丰富，垃圾种类复杂和杂乱堆积给自动化垃圾挑拣分类带来极大挑战。针对多种垃圾的目标检测和分类问题，本文提出一种基于改进的Cascade R-CNN算法的垃圾目标检测系统。首先，我们采用2020年华海AI挑战赛·垃圾分类比赛的公开数据集，使用Faster R-CNN构建目标检测系统，对多目标图片进行数据增强，将系统的mAP从78提升到了82。本文通过反复实验发现Faster R-CNN存在局限性，我们采用Cascade R-CNN算法进行模型训练，使用Focal loss调控正负样本比例，将NMS（非极大抑制）替换为Soft-NMS，引入在线难例挖掘算法，最终mAP达到了88.4，提升了14.6%。实验结果表明，使用改进的Cascade R-CNN可以有效的实现对垃圾目标进行检测，并且达到较高的精确度。

关键词**：**Cascade-RCNN; Faster-RCNN; 目标检测; 垃圾分类;

# 1. 引 言

随着居民生活质量的提升，城市垃圾产量迅速增长，为帮助垃圾回收，垃圾分类等环境保护工作有效展开，多个城市陆续出台垃圾管理条例[1,2,3]。对垃圾分类处理首先需要进行分类，而人工分类效率低，尤其在面对大量垃圾时，会由于分类环境恶劣、任务繁重对人的健康产生不利影响。如果垃圾数量过多，则可能无法完成对其准确分类，而将剩余部分视为其他垃圾处理掉，对环境造成染的同时，也浪费了可回收的资源。

如今，伴随着人工智能，尤其是卷积神经网络在图像处理方面跨越式的发展，智能机器通过摄像头自动识别和定位不同种类的垃圾成为可能，进而可以实现对垃圾的自动分类投放，在减轻人工劳动量、 提高分类效率的同时能够降低对环境的污染，减少对资源的浪费。 在智能制造战略逐步实施的背景下，结合计算机视觉相关技术，实现一种基于改进Cascade-RCNN的垃圾目标检测系统。

我们针对2020年华海AI挑战赛·垃圾分类比赛的数据集进行训练，首先使用了Faster-RCNN[4]这一成熟的目标检测算法，在对数据集进行增强以及模型调参后发现准确度并不能比市面上其他同类检测系统效果出色，我们放弃的了对Faster-RCNN的改进，从而选用[Zhaowei Cai](https://arxiv.org/search/cs?searchtype=author&query=Cai%2C+Z)和 [Nuno Vasconcelos](https://arxiv.org/search/cs?searchtype=author&query=Vasconcelos%2C+N)在2018年CVPR提出的Cascade-RCNN[5]，使用Focal loss作为损失函数，并且将之前较为“暴力”的NMS修改为soft-NMS，以及使用OHEM（Online hard example mining，在线难例挖掘）的策略来优化网络。最终，我们在测试集合上取得了较大的进步。

# 2. 相关工作

近年来，随着垃圾分类政策的推行，许多的国内期刊上发表了大量的关于垃圾分类研究成果。在特征提取、垃圾种类的增广和不同的模型方面人们进行了大量的实验寻求一种能够对垃圾进行有效分类的算法模型。

黄国维在他的论文中研究了基于一些深度学习技术的智能垃圾分类[6]，例如BP神经网络[7]、卷积神经网络[8]，以及使用特征提取技术对图片进行加强，最近一些学者提出新的卷积神经网络结构可以极大的提升垃圾识别的准确度，例如残差神经网络[9], DenseNet[[10](https://arxiv.org/pdf/1608.06993.pdf)]。

考虑到单纯的卷积神经网络进行垃圾分类只能一次对单目标进行分类，并且无法指定图片中垃圾具体位置。张方超等人提出基于图像视觉的垃圾分类机器人[11]，王铭杰在他的论文中提出目标检测对垃圾进行分类[12]，使用了无区域建议的目标检测方法YOLO V3[13]。此外，陶威远等人提出基于Faster R-CNN[14]的垃圾分类系统[15]也可以实现初步的垃圾分类与定位效果。

总结各种特行在垃圾分类数据库上的表现，我们可以发现基于区域建议的目标检测方法对于已知垃圾种类丰富的数据集上有更加出色的效果。本文选用Faster R-CNN和最近提出的Cascade R-CNN对垃圾进行分类和定位，并根据公开数据集特点，进行数据的增强和模型的改进。

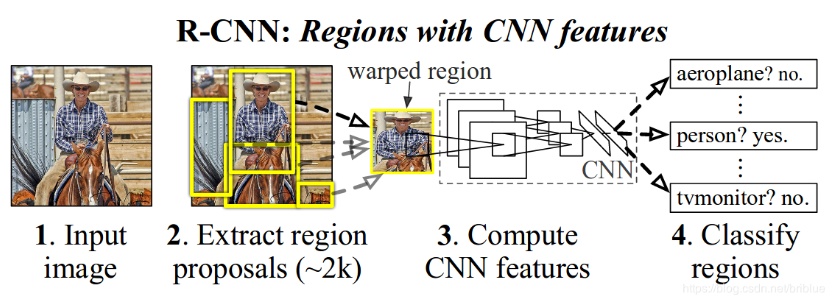
# 3. 方法

## 3.1 模型结构

### 3.1.1 R-CNN

2012年AlexNet在ImageNet举办的ILSVRC中大放异彩，Ross Girshick受此启发，发明了Region-CNN(RCNN)[16]。RCNN是第一个成功将深度学习应用到目标检测上的算法。R-CNN基于卷积神经网络(CNN)，线性回归，和支持向量机(SVM)等算法，实现目标检测技术。

**R-CNN目标检测的思路。**输入一张图片，采用 Selective Search 算法，从图片中选出多个独立的候选区域(Region Proposal)。得到所有候选区域后，将每个候选区域输入到预训练好的AlexNet中，提取一个固定长度的特征向量。利用提取到的候选区域的特征向量，对每个目标（类别）训练一个SVM分类器，识别该区域是否包含目标,是目标就标记positive，不是就标记negative。结构如图1.



图片1. R-CNN结构

当候选区域只是框出来了某个类的一部分时，在R-CNN中，设定一个IOU的阈值，如果该区域与Ground truth的IOU低于该阈值，就将给区域设置为Negative。上述操作只是得到了包含目标的候选区域，但有些候选区域框出的位置不是很准确。这里需要再训练一个线性回归模型判断，对于某个类别的SVM是Positive的候选区域，其框的目标区域是否很完美。

### 3.1.2 FASTER R-CNN

Faster R-CNN是何凯明等在2015年提出目标检测算法，该算法在Fast R-CNN基础上提出了RPN候选框生成算法，将候选区域的提取和Fast R-CNN中的目标检测网络融合到一起，这样可以在同一个网络中实现目标检测，生成建议框仅需10ms，使得目标检测速度和精度大大提高。并且真正实现了端到端的目标检测框架。

Faster R-CNN首先将样本图像整个输入到卷积层（Conv Layers）中，使用一组基础Conv+Relu+Pooling层提取输入图像的特征，最后得到Feature Map。

接下来，Faster R-CNN使用区域检测网络（RPN）生产候选区域。RPN以之前卷积层提取得到的Feature Map作为输入，Feature Map进入RPN后，先经过一次3\*3的卷积，得到的特征图尺寸和输入的特征图相同，这样做的目的应该是进一步集中特征信息。之后生成若干个Anchor box，一个Anchor就是Feature Map中的一个点，并有一个相关的尺度和纵横比。通过Softmax判断Anchors属于前景(foreground)或者后景(background)，即是物体或者不是物体，因为是２分类，所以它的维度是2k scores。同时，再利用另一分支bounding box regression 对anchor box进行修正，形成较精确的proposals,它的维度是4k coordinates。

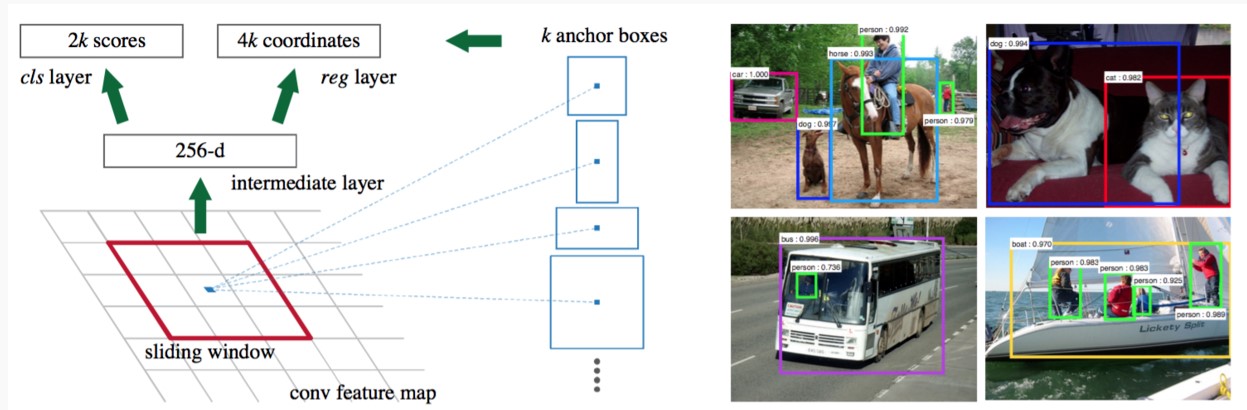


图2. RPN

然后由RoI池化层收集卷积层提取的feature map 和RPN生成的proposals，提取每个候选区域的特征图。最后将候选区域的特征图输入到用于分类的Softmax层以及用于边框回归全连接层。通过Softmax判断每个proposals的类别，同时再次使用边框回归（bounding box regression）获得建议框的最终精确位置，对边框进行修正。

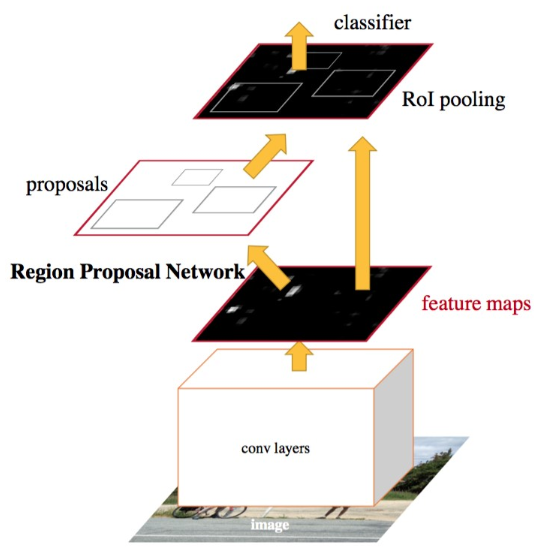


图3. Faster RCNN结构

在尝试中我们发现，如果使用Faster R-CNN网络的VGG16去提取特征，会导致存在一个很大的感受野，不利于检测小物体，但是在垃圾检测中里面，图片中的垃圾物体通常是比较小的，所以,用VGG16对于小物体进行检测的效果并不好。为了克服VGG16对小物体检测不好的缺陷,我们选用ResNet50和FPN相结合的网络结构去提取特征。 ResNet网络主要通过残差学习来解决深层网络训练出现的退化问题,能很好地解决梯度消失的问题。FPN网络主要是从网络不同层抽取不同尺度的特征进行预测，能提取足够底层的特征，而且底层的特征对于检测小物体是有帮助的。

### 3.1.3 Cascade R-CNN

Cascade R-CNN网络跟Faster R-CNN网络很相似，主要分成两步，首先是定位目标的位置，然后对目标进行分类。Zhaowei Cai等人发现通过设置的不同的IoU阈值带来不同的检测精度问题，并且认为是由过拟合和检测器最佳IoU与输入假说不匹配所引起的。

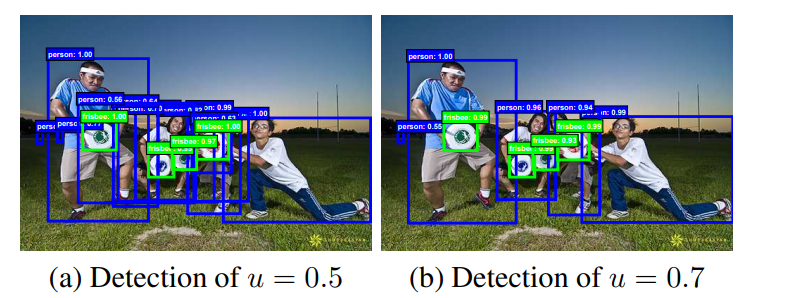


图4. 不同IoU阈值的识别结果

对于通常的Cascade R-CNN流程为：输入一张图片，通过特征提取网络得到图像的特征图; 然后，区域生成网络(Region Proposal Network, RPN)在特征图上定位候选目标，通过级联不同的IoU(Intersection over Union, 交并比)阈值对应质量的样本进行训练和回归。因此，Cascade R-CNN与Faster R-CNN的不同之处为前者级联了几个检测模型,后者只有一个检测模型。

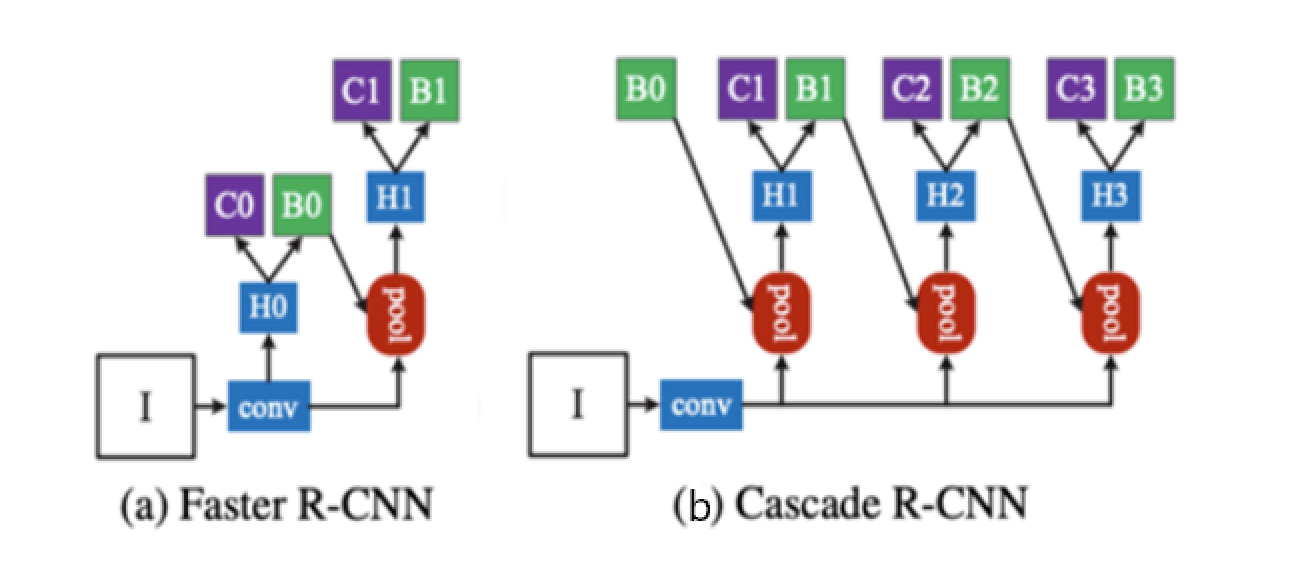


图5. Faster R-CNN与Cascade R-CNN结构对比

每个检测模型都是基于不同IoU的正负样本训练得到，并且前一个检测模型的输出是后一个检测模型的输入，越靠后的检测分支其检测的正负样本IoU阈值越高，导致每个检测模型的IoU尽量与输入proposal的IoU接近,让每一个检测模型都专注于检测IoU在某一范围内的proposal,从而提高检测的精度；反过来也能使RPN网络输出更优质的候选框，从而减少假正样本出现的次数，提高检测网络的精确度。更深层的检测器也就可以优化更大阈值的proposals。

### 3.1.4 创新点

**使用Focal损失函数。**在目标检测算法中，对于输入的一张图像，可能会生成成千上万的预选框（Region proposal），但是其中只有很少一部分是包含真实目标的，这就带来了类别不均衡问题。无用的易分反例样本会使得模型的整体学习方向跑偏，导致无效学习，即只能分辨出没有物体的背景，而无法分辨具体的物体。负样本数量太大，占总的Loss函数输入参数的大部分，而且多是容易分类的，因此使得模型的优化方向（即Loss函数的梯度下降方向）并不是我们所希望的那样。

为了解决类别不均衡问题，我们将损失函数由原本使用的Cross Entropy Loss修改为了Focal Loss。在各个类别的损失值之前, Focal Loss函数采取赋值权重的方式，可以通过减少易分类样本的权重,使得模型在训练时更专注于难分类的样本，从而解决解决训练样本的类别不均衡问题。

修改后发现，使用损失函数为Focal Loss进行训练后，识别准确率较之前有了明显的提升。

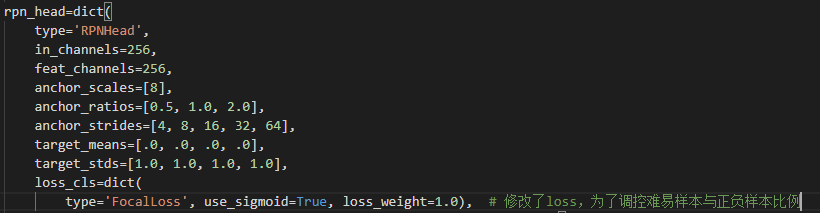


图6. 引入Focal loss

**将NMS(None maximum suppression, 非极大值算法) 修改为 Soft-NMS(软化非极大值算法)。**目标检测算法会输出多个检测边框，尤其是在真实目标周围会有很多置信度高的检测边框。为了去除重复的检测边框，达到每个物体有且只有一个检测结果的目的，我们采用非极大值抑制算法。

NMS的本质是搜索局部极大值，抑制非极大值元素。当 IoU 超过某个阈值后，直接删除该框。这就导致使用NMS时阈值的确定较为困难，设小了会出现重叠的物体容易被删掉的情况（如下图，绿色框因为和红色框重叠面积较大而被删掉），设置过高又容易导致误检。

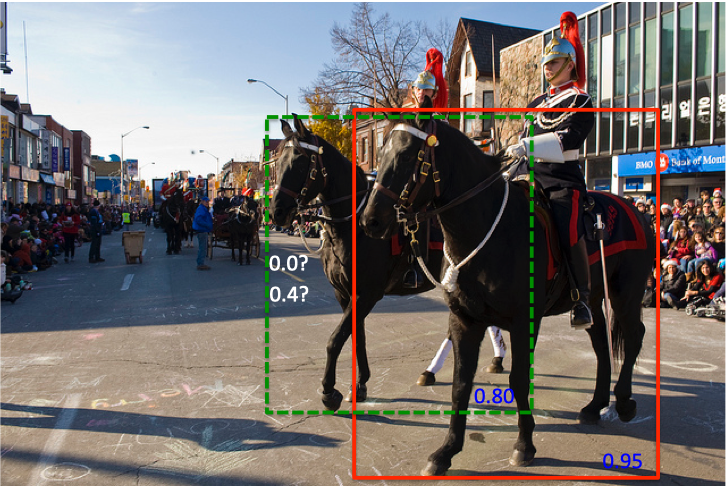


图7. NMS

为了解决这个问题，我们将原本使用的NMS算法修改为Soft-NMS算法。Soft-NMS 改进了之前比较暴力的 NMS，当 IoU 超过某个阈值后，不再直接删除该框，而是降低它的置信度 (得分)，如果得分低到一个阈值，就会被排除；但是如果降低后仍然较高，就会被保留。

原本NMS的描述表达如下：



Soft-NMS的描述表达如下：



从中我们很容易的可以看出,对于NMS来说当IoU小于一个值后,我们会直接删除,但对于Soft-NMS来说,我们只会降低他的置信度。使用Soft-NMS后，我们发现，重叠的物体较之前被更大程度的保留下来，模型得到更好的优化。



图8. 引入Soft-NMS

**OHEM(Online hard example mining, 在线难例挖掘)。**Fast R-CNN要依赖Selective Search算法生成的区域建议框，然后组合mini-batch。如果batch size是128的话，那么这128个预选框由2张图片，各生产64个区域。并且，128个预选框中，正负样本的比例是1:3，即25%的目标预选框和75%的背景预选框。正例的判定条件是，选择性搜索算法给出的预选框与Ground Truth的IoU要大于等于0.5。负例的判定条件是IoU要在[0.1,0.5)之间。这个IoU最小到了0.1是为了一种类似启发式的难例挖掘策略。

OHEM通过实验得到，mini-batch sampling人为定义的1:3的比例对于训练Fast R-CNN来说是个非常重要的超参数，如果更改该值，都会使mAP降低三个点左右。

但是OHEM认为，mini-batch sampling并不具有高效和最优的状态，因为Fast R-CNN中的mini-batch sampling还只是一种启发式的难例挖掘策略，它是在SGD算法优化之前就先验的组合好了mini-batch。而OHEM可以在每一次迭代中在线的从大量区域建议框中挖掘难例，它更加适合SGD这种迭代式的训练方式。OHEM可以取得更低的training loss和更高 mAP。

OHEM中有两套ROI Network(Region Of Interest Network)，其中(a)是绿的部分，它是一个只读(read-only)的ROI Network，每次只读取权重后进行forward，不参与反向传播和梯度更新，(b)是红色的部分，它是一个正常功能的ROI Network。

对于给定图像，经过选择搜索算法生成预选框，同样计算出卷积特征图，在(a)中会对所有的预选框全部计算forward，并计算出Loss，Loss的大小可以反映出对于当前的模型，这些样本的难易程度。随后会对loss进行排序，选择最前的Batch size个样本送入(b)中，进行forward和backward。

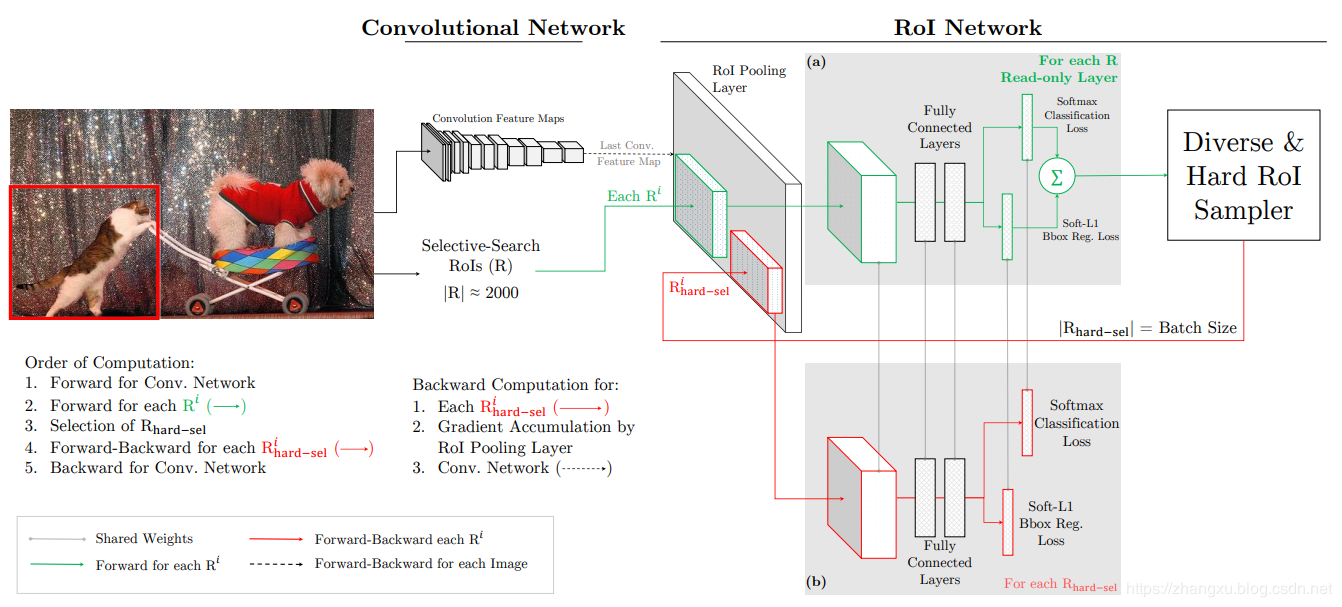


图9. OHEM结构

## 3.2 数据增强

本次2020年华海AI·垃圾分类数据集中，单类图片80000张，多类图片3000张。多类图片远少于单类图片，在训练过程中会导致模型对于多类图片学习出现偏差。因此我们使用多种图片的增强的方式对数据集进行扩充。

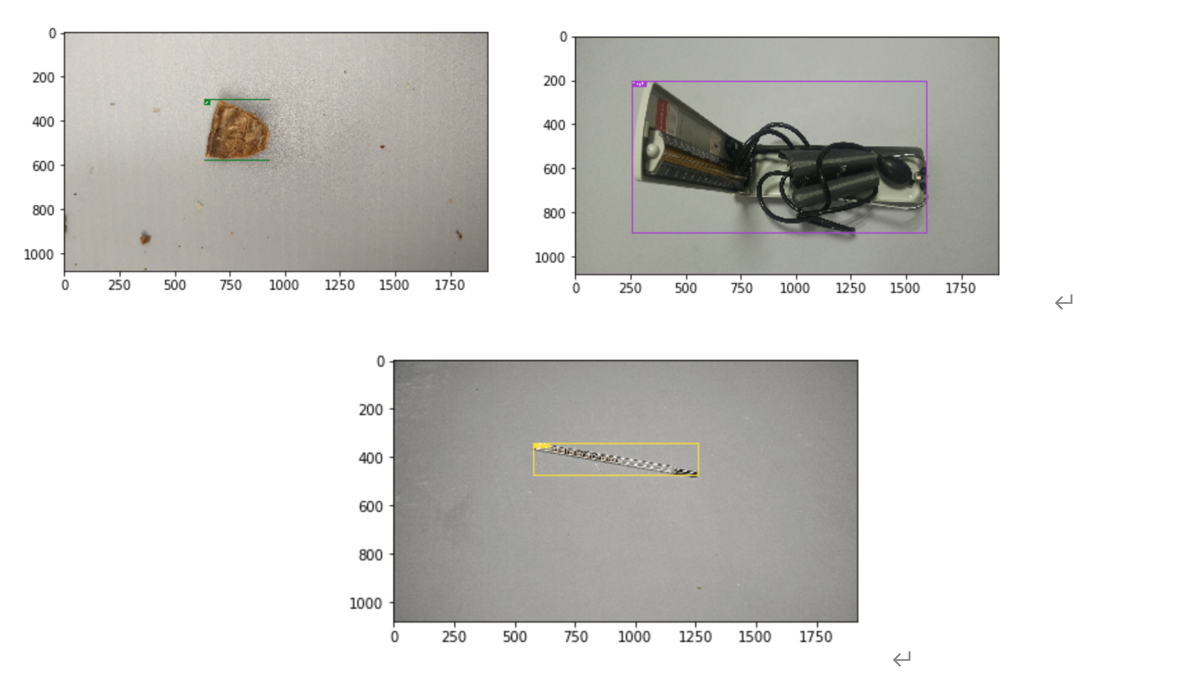


图10. 单分类图片

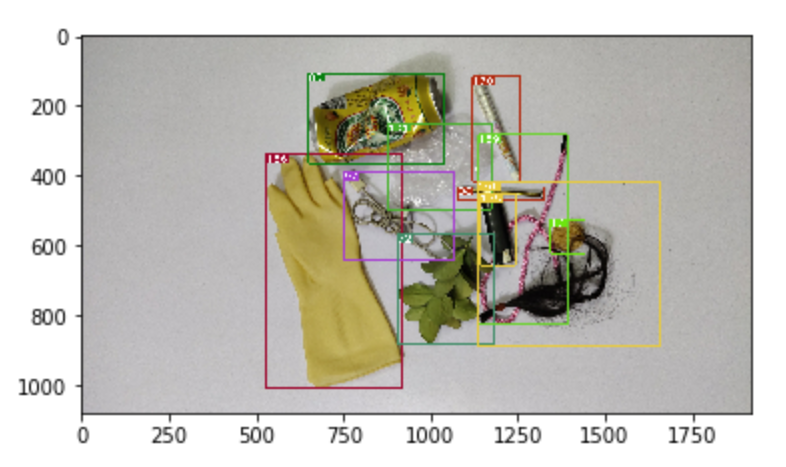


图11. 多类图片

由于单分类图片数量远多余多分类图片，在训练的时候引入单类训练集训练会导致分数降低。因此我们对单类图片进行重叠和填充，得到多类图片。

**图片重叠。**我们将两张图片同时读取，每张图片赋予不同的权重lamdb，按照不同的权重的将图片合成为一张图片。这种方法不但可以解决多类图片少的情况，还可以加强模型对于空间重叠的垃圾识别效果。

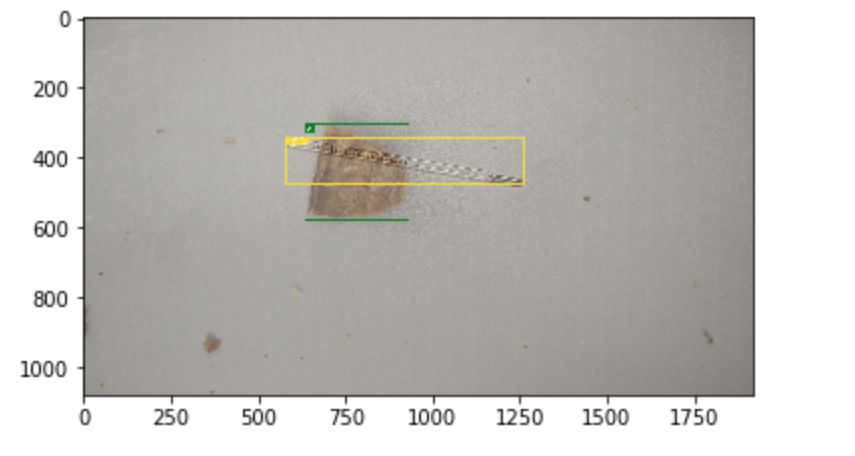


图12. 图片重叠结果

**图片填充。**我们将多种单类图片读取，压缩图片大小，并且构建一个新的像素空间，将多个图片随机插入不同位置，避免他们重叠后生成新的多类图片。

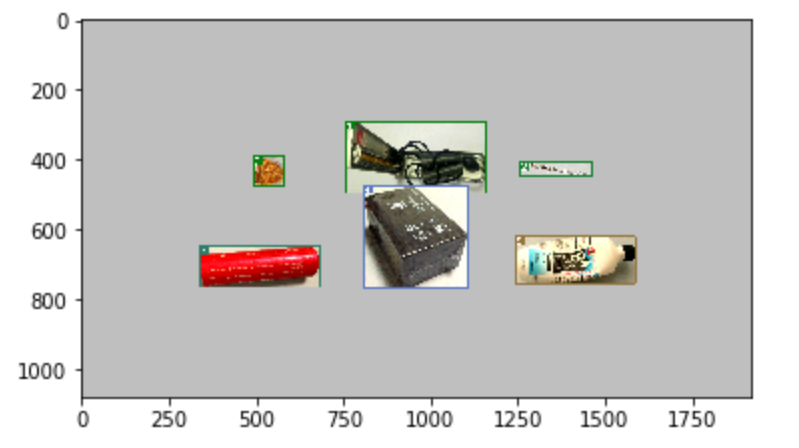


图13. 图片填充结果

最后，通过一些常用的图像数据增强的方法进行多类图片增强。我们使用了随机的对图片旋转和平移。

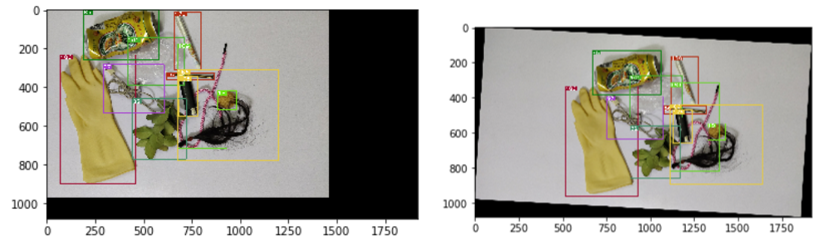


图14. 图片旋转和平移结果

# 4. 实验仿真及结果分析

## 4.1 数据集

本文使用华海AI大赛·垃圾分类的数据集，包括单类图片80000张，多类图片3000张。为了增加多类图片数量，我们使用20%的单类图片进行图片的堆叠，剩余80%的单类图片用于图片六合一填充。再使用通常的数据增强方式对30%的所有多分类图片进行旋转缩放。

|  |  |  |  |
| --- | --- | --- | --- |
| 数据集 | 类别数量 | 数量 | 训练集标注数量 |
| 多类图片 | 204 | 2999 | 28160 |
| 单类图片 | 204 | 80000 | 80000 |
| 数据增强-多类图片 | 204 | 21665 | 108160 |
| 测试集 | 204 | 1000 | 3425 |

## 4.2 模型构建

首先，我们直接使用基础的baseline torch中Modelzoo的Faster R-cnn+Resnet50+FPN。测试标准状态下该数据集的准确程度与mAP值。在初步训练后，我们发现单类,多类同时训练,就会导致测试集种无法识别一张照片中的多类,效果甚至远差于只有多类数据集。我们采用其他的数据增强方式解决这类问题，引入图片叠加和图片填充重新训练模型。

反复进行调参后，发现Faster R-CNN在我们有限的算力条件下难以得到令人满意的结果。我们选择使用Mmdetection构建Cascade R-CNN，并且使用上一步进行图像增强后的数据集进行训练。并且引入上文提到的Focal loss, OHEM, Soft-NMS进行模型的优化，完成最终垃圾目标检测系统的构建。

## 4.3 模型评估指标

使用与COCO竞赛相同的指标mAP进行模型的评估。

## 4.4 结果

首先我们使用测试集对于Faster R-CNN模型进行评估，结果如表1，发现使用新的数据增强能够明显提升模型准确度，提升约6.6%。但是由于算法本身的有限性，无法再提高模型准确度，最终模型评分为82.3。

表1. Faster R-CNN结果

|  |  |  |
| --- | --- | --- |
| 模型 | 未采用数据增强 | 采用数据增强 |
| Faster R-CNN | 78.2 | 82.3 |

进一步，我们使用测试集对于Cascade R-CNN进行测试，考虑到数据增强的有效性与实用性，我们延续了上一步使用数据增强的方法，针对使用不同分类模型的Cascade R-CNN分别进行模型评估。我们验证了优化后的模型可以更加贴合目前数据集，并且在评分上明显优于Faster R-CNN。

最后我们使用了SeNet101这一分类模型，通过对图片进行更好的分类来提升模型的准确度，最终模型的mAP达到了88.4，比最初模型提升了14.6%。

表2. Cascade R-CNN结果

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | 分类模型 | 未优化 | 优化 |
| Cascade R-CNN | Resnet50 | 84.7 | 87.3 |
| Cascade R-CNN | SeNet101 | \ | 88.4 |

# 5. 结 论

本文实现了基于改进Cascade R-CNN的垃圾目标检测系统。通过多次对比实验，发现Cascade R-CNN在目标检测上明显优于Faster R-CNN。并且引入多种模型优化策略，例如Focal loss和Soft-NMS，均使得模型得到优化提升，mAP达到88.4。

尽管在该数据集上本文模型表现出色，但是考虑到现实生活中垃圾种类多样，数量庞大，空间堆积复杂等问题，该模型实际表现可能会产生影响，导致评分降低。在未来工作中，我们考虑对数据集进行优化扩充，并且考虑到地区垃圾的种类和生活习惯等问题，因地制宜，训练适应不同地区的模型。此外，对预选框的修正以及图片特征提取进行进一步研究，减少预选框的误差，提升模型训练和识别速度。

# 参考文献

[1] 上海市生活垃圾管理条例[N]. 解放日报. 2019-02-20 (008)

[2] 沈阳市生活垃圾管理条例[N]. 沈阳日报. 2015-12-23 (012)

[3] 北京市生活垃圾管理条例[N]. 北京日报. 2019-12-18 (005)

[4] 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，2017（6）：1137-1149.

[5] Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.

[6] 黄国维. 基于深度学习的城市垃圾桶智能分类研究[D].安徽理工大学,2019.

[7] Jin W, Li Z J, Wei L S, et al. The improvements of BP neural network learning algorithm[C]//WCC 2000-ICSP 2000. 2000 5th international conference on signal processing proceedings. 16th world computer congress 2000. IEEE, 2000, 3: 1647-1649.

[8] Sahiner B, Chan H P, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images[J]. IEEE transactions on Medical Imaging, 1996, 15(5): 598-610.

[9] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Thirty-first AAAI conference on artificial intelligence. 2017.

[10] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[11]张方超,董振,张栋,武涛,李卫国.基于图像视觉的垃圾分类机器人识别控制系统算法设计[J].电子测量技术,2019,42(20):63-67.

[12]王铭杰.基于YOLO V3的垃圾自动定位及分类方法[J].无线互联科技,2019,16(20):110-112.

[13] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv preprint, 2018, arXiv:1804.02767.

[14] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[15] 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.

[16] Jiang Y, Zhu X, Wang X, et al. R2cnn: rotational region cnn for orientation robust scene text detection[J]. arXiv preprint arXiv:1706.09579, 2017.