

参赛队员姓名：杨学霖

中学：华南师范大学附属中学

省份：广东省

国家/地区：中国

指导教师姓名：杨晓安

报告标题：Faster R-CNN over Attention：
基于渐进区域提取的视频共享
单车检测

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知,除了文中特别加以标注和致谢中所罗列的内容以外,论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处,本人愿意承担一切相关责任。

参赛队员: 杨学森 指导老师: 杨晓宇

2017年 9月 12日

Faster R-CNN over Attention: 基于渐进区域提取的视频共享单车检测

杨学霖 华南师范大学附属中学

摘 要

随着共享单车的迅猛发展, 停放混乱已成为当前城市管理的最大痛点之一。常用手段人工巡查工作量大、目标性不强; 当前正在试点的电子围栏技术维护成本高、场地建设制约条件多。政府迫切需要并积极尝试新的管理方法与技术手段, 采用视频检测技术发现乱停放和藏匿的共享单车, 可以提供更加准确、快捷、直观的信息。

本文提出了 **Faster R-CNN over Attention** 算法, 它是一种基于渐进区域提取的视频共享单车检测方法, 按照从粗到细检测的视觉机理, 结合深度学习网络, 实现了对实际复杂城市环境下共享单车位置和种类信息的准实时检测。方法包括一个 **Attention Region** 提取模型与一个优化的 **Faster R-CNN** 网络。首先, 考虑共享单车的停放属性, 运用混合高斯模型, 通过迭代计算提取自行车可能存在的区域, 即 **Attention Region**。然后, 在优化的 **Faster R-CNN** 中通过多组卷积层与池化层得到特征图, 运用 **RPN** 得到候选区域, 通过 **RoI** 池化方式映射位置信息, 实现目标分类和位置精准提取。其中, 网络优化根据实际监控视频中共享单车停放特点, 运用 **k-means** 算法对共享单车长宽特征进行参数聚类, 从而提高检测准确性。

本文在包含 **4291** 张图片、**12697** 个已标注目标的共享单车分类数据集上进行了对比实验, 实验结果表明: 本文算法在检测能力与识别效果表现优秀, 对各类单车的检测准确率与识别准确率均超过 **90%**。

本文算法中, **Attention Region** 隔帧提取的模式与卷积神经网络的运算有机结合, 实现近似实时的检测; 算法具有广泛适用性, 学习到具有分辨性与鲁棒性的单车特征表达, 展现了一定的工程应用潜力。本文在运用与发展前沿技术的同时, 为管理共享单车提供了新思路、新方法。

关键词: 共享单车检测, **Attention Region**, 混合高斯模型, **Faster R-CNN**, 聚类, 目标检测

目 录

摘 要.....	1
1 研究背景与意义.....	3
1.1 共享单车发展现状.....	3
1.2 国内外研究现状.....	4
1.2.1 卷积神经网络研究现状.....	4
1.2.2 目标检测研究现状.....	5
1.3 图像处理技术在城市管理中的应用.....	6
2 研究目标与创新点.....	7
3 基于渐进区域提取的视频共享单车检测.....	8
3.1 总体技术路线.....	8
3.2 Attention Region 提取.....	9
3.3 Faster R-CNN.....	11
3.3.1 简述.....	11
3.3.2 卷积层、池化层与全连接层.....	11
3.3.3 候选区域生成网络.....	15
3.3.4 物体检测网络.....	16
3.3.5 共享计算训练.....	16
3.4 基于 k-means 聚类的网络优化.....	18
4 实验结果与分析.....	19
4.1 实验设置.....	19
4.2 结果分析.....	23
5 总结及展望.....	28
参考文献.....	29
致 谢.....	32

1 研究背景与意义

1.1 共享单车发展现状

互联网租赁自行车（俗称“共享单车”）是分享经济的一种新形态^[1]，起源于2016年，2017年呈现爆发式增长。其利用移动互联网技术，摆脱了固定停车桩的束缚，实现随时取用、随时停放、费用低廉，给市民出行带来了极大便利，深受老百姓欢迎，迅速成为社会热点及投资“风口”。据统计，截止至2017年7月，全国共有共享单车运营企业近70家，累计投放车辆超过1600万辆，注册人数超过1.3亿人次^[2]，使用摩拜、ofo等平台的人数比例最高，累计服务超过15亿人次，预计今后3年是共享单车用户人数高速增长期^[3]。

面对巨大的投放规模、庞大的使用人群以及各自为政的运营平台，如何引导共享单车有序发展、保障慢行交通通行秩序，是当下社会关注焦点，也是城市道路交通管理重要任务^[3]。

目前，停放混乱已成为共享单车管理最大的痛点之一。我国大部分城市道路上没有施划自行车停放区域、没有设置停放要求，共享单车随意停放、秩序混乱：经常堵塞地铁出站口、公交站进出口、人行道、盲道、应急安全出口等；有的直接占用机动车停车位、甚至停在车行道上，带来极大安全隐患。另一方面，有些使用者故意将车辆停放在小街小巷、小区或自家门口，甚至随意丢弃故障车辆，导致零散的共享单车越来越多，运营商归集的难度进一步增大。



图 1-1 晚高峰北京八王坟公交车站被上千辆共享单车“围困”数据来源：视觉中国



图 1-2 街头乱停放的共享单车数据来源：武汉晚报

对于共享单车停放管理难题，当前的常用手段仍局限于人工方式，如：运营商每日路面巡查，每1万辆共享单车投放市场至少需要50名管理人员；部分运

营商采用“信用分”方式鼓励用户自发举报违停；部分民众自发组成维护共享单车秩序的“单车猎人”，上报车辆故障，摆正乱停放车辆等。人工巡查的重点在主干道路和公共区域，工作量大，目标发散，且难以发现非主干道的零散单车^[4~6]。

电子围栏是当前正在试点的技术之一。2017年8月，经国务院同意，交通部等十部门联合出台了《关于鼓励和规范互联网租赁自行车发展的指导意见》^[1]，此“共享单车新政”鼓励：通过电子围栏等技术手段设定停车位，为自行车停放提供便利。为此，全国各主要城市和单车运营商都在加快电子围栏技术的研究和使用。正在试点的共享单车电子围栏技术主要有两类^[7]：第一类是基于卫星定位技术（GPS）的电子围栏。此方法正在北京通州试点，但受限GPS定位精度与通讯耗电量。第二类是基于蓝牙电子标签的电子围栏。此方法在北京朝阳区、上海等地试点，需要在停放点和车锁内加装蓝牙设备，提供相应的网络与供电服务，增加安装及维护成本，且挖坑立杆涉及市政规划管理问题。

可见，将信息化手段引入到管理中，既是对当前人工管理手段的重要补充，又是未来发展的趋势。视频图像处理技术作为信息化手段之一，能够充分利用现有的视频管理资源，有效弥补GPS及蓝牙电子围栏等技术手段的不足之处，为政府部门发现和解决问题提供一种简易、可行的方法。面对如何破解共享单车停放管理的困局，本文研究具有必要性和紧迫性。

1.2 国内外研究现状

1.2.1 卷积神经网络研究现状

随着深度学习的迅速发展，基于深度学习的物体检测识别研究已经相当深入。卷积神经网络（Convolutional Neural Network, CNN）是一种神经网络结构，作为一种高效检测方法，广泛应用于模式识别、图像分类、姿态估计等领域，在近年来引起重视。20世纪60年代，Hubel和Wiesel^[10]等对生物的大脑皮层进行研究，提出“感受野（Receptive Field）”的概念，并进一步发现了视觉皮层通路中对于信息的分类处理机制；Fukushima^[11]等受到此类研究的启发，在1980年提出一种基于感受野概念的神经认知机，可视为CNN的前身，也是第一个基于神经元之间层次结构组织以及局部连接性的人工神经网络。此后，研究人员开始尝试使用一种被称为“多层感知器^[12]”的人工神经网络来代替手工提取特征，并用随机梯度

下降方法训练,提出了用于计算误差梯度的反向传播算法(back propagation, BP),这一算法随后被证明十分有效。

在 1990 年, Lecun^[13]在一篇关于 CNN 的论文中,建立起经典的 CNN 模型架构,并在接下来的工作^[14]中对 CNN 做了改进,建立起名为 LeNet-5 的多层神经网络,用于手写数字的分类识别。LeNet-5 与传统的神经网络相同,通过梯度反向传播算法^[15]进行训练,通过模拟大脑视觉皮层,获取图像中具有代表性的特征,它无需对图像预处理,在手写数字的识别上的表现优于当时其他算法。此外,在同期工作中, Zhang^[16]提出了 SIANN (Shift-Invariant Artificial Neural Network) 来识别图像中的字符。但限于当时的计算能力及用于神经网络训练的大量标注数据的缺乏, CNN 在一些相对更复杂的问题上表现并不出色,因此未得到研究者的持续关注。

随着网络的快速发展,图像数据量急剧增大,各种公开的图像数据集具有齐全的人工标注;同时, GPU 的发展极大地减少了 CNN 的训练时间,使 CNN 能朝更深层次的网络架构发展。大规模图像分类问题上, Krizhevsky^[17]等在 2012 年提出了 AlexNet,并在经典的图像分类比赛 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 中取得第一名,识别错误率与第二名运用传统算法的队伍相比降低了近 10%。与 LeNet-5 比较, AlexNet 具有相似的架构组成,但网络层数更深。AlexNet 的成功后,越来越多有效的 CNN 模型出现,如 ZFNet^[18], VGGNet^[19], GoogleNet^[20]和 ResNet^[21]。这些网络的层数越来越深。其中, ILSVRC 2015 年的冠军 ResNet 的识别错误率为 3.57%,超越了人眼识别错误率 5.1%。ResNet 的网络层数为 152 层,几乎为 VGGNet 层数的 8 倍、AlexNet 层数的 20 倍。深层的网络能更好地学习输入、输出间复杂的非线性关系,得到更鲁棒的表达特征,然而,随之而来的网络的收敛困难以及过拟合问题也需要注意和解决。

CNN 是第一个真正意义上成功训练多层神经网络的学习算法模型,在其他计算机视觉问题上也有不俗的表现。目前,它已经成为人工智能领域的研究热点之一,用于语音分析、图像识别等研究。

1.2.2 目标检测研究现状

目标检测 (object detection) 指一个特殊目标 (或一种类型的目标) 从其它目标 (或其它类型的目标) 中被区分出来的过程。传统的物体检测方法大多通过

人工设计的特征进行特征提取，如 HOG^[22]、DPM^[23]、SIFT^[24]特征等。

2014 年，Ross Girshick 提出了 Region-based Convolutional Neural Network (R-CNN) 模型，它首先用图分割算法 Selective Search 提取可能存在物体的候选区域，然后提取候选区域的特征，送入支持向量机中分类；它在著名的 PASCAL VOC 公开物体检测数据集中表现得相对优异。此后，各种基于 CNN 的物体检测方法相继提出，检测效果越来越好，如 SSD^[25]，SPP-Net^[26]、YOLO^[27]等。

2015 年，Ross Girshick^[28]针对 R-CNN 物体检测耗时过长问题，提出了 Fast R-CNN，每幅待测图像只需进行一次特征提取，便可获得所有候选区域的特征向量；它在 PASCAL VOC 数据集上取得了更加优异的表现，但由于使用基于 CPU 的 Selective Search，候选区域的检测用时在算法的总用时中占有很大的比重，一定程度上影响了其工程应用性。2016 年，Girshick^[29]针对此问题提出 Faster R-CNN，与 Fast R-CNN 对比，Faster R-CNN 提取图像的候选区域时利用 RPN^[30] (Region Proposal Network) 代替 Selective Search。该算法能有效提高运算速度、检测效率与准确性，详细算法见章节 3.3。

在 R-CNN 系列目标检测的算法研究中，主要的方向为通过改变网络结构或训练方式提高对通用物体的检测效果，其要求图片像素较高、背景与目标较为简单。在实际场景，特别本文的应用场景中，监控视频图像会受到光照、天气、时间段等因素影响，视频中行人、树木、过往车辆等均会对作为目标的共享单车的检测造成干扰，在复杂环境下的目标检测具有一定技术难度。

1.3 图像处理技术在城市管理中的应用

近年来，各地政府依托“平安城市”、“智慧城市”建设，构筑了一张强大的安防网络来保证整个城市的安全，视频监控布点基本覆盖城市的大街小巷。以广东省为例，2011 年至 2015 年，政府部门共建成覆盖重点公共区域、重点行业、领域内涉及重要部位、易发案部位的公共安全视频图像采集点近 280 万个，采取 24 小时实时监控和录像。建设要求视频监控点位图像质量不低于 1080P，图像帧率 ≥ 25 帧/秒；图像存储时间 ≥ 30 天；图像在线率 $\geq 95\%$ ；在线图像完好率 $\geq 98\%$ ^[8,9]。计划到 2020 年，重点公共区域视频监控覆盖率达到 100%，新建、改建高清摄像机比例达到 100%；重点行业、领域的重要部位视频监控覆盖率达到

100%^[3]。可见，政府部门采集的视频资源内容丰富、图像质量好、可用性强。

通过图像监控系统，可以随时感知、回放城市的运行状态及各类事件，为城市管理提供了科学高效的手段。若要从海量、冗长、纷繁复杂的视频图像中提取出有用的信息，需要运用图像处理技术，针对不同的应用需求和场景进行分析、建模。其中的人脸识别、车辆识别、视频标签等领域的研究，是图像技术应用的优秀范例。

在共享单车停放管理问题上，发现违停的共享单车是首要及关键步骤。利用无处不在的监控视频采集到的信息，通过图像处理技术，进行自动检测，用于发现共享单车乱停乱放地点和状态，能极大减少人工路面巡查及视频巡查的劳动强度，从而为管理部门解决该问题提供一个简易、可行的方法。

在与城市管理相关的图像处理技术中，通用的目标检测旨在将人、机动车、自行车进行区分，自行车往往被作为一整类；目标检测领域对自行车的检测要求场景苛刻，而实际街景较为复杂。通过检索文献，目前，未发现有基于卷积神经网络、针对复杂条件下共享单车的精细识别研究。

2 研究目标与创新点

为解决共享单车乱停放问题，本文将对基于视频图像深度学习的共享单车检测方法进行研究。

本文研究目标如下：

(1) 为解决共享单车违停问题提供新思路，针对乱停放的共享单车，探索利用信息技术给出简易、可行、高效、科学的自动发现方法，弥补传统人工方法及其他技术手段的不足。

(2) 将视频监控系统采集到视频图像信息应用到共享单车的管理中，提高视频资源利用率，拓展视频监控系统的范围。

(3) 结合问题特性与前沿研究成果，在现有目标检测算法的基础上进行改进与优化，提出视频共享单车检测方法，实现在复杂环境下对共享单车的精细识别。

(4) 实际运用上，利用图像处理技术及深度学习算法建立共享单车检测模

型, 使其具备可扩展性, 能够通过训练与学习检测不同种类、不同停放方式的共享单车; 构造实用的视频共享单车停放检测系统, 实现准实时的快速检测。

本文创新点如下:

(1) 本文将图像处理技术运用于共享单车的自动检测, 通过训练卷积神经网络, 可实现从海量的视频监控图像信息中, 快速发现乱停放的共享单车的目的。该方法既突破了靠路面巡查或群众举报等人工方式耗时耗力、目标性不强等局限性, 又解决了视频图像的自动检测问题, 极大减少视频巡查的劳动强度, 为快速发现共享单车乱停放问题提供了科学、有效的手段。

(2) 本文提出了 **Faster R-CNN over Attention** 算法, 它是一种基于渐进区域提取的视频共享单车检测方法。区别于通用物体检测算法, 该方法可快速实现复杂环境下视频共享单车的位置检测与种类识别。它运用注意力模型与人眼视觉机理, 创新性地引入了 **Attention Region** 的概念, 并使用混合高斯背景模型对其进行提取。在 **Faster R-CNN** 的训练与学习中, 根据监控视频拍摄角度与共享单车停放特点, 采用 **k-means** 聚类算法对网络进行优化, 强化模型的针对性; 通过实验, 可以验证该方法检测效果良好, 优于改进前的网络。

(3) 政府部门建立的视频监控系统目前主要用于违法犯罪、城市治安、突发事件等监控, 其 7*24 小时全天候采集回来的图像信息覆盖面广、内容丰富, 但使用途径单一、利用率低。本文利用其作为载体进行共享单车停放识别, 拓展了视频监控系统的应用领域, 为视频应用方式从被动的事后取证到主动的事前防控, 应用模式从人工的搜索查看到智能化检索转变提供一个典型范例, 具有较强实用价值。

3 基于渐进区域提取的视频共享单车检测

3.1 总体技术路线

本文提出的 **Faster R-CNN over Attention** 是一种基于渐进区域提取的视频共享单车检测方法, 能检测复杂环境下的共享单车的位置与种类信息。该方法包括一个 **Attention Region** 提取模型与一个基于 **k-means** 优化的 **Faster R-CNN**, 它们通过提取渐进区域, 进行共享单车检测: 前者剔除运动目标干扰、提取共享单车的粗

略区域，Faster R-CNN 得到共享单车的精确定位。方法技术路线如下：

在 Attention Region 提取模型中，将外设监控摄像头采集的视频作为输入，运用混合高斯背景模型，用多个高斯分布描述图像像素的分布情况，进行迭代计算，提取自行车可能存在的区域，作为下一步的输入。

在 Faster R-CNN 中，利用多组卷积层与池化层对输入图像进行卷积与池化操作，得到图像的特征图，然后运用 RPN 对特征图进行滑窗检测与筛选，得到候选区域，其中引入了 k-means 进行优化。将候选区域的位置信息通过 RoI 池化方式映射回特征图，再用候选区域的特征对其进行分类和位置精修，得到最终目标类别与位置信息。

在 Faster R-CNN over Attention 的层次结构中，视频信息经过两层处理，渐进式地得到检测结果。算法技术路线图如下：

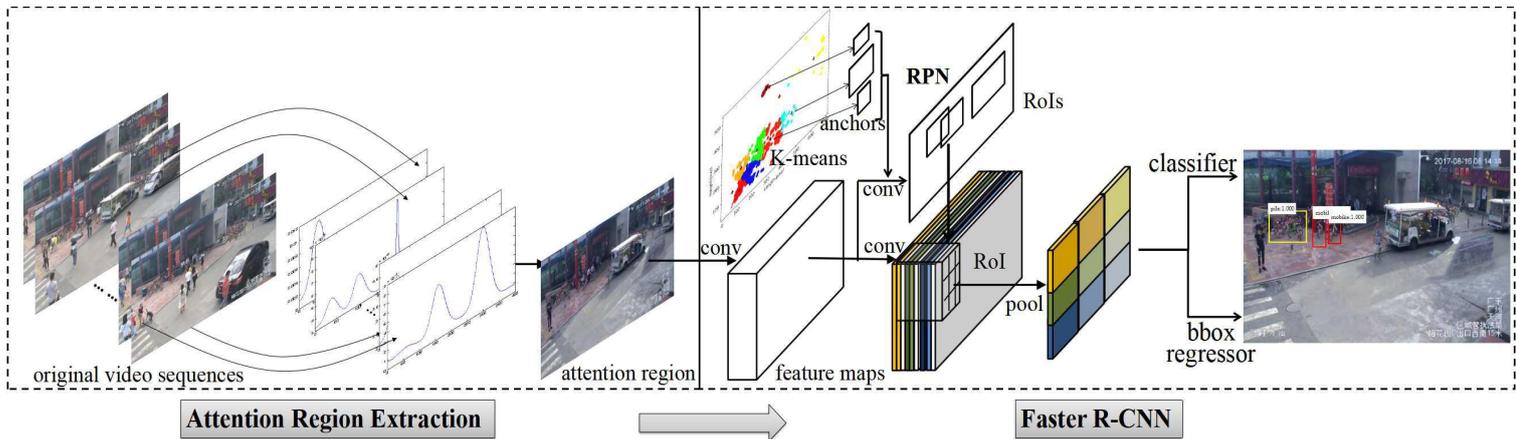


图 3-1 技术路线图

3.2 Attention Region 提取

深度学习中的注意力模型（Attention Model, AM）是一种由认知心理学中人脑注意力模型引入的资源分配模型，可描述为：在特定时刻，虽然具有全局视野，但注意力在画面某个焦点部分分配较多，在其他部分的注意力较少。人眼在寻找目标物体时，先寻找物体可能存在的区域（即感兴趣的区域），再从其中识别目标个体。例如，在视频中寻找停放的共享单车时，我们会先寻找可能存在共享单车的区域（如路面、街道两边），而不是先观察运动的物体（如行人、运动车辆）。

由注意力模型与人眼视觉机理，本文引入 Attention Region 概念。考虑需要检

测的（即感兴趣的）共享单车处于停放状态相对静止，并在一段时间内位置与状态保持不变，将 **Attention Region** 认为视频背景。

混合高斯背景建模是视频目标检测方法之一，目的是将前景和背景区分开来，通常用于提取视频中的动目标。鉴于本文中 **Attention Region** 的特点，决定采用鲁棒的混合高斯背景建模方法提取静目标，排除行人、运动车辆等动目标的遮挡干扰，提取 **Attention Region**。

混合高斯背景建模的过程主要包括匹配和更新两个关键步骤，匹配过程中，先为每个像素建立 K 个（ K 一般取 3~5）高斯分布来表征各像素在时间域上的变化规律，随后根据新一帧图像中的像素与背景模型进行比较，判断该点是否属于背景点；更新即对被赋予不同权值的高斯模型的参数将按照不同的机制进行更新。

具体实现为：设视频中的某个像素点在 t 时刻的像素值为 X_t ，在一定时间内 X_t 服从的混合高斯分布概率密度函数如式 (3-1) 所示。逐次读取视频序列的 N 帧图像，将图像当前像素值 X_t 与 k 个被赋予了权值的模型进行比较，如果它与某个高斯模型间的关系满足式 (3-2)，那么认为这个像素与该模型匹配，并按照式 (3-3) ~ (3-5) 更新该模型的分布参数，此时 $M_{k,t} = 1$ 。若现有的背景模型中没有可满足关系式 (3-2) 的，则权重值最小的模型会被新的模型替换，新模型的均值为当前像素值，标准差和权重分别赋予一个较大值和一个较小值。剩余的高斯模型保持均值和方差不变，权重按照式 (3-3) 衰减，此时 $M_{k,t} = 0$ 。之后各个模型按照 $\omega_{i,t} / \sigma_{i,t}$ 进行降序并重新排列。若前 B 个模型满足式 (3-6)，则这 B 个模型被认为是背景。

$$p(x_t) = \sum_{i=1}^k \omega_{i,t} \times \mathcal{N}(x_t, \mu_{i,t}, \sigma_{i,t}^2) \quad (3-1)$$

$$|X_t - \mu_{i,t-1}| \leq 2.5\sigma_{i,t-1} \quad (3-2)$$

$$\omega_{k,t} = (1 - \alpha) \times \omega_{k,t-1} + \alpha \times M_{k,t} \quad (3-3)$$

$$\mu_t = (1 - \rho) \times \mu_{t-1} + \rho \times X_t \quad (3-4)$$

$$\sigma^2 = (1 - \rho) \times \sigma_{t-1}^2 + \rho \times (X_t - \mu_{i,t})^2 \quad (3-5)$$

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right) \quad (3-6)$$

其中, k 为高斯分布的个数, \mathcal{N} 为 t 时刻第 i 个高斯分布, μ 为其均值, σ^2 为方差, ω 是第 i 个高斯分布的权重, α 为模型的更新率, 参数 T 表示背景所占比例。混合高斯用于背景的提取示意图如图 3-2 所示。

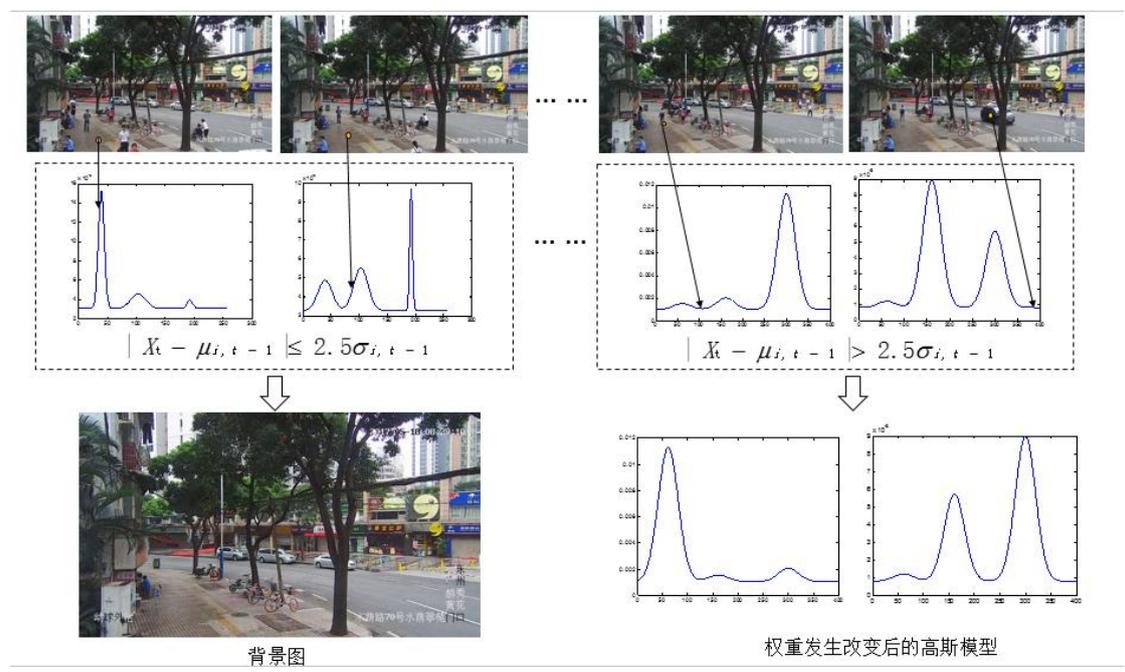


图 3-2 高斯混合背景提取示意图

3.3 Faster R-CNN

3.3.1 简述

Faster region-based convolutional neural network (Faster R-CNN) 于 2016 年提出, 是一种基于深度卷积神经网络的目标检测框架, 它包含了一个 Region Proposal Network (RPN) 与一个 Fast R-CNN 物体检测网络 (Object Detection Network)。该深度网络框架统一了目标检测的四个基本步骤 (候选区域生成, 特征提取, 分类, 位置精修); 为了快速检测, 两个网络训练时共用卷积层; 所有计算没有重复, 且都在 GPU 中完成, 运行速度大大提升。

3.3.2 卷积层、池化层与全连接层

作为 CNN 的一种, Faster R-CNN 使用参数共享的网络结构, 可以直接以原始

图像作为输入，节省了图像预处理步骤。Faster R-CNN 的网络结构包括了卷积层、池化层及全连接层。

卷积层

卷积层（convolutional layer, conv）用于特征提取，它的目标是学习输入图像的特征表达。对于图像的不同特征表达，一般卷积层包含多个卷积核。卷积核对图像的卷积运算如图 3-3 所示，对于给定的一张输入图像与一个 3×3 的卷积核，卷积运算时，卷积核与输入图像对应位置上的像素数值相乘，然后进行求和。移动卷积核，直到扫描完图像中的所有区域，每次求和得到的数值组成图像的卷积特征图（conv feature map）。

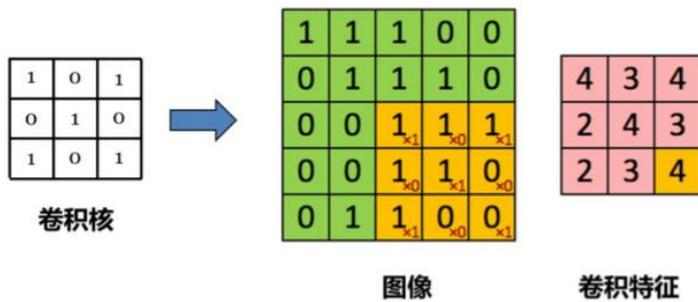


图 3-3 卷积运算示意图^[32]

在 CNN 中，卷积核的参数是待训练的，根据具体的任务目标，CNN 能够通过训练自适应地提取特征。

定义 x_i^l 为第 l 层的第 i 个特征图， k_{ij}^l 为第 l 层上当前其输入层第 i 个特征图和输出层第 j 个特征图的卷积核， b_j^l 为第 l 层第 j 个特征图的偏置项。则第 l 层第 j 个特征图计算如下：

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{3-7}$$

其中 $f(\bullet)$ 是激活函数，本文采用 ReLU (Rectified Linear Units) 函数，计算如下：

$$f(x) = \max(0, x) \tag{3-8}$$

ReLU^[31]函数是神经网络中常用的激活函数之一。如图 3-4，与传统激活函数 sigmoid、tanh 对比：ReLU 是单边函数，相比于反对称结构的 tanh，它被认为更

加符合神经元信号激活机理；sigmoid 与 tanh 在包含区域函数值接近 0，变化较平缓，ReLU 的梯度大于 0 时为常数，不会像 sigmoid 一样出现梯度消散的问题，能有效加快收敛速度；ReLU 只需比较、乘加运算，无需 pre-training，计算量较少；ReLU 强制某些输出为 0，输出有稀疏性，有效防止 CNN 过拟合问题；经实践证明，其训练后的网络具备适度的稀疏性，可视化效果和传统 pre-training 的效果相似，说明 ReLU 具备引导适度稀疏的能力。

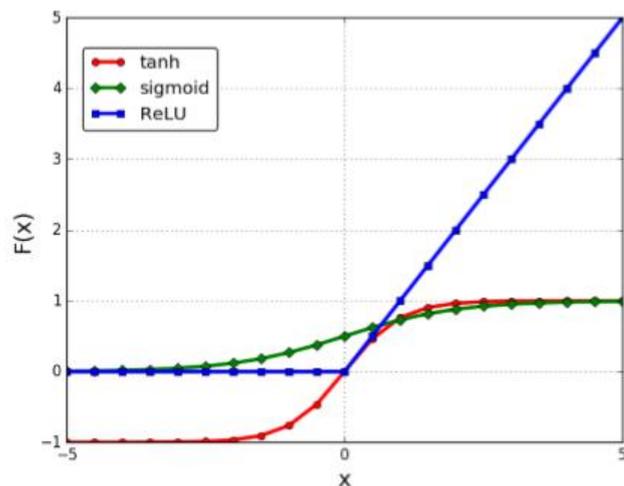


图 3-4 不同激活函数的曲线图^[32]

池化层

输入图像经过卷积层的处理后，可得到相应的特征图。对于尺寸较小的图像，直接将这此特征图输入分类器中进行训练和识别的计算量尚可接受；对于尺寸较大的图像，这种方法的计算量会显得十分庞大，且特征图构建的特征向量会存在大量冗余，出现过拟合的问题。因此，卷积层后通常连接一个池化层（pooling layer），在保留图像显著有效信息（即主要特征）的同时，减少计算量和参数（即降维），提高模型泛化能力，且具有不变性。

对于每个特征图 x_k^l ，经过相应的池化操作后输出的特征图为：

$$x_k^{l+1} = \text{pool}(x_k^l) \quad (3-9)$$

$\text{pool}(\bullet)$ 是池化计算函数，常用的是平均池化^[33]（Average-Pooling）和最大池化^[34]（Max-Pooling）。平均池化是计算池化区域中的特征平均值，而最大池化则是计算池化区域中的特征最大值。本文运用最大池化，以图 3-5 为例，进行最大池化操作时，给定大小为 4×4 的输入特征图，池化区域大小为 2×2 ，步长(stride)

为 2，计算每一步的池化区域中的最大值，得到池化后大小为 2×2 的输出特征图。平均池化同理。由于该池化运算对一定大小的区域进行特征统计，即便输入特征图发生了一定的平移，池化后的特征图也可能得到相同的输出特征，这使得特征对图像形变以及平移具有一定的不变性。

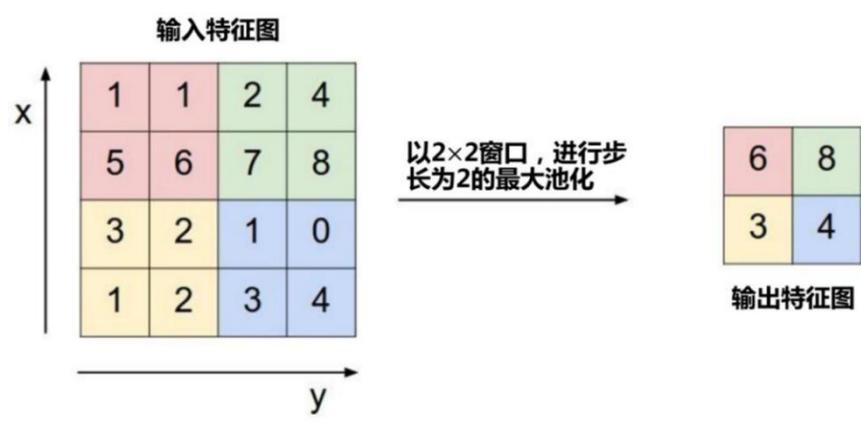


图 3-5 最大池化运算图^[32]

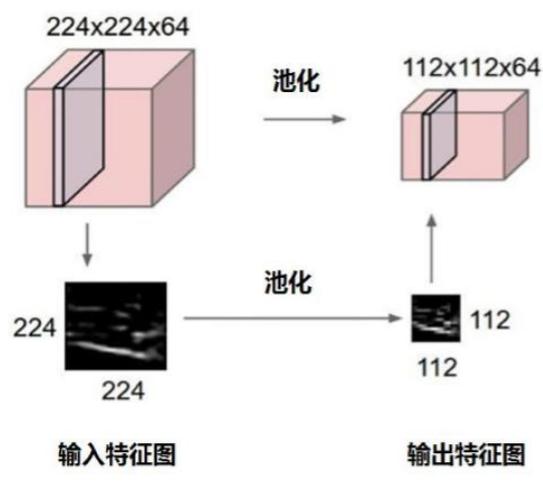


图 3-6 最大池化运算效果^[32]

全连接层

在经过多个卷积层与池化层的交替处理后，CNN 网络设置一个或多个全连接层进行特征的组合及网络的输出。全连接层（fully connected layer, fc）的每一个神经元都与输入层所有的神经元相连接，对所有的特征进行加权统计，计算公式如下：

$$x_j^l = f\left(\sum_i w_{ij}^l x_i^{l-1} + b_j^l\right) \tag{3-10}$$

其中 l 层是全连接层， x_j^l 是第 l 层的第 j 个神经元， w_{ij}^l 是第 l 层第 j 个神经元

与第 $l-1$ 层的第 i 个输入特征图中所有神经元连接的参数, b_j^l 是偏置项。全连接层可看作一种卷积层。若全连接层的神经元个数为 4096 个, 输入 512 个大小为 7×7 的特征图, 则输出为 1 个 4096 维的向量, 可看作全连接层利用 4096 个大小为 7×7 的卷积核对输入特征图进行卷积运算, 得到 4096 个大小为 1×1 的输出特征图。

当全连接层作为网络的最后一层 (即输出层), 它的作用是分类器。本文运用的分类器为 softmax, 它的输入为上一层的特征向量, 输出为各个类别的概率, 概率最高的类别为网络的预测结果。

3.3.3 候选区域生成网络

在候选区域生成网络 (Region Proposal Network, RPN) 是一个全卷积网络 (fully-convolutional network, FCN)。RPN 以一张任何大小的图片作为输入, 输出一组长方体的候选目标区域 (object proposals) 及它们的对象性分数。其中, 对象性 (objectness) 用于测量对象是物体还是背景; 同其他的目标检测方法^[35-37]一样, 所有“区域 (Region)”均是矩形的。

为使得 RPN 与 Fast R-CNN 物体检测网络计算时参数共享, 假设 RPN 与 Fast R-CNN 共享卷积层。为形成候选区域, 先使用一个小网络 (即滑动窗口, sliding window) 在最后的共享卷积层得到的卷积特征图上进行滑动扫描。滑窗每次与特征图上一个 $n \times n$ 的空间窗口全连接, 然后映射到一个低维向量, 最后将这个低维向量送入到 box-regression layer (reg) 与 box-classification layer (cls) 这两个兄弟全连接层中。

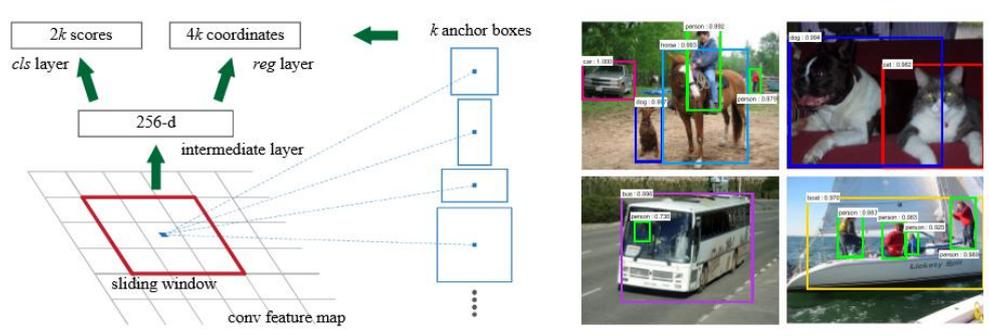


图 3-7 RPN 及检测实例^[30]
(右图数据来源: PASCAL VOC 2007 test)

图 3-7 中, 每次滑窗移动到一个位置, 我们由 k 个锚点框 (anchor boxes),

同时预测 k 个候选区域。这 k 个候选区域与给定作为参考的 k 个 `boxes` 的参数是相关的, 这些 `boxes` 可称为 `anchors`。相对应的, `reg` 层有 $4k$ 个输出, 为 k 个 `boxes` 的坐标编码; `cls` 层输出 $2k$ 个分数, 预测每个候选区域是否为目标的可能性。本文中, 每个 `anchor` 的中心是滑窗的几何中心, `anchor` 有大小 (`scale`) 与形状比例 (`aspect ratio`)。Faster R-CNN 的其中一个重要性质为平移不变性, `anchors` 以及计算与 `anchor` 相关的候选区域的方法都具有平移不变性。

3.3.4 物体检测网络

得到 RPN 的输出后, 送入物体检测网络。我们参考 Fast R-CNN 中多任务损失计算方式来使网络的损失函数最小化, 损失函数公式如下:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (3-11)$$

其中, i 是 `anchor` 的序号, p_i 是经预测得出的第 i 个 `anchor` 为目标物体的可能性。GT 标签 p_i^* 在 `anchor` 带正标签时为 1, 带负标签时为 0。 t_i 是预测得到的 `bounding box` 的 4 个坐标参数, t_i^* 是带一个正 `anchor` 的 `ground-truth box`。计算分类损失时, L_{reg} 是两个类别 (是或非目标) 的对数级损失; 用 $L_{reg}(t_i - t_i^*) = R(t_i - t_i^*)$ 计算回归损失, R 为[30]中定义的 `robust loss function`。 $p_i^* L_{reg}$ 指的是回归损失只当 $p_i^* = 1$ (即 `anchor` 为正) 时存在, 当 $p_i^* = 0$ 时无效。 `cls` 层与 `reg` 层的输出分别为 $\{p_i\}$ 与 $\{t_i\}$, 它们被 N_{cls} 与 N_{reg} 、以及平衡权值 λ 归一化。

对于回归, 本文运用 Girshick 提出的方法中对 4 个坐标的参数化^[38]:

$$t_x = (x - x_a) / w_a, \quad t_y = (y - y_a) / h_a, \quad t_w = \log(w / w_a), \quad t_h = \log(h / h_a),$$

$$t_x^* = (x^* - x_a) / w_a, \quad t_y^* = (y^* - y_a) / h_a, \quad t_w^* = \log(w^* / w_a), \quad t_h^* = \log(h^* / h_a);$$

式中 x 、 y 为 `box` 中心的坐标, w 、 h 分别为 `box` 的宽与高; 对于 x , x 、 x_a 、 x^* 分别为 `predicted box`、`anchor box`、`ground-truth box`, 对 y 、 w 、 h 同理。这可以被认为是从 `anchor box` 到邻近的 `ground-truth box` 的 `bounding-box` 回归。

3.3.5 共享计算训练

训练 RPN 时, 我们先将每个 `anchor` 用一个二态的种类标签赋值。有两种

anchor 被标记为正标签 (positive label), 原始设置如下, 本文参数设置见章节 4:

(i) 与一个 ground-truth box 的 Intersection-over-Union (IoU) 重叠最高的 anchor (或 anchors);

(ii) 与任意一个 ground-truth box 的 IoU 重叠大于 0.7 的 anchor。

其中, ground truth (GT) 为样本集中已标记的目标, IoU 为一种常用的检测评价函数, 数值上表示为模型产生的目标窗口和原来标记窗口的交叠率。这里, IoU 用于评价定位重合度, 为检测结果 (detection result, DR) 区域与 ground-truth box 区域面积的交集与并集之比, 公式如下:

$$IoU(GT_i, DR_j) = \frac{area(GT_i) \cap area(DR_j)}{area(DR_j) \cup area(GT_i)} \quad (3-12)$$

见式 3-12, 第 i 个 GT 与第 j 个 DR 的 IoU 为它们面积的交集与并集之比。可知, 一个 GT 可以对多个 anchor 赋值上多个正标签。实践证明, 采用规则 (ii) 基本上可以找到足够的正样本, 但是对于一些极端情况, 可采用规则 (i) 生成, 例如, 所有 anchor box 与 GT 的 IoU 不大于 0.7 的情况。

另外, 将与任意 GT 的 IoU 都小于 0.3 的 anchor 标记为负标签 (negative label)。一些非正负标签的 anchor 以及跨越图像边界的 anchor 对训练目标基本没有作用, 予以舍弃。

对于物体检测网络, 本文使用 Fast R-CNN, RPN 与 Fast R-CNN 的卷积层共享, 得到的卷积特征共享。以下是一个能够学习 RPN 与 Fast R-CNN 共享的卷积层的算法: RPN 与 Fast R-CNN 都是单独训练的, 他们训练时改变卷积层的方式不同, 且不会考虑未来需要利用候选区域。为此, 在训练过程中, 需要使用能够允许网络间共享卷积层的技术, 而非学习两个独立分离的网络。Faster R-CNN 的提出者 Girshick 在论文中提到, “Faster R-CNN 不是一个随意定义的 RPN 和 Fast R-CNN 的组合, 因为它们不能用共同反向传播优化。原本 Fast R-CNN 的训练依靠于确定 (fixed) 的候选区域, 在 Faster R-CNN 同时改变候选区域确定机制 (注: 即 RPN 中的参数) 的情况下是无法做到的。”

本文运用 Girshick 提出的交替训练法 (alternating training), 通过改变优化方法, 学习共享的卷积特征:

第一步, 先训练 RPN, 用 ImageNet 预训练的模型 (ImageNet-pre-trained model)

以及针对候选区域任务的端到端微调（fine-tuned end-to-end for region proposal task）对 RPN 初始化。

第二步，用第一步 RPN 生成的候选区域以及 Fast R-CNN 训练一个单独的检测网络，这个网络也由 ImageNet 预训练的模型初始化。此时，RPN 和 Fast R-CNN 没有共享卷积层。

第三步，用 detector network 初始化 RPN 的训练，但是我们固定了共享卷积层，只微调 RPN 特有的卷积层。现在 RPN 和 Fast R-CNN 有共享卷积层了。

第四步，保持共享的卷积层固定，微调 Fast R-CNN 的全连接层（fc）。这样，两个网络共享同样的卷积层，并且组成一个统一的网络。

3.4 基于 k-means 聚类的网络优化

Faster R-CNN 一般用于通用物体检测，对于视频图像共享单车检测，需对网络进行调整与优化以提高其针对性。本文提出运用 k-means 算法^[39]对网络进行优化，根据监控视频拍摄角度与共享单车停放特点，对视频中共享单车的长宽特征进行聚类，提高检测准确性。

K-means 算法是基于划分的聚类方法，属于无监督学习。对于参数 k，以空间中 k 个点为中心进行聚类，对最靠近他们的对象归类，通过迭代逐次更新各聚类中心的值，直至得到较为理想的聚类结果。该算法使同一聚类中的对象相似度较高，不同聚类中的对象相似度较低。聚类相似度由各类中的“中心对象”（引力中心）计算。该算法简洁快速，关键在于初始中心的选择和引力中心公式。本文随机选择初始中心，采用坐标均值公式计算点群中心。具体算法流程如下：

假设需要聚类的数据为 n 维数据，数据对象数目为 m，第 i 个 ($i = 1, 2, \dots, m$) 对象的坐标 $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}$ 是一个 n 维向量。

首先从 m 个数据对象任意选择 k 个对象 $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$ 作为初始聚类中心，根据其他对象与它们的相似度，逐个分配给与其最相似的（聚类中心所代表的）聚类。第 i 个数据对象所属类别 $c^{(i)}$ 计算如下：

$$c^{(i)} = \arg \min_j \|x^{(i)} - \mu_j\|^2, j = 1, 2, \dots, k \quad (3-13)$$

然后，计算更新后每个聚类的聚类中心，本文中聚类中心的坐标取该聚类中

所有对象对应坐标的均值:

$$\mu_j = \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}, \quad j = 1, 2, \dots, k \quad (3-14)$$

式中 $1\{c^{(i)} = j\}$ 表示的函数为:

$$1\{c^{(i)} = j\} = \begin{cases} 1, & c^{(i)} = j \\ 0, & c^{(i)} \neq j \end{cases} \quad (3-15)$$

不断重复这一过程直到标准测度函数达到最小值或者迭代次数达到指定的阈值。见公式 3-16, 本文采用所有数据对象与其所属的聚类中心点的距离之和作为标准测度函数。最终得到的标准测度函数如下:

$$f(x) = \min \sum_{i=1}^k \sum_{x \in \mu_i} dist(x, \mu_i) \quad (3-16)$$

算法中, 聚类中心的个数 K 需要预先确定, 实际中这个 K 值的选定是难以估计的。Fast R-CNN 网络中一般使用 9 种不同大小与比例的 anchor 类型, 故在 k-means 聚类时取 $K=9$, 聚类效果示意图 3-8, 得到参数见章节 4.1。

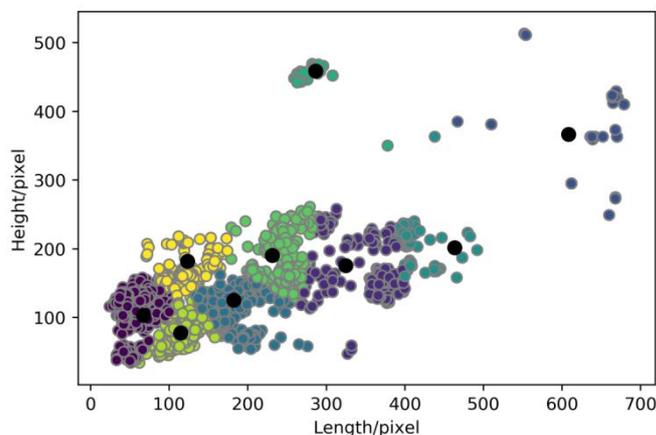


图 3-8 k-means 聚类效果示意

4 实验结果与分析

4.1 实验设置

为比较anchor优化前后的Faster R-CNN网络检测效果, 本文设置了一组对比

实验，即本文算法实验以及原算法的对比实验。本文算法先对标注目标的长宽进行k-means聚类，得到长宽比例较为理想的anchors，从而得到更准确的候选区域（proposals），以提升检测准确率及识别率；对比实验采用带有原始参数的Faster R-CNN算法（称为原算法）进行实验。网络输入均为经过attention region提取得到的图片。

实验环境

本算法应用的实验平台为 32GB 运行内存、3.6GHz 的 CPU 的 PC 机，同时配有 GEFORCE GTX GPU。操作系统为 Windows7 64 位，开发环境为 MATLAB 2014a，深度学习框架为 Caffe。

数据集设置

为了验证方法的有效性，本文构建了一个包含 4291 张图片、12697 个已标注目标的共享单车分类数据集。其中，选择 3861 张作为训练验证集，使训练得到的模型更加稳定、准确；选择 430 张作为测试集，验证模型检测效果。

数据来源主要为城市社会治安监控视频，拍摄角度为半俯拍，其原始场景主要包括以下几个典型场景：地铁站出入口处、路侧人行道、公交车站、医院门口等，数据集图片包含白天、夜间、晴天、雨天等各种环境，涵盖多样的光照、天气、时段以及车辆摆放姿势，具有较强的代表性，数据集部分图像如图 4-1 所示。根据实验目的，将数据集每张进行分类标注，标记示例如图 4-2 所示，以图像左上角为坐标原点。



图 4-1 数据集部分图像

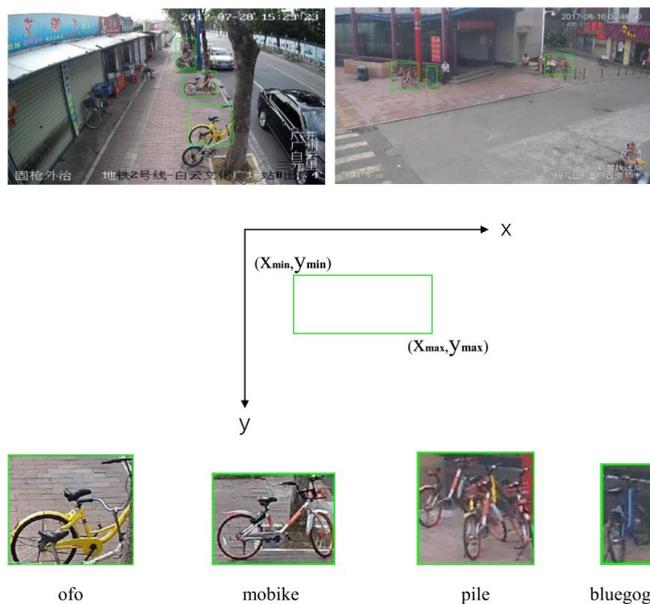


图 4-2 共享单车标记示例

经标注，实验数据集中的目标总量为 12697 个，目标类别共分为 4 类：ofo 单车、mobike 单车、bluegogo 单车以及多辆自行车堆积区域（pile）。各类别占比如图 4-3 所示，不同车辆的数量从 800 到 3600 不等，自行车堆积区域样本充足，与实际环境中车辆的分布不均匀等情况相符合。用于检测方法测试的目标总量为 1427，每类目标分布如图 4-4 所示。

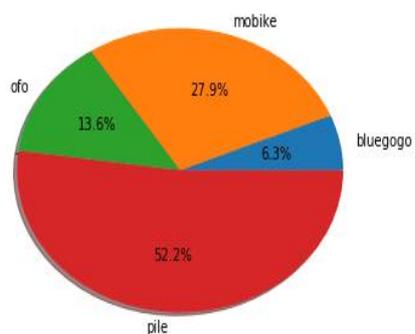


图 4-3 数据集各类目标分布比例图

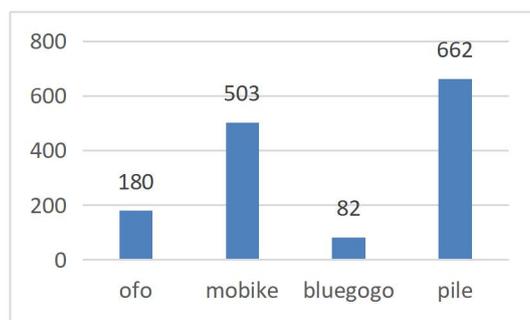


图 4-4 测试集各类目标分布图

评测指标设置

实验中，设置若干评测指标如下：

检测的判断基准为定位重合度 (IoU)，详细介绍见章节 3.3.5 式 3-12。

在评判算法检测效果时，评判基准为检测准确率 (accuracy)，如式 4-1 所示：

$$acc(i) = \frac{N_i(IoU \geq T)}{N_i} \quad (4-1)$$

其中， $acc(i)$ 为检测准确率， T 为认定物体检测成功的阈值， $N_i(IoU \geq T)$ 为第 i 类目标物体检测成功的数量， N_i 为该类目标总数量。

在评判算法识别效果时，评判基准为识别精准率 (precision)，计算公式如式 4-2 所示：

$$prec(i) = \frac{TP_i}{TP_i + FP_i} \quad (4-2)$$

其中， $prec(i)$ 为识别率， TP_i 为第 i 类目标识别正确的总数， FP_i 为第 i 类非背景区域判别为目标区域的总数。

参数设置

本文提出运用 k-means 算法对网络进行优化，将所有的目标数据的宽高作为变量，进行包含九种不同长宽比例类别的聚类，聚类效果如图 3-8 所示；聚类结果如图 4-5 所示。

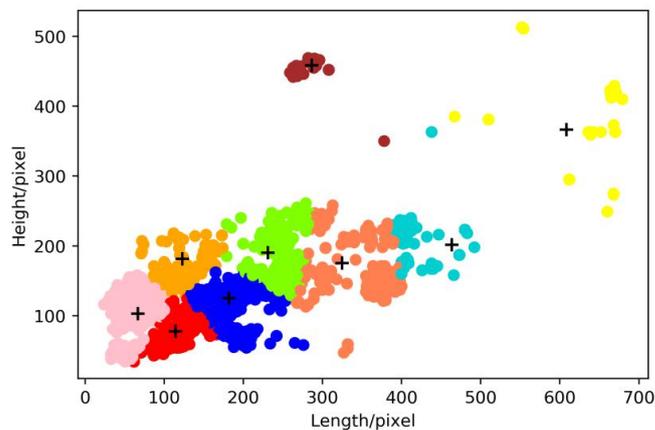


图 4-5 k-means 聚类结果

根据聚类结果, 本文算法实验对参数进行微调, 将与 anchors 产生有关的参数 scales 以及 ratios 修改为 scales=[6, 10, 20, 30]; ratios=[0.6, 1.5, 2.3]。共产生 12 个原始 anchors。

对比实验采用 Faster R-CNN 算法原始参数, anchors 生成指标为: scales=[8, 16, 32], ratios=[0.5, 1, 2]。聚类所得参数与两类 anchors 生成矩阵如表 4-2 所示。

表 4-1 k-means 聚类参数

Length/pixel	Height/pixel
231.3131313	189.496633
463.5333333	201.6044444
62.95156125	104.4059247
114.2843735	76.16564993
608.3362256	366.3210412
286.4364641	458.441989
115.733614	217.6512327
324.9880192	175.4816294
178.6647666	126.347177

表 4-2 anchors 生成矩阵

序号	原算法 anchors 参数/pixel				本文算法 anchors 参数/pixel			
1	-83	-39	100	56	-54	-30	71	47
2	-175	-87	192	104	-96	-56	113	73
3	-359	-183	376	200	-201	-121	218	138
4	-55	-55	72	72	-306	-186	323	203
5	-119	-119	136	136	-30	-51	47	68
6	-247	-247	264	264	-56	-91	73	108
7	-35	-79	52	96	-121	-191	138	208
8	-79	-167	96	184	-186	-291	203	308
9	-167	-343	184	360	-24	-66	41	83
10	/	/	/	/	-46	-116	63	133
11	/	/	/	/	-101	-241	118	258
12	/	/	/	/	-156	-366	173	383

4.2 结果分析

实验结果表明本文提出的 Faster R-CNN over Attention 算法相较于单纯的

Faster R-CNN 算法，能够剔除运动目标的干扰，更适用于共享单车检测的应用场景，在检测与识别方面效果表现优秀。通过将调整 anchors 后的算法与带有原始参数的算法进行对比，可知：本文算法能够较为完整、准确地检测复杂环境下共享单车停放位置，具有鲁棒性，它成功检测测试集中超过 90% 的目标，检测准确率较高；共享单车种类识别错误总量有了明显的降低，本文算法在识别效果上有了一定的提升。

共享单车检测与识别结果具体分析如下：

共享单车检测结果分析

部分共享单车检测结果如图 4-6 所示，图中绿色框代表的是人工标定的位置，红色框代表检测网络预测的位置。由图可见，本文算法检测到的目标位置与人工标定实际位置重合度较高，同时，即便是在光照条件较差或雨水环境下，算法能够较完整的检测出相应目标的位置，具有较好的鲁棒性。



图 4-6 部分共享单车检测结果示意图

为了更直观地分析 anchors 改变对各类检测目标检测准确率的影响，实验绘制出了各类目标在对比实验中的检测效果对比图（绿色为原始 anchors 的检测结果，红色为 anchors 修改后的检测结果）。如图 4-7 所示，在同一 IoU 阈值下，检测准确率越高，检测效果更好；在同一检测准确率下，IoU 阈值越大，检测效果越好。由图可得，4 类目标在任意 IoU 阈值下，检测准确率都有了一定的提升，尤其是 ofo，在阈值小于 0.9 时，检测准确率具有明显的增加。

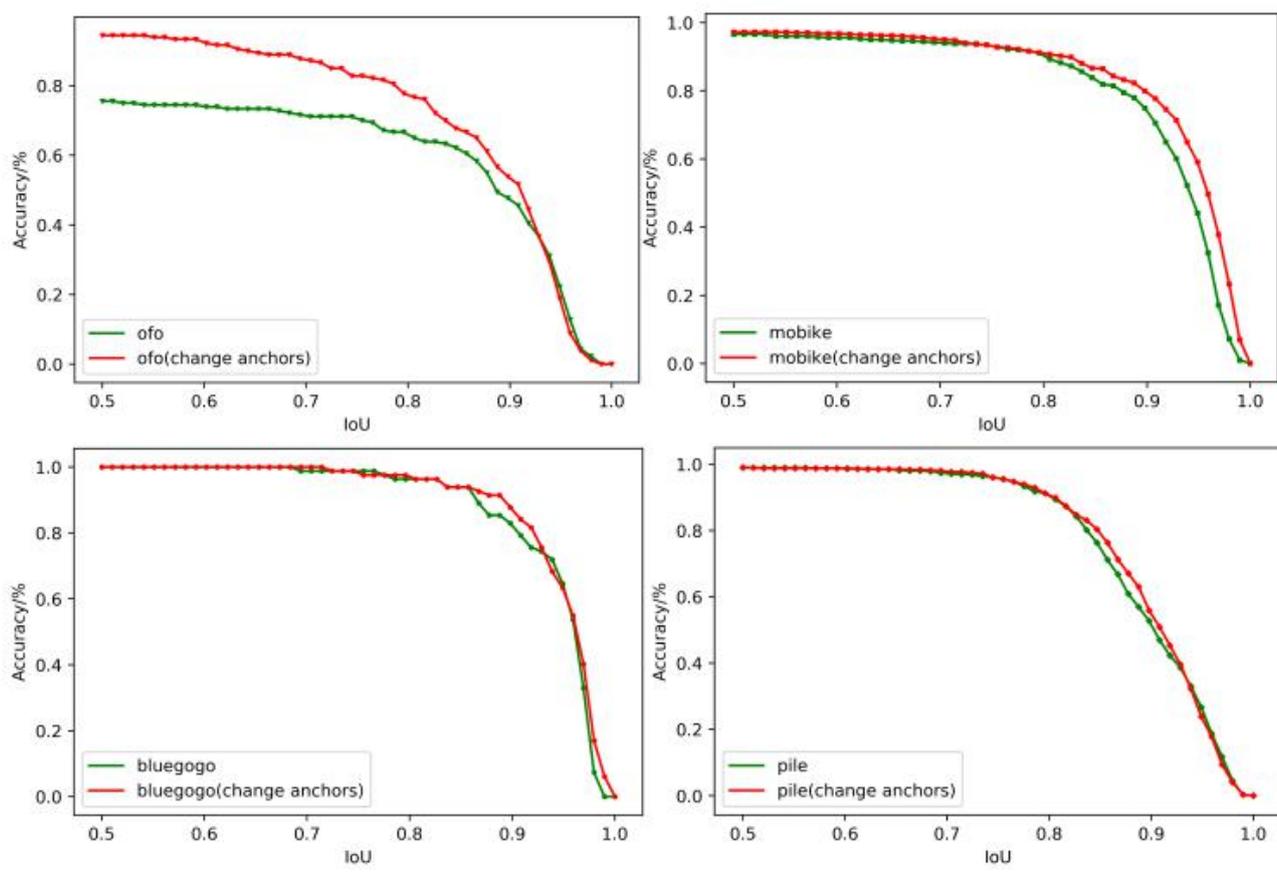


图 4-7 anchors 修改前后各类别物体检测结果对比图

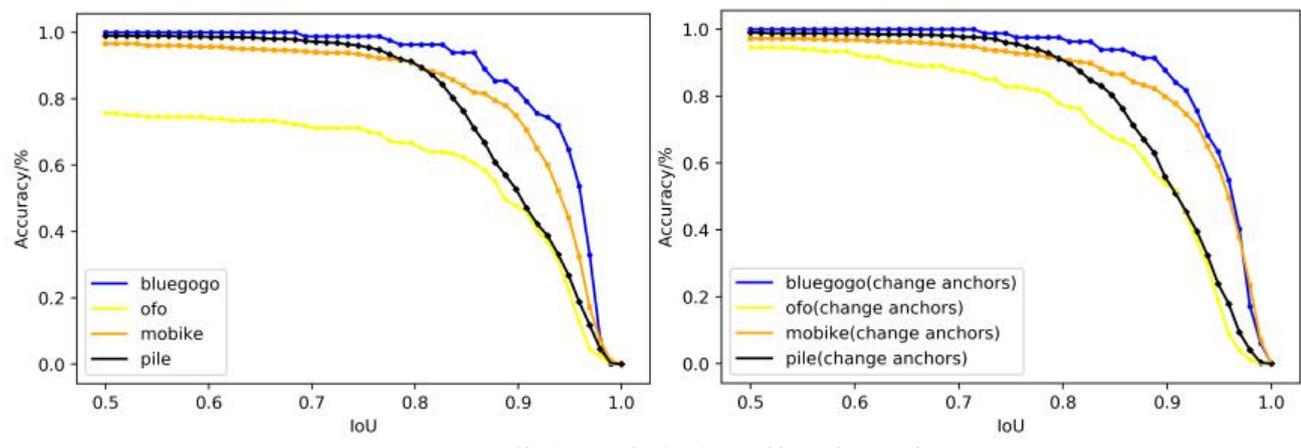


图4-8 anchors修改前后各类别物体检测结果示意图
 (左图: 修改前, 右图: 修改后)

如图4-8所示, 相对于原算法, 本文算法的实验结果中ofo类检测准确率有了较大的提升。从各类目标的横向比较可知, 在同一固定IoU阈值下, 当IoU阈值小于0.8时, bluegogo的检测准确率最高, 其次是pile和mobike, ofo的检测准确率最低; 当IoU阈值介于0.8与1之间时, bluegogo检测准确率依然最高, mobike的检测准确率逐渐高于pile, ofo检测准确率最低; 在同一准确率下, IoU阈值大

小依次是bluegogo、mobike、pile、of0, 说明bluegogo检测定位到的位置信息更为精细。

在物体检测中, 一般认为IoU大于0.5即视为物体检测成功。在anchors修改后, 当IoU=0.5时, 各类物体的检测准确率均大于0.9, 说明在测试集中有超过90%的待检测目标能够被成功的检测出来, 检测准确率较高。为了平衡物体的检测效果与识别效果, 物体识别中, 一般将IoU阈值取为0.5~0.8之间, 根据图4-8所作曲线图, 在IoU阈值大于0.7时, 检测准确率下降趋势越来越大, 故本文算法将IoU阈值取为0.7, 即认为IoU大于0.7为成功检测。

共享单车识别结果分析

共享单车识别对比实验结果如表4-3所示, 其中, 原算法指带有原始参数的Faster R-CNN, 本文算法是运用聚类优化后的方法。

表 4-3 共享单车识别结果

目标类别		of0	mobike	bluegogo	pile	平均值
识别精准率/%	原算法	72.2295	90.8895	99.4574	90.3351	88.2279
	本文算法	90.5216	90.9091	99.5669	89.9224	92.7300

在根据聚类结果对anchors进行改进后, of0的识别率有了明显增加, 识别率增加了约18.3%, 其余三类波动不大。本文算法对目标数据进行聚类, 根据聚类结果调整的anchors能够提取到更贴近于目标物体的proposals, 提高了检测重合度, 减少了两个距离相近的同类物体检测识别为同一物体的可能性, 从而提升目标物体的识别精准率。

在识别结果中, 主要识别错误情况有以下几类: 误检、错检和漏检。图4-9、图4-10、图4-11是相应的图例。每种类型错误示例图中, 左侧为原算法检测结果, 右侧为本文算法检测结果。从图中可以看到, 本文算法能够避免很多原算法中产生的错误, 提高了识别精准率。

- (1) 误检: 将部分背景区域的物体识别为目标物体类别。



图 4-9 误检情况示例

(左图: 原算法背景区域停放行人被识别为 ofo, 右图: 本文算法无误检)

(2) 错检: 将目标类别识别错误



图 4-10 错检情况示例

(左图: 原算法 bluegogo 识别为 mobike, 右图: 本文算法识别正确)

(3) 漏检: 目标区域未被检测到



图 4-11 漏检情况示例

(左图: 原算法漏检bluegogo, 右图: 本文算法未漏检)

为进一步分析本文算法识别效果, 对所有错误图像的结果进行统计。如图 4-12 所示, 错误识别结果中, 主要错误来源于将背景区域的文字识别为 pile 类别; 在原算法的试验中, ofo 产生了大量的漏检, 主要原因在于 anchors 设置参数尺寸过大, 产生的 proposals 也较大, 从而容易将两辆或多辆较小目标识别为一辆; 本文算法将 anchors 参数调整, 漏检情况明显减少。从整体上看, 识别错误总量有了一定的降低, 证明了本文算法在识别效果上有了一定的提升。

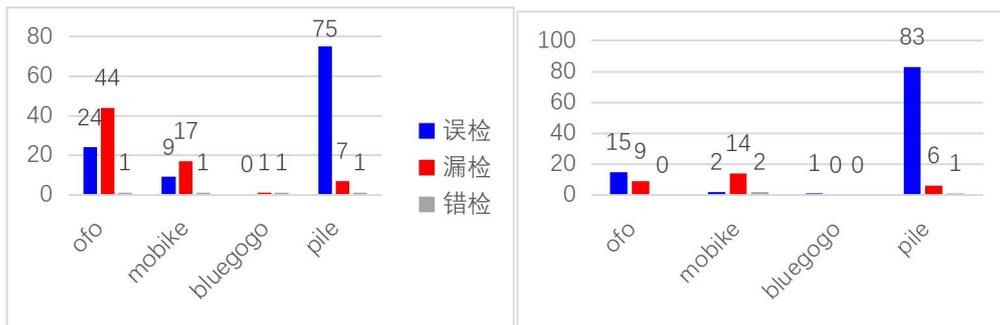


图 4-12 各类别物体识别错误统计图 (左图: 原算法实验, 右图: 本文算法实验)

5 总结及展望

为解决共享单车乱停放问题, 本文对基于视频图像深度学习的共享单车检测方法进行研究, 提出了 **Faster R-CNN over Attention** 算法, 并在构建的共享单车类别数据集上进行实验。结果表明, 相比原始的 Faster R-CNN 算法, 本文算法在此应用场景下, 具有更加优异的检测效果及识别能力。在此, 提出以下改进方向与展望:

(1) 本文算法在识别实验中存在将部分背景区域识别为目标区域的情况, 对此可以在 Attention Region 提取或网络训练方面进行优化。例如: 为防止背景中的文字区域被判断为共享单车堆叠, 除了消除运动中的干扰物, 还可去除文字区域, 从而提取更符合要求的 Attention Region; 也可以尝试在训练时增加相应负样本的数量, 优化识别效果。

(2) 由于本文中的数据集由真实视频间隔数帧提取, 同一个场景得到的 Attention Region 较为相似, 对此, 可以增加数据集的数量或增加视频图像的场景, 使得训练数据间变化更大、数据种类更丰富, 以建立一个更加完善的共享单车数据集。

(3) 实际运用中, 可基于本文算法建立共享单车管理系统, 对城市中共享单车的运行、停放情况开展集成化、标准化的监测、分析与管理。例如, 完善管理流程, 建立自动报警机制, 发生异常事件时 (如共享单车停放在禁停区域, 或零散停放在小街小巷等), 自动通知相关部门处置; 开放系统开源接口, 允许除公安监控外的社会单位、商户、居民社区或住宅小区等按照规范建设的公共视频图像采集点使用系统, 允许共享单车运营商增添单车品种, 及时获取数据, 调整

投放与管理策略；建立统一的互动式管理平台，鼓励群众自主监督，促进公众素质的提高。通过政府管理、行业自律、企业和民众自治三方合力，共同使共享单车这一新生事物更加健康、有序地发展。

参考文献

- [1] 交通运输部, 中央宣传部等.关于鼓励和规范互联网租赁自行车发展的指导意见[Z].2017-08-01
- [2] 公安部道路交通安全研究中心.共享单车发展对交通管理的影响分析及对策建议研究报告[Z]. 2017
- [3] 公安部道路交通安全研究中心.在共享经济视角下推进共享单车道路交通管理的思考及建议[Z]. 2017
- [4] 广州市交通委员会. 广州市中心城区城市道路自行车停放区设置技术导则[Z].2017-03-21
- [5] 成都市公安局交通管理局. 成都市中心城区公共区域非机动车停放区位技术导则[Z].2016-12-26
- [6] 厦门市市政园林局. 厦门市自行车停放区设置指引(试行) [Z].2017-02
- [7] 廖应成.共享单车电子围栏技术使用调研报告[Z]. 2017-08-08
- [8] GAT 669-2008, 城市监控报警联网系统技术标准[S].
- [9] GAT 367-2001, 视频安防监控系统技术要求[S].
- [10] Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex[J]. Journal of Physiology, 1968, 195(1):215.
- [11] K. Fukushima. A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological cybernetics, 1980.
- [12] Ruck D W, Rogers S K, Kabisky M . Feature selection using a multilayer perception[J]. Journal of Neural Network Computing, 1990, 2(2):40-48.
- [13] Lecun Y, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]. Advances in Neural Information Processing Systems. Morgan Kaufmann Publishers Inc. 1990:396-404.
- [14] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [15] Hecht-Nielsen. Theory of the backpropagation neural network[C]. International Joint Conference on Neural Networks. IEEE, 1989:593-605 vol.1.
- [16] Zhang W, Itoh K, Tanida J, et al. Parallel distributed processing model with local space-invariant interconnections and its optical architecture[J]. Applied Optics, 1990, 29(29):4790-4797.
- [17] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
- [18] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C].

European conference on computer vision. Springer International Publishing, 2014: 818-833.

[19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[20] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[21] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[22] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. computer vision and pattern recognition, 2005: 886-893.

[23] Felzenszwalb P F, Girshick R, 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.

[24] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[25] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. european conference on computer vision, 2015: 21-37.

[26] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[C]. european conference on computer vision, 2014: 346-361.

[27] Redmon J, Divvala S K, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. computer vision and pattern recognition, 2015: 779-788.

[28] Girshick R. Fast R-CNN[J]. Computer Science, 2015.

[29] Girshick R. Fast R-CNN[C]. international conference on computer vision, 2015: 1440-1448.

[30] 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, 2016:1-1.

[31] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In ICML, 2010.

[32] 袁敏贤. 基于图像理解的车辆类型精细识别研究方法研究[D]. 广东: 中山大学, 2017.

[33] Wang T, Wu D J, Coates A, et al. End-to-end text recognition with convolutional neural networks[C]. Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012: 3304-3308.

[34] Boureau Y L, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition[C]. Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 111-118.

[35] Zitnick C L, Dollar P. Edge Boxes: Locating Object Proposals from Edges[C]. european conference on computer vision, 2014: 391-405.

[36] Szegedy C, Reed S, Erhan D, et al. Scalable, High-Quality Object Detection[J]., 2014.

[37] Uijlings J, Sande K E, Gevers T, et al. Selective Search for Object Recognition[J].

International Journal of Computer Vision, 2013, 104(2): 154-171.

[38] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[39] Macqueen J. Some Methods for Classification and Analysis of MultiVariate Observations[C]// Proc. of, Berkeley Symposium on Mathematical Statistics and Probability. 1967:281-297.

致 谢

感谢在开展本次研究过程中给予帮助和指导的华师附中老师,他们鼓励我大胆进行课题研究,协助我确定课题的角度和方向。

感谢广东省公安厅及交通管理局的专家们。他们为我提供了很多最新的研究资料、管理思路及实验素材。他们渊博的知识与使命感使我能够充分了解当前政府对共享单车管理的重点与难点,找到研究的突破口。

感谢中山大学智能交通研究中心与计算机学院的教授、学长学姐们、以及我的朋友们。他们给了我很多有用的建议,教会我很多专业知识。他们的无私帮助让我少走了很多弯路。

感谢给予支持、帮助的父母,他们在物质上、精神上支持我,使我的研究得以顺利进行。

在此衷心地感谢各位!