## 1.1算法基础

目标检测算法可以完成对于街道垃圾的实时监控，通过采集大量的街道垃圾数据集训练良好参数的识别模型，可以将其应用在街道的摄像头实时识别垃圾。

### 1.1.1目标检测算法

**目标检测是计算机视觉领域中的一个重要研究方向，也是其他复杂视觉任务的基础。**作为图像理解和计算机视觉的基石，目标检测是解决分割、场景理解、目标跟踪、图像描述和事件检测等更高层次 视觉任务的基础。

传统的目标检测一般使用滑动窗口的框架，主要包括三个步骤：利用不同尺寸的滑动窗口框住图中的某一部分作为候选区域；提取候选区域相关的视觉特征。比如人脸检测常用的Harr特征；行人检测和普通目标检测常用的HOG特征等；利用分类器进行识别，比如常用的SVM模型。

传统的目标检测中，多尺度形变部件模型DPM（Deformable Part Model）是出类拔萃的，连续获得VOC（Visual Object Class）2007到2009的检测冠军， DPM把物体看成了多个组成的部件（比如人脸的鼻子、嘴巴等），用部件间的关系来描述物体，这个特性非常符合自然界很多物体的非刚体特征。DPM可以看做是HOG+SVM的扩展，很好的继承了两者的优点，在人脸检测、行人检测等任务上取得了不错的效果，但是DPM相对复杂，检测速度也较慢，从而也出现了很多改进的方法。

基于深度学习的目标检测发展起来后，很多之前研究传统目标检测算法的研究者也开始转向深度学习。其实效果也一直难以突破。比如文献中的算法在VOC 2007测试集合上的mAP只能30%多一点，文献中的OverFeat在ILSVRC 2013测试集上的mAP只能达到24.3%。2013年R-CNN诞生了，VOC 2007测试集的mAP被提升至48%，2014年时通过修改网络结构又提升到了66%，同时ILSVRC 2013测试集的mAP也被提升至31.4%。

深度学习相关的目标检测方法也可以大致分为两派：基于区域提名的，如R-CNN、SPP-net、Fast R-CNN、Faster R-CNN、R-FCN；端到端（End-to-End），无需区域提名的，如YOLO、SSD。

R-CNN是Region-based Convolutional Neural Networks的缩写，中文翻译是基于区域的卷积神经网络，是一种结合区域提名（Region Proposal）和卷积神经网络（CNN）的目标检测方法。Ross Girshick在2013年的开山之作《Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation》奠定了这个子领域的基础，这篇论文后续版本发表在CVPR 2014，期刊版本发表在PAMI 2015。

R-CNN是第一个真正可以工业级应用的解决方案， 在此之后又先后出现了SPP-net、Fast R-CNN、Faster R-CNN、R-FCN、YOLO、SSD等研究。这些创新的工作其实很多时候是把一些传统视觉领域的方法和深度学习结合起来了，比如选择性搜索（Selective Search)和图像金字塔（Pyramid）等。

本解决方案选择Faster R-CNN作为Baseline，在此基础上进行修改，形成最终的解决方案。

### 1.1.2算法可行性分析

在目标检测任务的分支中，行人识别任务是从大量的摄像头角度的复杂街道背景下将其中的行人识别并标注，具有极其广泛的应用：智能辅助驾驶，智能监控，行人分析以及智能机器人等领域，这个领域目前的研究非常活跃。

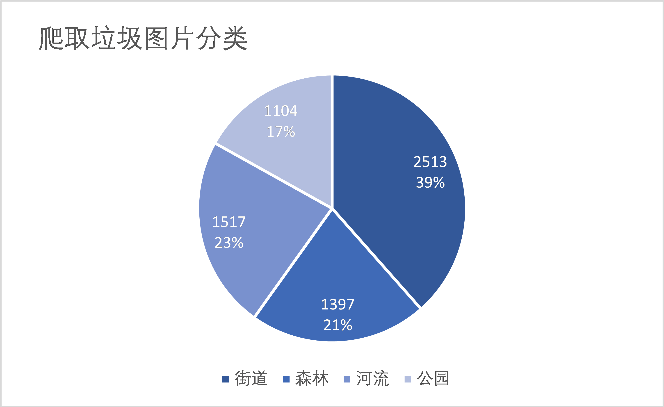
在现实应用中，可以认为多场景下的垃圾识别任务其背景、检测方式、任务复杂度与常规的目标检测任务（如自动驾驶、行人识别等）有较高的相似度，与此同时，由于街道场景相关的目标检测任务是大量存在的，在数据集方面可以通过爬虫大量获取与接到场景相关的目标检测任务数据集，如PASCAL VOC、MSCOCO、ILSVRC等，而目标检测算法在多个与本任务相关背景下，也有较高的识别准确率（在90%以上），因此认为对摄像头下的多场景垃圾目标检测任务也具有可行性。

## 1.2数据集

本项目的特点之一是采用弱监督学习来对数据进行预先标注，其中数据集的主要来源为自采集的街道垃圾数据集以及搜索引擎关键字爬虫获得，主要的场景有城市街道、河流、森林、公园等，预计在未来还可以将图像数据库扩充至更多场景如公共场所，基于多背景的图像可以保证在不同场景下保证采集到的图片的背景变化大，而垃圾的特征变化较小，可以有效地防止过拟合。将大量自采集图像和爬虫所获图像输入至基于弱监督学习模型的数据标注工具——WLM(Weak Label Master)，此工具为项目组自研发的高效数据自动标注工具，用于获取初代的垃圾检测数据库。在初代的机器标注数据库基础上，通过人工二次标注，建立可以用于模型训练的次代人工标注数据库，此数据库在用于训练弱监督学习模型的同时，也可用于训练垃圾目标检测的模型。采取这种方式可以获取大量的用于目标检测的垃圾图像数据，为后续扩充数据库，提供有力的技术支持。目前的数据库已有自采集图像1000张，网络爬取图像7000张，后续预计将扩充至20000张，强有力的数据支撑将大大增强模型的泛化能力。

### 1.2.1数据集采集

本项目的数据集主要有街道、河流、森林、公园、高山等背景下的垃圾图像。图像来源主要有两个部分：自采集，爬虫搜索引擎爬取。其中自采集的部分，由项目成员采集了403张城市街道背景下的垃圾图像；爬虫采集部分，通过Google, Baidu等搜索引擎爬取多背景下的垃圾图像6531张。图片的格式大多为.jpg或.png。其中图片中目标框的，同时采集数据的质量由于是爬虫采集，会出现过于模糊、遮挡严重或者目标大小不均匀，图像之间跨度较大，采集目标的大小无法根据使用的场景，接近应用场景的尺寸最佳，需要后续对图像进行处理，减小尺度不均匀对检测的影响。数据集的优点在于多场景下采集数据保证了图像变化的多样性，同时采集场景均为自然状态下的各种各样的照片，而不是人为地制造变化不大的图像数据，有利于防止模型过拟合。



图x：爬取垃圾图片类别

### 1.2.2数据集标注

首先通过探索性数据分析(Exploratory Data Analysis)将数据大致归类并对异常或难样本采取数据清洗，去除缺失值和噪声数据,得到训练可用的数据集。

1. **基于弱监督学习的数据自动标注**

监督学习通过学习大量训练样本来构建预测模型，其中每个训练样本都有一个标签标明其真值输出。尽管当前的技术已经取得了巨大的成功，但是值得注意的是，由于数据标注过程的高成本，很多任务很难获得如全部真值标签这样的强监督信息。因此，能够使用弱监督的机器学习技术是可取的。机器学习在各种任务中取得了巨大成功，特别是在分类和回归等监督学习任务中。预测模型是从包含大量训练样本的训练数据集中学习，每个训练样本对应一个事件或对象。训练样本由两部分组成：一个描述事件/对象的特征向量（或示例），以及一个表示真值输出的标签。在分类任务中，标签表示训练样本所属的类别；在回归任务中，标签是一个与样本对应的实数值。大多数成功的技术，如深度学习 ，都需要含有真值标签的大规模训练数据集，然而，在许多任务中，由于数据标注过程的成本极高，很难获得强监督信息，在图像分类任务中，训练数据的Groud-Truth标签由人类注释者给出; 虽然很容易从互联网上获取大量图像，而由于人工成本，只能注释一小部分图像。在面对大量数据集的情况下，传统的人工标注方式成本非常高，而且很耗时间，在某些数据集的标注中，需要领域专家才能够进行数据的精准、正确标注，这一问题则会更加明显。

为了解决这个问题，项目组基于弱监督学习，利用已经训练好的预训练模型，搭建一个可以用来预测垃圾标签信息的目标检测神经网络。则检测任务的类型类似于若监督学习中的不完整监督：部分样本label缺失。将获取的数据集均看作样本label的缺失，不完全监督考虑那些我们只拥有少量有标注数据的情况。

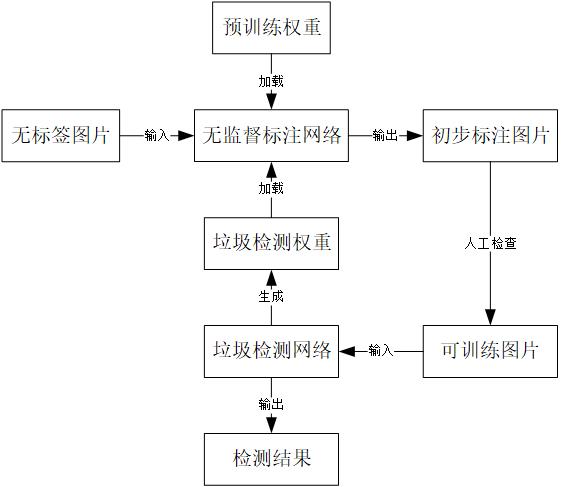
这些有标注数据并不足以训练出好的模型，但是我们拥有大量未标注数据可供使用。形式化表达为，模型的任务是从训练数据集中学习，其中训练集中有 n 个有标注训练样本（即给出的样本）和 u = m - n个未标注样本；其他条件与具有强监督的监督学习相同。将 n个有标注示例称为标注数据，将 u 个未标注示例称为未标注数据。

则由标注实例训练出的模型称为标注网络，其loss为对于多目标检测任务采用focal loss:

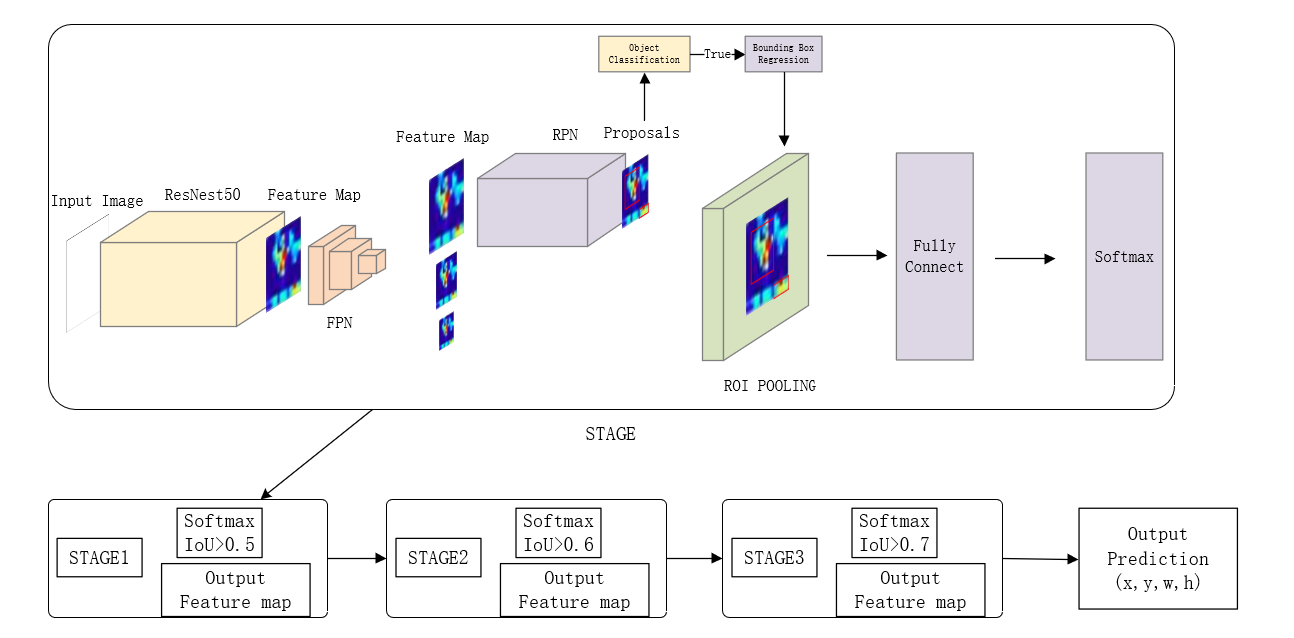
在本任务中为单目标检测任务，采用但目标检测常用的Triplet loss：

其中ρ是边距参数，是一个超参，d\_ij d\_ik :分别表示anchor sample 分别与positive sample、negative sample的欧式距离。

未标注示例u输入至标注网络，获得初代标签。



标注网络以ResNet50作为模型的骨干神经网络，设置三阈值Stage，逐步提升，不同阶段采用的IoU阈值不同，如果使用较低的IOU阈值，那么会学习到大量的背景框，产生大量的噪声预测。但是如果采用较高的阈值，这个检测器的表现往往会变得很差，两个主要的原因，第一就是随着IOU阈值的增加，正样本的数量会呈指数级的减小，因此产生过拟合。第二就是推理过程中出现于IOU的误匹配，也就是在训练优化感知器的过程中的最优IOU与输入proposal的IOU不相同，出现误匹配，这样很大程度上降低了检测精度。对于本任务，本项目经过实验，得到最优IoU阈值。



在初次使用时，加载第三方的预训练权重（如ImageNet），在训练垃圾目标检测网络后可以得到垃圾检测的权重，之后可以在下次使用标注网络标注时加载该权重，从而提高模型标注的准确率。

这一方法的三大优点是：

1、可以大规模标注，每个标注函数都可以用于成百上千个数据样本的标注。

2、可以利用海量的未标注数据，来构建大量虽然不完美但是足够好的大型训练数据集

3、这些标注可以用于训练一个具有大特征集的强大的判别分类器。即使只使用少量标注图像，每个数据样本依然可以有上千个特征。只经过测试，只需标注原数据集中30%的图片，就能让标注网络启动数据集预标注，能够一键标注剩余数据集。对比手工标注训练出的模型效果，其标注的准确率达到了95%以上。

（二）人工监督二次标注

通过弱监督学习的标注网络获得的初代标注数据并不能直接用于训练，但是已经完成了大部分的标注任务，大大减少了工作量。在这个基础上，项目组将对垃圾图像进行二次标注，确保图像是标注正确的，得到图像中异物标注框的四维向量信息(x, y, w, h)，其类别为garbage，将数据集按照 PASCAL VOC 格式进行4:1的比例划分为训练集和测试集。