

分类号:TP39

10710-2019124108



长安大学

# 硕士学位论文

基于深度学习的道路斑马线与行人运动检测算法研究与应用

付阳阳

导师姓名职称	王夏黎 副教授		
申请学位类别	工学硕士	学科专业名称	软件工程
论文提交日期	2022 年 3 月 28 日	论文答辩日期	2022 年 5 月 28 日
学位授予单位	长安大学		

# Research and application of road crosswalk and pedestrian motion detection algorithm based on deep learning

A Thesis Submitted for the Degree of Master

**Candidate: Fu Yangyang**

**Supervisor: Prof. Wang Xiali**

Chang'an University, Xi'an, China



## 摘 要

随着人工智能技术的发展,无人驾驶汽车与辅助驾驶汽车逐渐走入人们的日常生活。道路斑马线区域的交通场景中,行人与车辆是场景中的重要交互者,其中斑马线在保证穿行行人交通安全的重要设施,准确检测斑马线和行人的运动是无人车驾驶汽车与辅助驾驶汽车保证行人安全不可或缺的一项关键技术。因此开展道路斑马线与行人运动检测的研究对交通安全、无人驾驶汽车与辅助驾驶汽车的发展具有重要的实际意义。

由于斑马线与行人检测存在背景复杂、行人遮挡和姿态多样等问题导致行人行为识别困难,无法准确预测行人穿行斑马线意图,因此,本文针对道路斑马线与行人运动检测中存在的问题,采用深度学习的方法进行研究,具体研究内容如下:

(1) 为了改善现有斑马线检测精度低、检测不全面及实时性差等问题,提出了一种改进的 SegNet 语义分割模型。改进模型以语义分割中的原始 SegNet 模型为基础,先融入金字塔池化模块,减少斑马线细节信息丢失;后将模型改为编码不变、解码减少的不对称结构,减少网络参数,精准检测斑马线像素点位置。扩充斑马线图像数据集后进行实验,将四种模型进行对比,改进 SegNet 模型检测斑马线精度达到了 97.6%,与其他模型相比检测精度得到有效提高,且运行速度加快,满足当前实时性检测的需求。

(2) 针对行人检测中存在的耗时大与小目标检测困难问题,提出了一种改进 Faster-RCNN 算法。改进模型以 Faster-RCNN 算法为基础,先优化网络结构,替换特征提取网络模型,提高了提取特征的优越性;再优化了聚类算法,引入高斯聚类,减小计算量,提升了检测速度;最后将 SSD 算法、Faster-RCNN 算法与改进算法进行实验对比,改进算法检测精度更高、速度更快且实时性更好,能有效检测小目标行人,具有实际意义。

(3) 针对行为识别中存在的识别准确率低问题,在研究分析相关识别算法的基础上,给出了一种基于 CNN-LSTM 行人识别算法。该算法主要思想是使用 CNN 网络模型用于提取图像特征信息,以 LSTM 网络模型对行人行为分析时间序列进而完成行人行为识别分类。将 CNN-LSTM 模型与 ResNet-LSTM 模型进行实验对比,实验结果显示 CNN-LSTM 模型的损失值更小,且模型趋于稳定速度更快,模型波动较小,CNN-LSTM 模型的行为识别准确率更高。

(4) 针对行人穿行道路斑马线意图预测时间长与预测距离短的问题,构建了一种基于贝叶斯网络的行人穿行道路斑马线意图预测模型。该模型主要依据行人与车辆两个视

角分析的意图预测影响因素,设计本文交通场景,给出行人穿行道路斑马线的概率估计。训练分析得到最优预测模型,针对不同场景表现分析差异,本文模型能够提前 0.5s-0.8s 预测出行人穿行斑马线意图,相对距离为 15m-25m 时可完成行人穿行斑马线意图预测,结果显示,在各种行人穿行斑马线的场景中,本文构建的预测模型能够有效预测行人穿行斑马线意图。

(5) 基于本文以上研究成果,结合实际需求,设计实现了道路斑马线与行人运动检测原型系统,验证了本课题中研究以及改进算法的优越性。

**关键词:** 行人穿行意图预测,斑马线检测,行人检测,CNN-LSTM,贝叶斯网络

## Abstract

With the development of artificial intelligence technology, driverless vehicles and assisted driving vehicles have gradually entered people's daily traffic life. In the traffic scene of crosswalk area, pedestrians and vehicles are important interlocutors in the scene. Crosswalk is an important facility to ensure the traffic safety of pedestrians. Accurate detection of the movement of crosswalk and pedestrians is an indispensable key technology for driverless vehicles and auxiliary driving vehicles to ensure pedestrian safety. Therefore, the research on crosswalk and pedestrian motion detection has important practical significance for traffic safety and the development of driverless vehicles and assisted driving vehicles.

Due to the problems of complex background, pedestrian occlusion and diverse posture in crosswalk and pedestrian detection, it is difficult to identify pedestrian behavior and can't accurately predict the pedestrian's intention to cross crosswalk. Therefore, aiming at the problems existing in road crosswalk and pedestrian motion detection, this paper adopts the method of deep learning. The specific research contents are as follows:

(1) In order to improve the problems of low accuracy, incomplete detection and poor real-time performance of existing crosswalk detection, an improved SegNet semantic segmentation model was proposed. Based on the original SegNet model in semantic segmentation, the improved model first integrates the pyramid pooling module to reduce the loss of crosswalk detail information; After that, the model was changed to coding unchanged, decoding reduces the asymmetric structure, reduces network parameters, and accurately detects the pixel position of crosswalk. After expanding the crosswalk image data set, the experiment was carried out, and the four models were compared. The crosswalk detection accuracy of the improved SegNet model reaches 97.6%. Compared with other models, the detection accuracy is effectively improved, and the running speed is accelerated, which meets the current needs of real-time detection.

(2) Aiming at the problems of time-consuming and small target detection in pedestrian detection, an improved Faster-RCNN algorithm was proposed. Based on the Faster-RCNN

algorithm, the improved model first optimizes the network structure and replaces the feature extraction network model, which improves the superiority of feature extraction; Then the clustering algorithm was optimized and Gaussian clustering was introduced to reduce the amount of calculation and improve the detection speed; Finally, SSD algorithm and Faster-RCNN algorithm were compared with the improved algorithm. The improved algorithm has higher detection accuracy, faster speed and better real-time performance. It can effectively detect small target pedestrians, which has practical significance.

(3) Aiming at the problem of low recognition accuracy in behavior recognition, based on the research and analysis of relevant recognition algorithms, a pedestrian recognition algorithm based on CNN-LSTM is proposed. The main idea of the algorithm was to use CNN network model to extract image feature information, analyze time series of pedestrian behavior with LSTM network model, and then complete pedestrian behavior recognition and classification. Comparing CNN-LSTM model with ResNet-LSTM model, the experimental results show that the loss value of CNN-LSTM model is smaller, the model tends to stabilize faster, the model fluctuation is smaller, and the behavior recognition accuracy of CNN-LSTM model is higher.

(4) Aiming at the long time and short distance of pedestrian crossing crosswalk intention prediction, a pedestrian crossing crosswalk intention prediction model based on Bayesian network was constructed. The main idea of the model was to predict the influencing factors according to the intention analysis from the perspectives of pedestrians and vehicles, design the traffic scene in this paper, estimate the probability of travelers crossing the crosswalk, and obtain the optimal prediction model through training analysis. According to the performance analysis differences of different scenes, this model can predict the traveler's intention to cross the crosswalk 0.5s-0.8s in advance, and the pedestrian's intention to cross the crosswalk can be predicted when the relative distance is 15m to 25m. The results show that in various scenes where pedestrians cross the crosswalk, the prediction model constructed in this paper can effectively predict the pedestrian's intention to cross the crosswalk.

(5) Based on the above research results, the superiority of the pedestrian motion detection system is verified.

**Keywords:** Pedestrian crossing intention prediction, crosswalk detection, Pedestrian detection, CNN-LSTM, Bayesian network



# 目 录

第一章 绪论 .....	1
1.1 研究背景及意义 .....	1
1.2 国内外研究现状 .....	2
1.2.1 斑马线检测技术研究现状 .....	2
1.2.2 行人检测技术研究现状 .....	3
1.2.3 行人意图预测技术研究现状 .....	5
1.3 问题分析 .....	6
1.4 研究内容和论文结构 .....	7
1.4.1 研究内容 .....	7
1.4.2 论文结构 .....	8
第二章 基于语义分割的斑马线检测技术研究 .....	11
2.1 语义分割 .....	11
2.1.1 SegNet 网络模型 .....	11
2.1.2 U-Net 网络模型 .....	12
2.1.3 PSPNet 网络模型 .....	13
2.2 改进 SEGNET 模型 .....	14
2.2.1 融入金字塔池化模块 .....	14
2.2.2 结构改进 .....	15
2.3 实验分析 .....	16
2.3.1 数据集准备 .....	16
2.3.2 训练过程 .....	18
2.4 实验结果与分析 .....	19
2.4.1 评价指标 .....	19
2.4.2 检测结果分析 .....	20
2.5 本章小结 .....	21
第三章 行人检测算法研究 .....	23

3.1 传统行人检测方法.....	23
3.1.1 HOG 算法 .....	23
3.1.2 SVM 分类器模型 .....	25
3.1.3 实验结果分析 .....	26
3.2 深度学习行人检测算法.....	27
3.2.1 SSD 网络模型.....	28
3.2.2 Faster-RCNN 网络模型 .....	30
3.3 FASTER-RCNN 行人检测算法改进.....	33
3.3.1 网络结构优化 .....	33
3.3.2 聚类算法优化 .....	35
3.4 实验分析.....	38
3.4.1 数据集准备 .....	38
3.4.2 训练过程 .....	38
3.4.3 实验结果与分析 .....	40
3.5 本章小结.....	42
<b>第四章 基于深度学习的行人穿行意图预测算法研究 .....</b>	<b>43</b>
4.1 行人行为识别研究.....	43
4.1.1 CNN 网络模型.....	43
4.1.2 LSTM 网络模型 .....	45
4.1.3 数据集准备 .....	47
4.1.4 实验结果与分析 .....	48
4.2 基于贝叶斯网络的意图预测研究.....	49
4.2.1 影响因素分析 .....	49
4.2.2 交通场景设计 .....	50
4.2.3 贝叶斯网络原理与构建 .....	51
4.3 实验结果与分析.....	54
4.3.1 数据集准备 .....	54
4.3.2 实验分析 .....	54

4.4 本章小结.....	57
第五章 道路斑马线与行人运动检测系统设计与实现 .....	59
5.1 需求分析.....	59
5.2 开发环境与开发工具.....	60
5.2.1 开发环境 .....	60
5.2.2 开发工具 .....	60
5.3 系统设计与实现.....	61
5.3.1 系统设计 .....	61
5.3.2 功能设计与实现 .....	61
5.4 本章小结.....	65
总结与展望 .....	67
总结 .....	67
展望 .....	68
参考文献 .....	69
攻读学位期间取得的研究成果 .....	75
致谢 .....	77

## 第一章 绪论

### 1.1 研究背景及意义

人车交互场景中，行人是重要参与者，行人与车辆二者在交通系统中有着举足轻重的地位，行人一直是交通交互中的弱势群体，各种突发事故中行人受伤害概率往往比较高<sup>[1]</sup>，甚至会付出生命代价，并且驾驶员在驾驶过程中易出现驾驶错误情况，这种情况的发生会加剧交通事故出现的概率。2017年，深圳出台了条例，行人若在斑马线区域内，机动车需要停车避让，若发生冲突会进行罚款并扣分处理。如果汽车能够自动判断前方道路情况和行人穿行道路斑马线意图，然后依据判断情况做出相关应对处理，不仅可以减轻驾驶员的负担，而且可以保证行人的安全。因此，开展对道路情况的检测研究与行人穿行意图的预测研究，对于辅助驾驶员安全驾驶汽车以及无人驾驶汽车安全行驶、保证行人安全具有重要的实际意义。

随着人工智能技术的发展，无人驾驶汽车与辅助驾驶汽车正逐渐走入人们的日常生活中<sup>[2]</sup>。无人驾驶汽车和辅助驾驶汽车的行车安全一直是无人驾驶与辅助驾驶技术的重点研究内容。道路斑马线区域内的交通场景中，交通事故频繁发生，一方面是行人穿行意图的不确定性影响驾驶者的自主判断，另一方面是车辆本身行驶状况可能存在引发交通事故的条件，在此交通场景下，若能够完成车辆预测行人穿行斑马线意图工作，必将使得道路安全得到更有效的保障。因此，行人穿行道路斑马线意图预测是无人驾驶汽车与辅助驾驶汽车不可或缺的一项关键技术，开展行人穿行意图预测的研究对交通安全、无人驾驶汽车及辅助驾驶汽车的发展具有现实意义。

目前技术的不断进步以及各种软件的升级让人们对于无人驾驶技术的认识有了新的高度，不仅引起了国家的高度关注，而且各大汽车商研究中心也纷纷对其进行研究，使其一举成为了国家重要发展战略之一。但是人们的研究仍不够深入，技术不够成熟，仅仅实现的是低级无人驾驶级别，仍有很多困难需要解决。机器视觉作为无人驾驶中的重要技术，备受人们的关注，同时深度学习的发展也为技术的突破提供了基础。机器视觉使用车载摄像机去记录拍摄道路的周围环境，利用各种图像处理方法将前方道路情况进行检测识别，如斑马线检测、行人检测、行为识别等，这几项技术的研究在帮助无人驾驶技术的发展中起着至关重要的作用。深度学习的出现虽使得无人驾驶中的各项技术研

究取得了飞速进展，但是各方面仍存在不足。斑马线与行人检测工作中，现有方法不能精确有效完成检测，当前人们对于行人穿行意图研究较浅，方法不成熟，因此需要继续进行探究各方面技术。在人车交互系统中，道路斑马线周围区域内的行人更需高度关注，车辆若能够提前预知到行人的穿行斑马线意图，将会提前采取一定的规避措施，行人安全得到了更好的保障，使得斑马线区域内的交通事故发生概率大大降低，因此研究有效的道路斑马线与行人运动检测具有重要的现实意义与应用价值。

本课题采用车载摄像头采集道路图像信息，通过检测斑马线区域周围行人完成行人行为识别，进而实现行人穿行道路斑马线意图预测，并给出车辆提示信息，以实现车辆能够提前预知行人穿行意图后采取措施，达到减少交通事故发生的目的，基于此，本文将研究基于深度学习的道路斑马线与行人运动检测算法，并通过实验来验证研究中各算法的合理性与准确性，以满足当下无人驾驶和智能辅助驾驶的需求，减少交通事故发生率，提高行车效率，保障绿色出行。

## 1.2 国内外研究现状

本文主要研究基于深度学习的道路斑马线与行人运动检测算法研究与应用，利用语义分割技术检测斑马线区域，然后将斑马线区域的行人进行检测与动作识别，在此基础上实现行人穿行斑马线意图预测。因此本文主要论述的技术研究现状包含斑马线检测、行人检测和意图预测三方面。

### 1.2.1 斑马线检测技术研究现状

斑马线作为道路中安全标志之一，一直以来受到广大研究学者的关注，目前基于图像处理技术的斑马线检测与识别不断被完善，致力于提升斑马线检测的准确性，达到降低交通事故发生的概率。

上世纪 90 年代开始，已经很多学者投身于斑马线检测的研究中。Mohammad 等人<sup>[1]</sup>利用分析斑马线的方向、间隔和条数等特征信息来完成图像中斑马线的判别，该方式使用的是斑马线的基本特征信息，方法简单，但是该方法无法准确识别与斑马线具有相似特征的图像；Stephen 等人<sup>[4]</sup>提出了图像消逝点的方法来检测斑马线，利用坐标系来达到位置的提取标注，进而实现斑马线识别，该方法实现了检测精确率提升但是存在计算量大且实时性差的问题；双极系数法<sup>[5]</sup>通过提取的图片特征并利用斑马线具有的黑白相交

的特性去判断图像中是否存在斑马线，从斑马线的灰度特征出发，依据现象不同进行区域分块处理，成功识别出斑马线；伍珀塔尔大学的 Sebastian 学者<sup>[6]</sup>设计修复图片后进行偏航角补偿，依据角度变化提取特征区域后使用 Canny<sup>[7,8]</sup>边缘检测算子提取轮廓信息，采用傅里叶变化最终实现斑马线识别。伴随着卷积神经网络技术的发展以及语义分割<sup>[9-10]</sup>的提出，使用深度学习方法进行斑马线识别的技术在不断完善。Long J 等人<sup>[11]</sup>提出 FCN，它在进行语义分割时感受野固定且其分割物体时细节容易丢失或者被平滑；Badrinarayanan 等人<sup>[12]</sup>提出 SegNet，其采用了池化索引来保存图像的轮廓信息，降低了参数量，但是当对低分辨率的特征图进行去池化时，会忽略邻近像素的信息；Ronneberger 等人<sup>[13,14]</sup>提出了 U-Net 模型，它使用图像切块进行训练，但是其在卷积过程中没有加 pad，导致网络最后的输出与输入大小不一致；Zhang 等人<sup>[15]</sup>提出 PSPNet 模型，提出金字塔模块<sup>[16]</sup>来聚合背景信息，但采用 4 种不同的金字塔池化模块，对细节的处理要求较高。

综上所述，斑马线作为交通场景中的重要组成部分，关于斑马线检测技术研究从未停止，伴随着技术的发展斑马线检测技术虽然取得了一定成效，但仍然存在一些问题，如检测精度低、检测不全面、实时性差等问题，而且在无人驾驶与辅助驾驶系统中，斑马线检测识别更是发展意图预测的基础保障，因此需要投入更多精力研究斑马线检测技术，达到准确且快速识别出斑马线的目的，为无人驾驶系统与辅助驾驶系统提供基础服务。

### 1.2.2 行人检测技术研究现状

行人检测始于 20 世纪 90 年代，根据传感器的不同可将其分为两类：第一类是以光学摄像头为基础将图像进行处理，并且在技术发展的加持下不断进行算法优化，达到提高准确率的目的；第二类是汽车制造商们，他们利用各种传感器与光学摄像头进行结合，有效提高了检测效果。

目前人们关于研究行人检测后得到的检测方法达到几十种，具体以检测的不同目标为分类依据可分为基于运动信息的行人检测、基于模板匹配的行人检测、基于行人特征的行人检测<sup>[17]</sup>三大类。

(1) 基于运动信息的行人检测：多为处理视频中的行人检测，通过提取视频中行人运动位置、速度等的特征信息来对行人进行追踪与定位，进而实现检测行人。

(2) 基于模板匹配的行人检测：将检测要求提出，遍历行人标定模板，将检测图像

中提取到的行人信息与模板进行对比,然后将计算得到的重合度与设置的阈值比对,进而得到行人检测结果并得出结论。

(3) 基于行人特征的行人检测:作为应用于丰富场景的检测方法,此方法具有极强环境适应性,利用不同方法提取特征信息,进而完成行人检测。此类方法依赖于训练器的设置,需要将图像进行筛选,确定检测目标。

Ojala 等人<sup>[18]</sup>提出了基于局部二值(Local Binary Pattern, LBP)特征的检测方法,该方法利用 LBP 算子排除灰度因素,但是存在信息变化的问题,进而使得检测结果不准确,因此在此基础上 Meanpaa 等人提出将 LBP 算子特征值进行固化,解决了 LBP 算子改变情况,使得 LBP 算子适应各种场景,提高了检测精确度。Viola 等人<sup>[19,20]</sup>提出的特征提取方式使得检测有了更高的水准,起初其是应用于人脸特征检测,后经过改进,对特征标量的识别更加高效,同样适用于行人检测。Dalal 等人<sup>[21]</sup>提出了 HOG<sup>[22]</sup>即方向梯度直方图的特征描述子,利用图像中梯度变化提取特征向量,进而实现人体轮廓提取,完成行人检测,但是其方法只适用于无遮挡图像即图像效果很好的情况下,且此方法存在检测速度慢的缺点。基于这两种方法的优缺点,Wang 等人<sup>[23]</sup>提出了融合想法,将 HOG 与 LBP 算子<sup>[24]</sup>结合,处理遮挡图像的计算速度得到提升的同时检测精确度也较为可观。Felzenszwalb 等人<sup>[25]</sup>提出基于组件检测的可变形组件模型算法,与金字塔模型具有相似性,虽然具有检测精度高的优点,但缺点也非常明显即实时性很差。孙锐等人<sup>[26]</sup>将其改进,将 HOG-NMF 特征与显著性信息进行结合,将图像特征维数降低进而得以提升运算速度和精度。谭飞刚<sup>[27]</sup>对于韦伯特征的优点结合于局部二值化特征的优点,算法提取特征信息获得更好的检测效果。Dollar 等人<sup>[28]</sup>提出聚合通道特征(Aggrate Channel Features, ACF),其基于多特征融合,产生强分类器,具有良好的行人检测效果。

在多年的国内外研究学者的探索下,行人检测取得了巨大进步,并给后人总结记录了很多优秀方法,但是伴随着 2012 年深度学习方法的出现,传统行人检测技术被遗弃,深度学习方法开始应用于行人检测研究中。

Krizhevsky 等人<sup>[29]</sup>提出 CNN<sup>[30]</sup>,即卷积神经网络模型,CNN 网络被广泛应用。Sermanet 等人<sup>[30]</sup>提出基于非监督的多夹断特征学习方法进行行人检测,将一直存在的数据训练优先的问题解决,该方法的思想是首先初始化操作网络的每层结构,然后将网络结构进行细微的调整。王晓刚等人<sup>[30]</sup>提出的可转换行人检测模型使得行人检测迈上了新

的台阶，成为当时最优秀的行人检测模型，原因是此检测模型中应用了局限型的玻尔兹曼机，可以自动学习图像所传达的高层特征信息和底层特征信息。

关于深度学习的行人检测方法目前主流的为两种，一种是以生成检测区域为基础的检测方法，另一种是以回归为基础的检测。以 YOLO<sup>[32]</sup>、SSD<sup>[33]</sup>为基础利用回归思想的检测算法模型，王宇宁等人<sup>[34]</sup>提出了 YOLO 算法模型，解决目标检测中的行人检测速度慢的问题。以 Faster-RCNN 为基础生成候选区域的检测算法模型，其是在 RCNN<sup>[35]</sup>算法的基础上改进而来，全卷积网络的开始应用，使得特征信息提取速度问题得到有效解决。Girshick 等人<sup>[36]</sup>为了解决 Faster-RCNN 算法模型检测速度慢的问题，提出基于 Faster-RCNN 算法的通用目标检测算法，该算法模型提升速度方式为将候选区域传至 Faster-RCNN 网络中进而实现目标检测。薛超等人<sup>[37]</sup>的改进是将算法结构调整，特征进行融合，多尺度的再训练实现在人脸识别的技术领域。宋焕生等人<sup>[38]</sup>将算法改为著名的二分类问题，处理复杂背景的目标检测问题被优化。由于 Faster-RCNN 算法模型有着较好的检测能力，被多数研究者致力于实现目标检测，本文同样以此模型为基础进行优化，解决行人检测中存在的耗时大与小目标检测困难问题。

随着技术的不断发展，行人检测效果不断被完善，行人检测也取得了实质性的进展，但仍存在很多问题需要解决，如耗时大问题、小目标检测困难的问题等，更需要我们进行算法探究，从而完善无人驾驶系统与辅助驾驶系统，降低交通事故的发生率。

### 1.2.3 行人意图预测技术研究现状

行人与车辆作为交通系统的重要组成部分，人车交互场景普遍存在于真实生活场景中，对于无人驾驶系统与辅助驾驶系统来说，能够提前预测行人穿行道路斑马线意图具有重要研究意义，因此完成预测行人穿行斑马线意图能够较好的规避交通事故的发生。当前，关于行人意图预测的研究主要分为基于运动建模、基于规划建模、基于数据三大类。

(1) 基于运动建模：行人存在于各种不同的交通场景中，此方法为将场景中的行人实施运动动力学建模方法，预测行人运动。Nicolas 等人<sup>[39]</sup>完成短期内行人路径预测，在预测中将递归贝叶斯滤波器应用其中，并将卡尔曼滤波器与交互多模型(Interacting Multiple Models, IMM)进行对比，以典型过街行人的行为为基础，分析运动模型的性能；Keller 等人<sup>[40]</sup>将高斯过程动力学加入到研究中，解决短时间的行人意图预测和动作分类，



典型方法为：高斯过程动力学模型、概率分层轨迹匹配、卡尔曼滤波器以及其扩展的交互模型，完成方法对比，将行人行为意图预测研究到多个场景下，但是这些方法局限于应对短期内预测。

(2) 基于规划建模：当且仅当知道行人目的地所在，先将目的地假定完成，然后基于假定目的地完成运动或路径规划。Rehder 等人<sup>[41]</sup>提出基于动力学和环境的预测方法，这种预测是已经知道目的地，然后转化为规划问题，使用了粒子滤波器，与卡尔曼滤波器比较，加入粒子滤波器能够更好的完成预测；Karasev 等人<sup>[42]</sup>提出的规划路径完成长期行人运动预测，建模为跳跃马尔可夫过程，给定位置、方向等环境变量和目标，然后求解，获得随机次优策略，实现最短路径规划，结合拉奥-不来克韦尔化粒子滤波器，获得滤波密度，进而实现预测。

(3) 基于数据：深度学习的发展也扩展了预测方式。Volz 等人<sup>[43]</sup>提出基于数据驱动方法，本质是时间序列二分类问题，应用于路口场景，将行人过马路行为进行预测；Fang 等人<sup>[44]</sup>利用骨架特征信息估计意图，只使用图片信息，更加简便；李贤等人<sup>[45]</sup>借鉴前人思想，同样使用骨架特征，但将特征信息进行了选取，更易于分类；Saleh 等人<sup>[46]</sup>从行人头部和姿态入手，提取特征信息进行分类，实现预测；钟涛<sup>[47]</sup>利用行人轨迹以及行人定位，对序列行人进行特征组合，注意力机制结合于长短时记忆网络，实现预测；Hoy 等人<sup>[48]</sup>使用了变分递归思想，结合动态状态空间模型开发特定类别的跟踪与路径预测框架，预测行人穿行街道意图。

综上，行人意图预测研究仍不完善，需要更多研究者投入，针对行人穿行斑马线存在预测时间长、预测距离短的问题，构建了本文的行人穿行道路斑马线意图预测模型，给出行人穿行道路斑马线的概率估计，为无人驾驶技术与辅助驾驶技术发展提供帮助。

### 1.3 问题分析

通过对道路斑马线检测、行人检测与行人意图预测等方面的研究现状以及发展历程分析，在众多人车交互场景中，各方面仍存在大量问题尚未解决，影响无人驾驶与辅助驾驶发展进程，无法满足无人驾驶与辅助驾驶真正落实条件，具体各方面存在的问题如下：

(1) 斑马线检测问题：在日常交通场景中，由于使用车载摄像头的多样性导致拍摄

视频中的道路斑马线存在多样性且标志有不清晰现象出现, 斑马线也常存在污损、缺失问题, 更有背景的复杂性即拍摄存在光照强度、阴影变化等环境因素的影响导致的背景复杂性, 促使检测精度低、检测不全面、实时性差等多种检测问题的出现。

(2) 行人检测问题: 由于人车交互场景的多样性、复杂性, 导致行人背景复杂多样即可能存在树木、道路、汽车等的影响; 而且行人作为行动的个体, 存在各种姿态变化, 无法利用简单的人体特征来描述人体; 再者每个人的身形体态不同, 着装更是千变万化, 也同样给行人检测带来了一定挑战。

(3) 行人穿行意图预测问题: (1)行为识别, 行人行为存在多样性, 类之间存在各种差异, 即使是相同的动作, 不同的人也可能会有不同的表现; 环境存在差异即遮挡情况、视角多且不同、时间的变化等的背景影响问题; 人在进行各种动作的时候速度变化大, 导致很难确定动作的起始点, 视频提取特征存在干扰问题。(2)目前关于行人意图预测的研究较少, 实验完成度较低, 环境信息的提供没有真正应用到实验研究中, 且考虑不全面, 没有清楚了解各种特征的存在对行人意图预测的贡献; 然后通过行人轨迹实现预测的方法存在较大误差, 影响最终的预测结果; 再者通过行人行为识别方法基于行为识别效果, 动作识别不到位, 也会促使意图预测不准确; 现在已有的意图预测实验研究中, 也存在预测时长较长、预测不精准、预测距离短的问题。

因此, 本课题基于上述存在的问题, 找寻适宜的算法进行斑马线检测、行人检测、行为识别并最终实现预测行人穿行道路斑马线意图, 为无人驾驶与辅助驾驶研究提供帮助。

## 1.4 研究内容和论文结构

### 1.4.1 研究内容

在了解了斑马线检测、行人检测和行人意图预测的相关研究现状以及对其存在的问题分析后, 对本课题的道路斑马线与行人运动检测系统进行分析, 并基于深度学习方法研究系统中的各功能关键技术, 主要研究内容如下:

(1) 斑马线检测技术研究。道路斑马线区域作为本课题研究的框选区域, 行人穿行意图的预测限于道路斑马线区域的不同人车交互场景。针对复杂场景中斑马线检测的检测不全面、检测精度低、实时性差的问题, 选用语义分割中的 SegNet、U-Net 与 PSPNet 三种模型应用于斑马线检测研究中, 分析三种模型检测的效果, 并结合真实生活场景中

的斑马线特点,提出基于 SegNet 模型的改进模型,先融入金字塔池化模块,获取到全局信息,后将模型改为编码不变,解码减少的不对称结构,减少网络参数,精准定位斑马线像素点位置。将改进模型与三种语义分割模型进行实验对比分析,验证了改进 SegNet 模型完成斑马线检测的优越性。

(2) 行人检测技术研究。针对行人检测存在的耗时大与小目标检测困难的问题,选用深度学习的 SSD 算法与 Faster-RCNN 算法应用于行人检测研究中,分析实验效果,结合本课题要求与检测存在的问题,基于检测效果较好的 Faster-RCNN 算法提出改进,优化网络结构,替换特征提取部分,获得更全面的特征信息;优化聚类算法,减少计算量,提升训练时间,并将改进 Faster-RCNN 模型与 SDD 模型、Faster-RCNN 模型进行对比实验分析,验证本文改进模型的检测优越性。

(3) 行人穿行斑马线意图预测。针对行人识别中的识别准确率低的问题,研究分析识别方法后,给出 CNN-LSTM 模型进行行人行为识别,依据 CNN 网络模型提取图像行人行为的特征信息,LSTM 网络模型对行人行为分析时间序列进而完成行人行为识别分类,并将模型与其他识别模型对比,进行优劣分析,验证选用模型的有效性。将行为识别结果应用到后续的穿行道路斑马线意图预测中,依据意图预测影响因素,设计本文交通场景,针对行人穿行道路斑马线意图预测时间长与预测距离短的问题,构建适应本实验课题场景的贝叶斯网络模型,实现行人穿行道路斑马线意图预测。

#### 1.4.2 论文结构

依据研究内容,本文的结构如下:

第一章:绪论。本章主要介绍了关于道路斑马线与行人运动检测系统研究的背景与意义,阐述了关于斑马线检测技术、行人检测技术以及意图预测三个方面的研究现状,探讨了各方面存在的问题,最后将本文研究内容与结构进行概述。

第二章:基于语义分割的斑马线检测技术研究。介绍了语义分割技术,并将本文中应用的语义分割技术中的部分模型进行介绍,即 SetNet 模型、U-Net 模型、PSPNet 模型,接着完成各模型的实验验证,将检测效果良好的 SegNet 模型进行改进,通过实验分析,改进模型能够更好的检测斑马线,对污损、断续的斑马线有良好检测效果,且检测精度明显提升,检测更全面,实时性增强,具有更好的应用前景。

第三章:行人检测技术研究。本章首先使用传统方法进行检测,达到与深度学习方

法对比的作用。介绍 HOG 算法与 SVM 分类器，并将两者结合进行行人检测实验，验证传统检测算法中存在的检测不全面、检测精度低、训练耗时，计算量大、实时性差等问题，针对存在问题采用深度学习方法进行行人检测探究。利用 SSD 算法与 Faster-RCNN 算法进行实验，先详细介绍了两种算法的原理与结构，然后结合本课题要求以及 Faster-RCNN 算法中仍存在的耗时大与小目标检测困难的问题，提出改进的 Faster-RCNN 检测模型，最后进行实验对比分析，验证了改进模型的优越性。

第四章：行人穿行斑马线意图预测技术研究。本章在前文行人检测的基础上对行人行为进行识别技术研究，选取 CNN-LSTM 模型识别行人行为，详细介绍 CNN 模型与 LSTM 模型原理与结构，然后构建实验验证模型的实用性，与 ResNet-LSTM 模型进行实验对比，结果表现 CNN-LSTM 模型可以应用于本系统中。行为识别出后进行行人穿行斑马线意图预测研究，先分析意图预测影响因素，设计本文交通场景，在此基础上选取贝叶斯网络进行行人穿行道路斑马线意图预测，详细介绍了贝叶斯网络的原理与结构，进而构建本文的贝叶斯网络模型预测行人穿行道路斑马线意图，通过实验验证了本文构建的意图预测模型具有良好效果并可应用于实践中。

第五章：道路斑马线与行人运动检测系统设计与实现。这对本文的研究内容与技术，设计实现了无人车驾驶与辅助驾驶的系统，利用 Pycharm, OpenCV、QT 设计了道路斑马线与行人运动检测系统，该系统实现功能包括：斑马线检测、行人检测以及行人穿行斑马线意图预测。

第六章：总结与展望。对本文核心研究点进行概述，同时提出本文研究内容存在的局限，最后对本文研究内容的扩充与进一步研究进行了展望。

## 第二章 基于语义分割的斑马线检测技术研究

斑马线作为道路交通中重要的组成部分，其存在可以帮助行人安全通行。对于无人驾驶系统与辅助驾驶系统来说，行人穿行道路斑马线意图预测研究一直进行但未完成全面攻破。对于无人驾驶汽车或辅助驾驶汽车来说，分析斑马线区域内的行人穿行意图是一项必不可少的工作。道路斑马线作为行人穿行意图预测的框选区域，人们也一直研究斑马线检测技术，致力于寻找检测斑马线更精准更全面的方法。利用现代科技对斑马线进行检测为无人驾驶与辅助驾驶的预警系统提供了帮助，实现精准检测斑马线为车辆的无人驾驶与辅助驾驶提供了更加可靠的前提条件，使得行人的安全系数得以提升。关于斑马线检测技术现在人们的研究较多，其中以语义分割方法检测斑马线效果最为突出，本章选用语义分割的三种模型进行斑马线检测实验，并针对存在的问题进行改进。

### 2.1 语义分割

语义分割是将原始图像中具有不同类别的对象采取标注形式进行语义信息分割，最终实现目标是完成各种场景图像的像素分类，将背景与各种离散对象进行分割解析并正确的标记出语义信息。语义分割技术的发展随着人工智能的出现取得了显著成就，作为基于卷积神经网络的分割技术，其去除了人工操作，提升了分割的精准性和实用性。

#### 2.1.1 SegNet 网络模型

SegNet 是一种为了自动驾驶或者智能机器人的图像语义分割深度网络，SegNet 最吸引人的特点是对于分辨率低的输入图像，它的解码器使用了上采样的方式，SegNet 为了降低学习上采样的虚耗，将编码器获得的池化索引应用到上采样。SegNet 使用了卷积操作，将特征图进行增多，使其变密集，供特征提取。这样操作可以使模型具有良好分割性能的背景下揭示了内存与精度的权衡。

SegNet 模型主要针对于道路场景应用，在整体上运用了编码器-解码器(encoder-decoder)的结构。关于输入的数据存在的空间维度的增减问题，SegNet 模型使用编码器来进行，而对于编码器造成的问题，解码器使用反卷积来解决，恢复数据的空间维度以及目标的细节。编码器与解码器之间会存在直接的信息关联，这样解码器可以实现对目

标细节进行更好的恢复。

传统 SegNet 模型的网络结构如图 2.1 所示, Input 为输入图像, Output 为输出的分割图像, 不同颜色代表不同的分类。SegNet 是一个对称网络, 由中间绿色 pooling 层与蓝色 upsampling 层作为分割, 左边是通过卷积来提取高维特征, 并通过 pooling 使图片变小, 为 Encoder, 编码器网络由 13 个卷积层组成, 对应于 VGG16 网络中用于对象分类的前 13 个卷积层。因此, 对于初始化训练的权重, 可以从训练用于大数据集分类的权重中获得。而且为了在编码器的输出处保留更高的分辨率特征映射, 可以舍弃全连接层。这样 SegNet 模型中编码器的参数减少。编码器与解码器一一对应, 解码器也有 13 层。右边是反卷积与 upsampling, 通过反卷积重新获得图像分类以后的特征, 填充缺失的内容, upsampling 作为 pooling 的逆过程, 使图像变大, 为 Decoder, 最后一个卷积层会输出所有的类别, 网络的最后通过 Softmax 函数, 输出不同分类的最大值, 最终完成图像像素级别的不同分类, 得到每个像素属于斑马线的概率。Pooling Indices 为 pooling 前后的对应位置, 会被模型在 encoder 时使用 Pooling 记录, pooling 后的 indices 输出到对应的 upsampling。Convolutional Encoder-Decoder 为 Encoder 每一层卷积对应 Decoder 的一层卷积, 融合高分辨率信息。本文将 SegNet 网络作为斑马线检测的基础网络。

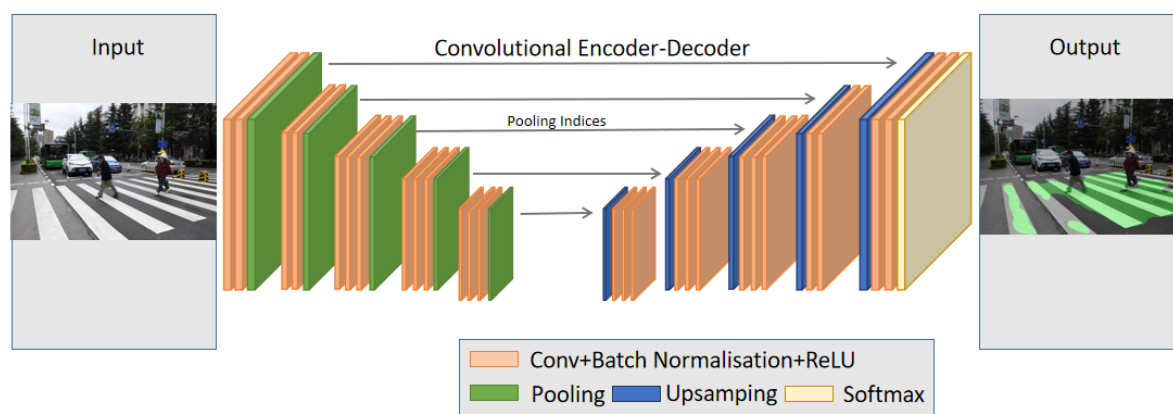


图 2.1 SegNet 模型结构图

### 2.1.2 U-Net 网络模型

U-Net 主要解决的是医学领域的图像分割问题, 由于其性能良好, 被应用于语义分割的各个方向。U-net 模型基于 FCN 的思想改进而来, 其网络结构为一个 U 型, 故名为 U-Net。其对 FCN 网络的改进处理为对 FCN 网络的上采样阶段进行了卷积加深, 具体结构如图 2.2 所示:

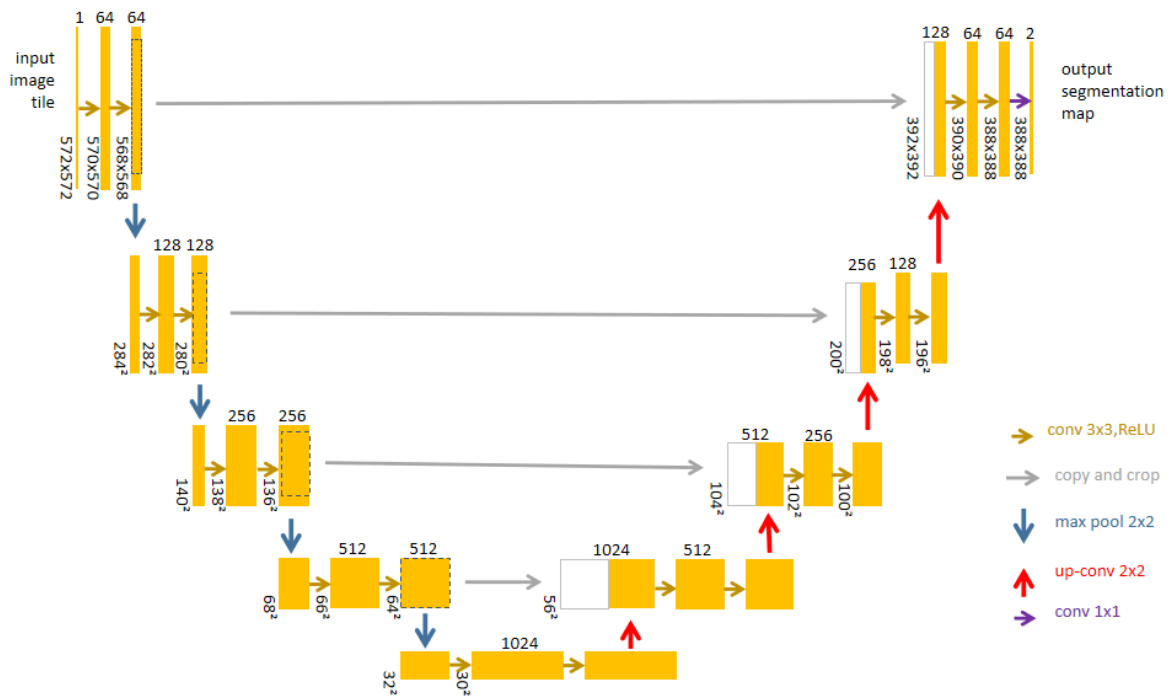


图 2.2 U-Net 模型结构图

U-Net 网络模型以其具有的经典卷积对称网络结构而被命名，左右分为下上采样，左侧下采样作用是将重要特征提取出，右侧上采样作用是将浓缩特征还原为图像。左侧是模型的压缩路径，完成输入镜像图片，主要采用的是卷积和 Max pooling，达到完成降采样功能的目的，主要有 4 个 block 组成，每个 block 的组成是 3 个有效卷积和 1 个 max pooling，降采样工程完成后得到的是特征图像。右侧是模型的扩展路径，也是 4 个 block 组成，在 block 之前进行反卷积和尺寸操作，同时个数减半，然后左右 feature map 合并。由于转置插值会使得输出图像不均匀，所以为处理此问题，就采用了临近插值。Concat 操作可以应用在跳跃连接中，然后再把特征拼接于 channel 维度中，以产生更厚特征，并将局部特征结合于全局特征，并不是纯粹的叠加。全程的特征卷积操作都用到 valid（包括 pooling），这个举措实施有两个原因。一是为实现保证分割的结果中大致是基于上下文特征得到的；二是实现信息不遗漏的目的，使得输出图像相较于输入图像要小，分割精度更高。网络的最后通过 1x1 的卷积层将粗分辨率提取的密集特征投影到原始图像的精确位置。

### 2.1.3 PSPNet 网络模型

PSPNet 使用了金字塔池化模块，是一种场景解析网络，具体结构如图 2.3 所示。首先将需要分割图片输入到网络模型中，接下来是 ResNet 模块，ResNet 模块作用是将全

连接层去掉，进而特征图像被压缩成为原本的  $1/8$ ，需要分割图像进入网络中经过此操作完成图像压缩，然后将压缩后的图像输出到金字塔池化模块中，图像完成全局上下文信息的提取，最后进行图像分割操作。

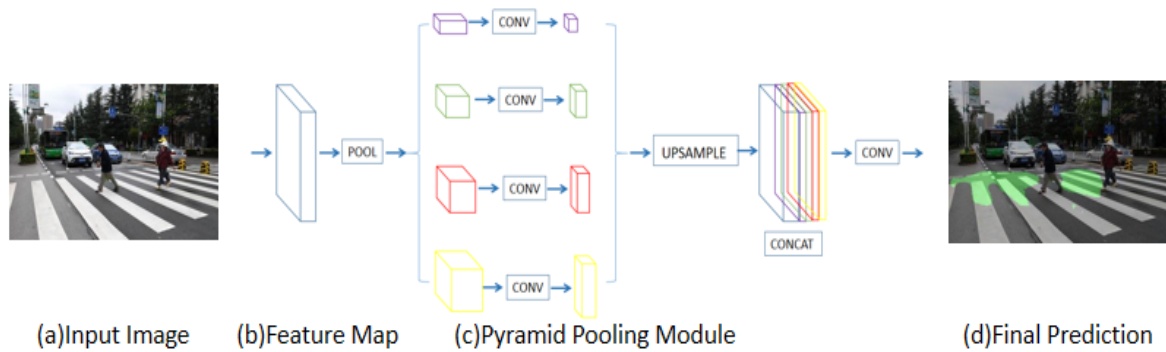


图 2.3 PSPNet 模型结构图

在本文进行斑马线检测研究时，融入了 PSPNet 的特征部分即将金字塔池化模块融入到本文改进网络模型中，更好的将上下文信息进行聚合，形成全局的上下文信息，更全面的定位到斑马线和完成斑马线像素的分类。

本文将上述 SegNet 模型、U-Net 模型与 PSPNet 模型等 3 种模型分别应用于道路斑马线检测中，以寻求最适合道路斑马线检测的算法，发现基于 SegNet 模型的道路斑马线检测方法在检测准确率和检测速度上均优于其他两种检测方法，但是 SegNet 模型检测斑马线仍存在检测精度低、检测不全面、实时性差的问题，同时结合真实生活场景中道路斑马线的特点，本章在 SegNet 检测模型的基础上，提出了一种融入金字塔池化模块并将其结构改变的改进检测模型，接下来将对改进检测模型做进一步详细介绍。

## 2.2 改进 SegNet 模型

以提高检测斑马线的精确度与实时性为目的，基于原始 SegNet 网络模型进行模型改进。首先将金字塔池化模块(PPM)融入，提高提取信息的全面性；然后进行结构改进，将对称结构改为编码不变，解码减少的不对称结构，提高检测速度，完成实时性要求。

### 2.2.1 融入金字塔池化模块

本文中将金字塔池化模块融入到原始 SegNet 模型中，将此模块放在最后的卷积层后面，构造一种独特的全局特征提取模块，达到改进 SegNet 模型<sup>[49,50]</sup>能够有效获得全局预警信息目的。



原始 SegNet 模型中的 CNN 存在输入图像大小固定的缺点,因此需要将此缺陷消除,而金字塔池化模块中的设计解决了此问题,设计为金字塔池化模块后会生成不同级别的特征图,各特征图最后会被展平且拼接起来,然后再将拼接后的特征图送入到全连接层中进行分类。

如图 2.4 所示,金字塔池化模块融合了 4 种不同金字塔尺度的特征。紫色表示最粗粒度的级别是全局池化。对特征图采用 4 个不同尺度池内核进行池化,对于其他的级别是将图像进行划分,然后将不同级别的包含位置信息的 bin 进行组合,池化后我们需要进行卷积,采用 1x1、2x2、3x3 和 6x6 这 4 种不同尺寸的卷积操作得到多个尺寸的特征图,并对这些尺寸的特征图再次选择 1x1 的卷积,降低每级通道数,然后将得到的特征图使用双线性插值模型上采样后得到与特征图相同的尺寸,最后将不同尺度金字塔池化并将上采样的特征斑马线图像与原来的斑马线特征图进行结合。

金字塔池化模块在卷积层的后面对每一张斑马线图片都进行了多方面的特征提取,有效地提取到斑马线的特征信息,而且 SegNet 模型经过与金字塔池化模块融合后,可以输入任意尺度的斑马线图像,可以实现从全局中提取出语境信息,提高了斑马线识别的精度。

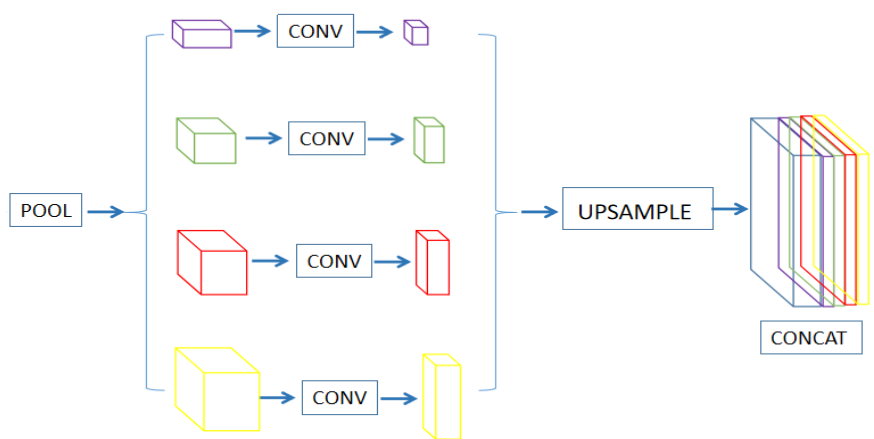


图 2.4 金字塔池化图

### 2.2.2 结构改进

为了减少网络模型的运行时间,提高斑马线检测的实时性,将模型对称的结构进行改进,改为编码不变,解码减少的不对称结构,结构的改进极大的提升了运行速度。

主要利用改进后的 SegNet 模型来完成斑马线像素点位置的检测,并在最终获得斑马线检测图像。改进的 SegNet 模型结构依然是分为编码和解码两部分,主要将解码部分

减少，编码部分不变，在解码部分利用  $1 \times 1$  的卷积核来减少网络参数，这样利用了减少的网络时间来提高了模型的运行速度，达到了提高斑马线检测实时性的目的。

在斑马线图像中存在大量的冗余信息，解决信息冗余的方式为压缩斑马线图像。我们利用模型的编码部分卷积层来获得各种不同的斑马线图像特征信息，将斑马线图像的感受野增大，这样做的目的是保证我们在进行斑马线检测时不会存在上下文的信息减少或者信息的可能增多。细节的处理也是不可避免，对细节的处理到位，才能减少边缘信息的丢失，防止信息丢失情况的发生，因此我们对细节进行微调，下采样中存在索引信息，上采样中的特点应用于解码器的 4 个上卷积层中需要索引信息的指引，并且在其中加入  $1 \times 1$  的卷积核实现网络参数的减少。改进后 SegNet 模型结构如图 2.5 所示。图 2.5 中 Input 为输入图像，output 为输出图像。PPM 为金字塔池化模块，是为了获取全局信息，从而达到减少斑马线细节信息丢失的目的。其中，Conv、Deconv、maxpool 和 ReLU 分别为卷积、反卷积、最大池化和激活函数，softmax+argmax 获取到斑马线图像特征点位置。

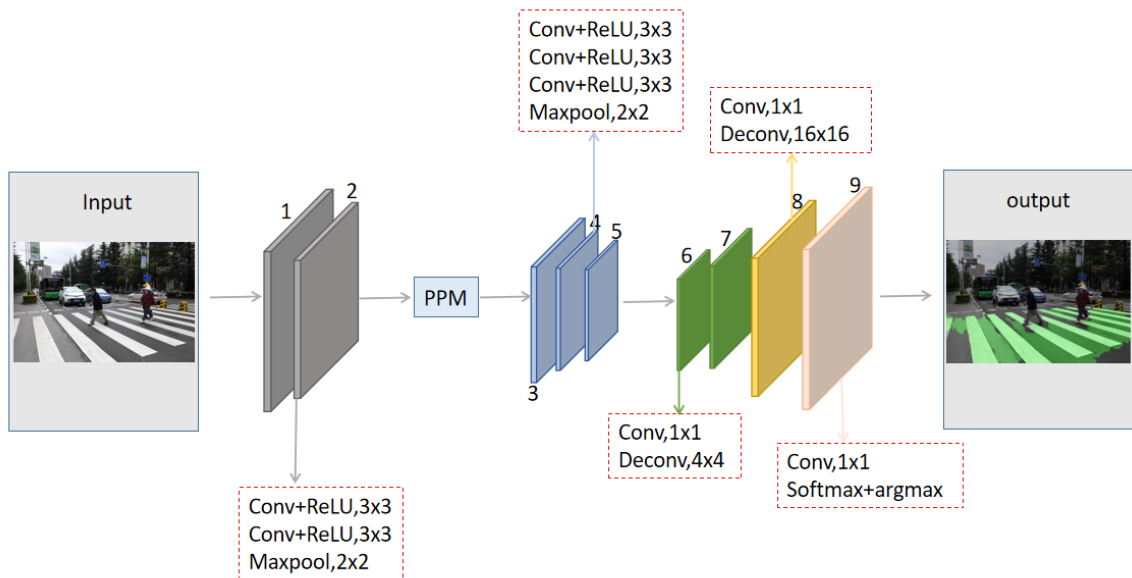


图 2.5 改进 SegNet 模型结构图

## 2.3 实验分析

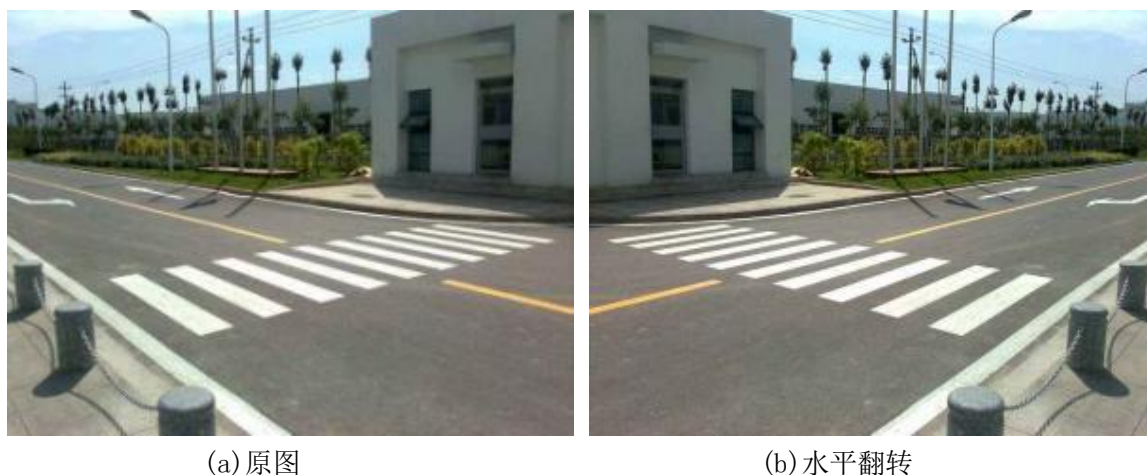
### 2.3.1 数据集准备

本文主要采集了 1150 张斑马线图像<sup>[51]</sup>样本作为本文数据集，主要来源于两部分，一部分是通过网络搜索到的网页图片，为了获得真实场景下的斑马线图像，另一部分通

过人工实景采集斑马线区域图像，尽可能得到各种干扰情况下的斑马线图像。由于采集到的图像数量较少，因此采用图像翻转、添加高斯噪声、图像增强等方法扩充数据样本，使样本数据能够满足训练需求，使得训练后的模型能够满足不同情况下的斑马线检测需求。将数据样本使用标注工具 labelme 标注后得到 json 文件，然后将 json 文件转换为用于斑马线检测训练的标签集。图像样本包含了不同方向不同场景下的拍摄，并且存有各种其他标线影响和各种不同的遮挡物遮挡及斑马线有污损等情况的干扰，将其随机划分出 80% 为训练集，20% 为测试集。测试集图片见图 2.6，样本扩充见图 2.7，图 2.7 中图（a）为原始图像，图（b）为将原始图像进行水平翻转，扩充图像样本；图（c）为在原图像上添加高斯噪声，通过改变系数、方差和均值等使图像发生改变，扩充样本；图（d）为图像增强，改善图像的视觉效果。



图 2.6 斑马线图片



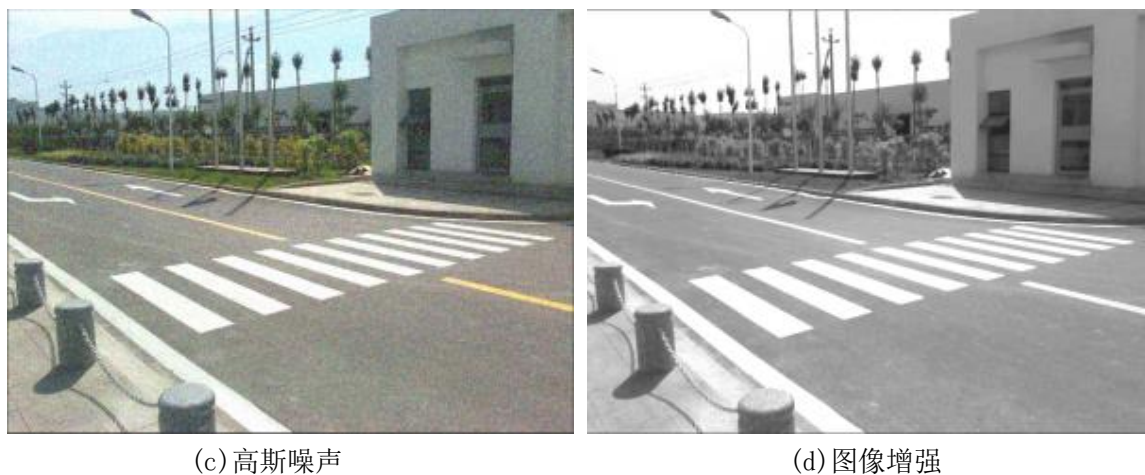


图 2.7 样本扩充

### 2.3.2 训练过程

训练参数设置 epoch 为 50，批处理量设置为 14，学习率的初始值设置为 0.01，动量配置为 0.9，权重衰减配置为 0.00055，在实验的 Accuracy 不在增加就停止训练。实验在迭代了 40 次后停止了训练，图 2.8 为精确度变化曲线，图 2.9 为训练过程的 loss 变化曲线。两图中横坐标都代表迭代次数，图 2.8 中纵坐标表示精确度，图 2.9 中纵坐标表示 Loss 值。迭代 35 次后，训练集的精确度收敛于 98.6%，Loss 值收敛于 0.046，测试集精确度收敛于 95.7%，Loss 值收敛于 0.205。

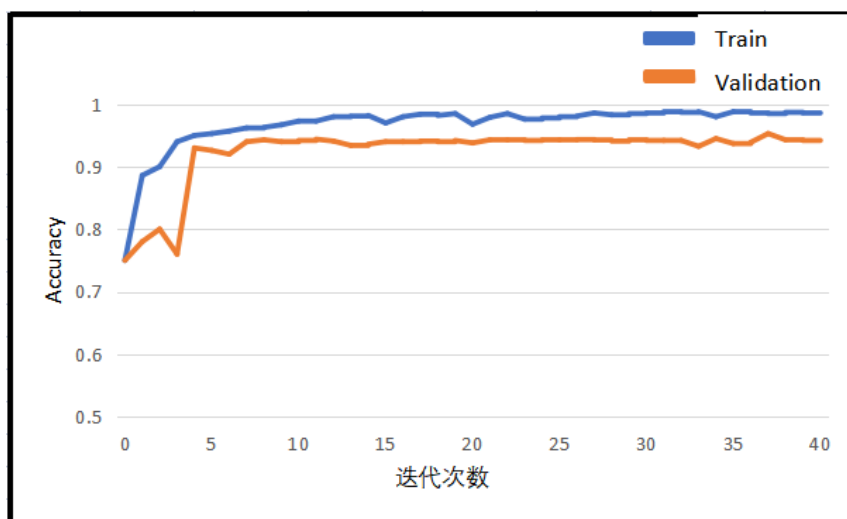


图 2.8 Accuracy 曲线图

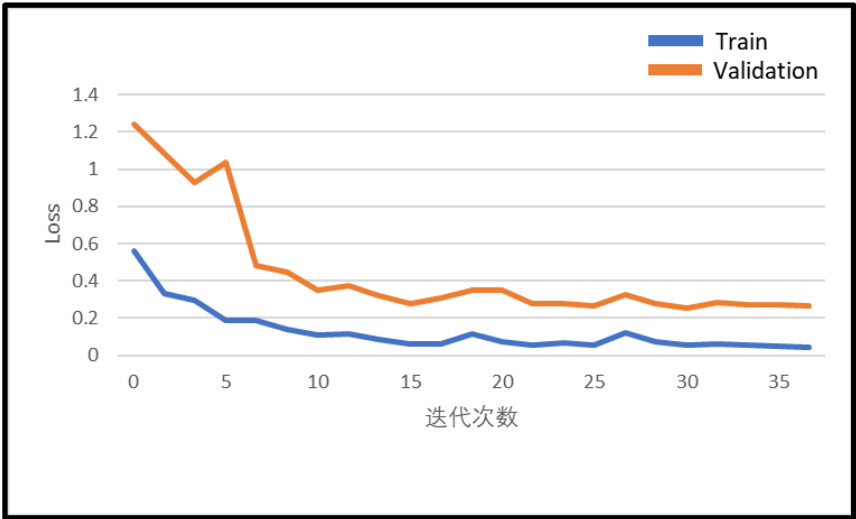


图 2.9 Loss 曲线图

2.4 实验结果与分析

2.4.1 评价指标

在实验的数据集上进行测试，本文使用相同的数据集进行改进 SegNet 模型、原始 SegNet 模型、U-Net 模型以及 PSPNet 模型等不同模型的训练，改进 SegNet 模型实现了斑马线检测<sup>[52,53]</sup>更加准确且清晰，即使遮挡因素影响的的斑马线图像，本实验检测效果依然良好，对图像中斑马线的分割检测准确，不会由于遮挡情况而受到很大的检测干扰。

为了准确客观的分析实验结果，本文使用准确率  $P$  (Precision)、召回率  $Re$  (Recall)、 $F1$  值三种评价标准来衡量模型的性能，其主要计算公式(2.1)，(2.2)，(2.3)

$$P = \frac{T}{T + F} * 100\% \tag{2.1}$$

$$Re = \frac{T}{T + Fg} * 100\% \tag{2.2}$$

$$F1 = \frac{2 * P * Re}{P + Re} \tag{2.3}$$

其中，正确检测出的斑马线目标数使用  $T$  (Truepositive)表示，错误检出的斑马线目标数使用  $F$  (Falsepositive)表示，没有被正确检出的斑马线目标数使用  $Fg$  (Falsenegative)表示。

当 Score\_threshold=0.5 时，其在斑马线数据集上性能对比结果如表 2-1 所示。改进 SegNet 模型在  $P$ 、 $Re$ 、 $F1$  值均更有优势。与其他模型对比，准确率与召回率的指标方面，改进 SegNet 模型的准确率明显提升，性能更高，不仅保证了检测数的有效性，且召



回率也有提升。在  $F1$  分数评价指标方面, 改进 SegNet 模型精确度有了显著提升, 且检测耗时也有所降低。实验数据表明, 本算法比其他算法更具有优势, 且综合评定结果更稳定。

表 2-1 4 种模型性能对比

方法	P	Re	F1
改进 SegNet	0.976	0.970	0.973
原始 SegNet	0.894	0.861	0.877
U-Net	0.838	0.760	0.797
PSPNet	0.786	0.652	0.716

#### 2.4.2 检测结果分析

本文为证明改进 SegNet 模型的检测效果, 依次对改进 SegNet 模型、原始 SegNet 模型、U-Net 模型以及 PSPNet 模型分别进行了实验, 选取了相同情况下的斑马线图片进行了测试, 结果如图 2.10 所示。可以明显看出, 改进 SegNet 模型的斑马线检测效果更好, 不仅表现于检测全面, 也表现于检测精度高。具体分析为: 在图 2.10 中 PSPNet 模型的检测效果特别差, 只检测出斑马线的一点区域, 而且还错将停止线检测为斑马线; U-Net 模型检测效果稍好, 但存在检测不均的问题, 可以大致检测出斑马线; 原始 SegNet 模型基本上可以将斑马线检测出来, 但是也存在检测不全面、不精准的问题, 斑马线的边缘部分并没有检测出来; 改进 SegNet 模型可以将斑马线清晰检测并且边缘部分检测效果良好, 检测精确度达到 97.6%。由检测结果分析可知相较于原始 SegNet 模型、U-Net 模型和 PSPNet 模型, 改进 SegNet 模型斑马线检测效果最好。

从表 2-2 中, 对比其他模型, 可以看出改进 SegNet 模型测试运行时间最短, 证明了改进模型的运行速度最快, 实时性得到提高。



(a) 原始 SegNet 模型实验结果图



(b) U-Net 模型实验结果图

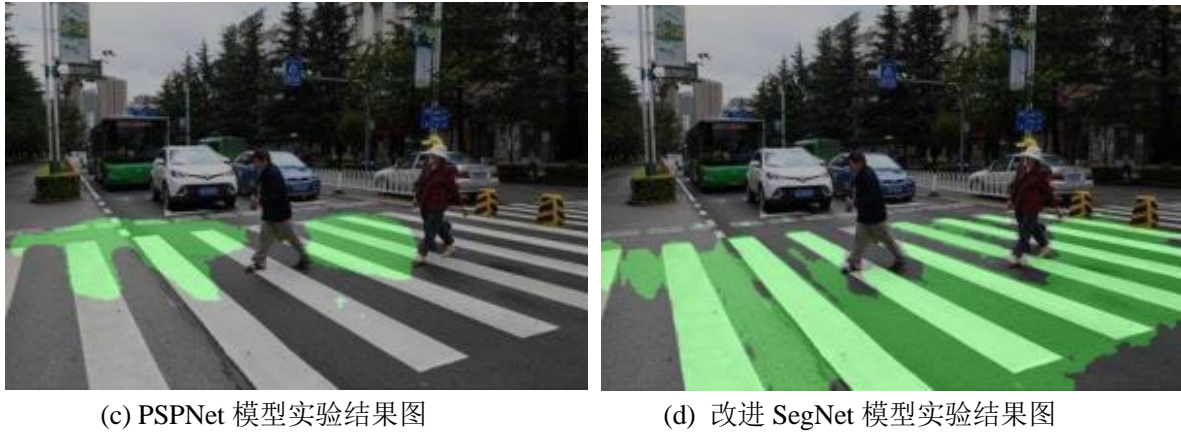


图 2.10 相同场景下的 4 种模型实验结果

表 2-2 各模型测试平均运行时间

方法	原始 SegNet	U-Net	PSPNet	Ours
时间	68.7ms	75.5ms	95.6ms	58.6ms

2.5 本章小结

本章首先介绍了应用于本文实验的语义分割模型，包括 SegNet 模型、U-Net 模型、PSPNet 模型，并对各模型进行了相关理论知识介绍，紧接着以 SegNet 模型为基础模型，提出改进 SegNet 模型。将模型的特征提取部分改进，融入金字塔池化模块，来获取全局信息；将模型对称的结构改为编码不变，解码减少的不对称结构，减少网络参数。本章在实证研究基础上，对获取到的斑马线图像进行亮度增强、添加噪声、图像翻转三种方式来扩充采集的斑马线图像样本，使斑马线图像样本满足各种语义分割模型检测需求。本文提出的改进 SegNet 模型与原始 SegNet 模型、U-Net 模型和 PSPNet 模型相比，改进 SegNet 模型检测斑马线精度更高，且运行速度加快，满足当前实时性检测的需求。识别斑马线使车辆获知道路信息，斑马线的精准检测实现为后续行人检测以及行人穿行斑马线意图预测提供了前提。

## 第三章 行人检测算法研究

在斑马线检测研究的基础上，将行人规范在斑马线区域范围内，致力于行人检测算法研究。行人检测在攻克无人驾驶与辅助驾驶方面是至关重要的一环，其是在图像或者序列中将计算机视觉技术进行应用判别行人的存在，并将行人的位置表述出来，行人被精准定位并被检测识别。

传统检测方法检测行人算法众多，较经典模型如以 HOG+SVM<sup>[54]</sup>方法为基础改进，其中 HOG 算法提取行人特征信息，SVM<sup>[55]</sup>分类器实现分类，从而使得检测速度与检测效果达到有效平衡。伴随着技术的进步，深度学习方法异军突起。该方法以人类视觉感知系统为模拟基础，直接提取输入图像的特征信息，提取到的特征经历层层传递的过程，进而获得图像想要传达的高维信息，避免了提取特征的复杂性以及数据分类的重建过程，精度提高，速度加快，被广泛应用于机器视觉中。深度学习方法中检测行人效果较好的有 SSD 检测算法与 Faster-RCNN 检测算法。

本章选用传统方法 HOG+SVM 进行问题研究，选用深度学习方法中的 SSD 算法与 Faster-RCNN 算法进行检测实验，并针对算法中存在的耗时大与小目标检测困难问题提出改进。

### 3.1 传统行人检测方法

本章传统行人检测主要研究 HOG(histogram of oriented gradient)+SVM(support vector machine)检测算法，将 HOG 提取到行人特征利用 SVM 进行分类，实现检测行人目的。方向梯度直方图（HOG）完成的功能是对物体完成轮廓、形状等的检测进而形成特征描述子，特征描述子通过计算与统计数字图像中可以被识别出来物体的形状、轮廓等边缘信息的梯度方向直方图的方式形成，主要应用于计算机视觉与图像处理中。法国的研究人员 Navneet Dalal 等人提出利用 HOG 对图像边缘特征进行提取，并利用支持向量机作为二分类器训练此模型，经过大量实验测试与对比，表明了 HOG+SVM 方法检测行人具有良好效果，在精度与速度方面，该检测方法相较于其他传统方法都处于较好水平，因此该方法广泛应用于计算机视觉。

#### 3.1.1 HOG 算法



HOG 特征是一种特定描述局部区域的符号，将检测物体的轮廓等信息用梯度分布进行描述，先对整体图像分割形成部分连接区域，然后各个连接区域形成一个方向梯度直方图，最后将形成的直方图进行组合后就可以表示出特征描述子。大致流程图如图 3.1 所示。

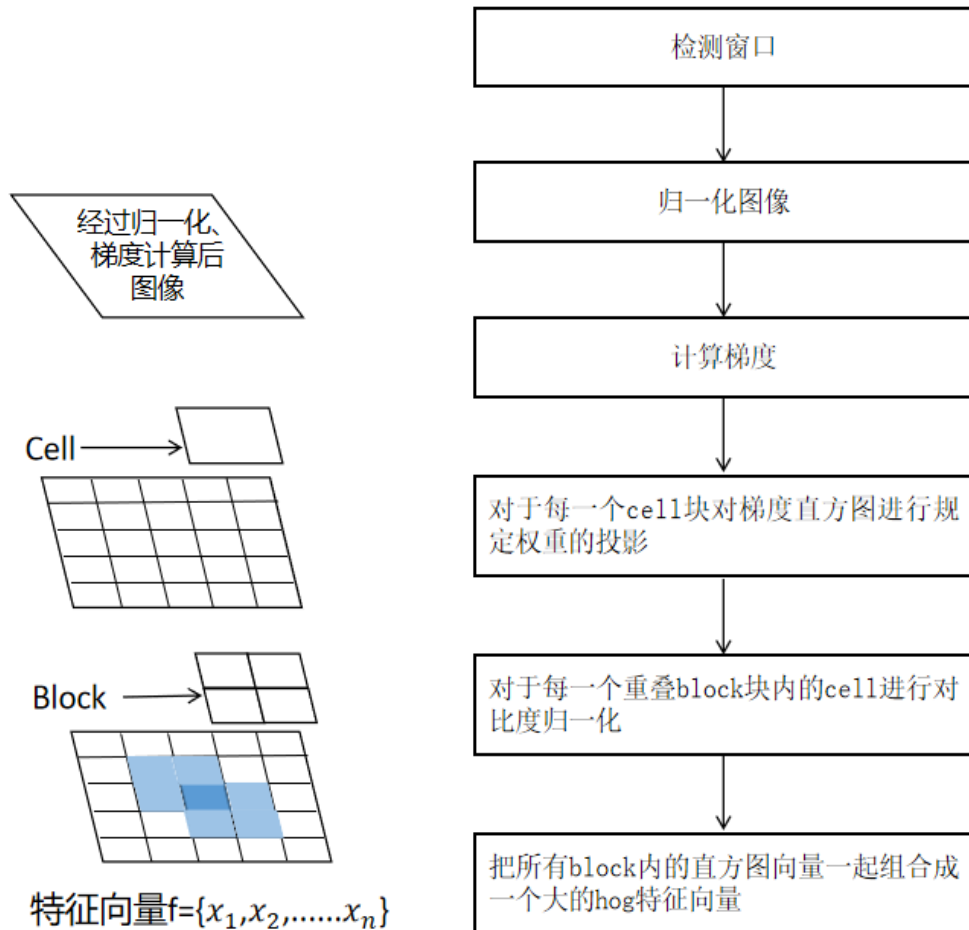


图 3.1 HOG 算法流程图

图像本身存在纹理特征程度不同的问题，以及有些局部区域表层也存在曝光概率，因此对于图像表面存在的信息阴影减低问题和光照变化产生的影响问题，解决此问题是通过图像进行压缩处理，原始检测图像被转化为灰度图后，将新得到的图像继续进行归一化处理，标准化 gamma 空间和颜色空间。gamma 压缩公式(3.1)：

$$I(x, y) = I(x, y)^{\gamma} \quad (3.1)$$

图像完成归一化处理后，需要将归一化后的图像各坐标方向的梯度信息进行计算，将计算出的梯度信息应用于图像中各个像素点位置的梯度方向的计算；存在求导操作，此操作不仅可以捕获到需要的图像形状、轮廓以及纹理的信息特征，且处理了光照影响，将影响降到了最低，图像中像素点  $(x, y)$  的梯度为式(3.2)，(3.3)：

$$G_x(x, y) = H(x+1, y) - H(x-1, y) \quad (3.2)$$

$$G_y(x, y) = H(x, y+1) - H(x, y-1) \quad (3.3)$$

式中  $G_x(x, y)$ ,  $G_y(x, y)$  分别表示输入图像中像素点(x,y)处的水平方向梯度、垂直方向梯度和像素值, 像素点(x,y)处的梯度幅值和梯度方向分别为式(3.4), (3.5):

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (3.4)$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (3.5)$$

通过对各像素点的梯度数据获取后完成建立梯度方向直方图的工作。通过梯度信息进行建立梯度方向直方图工作, 各个细胞单元不但要参与建立过程, 而且还要在参与的同时把编码信号提供给图像的局部域, 用来维持在图像中出现的人体的各种不同姿态以及对图像外观所存在的弱敏感性。Cell 即由单元格经过分割图像而获得, Cell 的图像中的梯度信息可以通过各方向块进行计算, 然后设置权值为梯度大小的加权投影, 将其应用于 Cell 中像素点的梯度方向上完成运算, 在进行了这些运算之后即可获得每个 Cell 即单元格的梯度方向直方图。

在上述操作后得到的梯度方向直方图存在范围不均问题, 图像上产生此问题的原因是光照、阴影以及边缘特征的影响, 通过对梯度强度进行归一化处理的操作, 完成降低影响的要求。在完成各项操作后形成不同的细胞单元, 组合各细胞单元, 形成区间(block), 各区间相互连通于空间中, 且区间满足足够大的要求。形成的区间之间存在重叠的问题即区间会出现重叠现象, 这就说明每个单元格可能存在出现多次的问题, 并且会出现最终的特征向量中各个单元格特征以不同结果展现的现象发生, 需要完成单元格处理操作, 获取各个区间单元格的特征向量信息, 再将得到的特征向量进行串联, 最后串联后得到的为区间 HOG 特征。

最后收集存在于检测窗口中的重叠的块的 HOG 特征信息, 结合它们来形成我们分类需要的特征向量。

### 3.1.2 SVM 分类器模型

支持向量机(support vector machines, SVM)是一种常见的判别方法, 通过二分类方法实现分类。分类原则是将特征空间间隔最大化, 在实验中找到最优超平面从而达到分

割样本数据的目的，SVM 最终形式化归为对一个凸二次规划问题求解。SVM 学习算法是完成最优求解，并寻找到最终解决问题的方案。数据集存在线性可分性的问题，因此有多个分离超平面，但其最大分离超平面有且只有一个。SVM 工作原理就是需要正确划分出提供的数据样本中的训练数据集，再将几何间隔最大的分离超平面进行求解。

输入：训练数据集计算式为(3.6)

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \text{ 其中, } x_i \in R^n, y_i \in \{+1, -1\}, i = 1, 2, \dots, N; \quad (3.6)$$

输出：分离超平面和分类决策函数

(1) 选择惩罚参数  $C > 0$ ，构造并求解凸二次规划问题计算如式为(3.7)，(3.8)：

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^N \alpha_i \quad (3.7)$$

$$s.t. \sum_{i=1}^N \alpha_i y_i = 0 \quad (3.8)$$

其中  $0 < \alpha_i \leq C, i = 1, 2, \dots, N$ ，得到最优解  $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ 。

(2) 计算  $w^* = \sum_{i=1}^N \alpha_i^* y_i x_i$  选择  $\alpha^*$  的一个分量  $\alpha_j^*$  满足条件  $0 < \alpha_j^* < C$ ，计算如式(3.9)

$$b^* = y_i - \sum_{i=1}^N \alpha_i^* y_i (x_i, x_j) \quad (3.9)$$

(3) 分离超平面： $w^* \cdot x + b^* = 0$ ；分离决策函数式(3.10)：

$$f(x) = \text{sign}(w^* \cdot x + b^*) \quad (3.10)$$

### 3.1.3 实验结果分析

HOG+SVM 算法对于特征明显的行人与非行人可以完成区分，背景简单即干扰因素较少的时候行人可以被检测到，且只有一个行人时可以将行人检测出来，对于情况复杂图像不能有效完成检测，包括图像中存在的人群较多和人体特征不够鲜明等问题出现的情况，且这种训练方法需要足够的样本来支撑图像的多样性，需要进行多次多样本训练。选取自己拍摄视频截帧图片检测情况如图 3.2 所示。



(a)单个行人检测效果图



(b)多个行人检测效果图

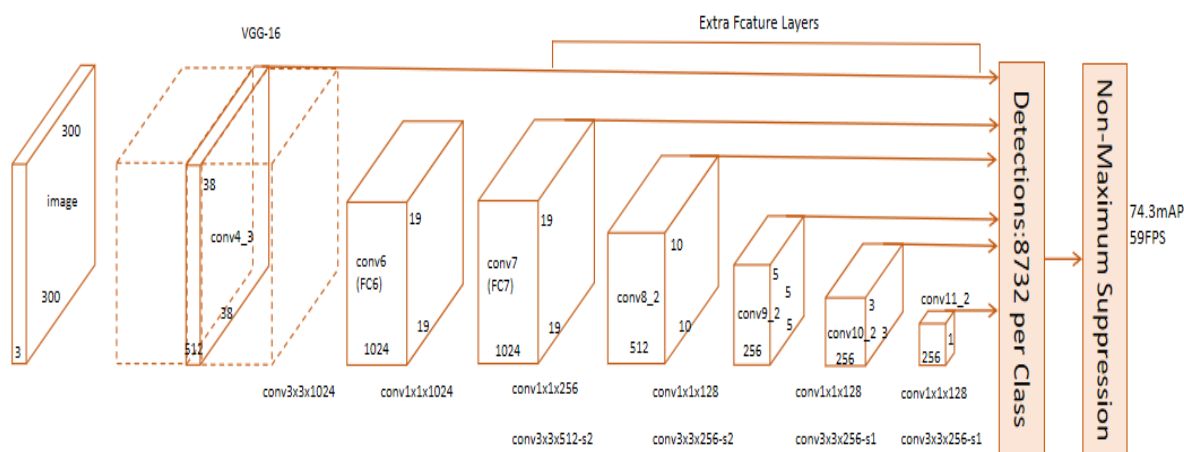
图 3.2 HOG+SVM 检测结果图

依据传统行人检测算法实验结果验证，传统检测方法中存在检测不全面、检测精度低、训练耗时、计算量大、实时性差等问题，为了达到行人检测效果更好的目的，转用深度学习方法进行行人检测探究，选用深度学习方法中目标检测效果良好的 SSD 算法与 Faster-RCNN 算法，接下来详细介绍两种算法模型。

### 3.2 深度学习行人检测算法

在上节中介绍了使用传统检测方法完成行人检测，存在一定缺陷，即使该检测方法存在一定有效性，但适用性不够广泛，并且行人具备多种影响因素，如背景复杂、光照影响、形体多样性等各种影响，图像往往存在更大的复杂性，传统行人检测方法并不能

### 3.2.1 SSD 网络模型



SSD 算法使用多尺度特征图完成检测。该算法模型以 VGG16 作为基础模型，其卷积层执行了通道数翻倍，需要训练的特征数量非常庞大，且全连接层使用较多，参数量大，使得模型在训练时收敛速度慢且导致了模型需要的存储空间较大，VGG16 的原始模型网络结构如图 3.4 所示，整个网络模型遵循规律性组合，在 3×3 的卷积层后紧接着是一个 2×2 的池化层，组合后模型末端是三个全连接层(FC6、FC7、FC8)和一个 softmax 分类层。

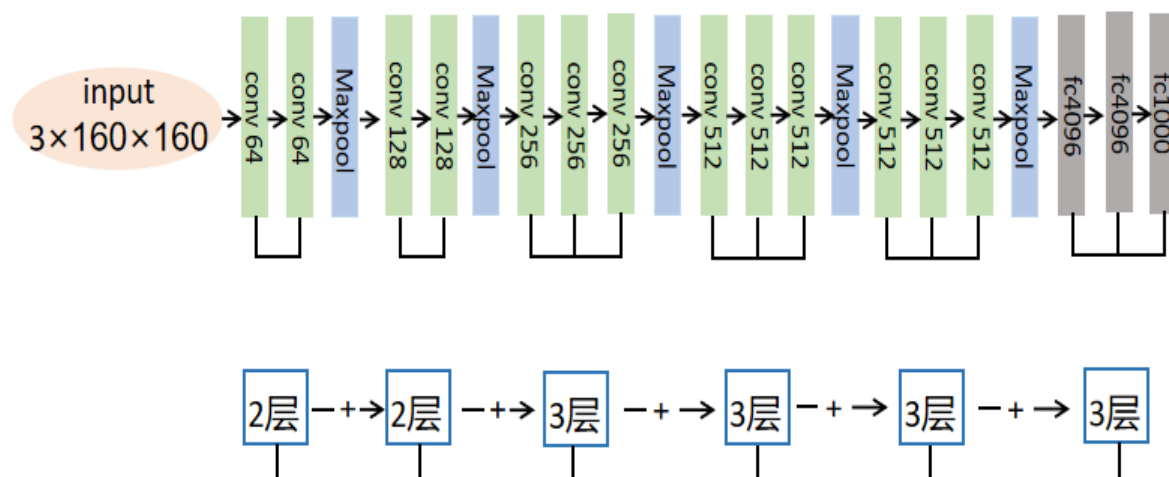


图 3.4 VGG16 网络结构图

而 SSD 模型将 VGG16 的基础模型进行了改进,首先改进的是全连接层,即将 VGG16 中的 fc6 和 fc7 转换为 conv6 和 conv7,两者组成是卷积层,卷积层数分别为 3×3 和 1×1,同时改变池化层,conv6 上的卷积层使用了空洞卷积,达到了保持参数不变的目的,满足了模型复杂度要求,卷积视野扩大,以扩张参数表示视野扩大大小,在视野扩大的背景条件下视野内的特征变得稀疏。然后移除 dropout 层和 fc8 层,同时去除了 softmax 分类层,增设卷积层提取目标图像特征信息,第一层单独设置特征图。

SSD 模型的多尺度预测网络由 Conv4\_3、Conv7、Conv8\_2、Conv9\_2、Conv10\_2、Conv11\_2 这 6 组特征图组成,在各组特征图上设置不同的先验框数目,先验框尺度的设置依据为大小和长宽比。特征图的大小是否降低与尺度增加密切相关,关系为特征图变小则尺度增加,其特征图变化对尺度影响公式为(3.11):

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1} (k-1), k \in [1, m] \quad (3.11)$$

其中  $m$  指的是特征图个数,  $s_k$  表示先验框大小相对于图片的比例,而  $s_{\min}$  和  $s_{\max}$  表示比例的最小值和最大值。

SSD 模型可以完成对大小目标的预测功能,完成预测大目标功能时使用的是高层特征图,如 Conv8\_2 等,这些特征图语义信息丰富能够完成大目标预测;而在完成小目标预测功能时使用的是底层特征图,如 Conv4\_3,这些特征图边缘信息足够丰富能够完成小目标预测。模型在 6 组特征图上设置默认框达到目标类别判断和位置预测的功能实现目的,默认框设置为不同尺寸并且长宽高各不相同,使用 3×3 和 1×1 的小卷积核作为检



测器实现。最后通过非极大值抑制（NMS）方法获得行人检测结果。

### 3.2.2 Faster-RCNN 网络模型

立足于深度学习的方法在不断发展过程得以优化，满足实用性需求，将深度学习用于行人检测中取得了良好效果。Girshick 团队提出了 RCNN(Regions with CNN features)，在目标检测难题面前以 CNN 攻破瓶颈，解决了检测中的一些问题，取得了良好成效，成为里程碑式的跨越，CNN 网络可以将特征信息提取较完善，且此网络可以实现较为良好的分类效果，因此完成了目标检测问题的转化，在转化中借助了 CNN 网络优点，且以 RegionProposal 方法为基础。但是 RCNN 方法存在空间不足，计算量大的问题，原因是 RCNN 算法需要提前完成提取多个候选框对应图像的提取工作，这样就使得此方法占用较大内存，且 CNN 网络需要对每个 ProposalRegion 进行计算操作，这样会出现大量重复提取特征信息现象，因此 Girshick 团队改进了 RCNN 算法，在 2015 年提出了 Fast-RCNN 算法，候选框映射工程的加入，促使算法网络实现反向传播，改善了 RCNN 算法存在的问题，即网络训练的问题，促进了计算时间的缩短与计算量的减少。尽管改进后的 Fast-RCNN 满足了一定实时性需求，但其在进行检测时仍需要几秒时间，主要时间耗费在提取候选区域上，提取速度仍然不能够满足人们要求，在 2016 年 Girshick 团队提出了 Faster-RCNN 算法，将区域提取网络（Region Proposal Network，RPN）融入算法，解决了实时性问题。Faster-RCNN 在前两种网络的基础上改进后，速度得到提升的同时检测效果也逐渐趋于更好状态。

Faster-RCNN 网络以 VGG16 网络模型为基础模型，结构如图 3.5 所示，主要分为四部分，第一部分是 Conv layers，在该部分所进行的主要是对目标图像的获取特征工作，通过 conv+relu+pooling 层，来达到获取在目标区域内的所有特征图像的目的，并把获取到的特征图像用于共享操作，然后提交给后续第二部分和第三部分，即 RPN 层和 ROI Pooling 层。第二部分是 RPN 网络，在此部分生成候选区域，此部分使用 Sofamax 函数，判别出由 Anchor 机制所生成的候选框中是否存在我们所需要的目标信息，进而通过回归机制，将生成的候选框粗略定位。第三部分的 ROI Pooling 层，把前部分输入的特征图和候选区域收集起来，然后完成提取步骤，获取到候选区域的特征图，送入到第四部分分类层，使用候选区域特征图来得到候选区域所属类别，最终获取行人检测框的精准位置。

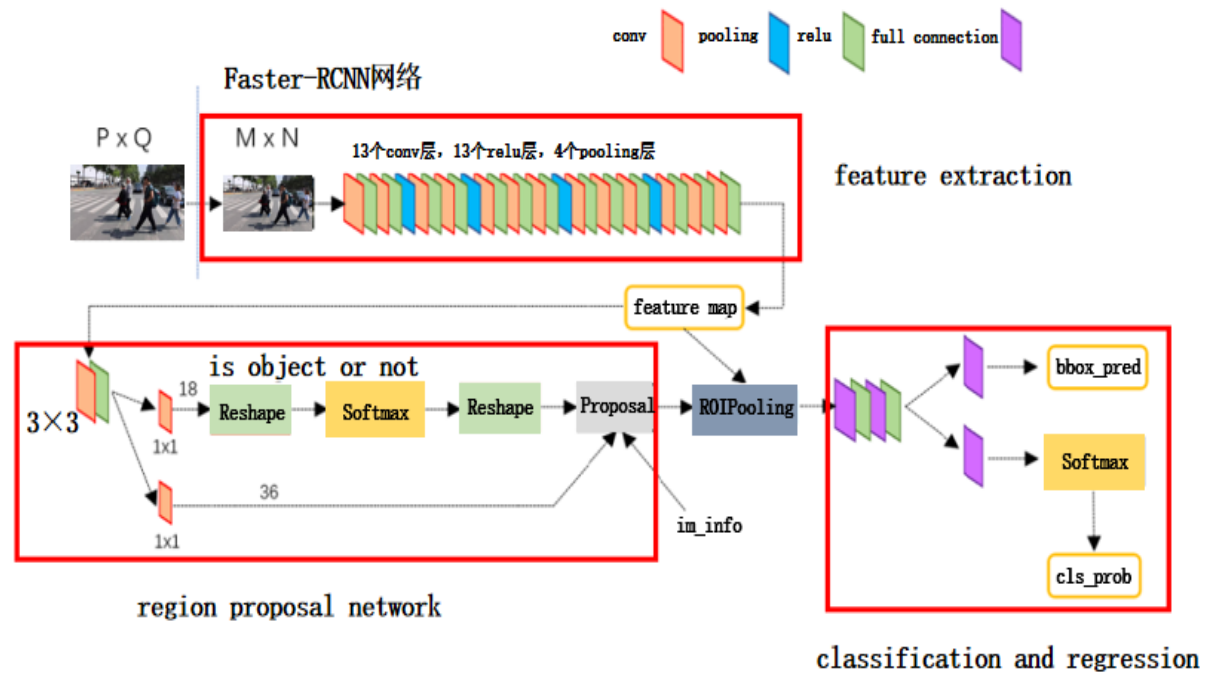


图 3.5 Faster-RCNN 模型结构图

## (1) Conv layer

由 13 个 conv 层+13 个 relu 层+4 个 pooling 层组成，完成原始图像的特征信息提取。Conv 层和 ReLU 层不会改变图像的大小，Pooling 层会使得图像改变大小。为了使深度卷积神经网络意义深远，ReLU 激活函数的使用增添了神经网络之间的非线性组合关系。另外 ReLU 激活函数使用具有良好成效，计算量较少，梯度消失现象得到有效解决，网络稀疏性得到提高，参数依赖性降低，过拟合问题缓解，深层网络训练的完成得到有效提升。其 ReLU 激活函数表达式如式所示，函数如图所示(3.12)。

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (3.12)$$

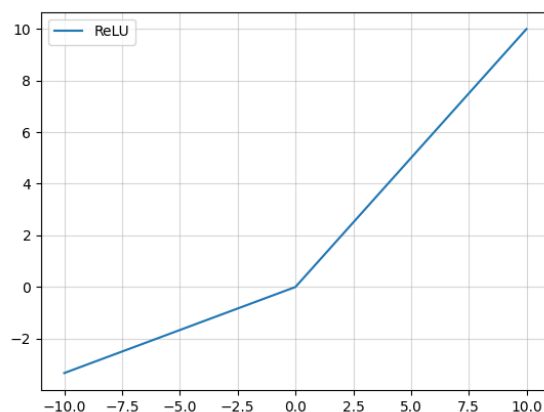


图 3.6 ReLU 函数曲线图



## （2）候选区域提取网络(RPN)

关于检测方法生成候选域功能方面，实验过程中会在此环节出现耗费大量时间的问题，例如 Adaboost 方法和 RCNN 网络，生成候选区域方法前者通过滑动窗口与图像金字塔实现，后者通过选择性搜索方法实现，两种方法面临的的就是时间消耗过多的问题，也更说明了耗时问题亟待解决。在 Faster-RCNN 以舍弃为解决办法，利用创新出的 RPN 网络结构去替代了原本的耗时方法，即替代了滑动窗口和选择性搜索方法，这样实现了端到端的候选区域提取，且在这个过程中只使用了卷积层网络，解决了生成候选区域的耗时问题，时间缩短后候选区域的生成速度自然得到有效提高。

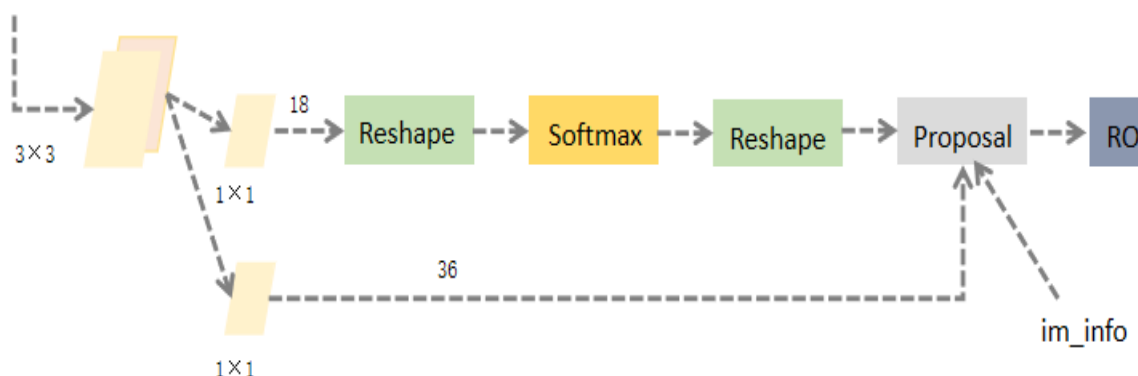


图 3.7 RPN 网络结构图

图 3.7 展示了 RPN 网络具体结构，特征图在经过  $3 \times 3$  的卷积层，将特征信息进一步集中，然后有两个走向，上面一个方向是经过 Softmax 函数来进行判断 anchors 中的检测目标，下面方向是计算 anchors 在进行对相同尺寸的图像倒推得到不同尺寸的图像产生的偏移量，将 Proposal 精准定位。

## （3）ROI Pooling

ROI Pooling 实质为通过映射将每个区域特征提取固定尺寸的特征图。这层主要将传输到的 proposal 收集起来，主要输入有两个：一是原始图像，二是上层 RPN 输出到这层的大小不同的建议框，然后计算出相同大小的输出图像到后续网络的建议特征图。具体操作：

①输入特征图，特征图为固定大小，不能改变，特征图的获取途径为深层网络中的多个卷积核池化，实现 ROI 映射，定位为特征图的对应位置；

②将映射后得到的区域分为大小相同的部分；

③然后对每个部分进行 max pooling 操作。

#### (4) 分类层

通过全连接层和 softmax 计算每个区域建议属于的类别，使用上层得到的建议特征进行计算，输出概率向量；同时完成偏移量的获取，获取途径为边框回归(bounding box regression)，使得回归操作使用的价值更高，获得更加精准的目标预测框。

本节将上述 SSD 算法与 Faster-RCNN 算法应用于行人检测实验中，以寻求最适合行人检测的算法，发现基于 Faster-RCNN 算法模型的行人检测方法在检测准确率和检测速度上均优于 SSD 检测算法，但是在真实场景中 Faster-RCNN 算法存在耗时大与小目标检测困难的问题，不适用于场景检测要求，因此本章在 Faster-RCNN 检测算法的基础上，提出了一种优化网络结构与聚类算法的改进检测模型，接下来将对改进模型做进一步详细介绍。

### 3.3 Faster-RCNN 行人检测算法改进

#### 3.3.1 网络结构优化

Faster-RCNN 算法在进行行人检测时，使用的 VGG16 的卷积层执行了通道数翻倍，需要训练的特征数量非常庞大，且全连接层使用较多，参数量大，消耗了更多的计算资源，原来的输入图片在经过 VGG16 网络后会变小，如果行人在背景复杂或者行人过小的图像中也有可能导致 Faster-RCNN 算法检测不到行人，因此改用 ResNet 网络结构，来解决计算和精确度问题。伴随着研究技术的发展，在实践中不断被证明的是网络层数的加深后，算法模型的精度也会提高，但是若无节制加深网络层数也会出现问题，即层数到达一定程度后，会导致梯度消失与爆炸现象发生，而这种现象的出现会不利于检测，影响检测效果。将原本的 VGG16 替换 ResNet 就可以将此问题进行解决，ResNet 网络具有较深的网络层数是解决此现象出现的关键。

ResNet 网络创新之处在于其连接方式，使用了“shortcut connection”连接方式，其网络结构如图 3.8 所示。每层在完成输入时同时进行学习，学习上层网络的残差进而形成更易优化的残差，网络层数加深的目的成功。图中展示了残差块，具体展示了两层，表达式为(3.13)：

$$F = W_2 \sigma(W_1 x) \quad (3.13)$$

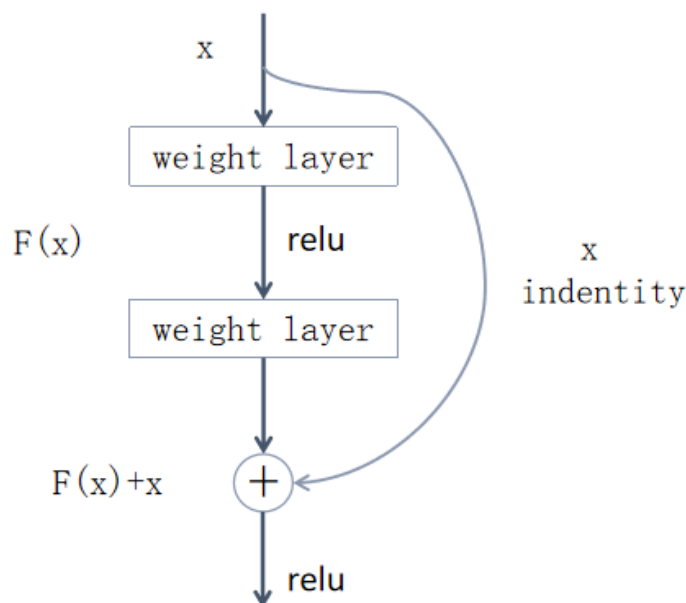


图 3.8 shortcut connection 网络结构图

残差块结构是 ResNet 的核心，通过这种结构的构建，网络层数加深，使得分类效果非常好。残差块中采用恒等映射的方式，将上一层的网络输出权重进行叠加后输出到激活函数传递给下一层，二次叠加操作即将权值第二次叠加后再把输入信息与此时输出的信息进行叠加，再通过激活函数，同时反向传播时下一层网络计算得到的梯度值传递给上一层，解决了梯度消失引起的检测效果不好的问题。

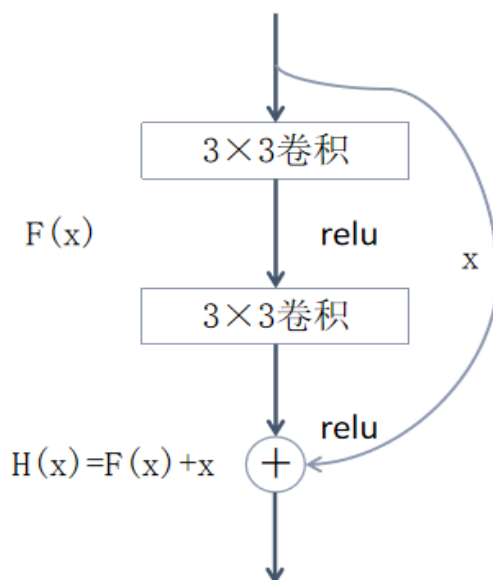


图 3.9 残差块图

残差块形式与常规的神经网络形式相比，其特别之处在于多了一条连接到输出位置的连线。通过这种形式，在每层输出时，可以完成浅层特征与深层特征的结合，极大地

增强了特征提取能力，同时梯度也不容易消失。残差块的输出可表示为(3.14)：

$$H_{(x)} = F_{(x)} + x \quad (3.14)$$

其中， $x$  代表残差块单元中的输入， $F(x)$ 表示分支的输出， $H(x)$ 是整个残差块的输出。

### 3.3.2 聚类算法优化

在行人检测上 Faster-RCNN 算法有检测速度慢、实时性差等弊端，针对该问题，本文提出一种聚类优化算法。本文以行人检测角度看待问题，基于 Faster-RCNN 算法中存在的实时性不能满足实际需求的情况下，提出改进。一般使用的 K-means<sup>[59]</sup>聚类算法，此算法为聚类基本算法，具有速度快，计算简便的优点，但在真正实践中 K 值难以估计，数据集开始无法正确分类，且在最初开始所选择使用的聚类中心具有较强影响性，因此会产生无法获得有效聚类效果的问题，且在实验过程中使用该算法会导致我们需要不断进行样本分类调整，同时还需完成不断重新计算调整样本后的聚类中心，所以当进行行人检测所需的样品量非常大时，这种算法在完成检测实验时需要耗费大量时间，因此选择引入高斯混合聚类算法(GMM)，此算法简短易用且计算量小，能够大大节省时间。GMM 算法的核心思想是将数据集中的样本点看作多个单高斯分布线性叠加的结果。算法假定的是以多个高斯分布函数为基础进而逼近其他形状的概率分布。基于各个高斯函数的均值与协方差的估计值，完成数据集中点分布的拟合，进而将各个点相对应的高斯函数中心确定下来，划定为相应的簇。

高斯分布的定义为式(3.15)，对  $n$  维样本空间  $X$  中的随机向量  $x$ ，如果  $x$  服从高斯分布，其概率密度函数为：

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} / \sum^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)} \quad (3.15)$$

其中  $\mu$  为  $n$  维均值向量， $\sum$  为  $n \times n$  维的协方差矩阵，因为高斯分布完全由平均向量与协方差矩阵的这二种参数所决定，因此为了清楚地表示高斯分布与相应参数之间的相互依赖关系，将概率密度函数记为： $p(x; \mu, \sum)$

高斯混合模型是指具有  $P(X; \theta) = \sum_{j=1}^K \alpha_j p(x; \mu_j, \sum_j)$  形式的概率分布模型。

其中  $\alpha_j$  是满足  $\alpha_j \geq 0$ ,  $j=1,2,\dots,K$ , 且  $\sum_{j=1}^K \alpha_j = 1$  的系数; 该分布的组成为混合成分, 共

有  $k$  个, 各混合成分与高斯分布一一对应。其中  $\mu_i$  与  $\sum_i$  是第  $i$  个高斯混合成分的参数。

EM 算法完成 GMM 模型的参数估计, 具体步骤为下:

1. 明确隐变量, 写出完全数据的对数似然函数。

上文中表述了高斯混合模型的概率密度函数式, 则对数似然函数为式(3.16):

$$L(\alpha, \mu, \sum) = \sum_{i=1}^n \log \left[ \sum_{j=1}^K \alpha_j p(x_i; \mu_j, \sum_j) \right] \quad (3.16)$$

2. “E”步: 确定 Q 函数。

高斯混合模型产生  $m$  个样例, 每个样例  $x_i$  的隐含类别  $Z = z_j$  的概率, 即  $Z$  的后验概率为式(3.17):

$$\begin{aligned} Q_i(Z = z_j) &= p(Z = z_j | X = x_i; \theta) = \frac{p(Z = z_j, X = x_i; \theta)}{p(X = x_i; \theta)} \\ &= \frac{p(Z = z_j, X = x_i; \theta)}{\sum_{r=1}^k p(Z = z_r, X = x_i; \theta)} = \frac{\alpha_j \phi(X = x_i; \theta_j)}{\sum_{r=1}^k \alpha_r \phi(X = x_i; \theta_r)} \end{aligned} \quad (3.17)$$

代入  $\phi$  的表达式, 即可得到  $Q_i(Z = z_j)$

3. “M”步: a: 求解参数  $\mu_1$ ; b: 求解参数  $\sum_1$ ; c: 求解参数  $\alpha_1$ 。

根据“M”步公式, 写出变形后的  $L(\theta)$  为式(3.18):

$$\begin{aligned} L(\theta) &= \sum_{i=1}^m \sum_z Q(Z; \theta_z) \log \frac{p(X, Z; \theta)}{Q(Z; \theta_z)} \\ &= \sum_{i=1}^m \sum_{j=1}^k Q_i(Z = z_j) \log \frac{p(X = x_i; \mu, \sum) p(Z = z_j; \lambda)}{Q_i(Z = z_j)} \end{aligned} \quad (3.18)$$

为了表达方便, 令  $Q_i(Z = z_j) = w_{i,j}$ ,  $p(Z = z_j; \lambda) = \alpha_j$ , 因此  $L(\theta)$  可变形为式(3.19):

$$L(\theta) = \sum_{i=1}^m \sum_{j=1}^k w_{i,j} \log \frac{p(X = x_i; \mu, \sum) \cdot \alpha_j}{w_{i,j}}$$

$$= \sum_{i=1}^m \sum_{j=1}^k w_{i,j} \log \frac{\frac{1}{(2\pi)^{\frac{n}{2}} |\sum_j|^{-\frac{1}{2}}} e^{-\frac{1}{2}(x_i - \mu_j)^T \sum_j^{-1} (x_i - \mu_j)} \cdot \alpha_j}{w_{i,j}} \quad (3.19)$$

a. 求解参数  $\mu_1$  公式为(3.20)

$$\begin{aligned} \frac{\partial L(\theta)}{\partial \mu_l} &= \nabla_{\mu_l} \sum_{i=1}^m \sum_{j=1}^k w_{i,j} \log \frac{\frac{1}{(2\pi)^{\frac{n}{2}} |\sum_j|^{-\frac{1}{2}}} e^{-\frac{1}{2}(x_i - \mu_j)^T \sum_j^{-1} (x_i - \mu_j)} \cdot \alpha_j}{w_{i,j}} \\ &= \sum_{i=1}^m w_{i,l} (\sum_l^{-1} x_i - \sum_l^{-1} \mu_l) \end{aligned} \quad (3.20)$$

在公式中，由于  $j=1\dots k$ ，即共有  $k$  个参数，而  $\nabla_{\mu_l}$  只是求解  $k$  个参数中第  $l$  个，因此在此  $\nabla_{\mu_l}$  的过程中，含参数  $\mu_{1 \sim l-1}$  和  $\mu_{l+1 \sim k}$  的项均为零，只剩下含  $\mu_l$  的项。

令上述公式等于零，解得：
$$\mu_l = \frac{\sum_{i=1}^m w_{i,l} x_i}{\sum_{i=1}^m w_{i,l}}$$

b. 求解参数  $\sum_l$  公式为(3.21):

$$\begin{aligned} \frac{\partial L(\theta)}{\partial \sum_l} &= \nabla_{\sum_l} \sum_{i=1}^m \sum_{j=1}^k w_{i,j} \log \frac{\frac{1}{(2\pi)^{\frac{n}{2}} |\sum_j|^{-\frac{1}{2}}} e^{-\frac{1}{2}(x_i - \mu_j)^T \sum_j^{-1} (x_i - \mu_j)} \cdot \alpha_j}{w_{i,j}} \\ &= -\nabla_{\sum_l} \frac{1}{2} \sum_{i=1}^m w_{i,l} [\log |\sum_l| + (x_i - \mu_l)^T \sum_l^{-1} (x_i - \mu_l)] \end{aligned} \quad (3.21)$$

微分处理，解得 
$$\sum_l = \frac{M_l}{\sum_{i=1}^m w_{i,l}} = \frac{\sum_{i=1}^m (x_i - \mu_l)(x_i - \mu_l)^T w_{i,l}}{\sum_{i=1}^m w_{i,l}}$$

c. 求解参数  $\alpha_1$  公式为(3.22):

$$\frac{\partial L(\theta)}{\partial \alpha_l} = \nabla_{\alpha_l} \sum_{i=1}^m \sum_{j=1}^k w_{i,j} \log \frac{\frac{1}{(2\pi)^{\frac{n}{2}} |\sum_j|^{-\frac{1}{2}}} e^{-\frac{1}{2}(x_i - \mu_j)^T \sum_j^{-1} (x_i - \mu_j)} \cdot \alpha_j}{w_{i,j}} \quad (3.22)$$

借用朗格朗日，解得  $\alpha_l = \frac{1}{m} \sum_{i=1}^m w_{i,l}$

## 3.4 实验分析

### 3.4.1 数据集准备

由于本文行人检测研究是无人驾驶与辅助驾驶研究的基础，因此本章研究中数据集选用 Caltech 数据集<sup>[60]</sup>，相对来说，此数据集样本足够丰富，能够满足实验需求，数据集通过车载摄像头拍摄而成，总体数据集图片质量相对于静态照片分辨率略低，频率为 30Hz，视频拍摄完成于洛杉矶，共有 5 个场景同时被分成了 11 个片段，数据集被分为训练集和测试集两部分。在数据集中使用矩形框标注了行人，同时也标注了矩形框存在的时间关系与遮挡情况。针对 Faster-RCNN 中存在的问题，选用此数据集，具体可解释为两个原因：一是数据集中存在行人背景复杂图像，能够有效突出 Faster-RCNN 算法中存有的耗时问题，进而彰显出本文改进模型的速度提升优越性；二是数据集中存在行人目标较小的特点，行人密集，突出了 Faster-RCNN 算法中存有的小目标检测问题，也彰显了本文改进模型可以更好完成小目标检测。

由于 SSD 算法和 Faster-RCNN 算法检测行人实验需要使用 Pascal VOC 数据集进行训练，因此实验前先将 Caltech 数据集转换为 VOC 格式。使用算法将 Caltech 数据集产生为训练集、验证集、测试集和验证测试集 4 部分。

### 3.4.2 训练过程

SSD 算法、Faster-RCNN 算法与改进 Faster-RCNN 算法完成实验后得到变化图像，在完成训练后得到损失值与迭代次数的关系为图 3.10，图中展示了三种算法在迭代次数增加的情形下，损失值如何完成变化的，从图中可以得到，在于其他两种算法的比较中，改进后的 Faster-RCNN 算法损失值收敛速度更快，迭代次数不断增加使得改进模型完成训练效果更好，损失值也在此过程中趋于稳定，改进 Faster-RCNN 算法在迭代次数 epoch=50 时损失值趋于 0.03，然后随着迭代次数增加损失值有小幅度的增加，之后当 epoch=60 时，损失值趋向稳定，趋于 0，即改进 Faster-RCNN 模型趋向稳定。而 Faster-RCNN 算法在迭代次数 epoch=80 以后模型趋于稳定，收敛速度明显低于改进模型，SSD 算法损失值一直保持最高且在 epoch=75 趋于稳定，但损失值较高，收敛于 0.1 附近。

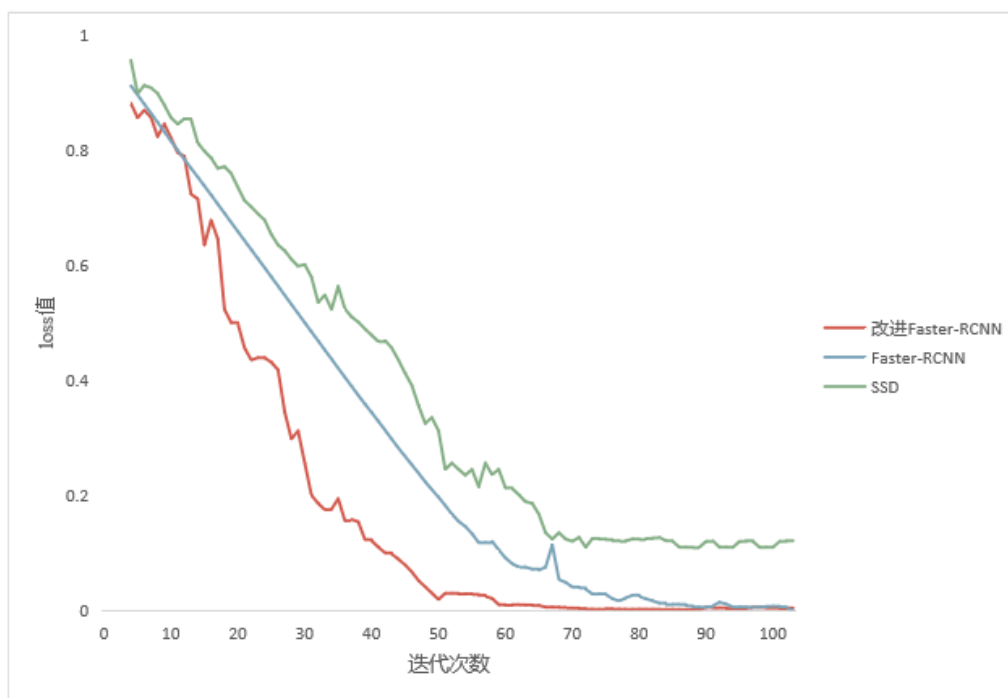


图 3.10 loss 曲线图

图 3.11 是 SSD 算法、Faster RCNN 算法与改进 Faster RCNN 检测算法在平均精确度的实验结果对比图。横坐标是迭代次数 epoch，纵坐标为平均精确度，图中展示的是伴随着迭代次数的增加，SSD 算法、Faster-RCNN 算法以及改进后 Faster-RCNN 算法的精度变化情况。图中对比可知，三种算法完成行人检测的平均精确度的变化曲线中，改进模型完成行人检测的平均精确度更高，且平均精确度不断变化中可以看出，改进 Faster-RCNN 模型完成行人检测效果更好。改进后 Faster-RCNN 模型在开始时平均精确度不及其他模型，但伴随着迭代次数的不断增加，改进模型的平均精确度也在不断提高，在 epoch 到达 50 时，改进模型的平均精确度达到三种模型的最高且开始稳定上升；当 epoch=80 时，改进模型平均精确度达到了 94%，之后迭代次数的增加不会致使改进模型平均精确度发生大幅度变化，改进 Faster-RCNN 模型平均精确度趋向稳定，为 97% 左右。



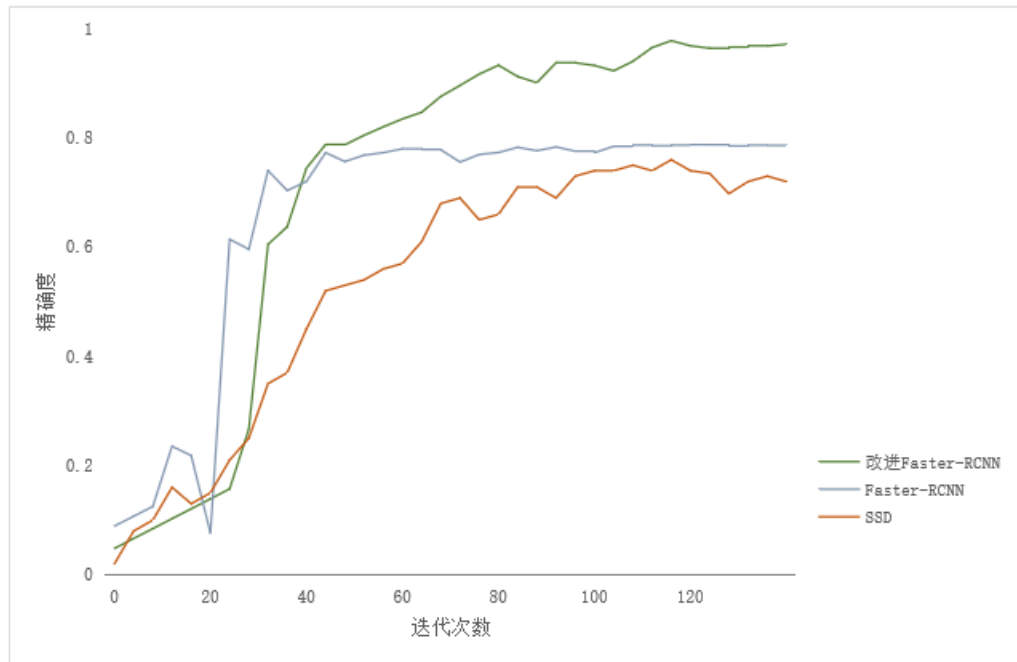


图 3.11 精确度曲线图

### 3.4.3 实验结果与分析

本文行人检测<sup>[61,62]</sup>实验使用的三种算法统一在 Caltech 数据集上进行,将三种算法即 SSD 算法、Faster-RCNN<sup>[63]</sup>算法与改进 Faster-RCNN 算法<sup>[64]</sup>的实验结果分别进行对比。实验结果如表 3-1 所示。

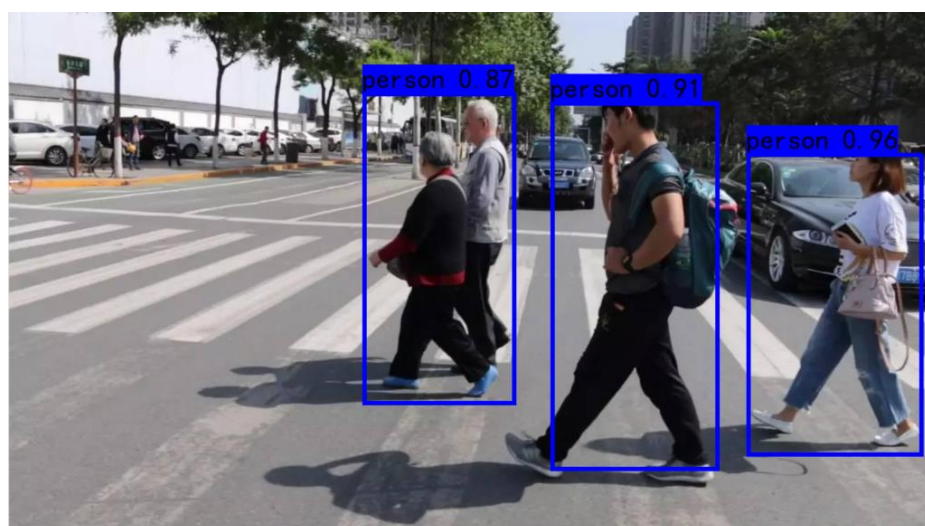
表 3-1 实验结果

算法模型	召回率 (recall)	准确率 (Accuracy)	平均精确度 (AP)
SSD 算法	85.1%	76.5%	86.5%
Faster RCNN 算法	97.9%	89.4%	94.2%
改进 Faster RCNN 算法	97.6%	96.8%	97.8%

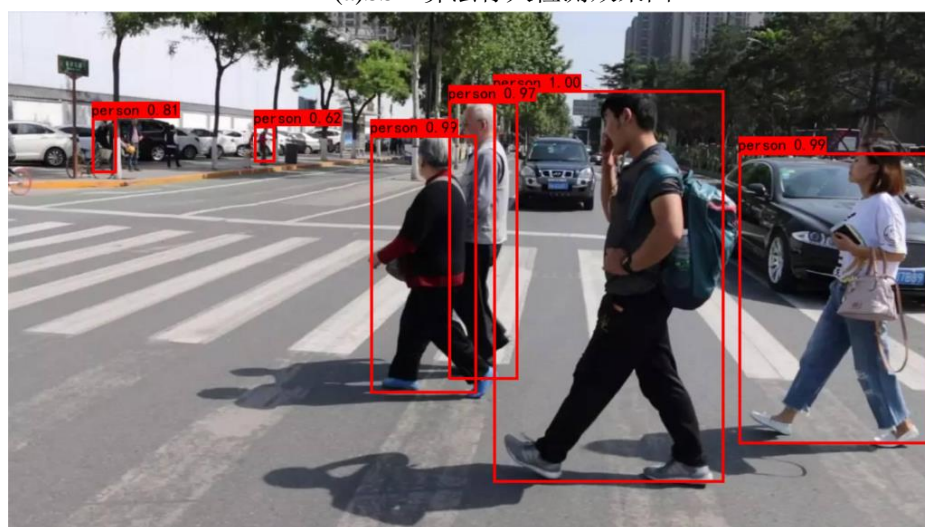
从表中展示的实验结果可以得到, SSD 算法进行行人检测效果比较差,准确率和平均精确度都比较低。Faster-RCNN 算法在行人检测实验中相对于 SSD 的算法来说,其检测精度更高,且在精度提升的同时召回率并没有下降,因此在完成行人检测方面, Faster-RCNN 算法检测行人明显优于 SSD 算法。对于 Faster-RCNN 算法在替换了特征提取网络结构和引入高斯聚类算法后,相对于原始 Faster-RCNN 算法准确率提升了 7.4%,精度提高了 3.6%,召回率由微下降趋势。精度显著提升是因为将特征提取网络进行了替换,即将 VGG 替换为 ResNet 网络,解决了 Faster-RCNN 算法中存在的网络层数越多导致梯

度消失的问题，以及引入高斯聚类算法使得行人小目标检测能力得以提升。召回率与准确率之间存在影响关系，准确率提升的同时会导致召回率得下降，但是改进后的 Faster-RCNN 检测算法在准确率显著提升的同时召回率仅仅下降了 0.3%，可以看出优化后的 Faster-RCNN 检测算法在完成准确率有效提升的同时召回率并不会受到很大影响。

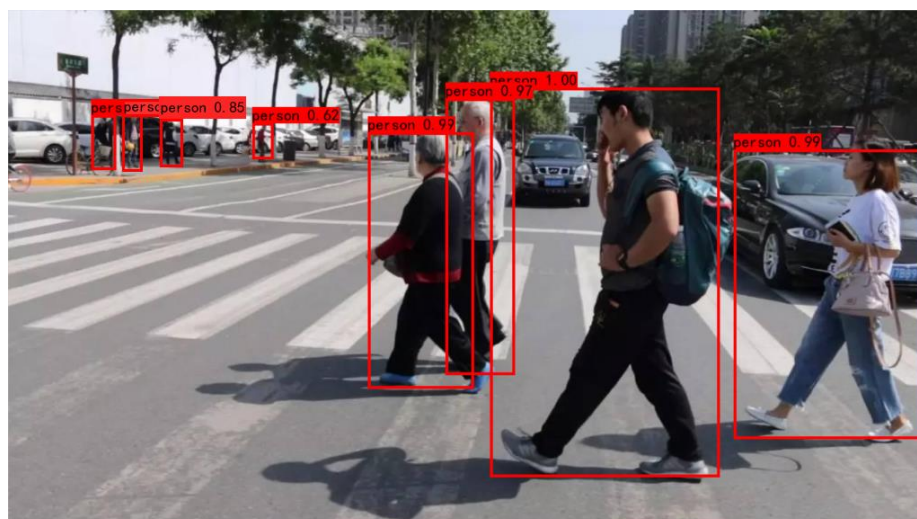
图 3.12 展示的三种检测算法检测效果对比图，从图 3.12 的检测效果来看，SSD 检测算法对于目标明显图像检测效果良好，但是对于背景复杂的检测图像不能够很好检测，特征提取不够完善，Faster-RCNN 对比 SSD 算法检测效果优越一点，但是对于小目标检测仍然不够，不能够很好检测出来小目标行人，而改进后的 Faster-RCNN 完美解决这些问题，在背景复杂且行人密集、目标较小的情况下能够进行检测，且效果良好。在各模型进行训练后与测试后，在自己拍摄的视频中截帧图片验证检测效果，效果如图 3.12 所示。



(a)SSD 算法行人检测效果图



(b)Faster-RCNN 算法行人检测效果图



(c)改进 Faster-RCNN 算法行人检测效果图

图 3.12 行人检测结果对比图

### 3.5 本章小结

本章首先对传统行人检测方法相关知识进行了介绍，然后针对传统检测方法存在的不足改用深度学习方法进行行人检测实验。先介绍 HOG 算法原理以及 SVM 分类器，然后利用传统方法完成行人检测实验，说明了传统检测存在检测不全面、检测精度低、训练耗时、计算量大、实时性差等问题。接着转用深度学习方法进行实验，选用了 SSD 算法与 Faster-RCNN 算法完成行人检测实验，先分别介绍了两种算法的基础知识以及算法相关演化史与算法的网络结构，然后进行实验对比，选用了两者中检测表现较好的 Faster-RCNN 算法为基础算法进行改进，依据深度学习算法中存在的耗时大的问题与小目标检测困难的问题，改进模型先优化网络结构，替换特征提取网络模型，提高了提取特征的优越性；再优化了聚类算法，引入高斯聚类，减小计算量，提升了检测速度；最后将 SSD 算法、Faster-RCNN 算法与改进算法进行实验对比，改进算法检测精度更高、速度更快且实时性更好，能有效检测小目标行人，本章为接下来的行人行为识别和行人穿行道路斑马线意图预测提供了基础。

## 第四章 基于深度学习的行人穿行意图预测算法研究

行人行为识别和意图预测<sup>[65]</sup>一直是无人驾驶系统与辅助驾驶系统中的研究重点。近些年, 由于深度学习的成功, 各种神经网络设计的算法也普遍应用于无人驾驶与辅助驾驶过程中, 在进行了行人检测后, 以行人行为为依据, 完成行人行为识别<sup>[66]</sup>后进而预测行人穿行道路斑马线意图, 帮助无人驾驶系统与辅助驾驶系统进行决策。行为识别以 CNN 网络模型用于提取图像特征信息, 以 LSTM 网络模型对行人行为分析时间序列进而完成行人行为识别分类; 行人穿行道路斑马线意图预测<sup>[67]</sup>实验从行人视角和车辆视角分析影响意图预测因素, 完成本文交通场景设计, 构建本文贝叶斯网络模型, 实现行人穿行道路斑马线意图预测。

### 4.1 行人行为识别研究

通过行为识别可以获知行人的动作信息, 为行人意图预测铺垫。由于深度学习的快速发展, 行为识别的研究也受到了越来越多国内外学者的关注与研究。在行为识别中, 相比于研究静态图像, 视频中存在的行人行为研究更倾向于研究人体在视频中的序列时空变化, 根据人在视频序列中运动的图像视频帧或者图像序列, 利用计算机处理和分析视觉信息, 将人体动作自动识别出来。行为识别技术应用于智能监控、智能家居、虚拟现实以及视频检索等各种智能场景中。人体行为具有多样性, 影响着人体行为识别的难度, 神经网络的快速发展, 不断冲击并解决这些难题。本文在众多行为识别算法中选取 CNN-LSTM 算法进行实验解决识别精确率的问题, 完成行人行为的识别工作。

#### 4.1.1 CNN 网络模型

CNN 网络模型得主要作用是完成空间关系的提取, 空间关系者为多维时间序列数据, 结构构成部分是卷积层与池化层, 以完成局部连接操作与权值共享操作得以著名, 使得 CNN 模型有了降低参数量、提取到特征数据、训练速度加快、泛化性能提高的功能, 网络结构图如图 4.1 所示。

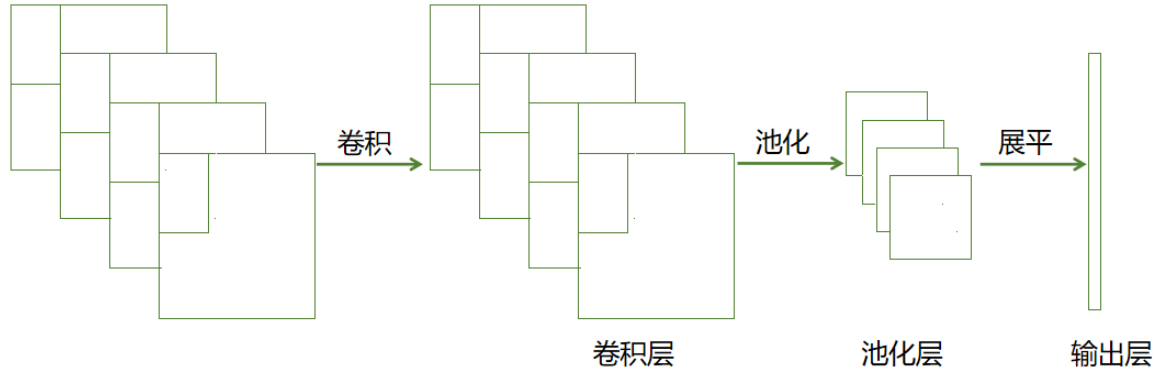


图 4.1 CNN 网络结构图

CNN 卷积神经网络结构主要是展示的卷积层和池化层，这两部分构成了其网络模型结构，卷积层的最主要功用自然是卷积运算操作，将在各个时间顺序下的各种特征信息加以提取，卷积运算层的主要组成元素是特征面，在各个特征面上的主要组成元素又是若干个神经元，卷积核在将上层局部的特征面区域进行相互连接之后就形成了神经元的存在，这样也就实现了卷积的运算功能。进行卷积计算操作之后进行池化层的池化操作，池化操作并不会导致特征面的数量发生变化，原因是池化层的主要构成元素同样也是多个特征面，且各个特征面与卷积层特征面对应。池化层执行的是将特征信息进行二次提取的工作，达到数据降维、参数量减少的目的，实验中最常使用的池化方法为最大池化、均值池化等。

卷积神经网络由三维卷积构成，每一类卷积都有自己的应用场景。在时间序列数据上主要应用一维卷积提取数据特征，一维卷积计算式为(4.1)

$$x_k^l = f\left(\sum_{i=1}^N x_i^{l-1} * w_{ik}^l + b_k^l\right) \quad (4.1)$$

式中  $x_k^l$  是  $l$  层第  $k$  此卷积映射， $f$  是激活函数， $N$  是输入做卷积映射的数量， $*$  是卷积操作， $w_{ik}^l$  是第  $l$  层第  $k$  个卷积核做第  $i$  层运算的权值， $b_k^l$  是  $l$  层相对应的第  $k$  个卷积核的偏置。

池化层采用的是最大池化法。式(4.2)表示取从向量  $x_k^l$  到  $x_{k+r-1}^l$  的最大值。对于序列  $x$ ，重复对每个窗口为  $r$  的连续向量进行最大池化操作，即可得到最大特征序列。

$$\hat{x}_k^l = \max(x_k^l : x_{k+r-1}^l) \quad (4.2)$$



## 4.1.2 LSTM 网络模型

长短期记忆网络(Long Short-Term Memory, LSTM)<sup>[68]</sup>作为改进的 RNN<sup>[69]</sup>网络模型,与传统神经网络相比,RNN 网络不仅只是将前一时刻的输入进行考虑,更是对前面的信息进行记忆,能够将记忆信息应用,更好的处理时间序列问题。RNN 结构图如图 4.2 所示,图中展示出 RNN 由输入层、Hidden Layer 和输出层组成,在 Hidden Layer 层上存有循环更新,以箭头表示,箭头所指代的循环路径表达为模型记忆功能的实现。

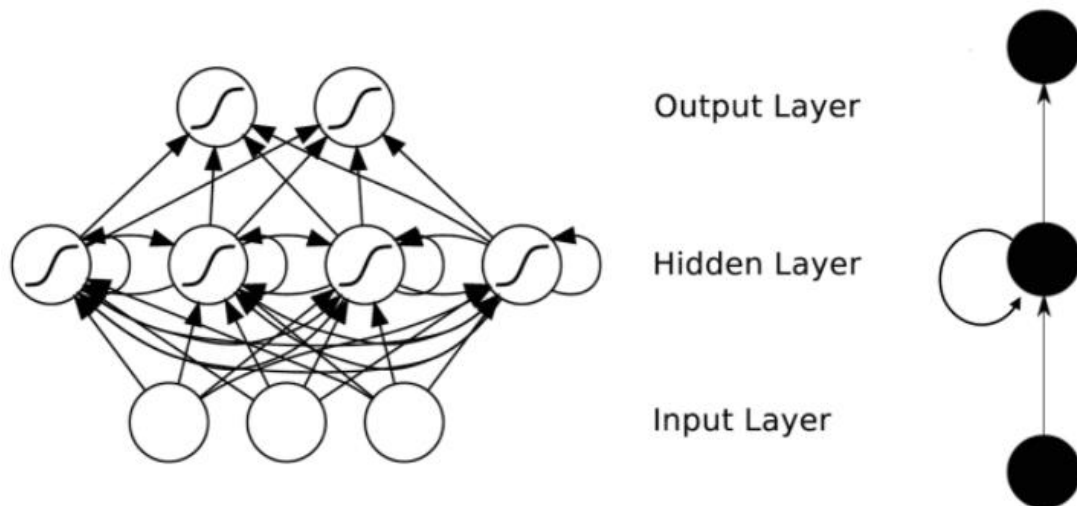


图 4.2 RNN 网络结构图

RNN 网络被称为循环网络的原因是网络的当前输出关联于前面输出,具体表现为网络能够记忆前面信息并将记忆信息应用于当前需要进行的计算中,而 LSTM 网络是 RNN 网络的改进,应用领域为时间序列预测领域。和 RNN 网络不同的是 RNN 网络是简单的先行求和,而 LSTM 网络则是利用自己特殊的“门”结构处理重要信息的增删问题,实现重要内容的存留功能,LSTM 网络在隐含层加入了输入门、输出门、遗忘门,并且将用于存储的单元增加。LSTM 中的输入门执行的是进行非线性变化,当输入量被输送入 LSTM 的网络的隐含层后,输入门会对输入量完成此执行功能,此功能完成后形成新的输入量,新输入量叠加于记忆单元,且记忆单元已经被遗忘门完成处理,进而形成新的记忆单元状态,然后完成点乘操作,点乘对象为均被非线性处理后的当前信息状态与记忆单元状态,记忆单元被非线性函数处理,完成点乘后得到隐含层的输出量。LSTM 结构图如图 4.3 所示,Sigmoid 层完成了输出功能,将 0 到 1 之间的概率值输出,描述出来每个部分的通过量,以 0 代表任务变量不被允许通过,以 1 代表所有变量可以通过。

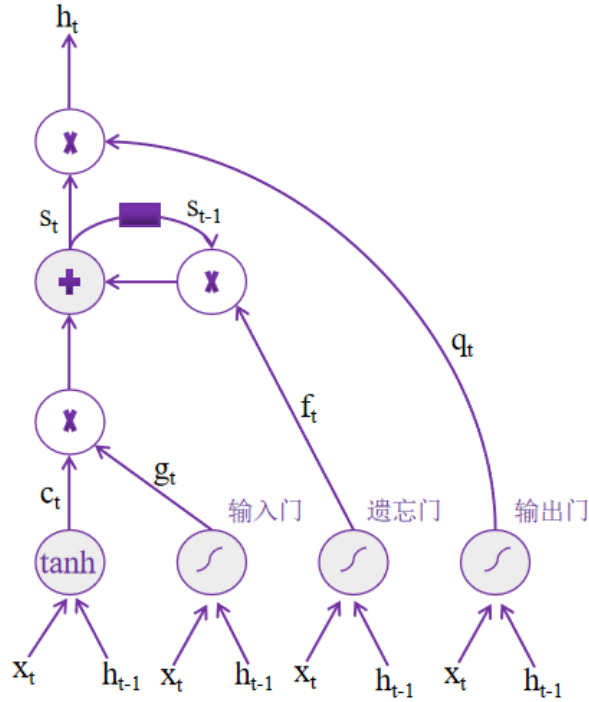


图 4.3 LSTM 网络结构图

图中  $x_t$  为  $t$  时刻数据输入， $h_t$  为  $t$  时刻的隐藏层输出。所有门控的输入  $x_t$  和  $h_{t-1}$  的组合， $h_{t-1}$  是  $t-1$  时刻的隐藏层输出。

sigmoid 函数  $\sigma(x)$  的计算公式为(4.3)

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (4.3)$$

遗忘门控  $f_t$  的计算公式为(4.4)

$$f_t = \sigma(b_f + U_f x_t + W_f h_{t-1}) \quad (4.4)$$

其中  $b_f$  为输入门的偏置， $U_f$  为输入门的输入权重， $W_f$  为输入门的循环权重。

采用的是  $\tanh$  函数达到非线性压缩输入数据，压缩后输入  $c_t$  计算公式为(4.5)

$$c_t = \tanh(b_c + U_c x_t + W_c h_{t-1}) \quad (4.5)$$

其中  $b_c$  为输入的偏置， $U_c$  为输入的输入权重， $W_c$  为输入的循环权重。

$s_t$  为网络的内部状态，是遗忘门控制的  $t-1$  时刻的内部状态和输入门控制的非线性

压缩输入  $c_t$  的和，其更新的公式为(4.6)

$$s_t = f_t s_{t-1} + g_t c_t \quad (4.6)$$

输出门为  $g_t$ ，其计算公式为(4.7)

$$g_t = \sigma(b_q + U_q x_t + W_q h_{t-1}) \quad (4.7)$$

其中  $b_q$  为输出门的偏置， $U_q$  为输出门的输入权重， $W_q$  为输出门的循环权重。

隐藏层输出为  $h_t$ ，隐藏层在输出前会对内部状态  $s_t$  进行非线性压缩，则  $h_t$  的计算公式为(4.8)

$$h_t = q_t \tanh(s_t) \quad (4.8)$$

最后利用 softmax 函数处理隐藏层输出，得到行人行为分类的概率。

LSTM 被广泛应用于机器视觉的识别预测中，其完成参数梯度、网络参数训练等方面具有良好效果，LSTM 利用通过时间的反向误差传播算子(BPTT),估计参数梯度,并训练网络参数。

#### 4.1.3 数据集准备

本文使用 Weizmann 数据集<sup>[70]</sup>完成行人行为识别实验，数据集中的动作共有 10 个，每个动作的完成会以多种不同样本形式表达，而且视频拍摄是以车辆视角拍摄而成，符合以车辆为主体的实际场景。本文完成的道路斑马线上的行人行为包含行走、奔跑、弯腰以及挥手等常规行为识别。视频经过预处理分成两部分，即训练集和测试集，训练集功能主要为构建模型和训练参数，而测试集功能则是检测已构建好的模型的识别精度。

图 4.4 所示为视频中的行走动作的连续视频帧。



图 4.4 行走动作图



#### 4.1.4 实验结果与分析

为了验证本文构建的 CNN-LSTM 模型能够有效完成行人行为识别功能以及模型的实用性,设计对比实验,对比模型为 ResNet-LSTM<sup>[71]</sup>模型,实验条件相同:采用同一个数据集,同样的数据量,在进行模型测试时采用数据集划分出的相同的随机动作视频,选用 50 帧,在各模型中输入相同的图像数据,对比结果如图 4.5 和表 4-1 所示。

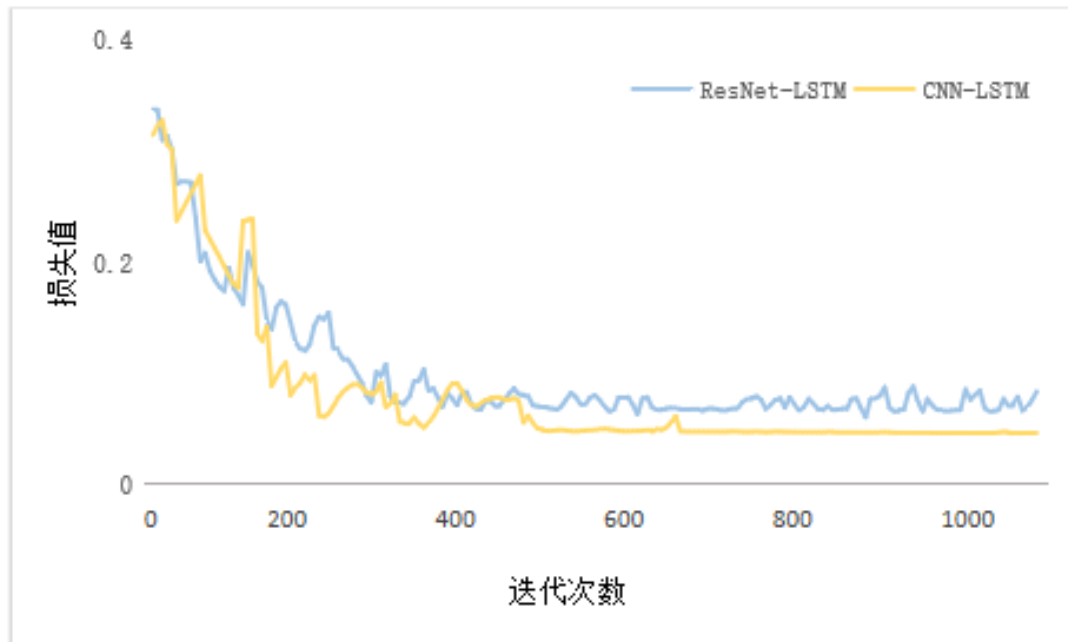


图 4.5 loss 曲线图

如图 4.5 所示为损失变化图,从图上可以比较两个模型之间得损失与精度之间存在的差异。从图中可以得出,训练过程中的 CNN-LSTM 模型在进行迭代了 450 次左右后趋于稳定,趋于稳定的速度也优于 ResNet-LSTM 模型,且 CNN-LSTM 模型波动较小,ResNet-LSTM 模型训练过程中波动较大,速度较慢,损失值相对 CNN-LSTM 模型较大一点,且在进行了相同迭代次数后,ResNet-LSTM 模型仍存在波动。

表 4-1 模型精确度表

模型	精确度
ResNet-LSTM	97.6%
CNN-LSTM	98.1%

表 4-1 中展示的是 CNN-LSTM 模型与 ResNet-LSTM 模型在 Weizmann 数据集上进行行为识别的精度对比,在经过训练后,将训练后的权重应用于测试集中,进行预测。从表中可以明显看出 CNN-LSTM 模型的识别准确率优于 ResNet-LSTM 模型的识别准

确率。因此, CNN-LSTM 模型可以应用于本实验的行为研究识别中, 为后面行人穿行斑马线意图预测提供基础。

## 4.2 基于贝叶斯网络的意图预测研究

行人的穿行斑马线意图影响着车辆的决策, 更影响交通安全。在智能交通中, 智能汽车若能够预知行人的穿行意图会大大减少事故的发生, 且通过各种技术实现预测行人穿行道路斑马线意图使汽车能够更加安全的驾驶, 增强交通安全性。本文将行人规定在斑马线区域内, 在完成行人检测与行人行为识别研究的基础上, 探究行人穿行道路斑马线意图预测。首先分析行人穿行道路斑马线意图预测的影响因素, 从行人与车辆两角度出发进行阐述, 然后设计本文的交通场景, 在完成贝叶斯网络的介绍后, 利用贝叶斯网络搭建本文行人穿行道路斑马线意图预测模型, 预测行人穿行道路斑马线意图。

### 4.2.1 影响因素分析

行人在穿行斑马线时受到多种因素的影响, 一方面, 行人存在自己的思想, 在穿行斑马线时会根据实时交通情况来决定自我行走要求; 另一方面, 智能车辆作为行人运动的旁观者, 不能直接通过自身感知获得行人行走意图, 只能借助于他物即行人姿态、动作以及行人周围的环境来获得行人可能的行走意图, 进而做出行驶判断。因此从行人与车辆两个角度完成行人意图预测的因素分析。

#### (1) 行人视角

从行人与车辆的交互场景中考虑得到影响行人穿行斑马线意图的三个因素:

①行人去往目的地, 行人目的地在斑马线对面时, 行人才会采取穿行斑马线的举措, 目的地时影响行人穿行的最重要因素。

②行人穿行时的危险等级, 行人在进行穿行时会对周围环境进行观察, 包括信号灯、车辆等环境, 确定没有任何危险才会进行穿行。

③行人当前时刻所处位置, 如果行人在斑马线周边即马路边上时, 会根据第二条中的危险等级穿行, 如果行人处于斑马线区域内, 行人会按照原本穿行模式继续运动。

#### (2) 车辆视角

车辆是行人的旁观者, 无法直观掌握行人的穿行斑马线意图, 但可以根据行人特征信息以及周围环境信息判断步行人穿行意图, 包括行人的具体动作, 比如奔跑、走动、

弯腰等；另外行人是否在斑马线区域内，若是在斑马线区域内注意避让行人；再者交通环境影响，例如信号灯、其他车辆等环境变量。

#### 4.2.2 交通场景设计

本文研究面向的是斑马线区域内的行人穿行场景，依据行人与车辆的因素影响设计本文交通场景。各场景设计如图 4.6 所示。

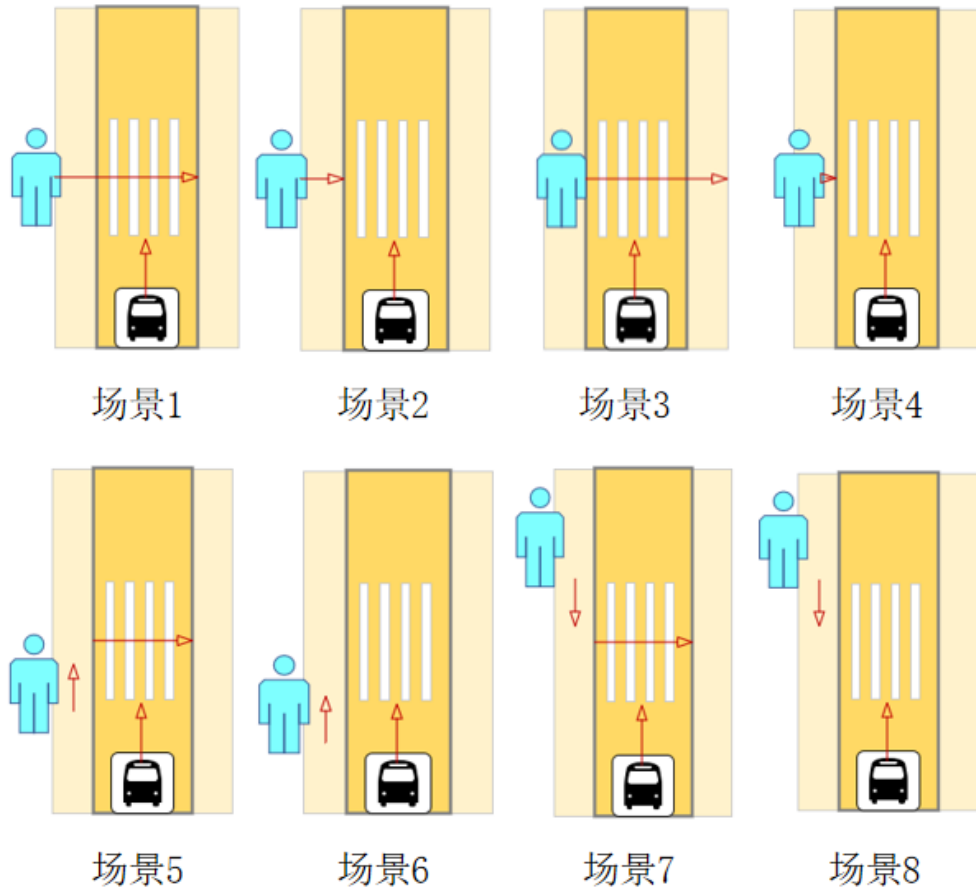


图 4.6 人车交互场景设计图

场景 1：行人先在马路边上运动，斑马线区域向斑马线对面方向运动，并穿行；

场景 2：行人先马路边上运动，斑马线区域向斑马线对面方向运动，未穿行；

场景 3：行人原本站在路边等待，突然穿行；

场景 4：行人只是等候在路边，并未穿行；

场景 5：行人与车辆同向而行，靠近斑马线，转向斑马线，穿行；

场景 6：行人与车辆同向而行，靠近斑马线，未穿行；

场景 7：行人与车辆相向而行，靠近斑马线，转向斑马线，穿行；

场景 8：行人与车辆相向而行，靠近斑马线，未穿行。

## 4.2.3 贝叶斯网络原理与构建

基于上文中提到意图预测影响因素,使用贝叶斯网络建立变量关系,推断概率分布,完成行人意图估计。贝叶斯网络 (Bayesian Network, BN) [72]作为一种概率图模型 (Probabilistic Graphical Model, PGD), 可以通过有向无环图(Directed Acyclic Graph, DAG)来表现。概率图模型以图表示变量概率检测依赖关系,将模型与概率论、图论结合,使用图表达与模型相关变量的联合概率分布。实际问题中,将数据中引号的知识进行挖掘,可通过模型构建图实现,具体实现方式为以观测结点表示观测到的数据,用隐含结点表示潜在的知识,用边来描述知识与数据的相互关系,最后基于这样的关系图来获得一个概率分布。在概率图中的节点分为隐含节点和观测节点,边可分为有向边和无向边。从概率论的角度,节点对应于随机变量,边对应于随机变量的依赖或相关关系,其中有向边表示单向的依赖或者说存在因果关系,无向边表示相互依赖关系。

贝叶斯网络的节点表示随机变量,有的可观测到,有的不能观测到。无环表示在信息流动过程中有确定方向的。当一事物发生促使另一事物更易发生,这时选用概率表达,且概率表达的是因果关系。贝叶斯网络主要模拟人推理过程中因果关系的不确定处理模型,贝叶斯网络结构图,作为有向无环图,其中节点表示变量,节点之间存在连接关系,表示网络的添加独立语义;二是节点之间的条件概率表,即概率值。若网络提供足够的条件概率值,满足所有联合概率计算,即网络可推理。

数学定义为令  $G=(I,E)$  表示一个有向无环图(DAG),其中图形中的节点的所有集合以  $I$  表示,有向连接线段的集合以  $E$  表示,令节点  $X=(X_i)_{i \in I}$  表示存在于有向无环图中的某一个节点  $i$  所能代表的随机变量,若将节点  $X$  所有的联合概率分配表示为式(4.9)

$$p(x) = \prod_{i \in I} p(x_i | x_{pa(i)}) \quad (4.9)$$

则  $X$  被称为是相对于有向无环图  $G$  的贝叶斯网络,其中的  $pa(i)$  则代表父节点即为“因”, $i$  则为子节点即为“果”。

如图所示为贝叶斯网络的 3 种结构形式,图 4.7 中各结构分别为 *head-to-head*, *tail-to-tail*, *head-to-tail*。

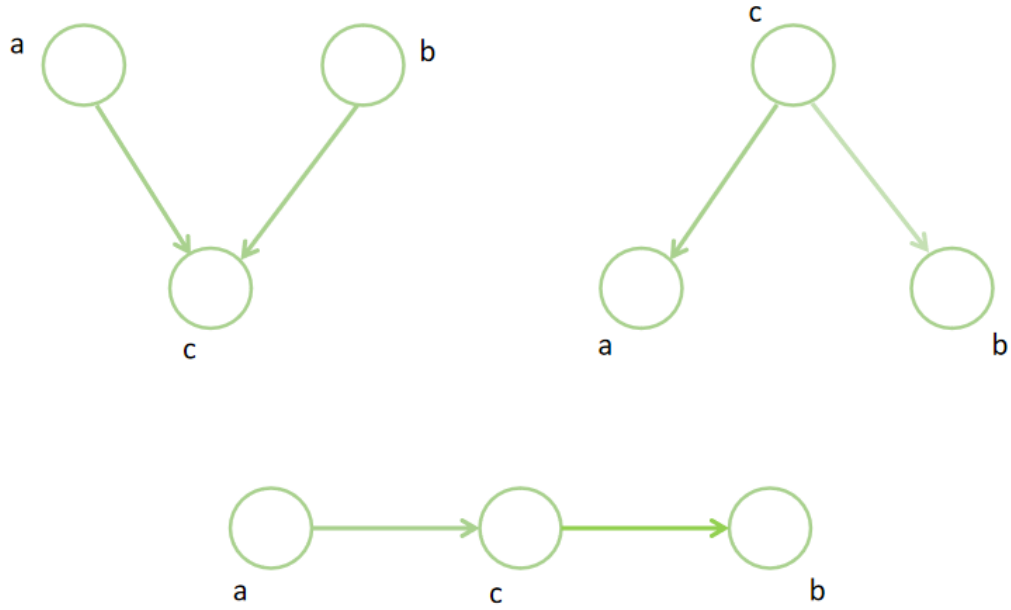


图 4.7 贝叶斯网络结构图

(1) *head-to-head* :  $P(a,b,c) = P(a) * P(b) * P(c|a,b)$  成立且  $c$  未知的情况下,  $a$ 、 $b$  被阻断,  $c$  为独立, 称之为 *head-to-head* 条件独立。

(2) *tail-to-tail* :  $c$  存在两种情况, 分别为已知和未知, 当  $c$  已知时  $a$ 、 $b$  为独立, 即  $P(a,b,c) = P(c) * P(a|c) * P(b|c)$  代入到  $P(a,b|c) = P(a,b,c) / P(c)$  中, 得出  $P(a,b|c) = P(a|c) * P(b|c)$ ; 当  $c$  未知时  $a$ 、 $b$  为不独立, 即  $P(a,b,c) = P(c) * P(a|c) * P(b|c)$  时无法得到  $P(a,b) = P(a)P(b)$ 。

(3) *head-to-tail* :  $c$  存在两种情况, 分为已知和未知, 当  $c$  未知时, 如上, 依然无法推出  $P(a,b) = P(a)P(b)$ ; 当  $c$  已知时, 存在  $P(a,b|c) = P(a,b,c) / P(c)$ , 将  $P(a,c) = P(a) * P(c|a) = P(c) * P(a|c)$  代入, 通过简化可以得到  $P(a,b|c) = P(a|c) * P(b|c)$ , 即可理解为  $c$  已知的情况下,  $a, b$  被阻断, 为独立, 称 *head-to-tail* 条件独立。

了解贝叶斯网络相关原理知识后, 构建贝叶斯网络模型进行行人穿行斑马线意图预测。本文研究中将行人视角与车辆视角的决策变量应用, 分别作为隐变量与观测变量, 构建贝叶斯网络, 如图 4.8 所示。

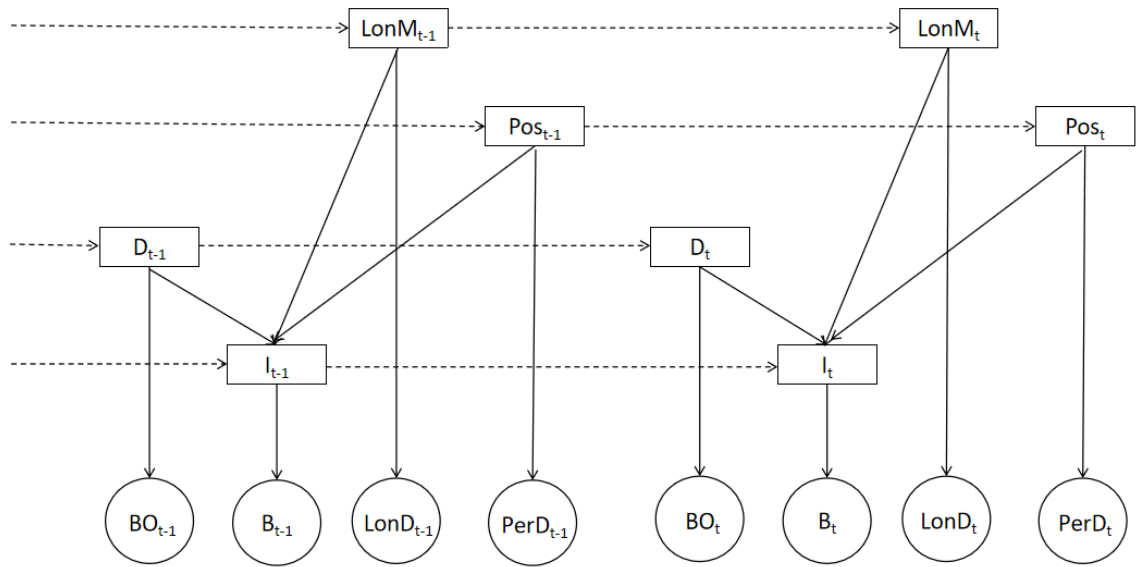


图 4.8 构建的贝叶斯网络模型图

图 4.8 中隐变量以矩形框表示，观测变量以圆表示。图中各符合存不同含义，表达如下：

$D$  表示目的地，定义行人目的地，隐变量表示行人的目的地不是道路对面即行人不到达所在斑马线尽头，离散变量表示行人目的地是对面即行人需到达所在斑马线的尽头，集合取值为 $\{0,1\}$ 。

$LonM$  表示纵向危险，定义存在于车辆的纵向运动方向上事故发生的概率，以行人可能遭遇事故为基础情况，隐变量表示不存在发生事故情况，离散变量表示存在发生事故情况，集合取值为 $\{0,1\}$ 。

$Pos$  表示位置，定义行人是否处在斑马线区域内，隐含量表示处在斑马线区域内，离散变量表示处在斑马线区域内。

$I$  表示穿行意图，定义行人是否要穿越斑马线，目的地到对面，隐变量表示行人要穿越斑马线，离散变量表示行人不穿越斑马线。集合取值为 $\{0,1\}$ 。

$BO$  表示行人身体朝向，定义行人身体朝着的方向，以角度表达，局限在车辆的坐标系中，变量取值区间为 $[0^\circ, 360^\circ)$ 。

$B$  表示行人行为， $I$  代表的穿行意图取值为 $\{S, W, R\}$ ， $S$  表示行走， $W$  表示站立， $R$  表示奔跑。

$LonD$  表示人车之间纵向相对距离， $LonM$  变量取值区间为 $[0, 40]m$ 。

$PerD$ 表示行人与车辆之间横向存在的相对距离,行人位置  $Pos$  的变量取值区间为 $[-10,10]m$ 。

## 4.3 实验结果与分析

### 4.3.1 数据集准备

目前对于行人意图预测<sup>[73]-[74]</sup>的研究较少,数据集不够丰富,目前致力于研究意图预测的数据集中使用较多的为Daimler<sup>[39]</sup>数据集和JAAD<sup>[73]</sup>数据集,本文选用后者,即JAAD数据集,此数据集在真实场景下由摄像机拍摄而成,在充分自然驾驶条件下获得,具有一定有效性,并且数据集中标有框来进行行人是否穿行识别,数据集中含有不在斑马线区域内的帧像,丰富了本实验的实用性。本数据集数据视频以时长为5s-10s为主,大约包含346个视频,速度为30FPS。总体来说,数据集中数据样本足够充分,以适应不同情况下的行人意图预测。数据集数据图片如图4.9所示。



图 4.9 数据集部分图片

### 4.3.2 实验分析

#### (1) 训练过程

以数据集中的样本完成训练,研究行为意图预测问题,使得模型符合本文要求。训练后得到如图的损失函数图,从图4.10可以得到,训练多次后损失趋于稳定,大约在0.1左右,模型识别效果达到最优。

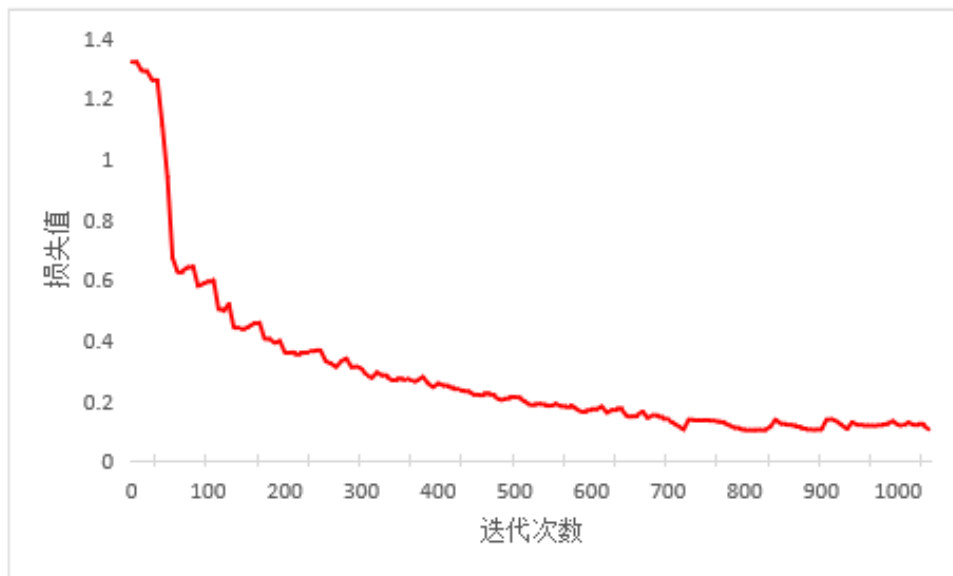


图 4.10 loss 曲线图

## (2) 实验结果分析

行人穿行斑马线意图预测算法检测行人穿行斑马线的不同场景存在表现差异，因此设计实验来对表现差异进行对比，对不同场景均选择一组实验数据且实验数据具有一定代表性，然后将数据下存在的行人穿行斑马线意图存在的表现差异变化进行总结。为达到更好进行在时间域比较的目的，可以定义零时间，零时间有两种：一种是行人在穿越斑马线的所有情景中行人进入斑马线的最前一个时间，另一种则是行人不穿越斑马线的所有情景中的最后一刻，即是车辆能在道路斑马线范围内采集到行人的最后一刻。在零时间相对时间的前后为正负，即前为负、后为正。如图 4.11 所示，行人在穿行斑马线的意图变化，随着时间的变化不同场景下存在不同的概率变化。

图 4.11 中分析得出，在设计不同场景中，行人穿行斑马线的所有场景行人穿行概率均收敛到 1，行人未穿行斑马线的所有场景穿行概率均可收敛到 0，明显表示所使用的行人穿行斑马线意图预测算法取得良好效果，大体上将行人穿行斑马线行为给予了准确分类。在行人未穿行斑马线的不同场景中，场景 4、场景 6、场景 8 在整个实验过程中穿行斑马线的意图概率都小于 0.5，场景 2 展示的是行人在最初有较高的穿行斑马线意图，即在开始时有较高的穿行概率，概率高的原因是最初时该场景中的交通条件较为安全，行人有向斑马线进行运动的趋势。在行人穿行斑马线的不同场景中，场景 3 展示的是行人具有不想穿行斑马线的最低意图，即有着最不理想的起步状态，但是本文行人穿行斑马线意图预测模型仍然给出了提前了 0.5s 的预判，预测了行人穿行斑马线的意图，而在其



他各种场景中，本文构建测行人穿行斑马线意图预测模型具备更好的行人穿行斑马线意图预判预测，预判时间均为 0.8s 以上。

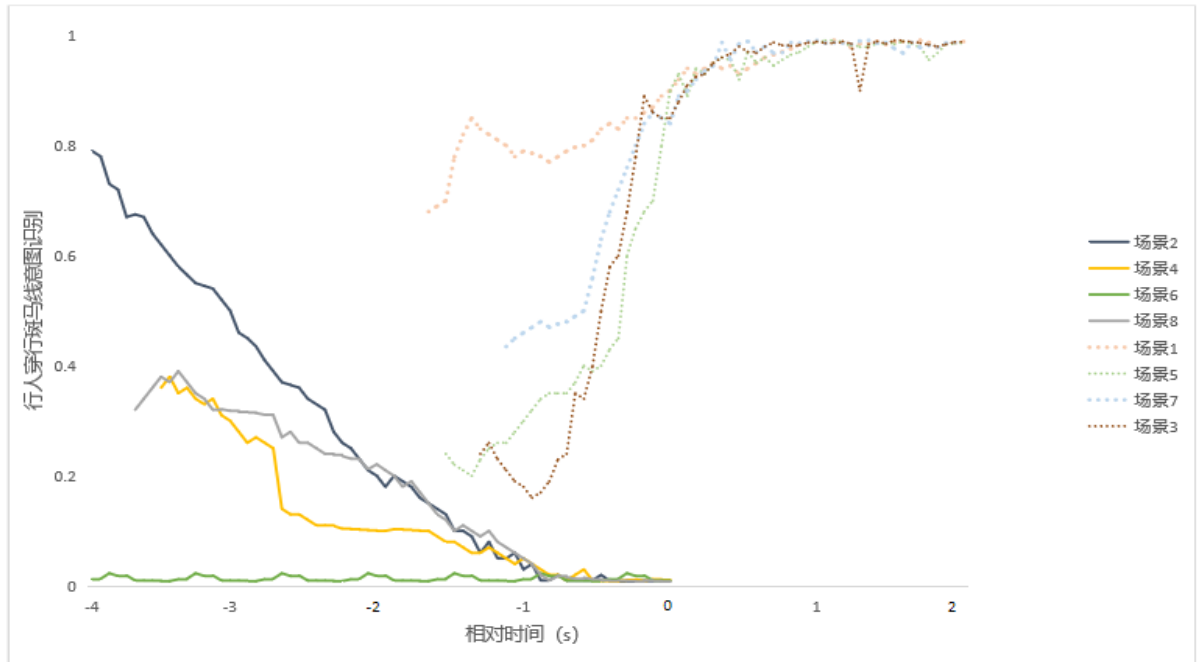


图 4.11 随相对时间行人穿行斑马线意图变化图

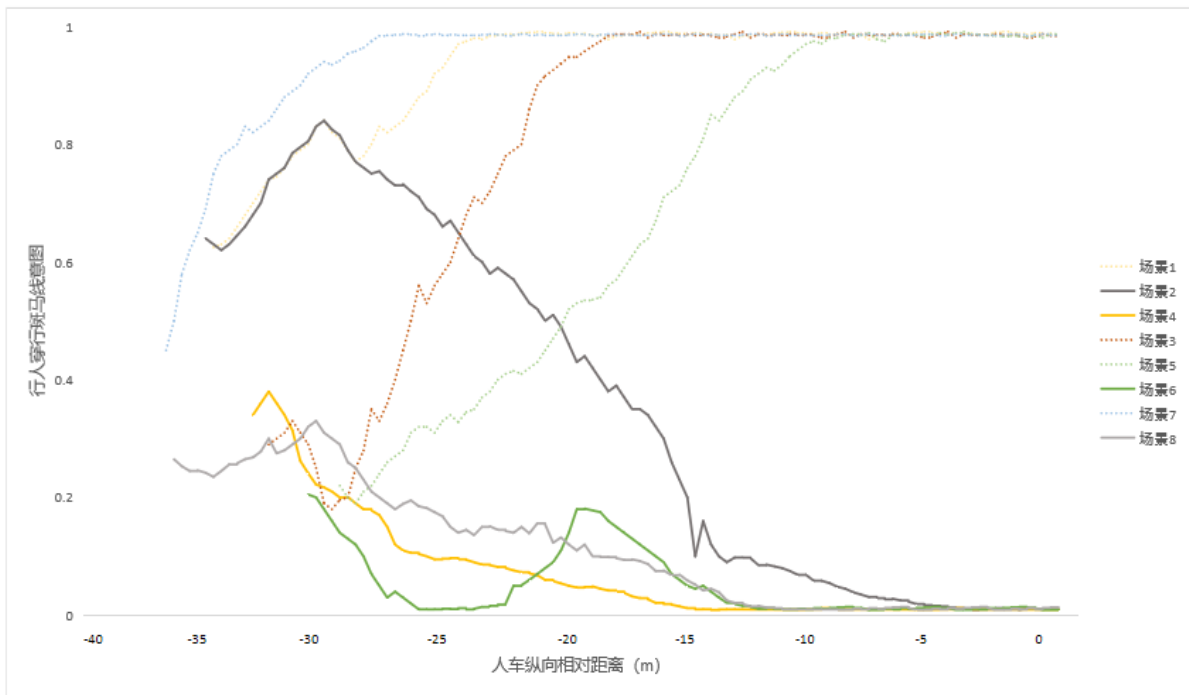


图 4.12 随纵向相对距离行人穿行斑马线意图变化图

图 4.12 中表达的是随着纵向相对距离在不断变化行人的穿行斑马线意图也在不断变化。不同场景中存在行人行为相似的场景，即场景 1、场景 2 的初始阶段，在两场景中的行人拥有极为相似的穿行意图，穿行概率均为 0.65 左右，因此接下来两人会有基本

相同的穿行意图变化曲线。模型预测行人穿行斑马线意图出现分歧的是相对距离为 27m 的时候，在这之后场景 1 穿行意图变化开始上升，场景 2 穿行意图变化继续下降，对照图 4.11，这时与行人实际穿行斑马线时间相比，相对时间约为 0.6s。对于模型预测的场景 3、场景 4，两场景在图中显示是在相对距离为 28m 时产生分歧，行人穿行斑马线意图预测提前了 0.5s，即前文所说提前时间。对于其他设计场景中的行人意图情况，行人的穿行斑马线意图概率均在 0.5 以下并一直维持，但是场景 5、场景 7 在实验中模型预判时间均为 0.8s 给出，预判行人的穿行概率在 0.5 以上。

结合实验结论，在各种行人穿行斑马线的场景中，本文构建的贝叶斯网络行人穿行斑马线意图预测模型能够有效预判行人穿行斑马线意图，且会提前 0.5s-0.8s 将意图预测，相对距离大约是 22m 的时候就可以将行人穿行斑马线的穿行意图预测出来；在各种行人不穿行斑马线的场景中，本文的预测模型也能够完成行人的不穿行斑马线意图预测，相对距离为 15m 到 25m 内预判出来。通过实验证明，该设计模型具有一定的实用性，可用于构建道路斑马线与行人运动检测系统，给人车交互场景中的车辆保障行人安全提供了决策时间。

#### 4.4 本章小结

本章首先进行行人行为识别研究，选取使用 CNN-LSTM 模型进行行为识别，先分别介绍了 CNN 算法与 LSTM 算法，然后使用 Weizmann 数据集进行行人行为识别实验。使用 CNN-LSTM 模型与 ResNet-LSTM 模型进行了实验对比，CNN-LSTM 模型的损失值较小，趋于稳定的速度优于 ResNet-LSTM 模型，模型波动较小，对于行为识别准确率 CNN-LSTM 模型更高。接着在行人行为识别研究的基础上，构建本文贝叶斯网络预测行人穿行斑马线意图模型，完成行人意图预测研究。先分析了影响意图预测的因素，然后设计了本文斑马线区域的人车交互场景，在详细介绍贝叶斯网络原理的基础上构建了本文的贝叶斯网络模型，使用 JAAD 数据集进行训练与推理，得到行人穿行道路斑马线意图预测结论。实验采用的算法均能进行良好实验并能应用于实际中，具有更好的精确性与实用性。

## 第五章 道路斑马线与行人运动检测系统设计与实现

本章在上述章节研究的基础上设计了道路斑马线与行人运动检测系统，进而更好的验证本文算法的实用性。该系统主要包含斑马线检测模块、行人检测模块、行人穿行斑马线意图预测模块以及同时进行信息提示。为了系统能够更好的应用于真实生活场景中，将本文中的改进算法融入系统，完成并展示道路斑马线与行人运动检测系统的功能，为无人驾驶与辅助驾驶提供帮助。

### 5.1 需求分析

人车交互场景中常常伴有未知的危险，行人在交通场景中的行为往往具有不可预知性，且行人在行走过程中不会严格遵守交通规则和既定规范，具有发生各种交通事故的危险性，因此识别行人穿行意图在智能交通的发展中尤为重要。随着科技的发展，无人车驾驶与辅助驾驶的出现可以更好的帮助避免交通事故的发生，帮助人们避免安全问题。目前关于行人穿行斑马线意图预测的研究较少，距离应用于智能交通还有一段距离。当然，道路斑马线与行人运动检测研究在智能驾驶是非常重要的一部分，因此，研究道路斑马线上的行人运动检测对于无人车驾驶与辅助驾驶具有非常重要的现实意义。

目前，斑马线检测和行人检测技术受到各种复杂因素的困扰，如背景复杂、行人数量多且目标小、各种遮挡情况以及光照环境的变化等因素，行人行为识别以及意图预测更是受到行人姿态的多样性、行人意图不可知性的困扰，现实场景中更是容易发生各种不可预知的变化，这些因素大大影响了道路斑马线上行人运动检测研究的难度。因此，研究出更高水准的斑马线检测与行人检测技术、能够应对各种行人行为变化以及不可预知的意图行为预测系统是向无人驾驶走出的巨大一步。

最后实现的道路斑马线与行人运动检测系统主要功能实现是以下三个，分别为斑马线检测功能、行人检测功能、行人行为识别后的行人穿行斑马线意图预测功能并将各功能中的识别信息进行文字提示。视频导入到本文设计的道路斑马线与行人运动检测系统中时，可以同时完成对视频的截帧处理，截取后的图像进行相应功能实现，即进行斑马线检测以及行人检测，对目标行人进行动作分析，进而预测行人穿行斑马线意图，最后将得到的穿行斑马线意图预测信息传递到车辆中，辅助车辆完成驾驶。

## 5.2 开发环境与开发工具

### 5.2.1 开发环境

本系统基于 Pycharm 平台使用 Python 语言进行开发，系统中图像处理使用 OpenCV 库算法，各页面使用 QT 设计，实现各网络模型关于斑马线检测、行人检测、行为识别以及行人穿行斑马线意图预测的功能。

本文所有实验均在如表 5-1 所示的实验平台中进行。

表 5-1 实验平台

名称	相关配置
CPU	Intel(R) Core(TM) i5-10400F CPU @ 2.90GHZ 2.90GHZ
内存	16GB
GPU	NVIDIA GeForce GTX 1660 SUPER
GPU 加速库	CUDA10.0+CUDNN7.4.1
操作系统	Win10
数据处理	Python3.6 + tensorflow1.14+pytorch1.2
系统界面	PyQt5

### 5.2.2 开发工具

#### (1) Python

Python 语言具有简单容易掌握学习、软件支持库丰富等优点，被广泛应用于各种编程序设计，例如 Web 开发、数据分析与计算、人工智能等多个领域发展中，而且 python 的可扩展性大，非常全面的应用于各种场景中，另外 Python 语言中提供了非常多的工具和接口，方便了程序员进行语言的扩充与使用。

#### (2) OpenCV

OpenCV 是使用 C++或者 C 语言开发致力于处理图像和计算机视觉的各种算法的集合，它不依赖于外部库，拥有多数量的 API，为用户提供了透明接口，方便用户使用。OpenCV 在当前致力于实时应用，其执行速度满足用户需求且在不断提升，相比其他函数库具有更好的性能。

#### (3) QT

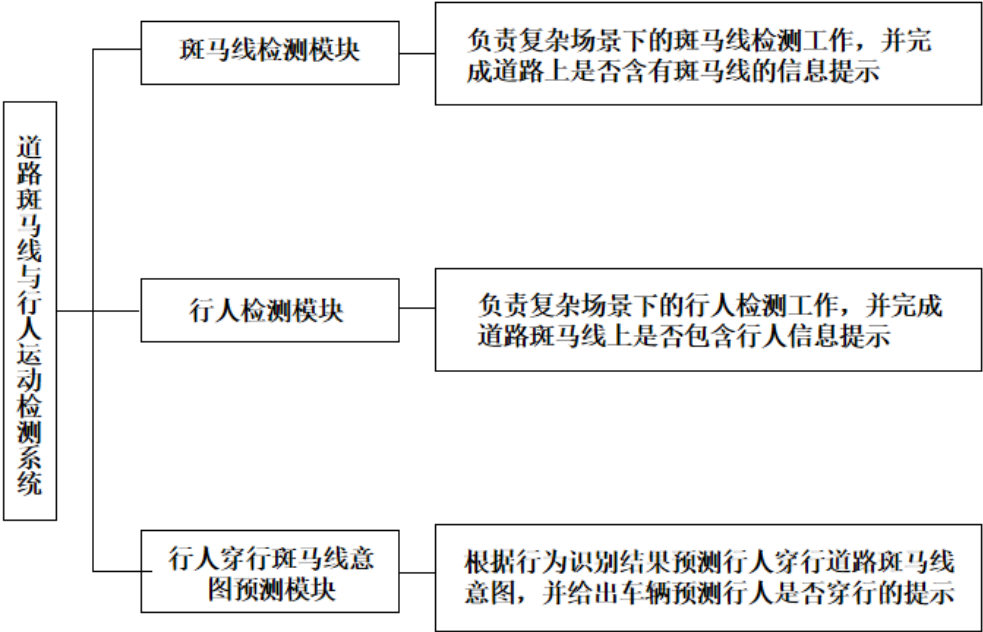
QT 致力于构建快速响应、性能高的应用程序，其不光是 GUI 工具包更提供了能在多种领域跨各种平台进行开发的模块，支持所有主流软件开发平台，并且将平台的特有图形界面设计风格体现得淋漓尽致。另外，QT 的封装机制较好，拥有较高的模块化，各

元件之间协同配合，顺利完成用户的指定操作，使得用户使用方便简单。

5.3 系统设计与实现

5.3.1 系统设计

本系统根据上述研究基础，开发设计了基于深度学习的道路斑马线与行人运行检测系统。该系统首先将选择好的视频导入，然后视频截帧处理并对截取到的图像进行图像处理，对处理后的先进性斑马线检测，归属于斑马线区域后进行行人检测，无论存不存在行人都会进行驾驶提示，当检测到行人后对行人行为进行识别判断行人穿行斑马线意图。基于深度学习的道路斑马线与行人运动检测系统架构功能如图 5.1 所示。



5.3.2 功能设计与实现

本系统依据设计的功能，结合图像处理算法，使用深度学习的网络训练模型，经过实验对比分析后，将各算法设计实现到系统中，用户可以根据自己的需求，完成处理步骤，进行程序验证。

（1）登陆注册界面

如图 2.5 所示为道路斑马线与行人运动检测系统登陆界面与注册界面，保证了用户使用系统的安全性。



图 5.2 登陆注册界面

## (2) 系统主界面模块

基于深度学习的道路斑马线与行人运动检测系统主界面如图 5.3 所示，基于深度学习的道路斑马线与行人运动检测系统主要分为两个区域，左侧区域为操作区域，用户可根据自己的需求进行操作，可操作视频导入，然后视频经过处理后选择其他目的操作，操作区域主要有三项操作按钮，分别为斑马线检测、行人检测、行人穿行意图预测，因此该系统主要有三个主要功能：斑马线检测、行人检测以及穿行意图预测。点击按钮时会完成相应视频的截图，将截帧图像显示在右侧上方，即作为输入图像，功能的实现图像显示在右侧下方即作为输出图像输出结果，且同时将结果进行分析在提示栏里对应位置提醒车辆，通过点击各功能按钮可完成相应功能的实现。



图 5.3 道路斑马线与行人运动检测系统主页面

### （3）斑马线检测模块

如图 5.4 所示，斑马线检测按钮实现的主要是对视频中的斑马线进行检测，主要采取点击“斑马线检测”功能按钮的方式，将视频截图图像输入到右侧原始图像区域框中，进行斑马线区域检测，结果为原始图像显示在右侧区域的上方，检测斑马线图像显示在右侧区域的下方，且在检测到斑马线时系统的最右侧提示框中会给出提示“请注意，前方有斑马线！”的字样，提醒车辆减速慢行，注意周围环境安全。



图 5.4 斑马线检测功能实现界面

### （4）行人检测模块

如图 5.5 所示，行人检测按钮功能的实现主要是对视频中的行人进行检测，采取点击“行人检测”功能按钮的方式，视频截取图像同样显示在右侧区域上方，进行行人的检测，且对截屏图像进行检测时可检测全部行人，即若行人不在斑马线区域内也被精准检测，若有行人在斑马线周围区域内系统会给出提示信息，在相应提示框中显示“请注意，斑马线周围有行人！”的字样，达到保证行人安全的目的。



图 5.5 行人检测功能实现界面

#### (5) 行人穿行道路斑马线意图预测模块

如图 5.6 所示，为行人穿行道路斑马线意图预测模块界面图。



图 5.6 行人穿行意图预测界面

行人穿行斑马线意图预测的功能按钮是“穿行意图预测点击”，采用点击穿行意图预测按钮的方式，经截取图像先完成行为识别，再依据行为识别结果判断行人穿行意图，



输出时在检测到斑马线与行人的基础上再给出其他提示,其中一种情况输出为即“请注意,行人正在穿行斑马线!”,使车辆提前采取相应安全措施。

## 5.4 本章小结

本章首先介绍了基于深度学习的道路斑马线与行人运动检测系统的需求,然后详细介绍了该系统开发所需的相关技术,最后设计实现相关功能,通过导入目的视频,将视频处理后进行行人穿行斑马线意图预测,系统展示了关于斑马线检测、行人检测和行人穿行意图预测等功能,最终通过行为识别将行人穿行斑马线意图进行提示,系统的实现帮助无人驾驶与辅助驾驶汽车在实际交通场景中规避交通事故的发生。

## 总结与展望

### 总结

本文基于斑马线检测、行人检测与意图预测三方面的研究现状基础上,对其存在的问题进行了分析,为了解决三方面各自存在的问题,分别对斑马线检测、行人检测以及行人穿行斑马线意图预测进行研究。分析研究传统检测算法与深度学习方法,根据使用在斑马线检测与行人检测实验中的算法模型存在的问题缺陷,对算法模型进行了改进,为后文中行为识别与行人穿行斑马线意图预测做铺垫,利用 CNN-LSTM 模型完成了行人行为识别研究,意图预测研究中设计了本文的人车交互场景,综合考虑场景中存在的影响因素,构建了本文的贝叶斯网络模型预测行人穿行道路斑马线意图,具体工作完成如下:

(1) 关于斑马线检测研究。主要采用语义分割方法完成斑马线检测工作,分别使用 SegNet 模型、U-Net 模型以及 PSPNet 模型完成检测实验,选取检测效果较好的 SegNet 模型完成改进工作,改进模型的特征提取网络部分融入金字塔池化模块,达到获取全局信息的目的;结构改为编码不变,解码减少的不对称结构,减少网络参数。本文基于实际要求,对获取到的斑马线图像进行亮度增强、添加噪声、图像翻转三种方式来扩充采集的斑马线图像样本,使斑马线图像样本满足各种语义分割模型检测需求。本文提出的改进 SegNet 模型与原始 SegNet 模型、U-Net 模型和 PSPNet 模型相比,改进 SegNet 模型检测斑马线精度更高,且运行速度加快,满足当前实时性检测的需求,解决原始语义分割方法模型中存在的检测精度不高、检测不全面、实时性差的问题。

(2) 对于行人检测研究。从传统行人检测与深度学习方法检测两方面完成研究工作,传统行人检测使用 HOG+SVM 方法完成检测工作,但是传统行人检测中对于背景复杂情况无法完成检测并且存在检测不全面、检测精度低、训练耗时,计算量大、实时性差等问题,因此引入深度学习方法,选用 SSD 算法与 Faster-RCNN 算法完成实验,对比两算法的行人检测效果后,在行人检测效果更好的 Faster-RCNN 算法上完成改进,替换特征提取结构,将 CNN 改为 ResNet 网络,引入高斯聚类算法,节省了算法计算时间,解决原始深度学习算法中存在的训练耗时、小目标检测困难的问题。

(3) 对于行为识别研究。本文使用 CNN-LSTM 模型完成行人行为识别实验,与

ResNet-LSTM 模型进行实验对比,验证本文选用模型的实用性,从实验中可以得到 CNN-LSTM 模型的识别准确率优于 ResNet-LSTM 模型的识别准确率,可以使用于本文系统设计中。

(4) 对于行人穿行斑马线意图预测研究。分析了意图预测的影响因素,设计了本文交通场景,在介绍了贝叶斯网络模型的原理知识后,构建本文的贝叶斯网络完成行人穿行斑马线意图预测研究实验结果显示,本文构建的贝叶斯网络能够很好的预判行人的穿行斑马线与不穿行斑马线意图,在各种穿行场景中可达到 0.5s-0.8s 的预判时间,在相对距离为 15m-25m 内可完成预判。

(5) 道路斑马线与行人运动检测系统设计。综合斑马线检测、行人检测、行为识别以及行人穿行斑马线意图预测的算法研究,构建了本文道路行人斑马线与行人运动检测系统。系统通过 QT 展示了本文设计系统的各界面功能,并将实验结果进行了展现,表明本系统具有一定应用能力。

## 展望

本文以斑马线检测与行人检测为基础,以行为识别为辅助,预测了行人穿行斑马线意图,研究了关于这四个方面工作的算法,并取得了实质性进展。但自己学问水平有限,关于基于深度学习的道路斑马线与行人运动检测算法的研究存在诸多不足,需要继续研究与改进,主要为下面几个方面:

(1) 由于本文研究的框选性,只能应用于道路斑马线区域范围内的行人穿行斑马线意图预测,对于其他场景预测具有很大局限,应尽量将其完善,应用于多场景下的行人穿行行为预测。

(2) 本文研究的行为识别的基于行人检测基础上进行识别实验,但是行人行为识别与行人检测是两个相互独立的个体,并没有真正进行融合,可以在以后研究两者的结合以便于达到更精准的对应检测行人的行为识别的目的。

(3) 本文研究的行人穿行斑马线意图预测是基于个人的穿行预测,并不能够广泛应用于多人场景,如何将个人的穿行意图预测推行至多人穿行意图预测仍需要继续探究。

## 参考文献

- [1] 中华人民共和国道路交通事故统计年报（2020 年度），公安部交通管理局（内部资料）。
- [2] 韩晔彤. 人工智能技术发展及应用研究综述[J]. 电子制作, 2016(12): 95.
- [3] Uddin M S, Shioyama T. Measurement of the length of pedestrian crossing anavigational for blind people[C]. The 7th International IEEE Conference on Intelligent Transportation Systems. IEEE, 2004: 690—695.
- [4] Uddin M S, Shioyama T. Measurement of the length of pedestrian crossing length using vector geometry-an image based technique [C]. Symposium on Circuits & Systems. IEEE, 2004: 229—232.
- [5] Sichelschmidt S, Haselhoff A, Kummert A, et al. Pedestrian crossing detecting as a part of an urban pedestrian safety system[J]. IEEE, 2010: 840—844
- [6] 曹玉珍, 刘刚, 杨海峰. 导盲系统中的道路斑马线识别方法[J]. 计算机工程与应用, 2008, 44(15): 176-178.
- [7] 戴燕. 图像边缘检测与应用[D]. 西安: 西安科技大学, 2010.
- [8] 黄时杰, 曾建华. 图像边缘检测与分割的 MATLAB 实现[J]. 赤峰学院学报(自然科学版), 2019, 35(9): 47-49.
- [9] 卢旭, 刘钊. 基于深度学习的图像语义分割技术综述[J]. 软件导刊, 2021, 20(1): 242-244.
- [10] 王嘉, 张楠, 孟凡云, 等. 基于金字塔场景分析网络改进的语义分割算法[J]. 计算机工程与应用, 2021, 57(19): 220-227.
- [11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 3431-3440.
- [12] Badrinarayanan V, Endall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. IEEE, 2017, 39(12): 2481-2495.
- [13] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. International Conference on Medical Image Computing and Computer-

- Assisted Intervention. IEEE, 2015, 93(51): 234-241.
- [14]黄生鹏, 柳海南, 周克帅, 等. 基于改进 Unet 的斑马线分割[J]. 智能计算机与应用, 2020, 10(11): 61-64,69.
- [15]Long X, Zhang W, Zhao B. PSPNet-SLAM: A Semantic SLAM Detect Dynamic Object by Pyramid Scene Parsing Network[J]. IEEE, 2020, 8:214685-214695.
- [16]施浩. 基于金字塔池化和注意力机制的图像语义分割[D]. 南昌: 南昌大学, 2021.
- [17]崔剑, 侯晓荣. 基于多尺度特征近似计算的行人检测方法[J]. 电脑与电信, 2015(4): 50-52+65.
- [18]Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns[J]. IEEE Transactionson Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
- [19]Viola P, Jones M J, Snow D. Detecting pedestrians using patterns of motion and appearance[J]. International Journal of Computer Vision, 2005, 63(2): 153-161.
- [20]Viola P, Jones M J. Robust real-time face detection[J]. International journal of computer vision, 2004, 57(2): 137-154.
- [21]Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2005: 886-893.
- [22]刘方园, 王水花, 张煜东. 方向梯度直方图综述[J]. 计算机工程与应用, 2017, 53(19): 1-7+25.
- [23]Wang X Y, Han T X, Yan S C. An HOG-LBP human detector with partial occlusion handling[C]. 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 32-39.
- [24]姚立平, 潘中良. 基于改进的 HOG 和 LBP 算法的人脸识别方法研究[J]. 光电子技术, 2020, 40(2): 114-118+124.
- [25]Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2008, 1-8.
- [26]孙锐, 陈军, 高隼. 基于显著性检测与 HOG-NMF 特征的快速行人检测方法[J]. 电子与信息学报, 2013, 35(08): 1921-1926.

- [27] 谭飞刚. 基于特征融合的行人检测研究[D]. 长沙: 长沙理工大学, 2013.
- [28] Dollár P, Tu Z, Perona P, et al. Integral channel features[C]. British Machine Vision Conference. BMVA, 2009: 91.
- [29] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2): 365-386.
- [30] 王晓刚. 对机械安全技术现状的分析及发展趋势展望[J]. 科技创新与应用, 2014, 000(007): 94-94.
- [31] Sermanet P, Kavukcuoglu K, Chintala S, et al. Pedestrian detection with unsupervised multi-stage feature learning[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2013: 3626-3633.
- [32] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 779-788.
- [33] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]. Computer Vision – ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [34] 王宇宁, 庞智恒, 袁德明. 基于 YOLO 算法的车辆实时检测[J]. 武汉理工大学学报, 2016, 038(10): 41-46.
- [35] 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. IEEE, 2014: 580-587.
- [36] Ren S Q, He K M, 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, 2017, 39(6): 1137-1149.
- [37] 薛超, 于宏志, 王景彬. 基于卷积神经网络的级联人脸检测[J]. 中国安防, 2017(11): 93-96.
- [38] 宋焕生, 张向清, 郑宝峰, 等. 基于深度学习方法的复杂场景下车辆目标检测[J]. 计算机应用研究, 2018, 35(04): 1270-1273.
- [39] Nicolas S, Gavrilă D M. 35th German Conference on Pattern Recognition, Saarbrücken, 2013[C]. German: Springer, 2013: 174–183.

- [40] Keller C G, Gavrilu D. Will the Pedestrian Cross? A Study on Pedestrian Path Prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 15(2): 494-506.
- [41] Rehder E, Kloeden H. Goal-Directed Pedestrian Prediction[C]. 2015 IEEE International Conference on Computer Vision Workshop. IEEE, 2015.
- [42] Karasev V, Ayvaci A, Heisele B, et al. Intent-aware long-term prediction of pedestrian motion[C]. IEEE International Conference on Robotics and Automation. IEEE, 2016:2543-2549.
- [43] Völz B, Behrendt K, Mielenz H, et al. A data-driven approach for pedestrian intention estimation[C]. international conference on intelligent transportation systems, IEEE, 2016, 2607-2612.
- [44] Fang Z J, David V, Antonio M L. On-Board Detection of Pedestrian Intentions[J]. Sensors, 2017, 17(10), 2193.
- [45] 李贤. 基于行人意图识别的智能车辆决策研究[D]. 大连: 大连理工大学, 2019.
- [46] Saleh K, Zeineldin R A, Hossny M, et al. Navigational path detection for the visually impaired using fully convolutional networks[J]. 2017 IEEE International Conference on Systems, Man, and Cybernetics. 2017:1399-1404.
- [47] 钟涛. 基于视频处理的行人过街意图识别技术研究[D]. 长春: 吉林大学, 2019.
- [48] Hoy M, Tu Z, Dang K, et al. Learning to Predict Pedestrian Intention via Variational Tracking Networks[C]. 2018 IEEE International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018: 3132-3137.
- [49] 程焕新, 蒋泽芹, 成凯. 基于改进 SegNet 模型的斑马线图像语义分割[J]. 电子测量技术, 2020, 43(23): 104-108.
- [50] 邓天民, 王琳, 杨其芝, 等. 基于改进 SegNet 算法的车道线检测方法[J]. 科技与工程, 2020, 20(36): 14988-14993.
- [51] 周豹, 李振伟, 李彦玥, 等. 不均匀光照下斑马线的处理与识别[J]. 电子设计工程, 2020, 28(21): 168-172.
- [52] 赵娟. 斑马线检测与识别算法研究[D]. 武汉: 武汉轻工大学, 2015.
- [53] 白松让. 无信号灯斑马线检测及行人识别研究[D]. 辽宁: 辽宁工业大学, 2021.
- [54] 任小康, 陈鸿享. 基于 HOG+SVM 实现行人检测[J]. 科学技术创新, 2021(06): 73-74.
- [55] Platt J. Sequential minimal optimization: a fast algorithm for training support vector

- machines[J]. Advances in Kernel Methods-Support Vector Learning, 1998, 208: 1-21.
- [56]胡洁. 基于深度学习的车辆检测方法研究[D]. 广西: 广西大学, 2021.
- [57]冯婷婷. 基于 SSD 的行人检测算法研究[D]. 上海: 东华大学, 2021.
- [58]刘盼. 一种改进 SSD 的行人检测方法研究[D]. 西安: 西安理工大学, 2020.
- [59]Hartigan J A, Wong M A. Algorithm as 136: a K-means clustering algorithm[J]. Applied Statistics, 1979, 28(1): 100.
- [60]Dollar P, Wojek B C, Schiele, et al. Pedestrian Detection: An Evaluation of the State of the Art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(4): 743–761.
- [61]谭宇志. 基于深度学习的行人检测算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.
- [62]刘曦. 复杂道路场景下的行人检测方法研究[D]. 重庆: 重庆邮电大学, 2020.
- [63]杨萌月. 基于 Faster RCNN 的道路行人检测方法研究[D]. 西安: 西安科技大学, 2020.
- [64]黄继鹏, 史颖欢, 高阳. 面向小目标的多尺度 Faster-RCNN 检测算法[J]. 计算机研究与发展, 2019, 56(02): 319-327.
- [65]车领. 智能车辆对行人的运动感知和过街意图预测研究[D]. 西安: 西安理工大学, 2019.
- [66]吕天齐. 基于智能感知的行为识别分析研究[D]. 北京: 北京邮电大学, 2021.
- [67]高信凯. 基于图卷积网络的人体行为识别和预测研究[D]. 北京: 北京邮电大学, 2021.
- [68]Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Comput, 1997, 8: 1735-1780.
- [69]Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the national academy of sciences, 1982, 79(8): 2554-2558.
- [70]Blank M, Gorelick L, Shechtman E, et al. Actions as Space time Shapes[J]. Proceedings of the International Conference On Computer Vision, 2005, 2: 1395-1402.
- [71]窦雪婷. 基于 ResNet-LSTM 的行人过街行为识别方法[J]. 计算机与数字工程, 2021, 49(09): 1872-1877.
- [72]Pearl J. Fusion, propagation and structuring in belief networks[J]. Artificial Intelligent, 1986, 29(03): 241-288.



- [73] 蒋海洋. 基于机器视觉的行人检测与行为意图识别研究[D]. 武汉: 武汉理工大学, 2020.
- [74] 王礼坤. 路侧行人运动预测方法研究[D]. 北京: 清华大学, 2018.
- [75] Rasouli A, Kotseruba I, Tsotsos J K. Are They Going to Cross? A Benchmark Dataset and Baseline for Pedestrian Crosswalk Behavior[C]. IEEE International Conference on Computer Vision Workshop. IEEE, 2017: 206-213.

## 攻读学位期间取得的研究成果

### 攻读学位期间公开发表的论文:

- [1] 付阳阳,陶建军,王夏黎,李妮妮,袁绍欣.基于改进 SegNet 模型的斑马线检测方法研究[J].公路交通科技,2022,39(04):117-122+149.
- [2] 李妮妮,王夏黎,付阳阳,郑凤仙,何丹丹,袁绍欣.一种优化 YOLO 模型的交通警察目标检测方法[J].图学学报,2022,43(02):296-305.
- [3] 郑凤仙,王夏黎,何丹丹,李妮妮,付阳阳,袁绍欣.单幅图像去雾算法研究综述[J].计算机工程与应用,2022,58(03):1-14.

### 攻读学位期间参与的科研项目:

- [1] 国家重点研发计划项目.自主式交通复杂体系架构研究.项目编号:2020YFB1600400.
- [2] 浙江省交通运输厅.浙江省交通运输厅科技计划项目.项目编号:2020026.

## 致谢

行文至此，百感交集，既往种种，历历在目，学生生涯落下帷幕，一路跌跌撞撞走到现在，些许遗憾，些许不舍，些许憧憬，但依然对未来充满热情，重新开始不一样的旅程。回首在长安大学的三年，感恩出现在我身边的每一个人，善意与温暖始终包围着我，拼凑了我的幸福画卷。祝福你们的未来如花般灿烂。

一朝沐杏雨，一朝念师恩。感谢我的指导教师王夏黎老师，从最终的选题到最终成文，感谢您陪我字斟句酌，倾尽所能的点播和指导我。三年的求学之路，不论是传道授业、未来规划还是生活琐事，感谢您的体谅、包容与关爱。回首开会的点点滴滴、谆谆教诲，都是您在学术上对我的辛勤栽培与耐心教导，让我养成了一丝不苟的研学习惯，使我面对未来的挑战，愈发严谨、稳重、坚韧。感谢学院的每一位老师呕心沥血的教授，给予我们追逐理想的勇气。

愿岁并谢，与友长兮。感恩我的同门妮妮、丹丹和凤仙，是他们给了我很多帮助，陪我走过了这一段青春岁月。时间飞逝，一起学习、一起做实验、做项目的情景仿佛就在昨天，感谢你们对我的帮助，让我这种学渣也能顺利毕业，完成了各项试验、各项任务，那些欢声笑语依然回荡在耳边。舍友壮壮也是我的感谢对象之一，当我在生活上遇到问题的时候，壮壮会帮助我，给我支持与建议，虽然会有矛盾存在，但依然相爱，两人宿舍的快乐提前体验。

在此，更感谢父母，无需多说，感激涕淋。

最后，文章至此，所有都已画上句号，面对未来更加不知所措，但依然保持着一份热忱，一份勇气，一份纯真，走在生活的路上，踏上人生新的征程。