硕 士 学 位 论 文

复杂交通场景中各智能体多模概率化轨迹预测方法研究

Study on Vehicle Environment Modeling Method Based on

Stereo Vision

作 者 姓 名： 周彬

学科、 专业： 车辆工程

学 号： 21803135

指 导 教 师： 李琳辉

完 成 日 期： 2021.4.29

大连理工大学

Dalian University of Technology

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目： 复杂交通场景中各智能体多模概率化轨迹预测方法研究

作 者 签 名 ： 日期： 年 月 日

摘 要

现阶段人工智能的迅速发展对汽车行业产生了巨大的影响，汽车行业正在向更加智能，更加清洁、更加安全的方向发展。其中无人驾驶车辆发展对于提高交通安全、社会效益、经济效益具有重要意义。本文针对车辆的交通场景，基于立体视觉传感器对车辆环境的几何信息、语义信息、状态信息进行建模计算，并且将其三者融合构建语义地图，为无人驾驶车辆行为决策奠定基础。

首先，针对环境中语义信息，基于紧凑型全卷积网络对交通场景进行像素级别的语义分割。同时，为提高语义分割的精确率，将深度图与彩色图相融合构建RGB-D四通道图像作为卷积神经网络的输入。语义分割的结果表明，所提出的网络具有良好的实时性和精度。所获得的语义图像为后续构建语义地图提供语义信息。

然后，针对ORB-SLAM2中地图信息不够丰富、地图可用性不足的缺点，提出一种语义SLAM方法，在原ORB-SLAM2的基础上提出了稠密地图构建线程，基于立体视觉对车辆进行准确定位的同时，对环境进行几何建模，将几何信息与之前获得语义信息相融合构建语义地图，实现可行驶区域和障碍物区分，并且通过数据集实验以及实车实验验证了算法的有效性和普适性。

最后，在构建语义地图的基础上，估计语义地图中的车辆、行人等移动物体的运动状态。提出Tiny-Yolo与光流法融合的车辆检测与跟踪算法，引入时间序列，在检测跟踪的基础上使用ICP方法对检测物体的运动状态进行估计。相比较Tiny-Yolo，实验验证本文所提出的车辆检测算法在能够有效的改善Tiny-Yolo漏检、重检的问题。

关键词：车载环境建模；语义SLAM；车辆检测；光流跟踪

Study on Vehicle Environment Modeling Method Based on Stereo Vision

Abstract

Booming development of Artificial Intelligence have a huge influence on vehicle industry during this period, the vehicle is becoming more and more intelligent, clean and safe. And self-driving vehicle is meaningful for improving the transportation safety, social benefits and economic benefits. In the paper, we explore and calculate the geometrical information, semantic information and pose information for transportation environment based on stereo vision sensor and we fuse these information together to build semantic maps to establish foundation for self-driving car behavior decision.

At first, aiming to environment semantic information we propose a novel and compact fully convolutional neural network to obtain the pixel-wise semantic information. Simultaneously, we fuse the depth map and RGB image to obtain the 4-channel RGB-D images, we take RGB-D image as input of network to improve the accuracy of segmentation. The experiment results show the depth map is helpful for improve segmentation accuracy, and our network is competitive in speed and accuracy. The semantic information will be used in the chapter three to build semantic map.

Secondly, we propose a novel semantic SLAM system to improve the ORB-SLAM2. Original ORB-SLAM2’s map has no semantic information and just have sparse geometry information. We present a fourth dense point cloud map thread, accurate localization and simultaneously build geometrical model for vehicle by fusing geometry and semantic information to help vehicle distinguish the driving areas and obstacles. We also test the effectiveness and university of the algorithm by using the KITTI dataset and vehicle test.

Finally, we calculate the pose of moving objects such as vehicle and people. In the chapter we present a novel vehicle detection way by fuse Tiny-Yolo and optical flow, and then calculate the pose of detection object by using ICP on the foundation of optical flow. We also prove vehicle detection way that we present is helpful for avoid leak detection and duplicate detection compared to Tiny-Yolo.

Key Words：Vehicle Environment Modeling ; Semantic SLAM; Vehicle Detection; Optical Flow Tracking

目 录

[摘 要 I](#_Toc67470505)

[Abstract II](#_Toc67470506)

[1 绪论 1](#_Toc67470507)

[1.1 研究背景和意义 1](#_Toc67470508)

[1.2 轨迹预测研究动态与现状 2](#_Toc67470509)

[1.2.1 基于浅层学习的轨迹预测方法的研究现状 3](#_Toc67470510)

[1.2.2 基于深度学习的轨迹预测方法的研究现状 4](#_Toc67470511)

[1.3 本文主要研究内容 10](#_Toc67470512)

[2 基于State-Anchor 的多模概率化车辆轨迹预测模型 13](#_Toc67470513)

[2.1 引言（意图不确定性/控制不确定性） 13](#_Toc67470514)

[2.2 卷积神经网络 13](#_Toc67470515)

[2.3 预测模型网络设计 16](#_Toc67470516)

[2.3.1 MobileNet-V2 20](#_Toc67470517)

[2.3.2 仿射变换网络 20](#_Toc67470518)

[2.3.3 Sate-Sequence Anchor 的提取 20](#_Toc67470519)

[2.4 数据集的选用 20](#_Toc67470520)

[2.5 网络训练（训练参数，损失函数） 20](#_Toc67470521)

[2.6 实验结果与分析 22](#_Toc67470522)

[2.6.1 车辆轨迹预测模型验证分析 22](#_Toc67470523)

[2.6.2 车辆轨迹预测模型改进与优化 23](#_Toc67470524)

[2.6.3 车辆轨迹预测模型收益和可视化分析 23](#_Toc67470525)

[2.7 本章小结 24](#_Toc67470526)

[3 基于State-Anchor 的多模概率化非机动车轨迹预测模型 25](#_Toc67470527)

[3.1 引言 25](#_Toc67470528)

[3.2 非机动车轨迹预测模型验证 25](#_Toc67470529)

[3.3 网络结构设计 26](#_Toc67470530)

[3.3.1 ORB特征点 27](#_Toc67470531)

[3.3.2 追踪线程 28](#_Toc67470532)

[3.3.3 局部地图构建线程 29](#_Toc67470533)

[3.3.4 回环检测线程 29](#_Toc67470534)

[3.4 网络训练 31](#_Toc67470535)

[3.5 实验结果分析及非机动车预测轨迹可视化 32](#_Toc67470536)

[3.5.1 构建数据集 32](#_Toc67470537)

[3.5.2 ORB特征点匹配验证 34](#_Toc67470538)

[3.5.3 SLAM以及回环检测结果 35](#_Toc67470539)

[3.6 本章小结 37](#_Toc67470540)

[4 基于Anchor-Free的多模概率化行人轨迹预测模型 38](#_Toc67470541)

[4.1 引言 38](#_Toc67470542)

[4.2 行人轨迹预测模型验证 38](#_Toc67470543)

[4.3 网络结构设计 39](#_Toc67470544)

[4.4 网络训练 40](#_Toc67470545)

[4.4.1 光流跟踪 41](#_Toc67470546)

[4.4.2 融合算法 42](#_Toc67470547)

[4.5 实验结果分析 43](#_Toc67470548)

[4.6.1 行人轨迹预测结果的定量分析 43](#_Toc67470549)

[4.6.2 行人轨迹预测结果的定性分析 44](#_Toc67470550)

[4.7 本章小结 46](#_Toc67470551)

[5. 复杂交通场景中多任务多模概率化轨迹预测模型 47](#_Toc67470552)

[5.1 引言 47](#_Toc67470553)

[5.2 多任务网络模型结构设计 47](#_Toc67470554)

[5.3 多任务网络模型训练 47](#_Toc67470555)

[5.3.1 多任务模型损失函数设计 47](#_Toc67470556)

[5.3.2 训练参数的设定 47](#_Toc67470557)

[5.4 实验结果分析 47](#_Toc67470558)

[5.5 本章小结 47](#_Toc67470559)

[6 总结与展望 48](#_Toc67470560)

[6.1 总结 48](#_Toc67470561)

[6.2 展望 48](#_Toc67470562)

[参 考 文 献 50](#_Toc67470563)

[攻读硕士学位期间发表学术论文情况 55](#_Toc67470564)

[致 谢 56](#_Toc67470565)

[大连理工大学学位论文版权使用授权书 57](#_Toc67470566)

1 绪论

## 1.1 研究背景和意义

近年来，随着人工智能（Artificial Intelligence，AI）领域的蓬勃发展，AI技术在各个领域的应用也取得了突发猛进的进展，智能算法也日渐称为目前人民日常生活不可或缺的一部分。医院使用AI技术来对疾病进行诊断；自媒体短视频APP通过推荐算法获得了巨大的流量，也在日益改变人们的日常生活，闲暇时间刷会视频成为社会生活中及其普遍的存在；短短的几个例子，足以发现AI技术已经深深的存在我们的日常生活中。尽管AI技术已经取得了很大的进步，但是人工智能的革命却还远未结束，而且可能在未来几年获得更突飞猛进的发展[ ]。值得推敲的是，其中一个未完全受到人工智能影响的领域是汽车领域，大型的整车汽车制造商，如大众，通用等，虽然主机厂通过ADAS系统逐步提高人工智能技术在汽车领域的应用。但是根据目前车辆自动化水平的级别划分，目前应用到量产实车上的技术大多属于L2或L2.5级别，人工智能的全部力量应用到自动驾驶领域仍然有待于新的智能技术的出现。在交通中驾驶汽车是一项非常危险的任务，因为这是许多人的共同活动，甚至对有多年经验的人类驾驶员来说也是如此，虽然汽车制造商正努力通过更好的设计和ADAS系统来提高汽车的安全性，但严峻的统计数字表明，还有很多工作要做。

无人驾驶技术已经开发了很长一段时间，关于无人驾驶技术最早的尝试可以追溯到上世纪80年代ALVINN[13]的工作。然而，直到近10年，无人驾驶技术进步才达到了可以广泛应用的程度，在2007年DARPA城市挑战赛[14]和[15]中，参赛队伍被要求在复杂的城市环境中航行，处理公共道路上常见的情况，并且与人类和机器人驾驶的车辆进行互动。这些早期的成功激发了人们对于自动驾驶领域的极大兴趣，一些行业内的顶尖科研机构、自动驾驶初创公司(如Uber或Waymo)以及各国政府机构正相继建立自动驾驶技术和法律框架，让自动驾驶技术可持续的向前发展成为现实。无人驾驶平台作为自动驾驶汽车的终端，在国家战略意义、经济效益、社会效益都具有举足轻重的作用。在国际上，谷歌（Google）公司推出无人驾驶汽车（图1.1（c）），Uber自动驾驶初创公司（现已被收购）在亚利桑那州推出了自动驾驶汽车租赁服务（图1.1（b）），日本丰田推出了无人巴士（图1.1（a））；在国内，无人驾驶领域的发展，则以互联网造车势力及新造车势力为代表，其中百度推出的Apollo-Robotaxi无人出租车的服务，现在已经在北京市市区运营；智行者科技推出蜗小白系列，瞄准低速配送等业务市场；以禾多科技和主线科技为代表的自动驾驶卡车派，也已经推出了各自的产品，并且已经在天津港等卡车应用场景内运营，京东推出了无人配送车（图1.1（d））。蔚来汽车，小鹏汽车和理想汽车作为新造车势力的代表，则在L4级别自动驾驶汽车上投入甚多，并陆续推出各自的L3级别的量产车型。智能汽车进入了蓬勃发展的时代。



图1.1 各类型无人驾驶车辆

Fig. 1.1 Different kinds of Self-driving cars

谷歌旗下的Waymo自动驾驶团队在智能车安全报告[1]中指出，智能车的全自动驾驶需解决“我在哪里？”、“我周围有什么”、“我周围会发生什么”，“我接下来怎么做”四个问题，其中“我周围会发生什么”，意味着自动驾驶车辆需要了解车辆周围的环境状态，并基于所了解到的信息预测周围环境的状态变化，并以此指导自动驾驶汽车接下来的行为。

## 1.2 轨迹预测研究动态与现状

20世纪90年代以来，大量基于浅层学习的轨迹预测模型率先被提出，这些方法对计算设备的算力要求比较大，且缺乏统一的评价标准，用于测试模型的数据集同时也鱼龙混杂，质量参差不齐。近十几年来，随着机器学习技术（尤其是深度学习技术）的兴起，由于递归神经网络在处理时序数据上优异的性能，基于递归神经网络的时序预测模型层出不穷。下面本章节将对国内外主要的轨迹预测算法研究动态与现状进行分类阐述。

根据预测模型的建模方式，将轨迹预测算法分为基于浅层学习的方法和基于深度学习的方法。浅层学习中基于运动学的方法是最早应用在轨迹预测领域的，这类方法一般需要对智能体的运动学特征（速度，位置和角速度等）进行建模并将其与贝叶斯滤波器、马尔卡夫网络，卡尔曼滤波以及贝叶斯网络等模型结合起来，将当前状态传播到未来状态做预测。在基于深度学习的轨迹预测方法中，根据是否考虑智能体之间的交互影响，可将其划分为单轨迹预测模型和交互轨迹预测模型；根据是否预测生成确定性的轨迹，又可划分为确定性轨迹预测模型和可接受的轨迹预测模型。随着近十几年模型的提出和优化，目前基于深度学习的轨迹预测方法日渐趋于完善和模块化，考虑和利用的特征信息也逐渐完善，预测的精度和实时性也逐渐提高。综上所述，轨迹预测方法分类如图3所示。



图3 轨迹预测方法分类

实际上，一个好的轨迹预测模型往往需要综合利用不同的信息（轨迹信息，语义信息，社交信息）等，因此很难从其诸多信息中的单一角度去区分预测模型的好坏。为了对现有的轨迹预测方法进行更加细致的梳理和归纳。接下来将以预测模型的建模方式为主干，对轨迹预测方法进行总结，同时以发现问题为导向，解决问题为目标，对当前业界主流的轨迹预测前沿算法的研究动态进行详细地介绍。

### 1.2.1 基于浅层学习的轨迹预测方法的研究现状

最基本、最早出现的轨迹预测算法是利用基本的运动学模型（恒定速度/加速度/转弯）和贝叶斯滤波器或其扩展组合起来，将当前状态传播到未来状态[5]。Schneider等[6]将基于单个运动学模型的方法与基于多个基本运动学模型的多模型交互的方法进行了比较，结果表明，多模型交互的方法可以将行人智能体橫向位置预测提高30 cm。但是多模型的方法以恒定速度为假设，这显然不符合实际情况，还使得基于贝叶斯滤波器的模型难以捕获智能体的切换动态，且其用来测试的数据集样本数量和运动类型有限，不足以支撑更为复杂的运动模型的建立与预测。Pavlovic等[7]釆用切换线性动力学系统（Switched linear dynamical system, SLDS）模型[8]来描述非线性和时变动力学，基于马尔科夫链进行概率转移，在多种线性运动学模型间进行切换，从而处理实际情况中非线性运动情况的预测。SLDS每次都具有不同的动态特性，并以指定的先验和转移概率对切换状态的隐马尔可夫链进行调节。然而，基于该方法的运动特征信息有时不足以支持模型进行状态的切换，且其对于一些更加复杂的运动模型来说效果有限，需要通过构建更大的运动捕捉数据集来满足更加复杂的运动模型测试的精确性。Kooij等[9]建立了基于上下文的动态贝叶斯网络（Dynamic Bayesian Network, DBN），该行人路径预测模型将上下文信息（即行人头部方向，情况紧急程度和环境空间布局）作为潜在状态合并到SLDS的顶部，从而控制SLDS的切换状态，并且比使用SLDS能够生成更为准确的预测。但是，无论是基于SLDS的预测方法，还是基于DBN模型的预测方法在进行模型预测推理和数学模型搭建的过程中始终需要大量的计算，这将消耗计算设备很大的算力，而且额外的场景（例如红绿灯、人行横道）、基于SLDS的基本运动类型的扩展（例如转弯情况）等信息体现不足。Helbing等[10]提出了一种具有吸引力和排斥力的行人运动模型，称为社会力模型。该模型现广泛应用于机器人和活动理解等[11]领域。Alahi等[12]通过从人群里的人类轨迹中学习他们的相对位置来呈现社会亲和力特征，而Yi等[13]提出使用人类属性来改善人群中的预测。

随着机器学习的快速发展，基于运动学的方法在做预测时有时会使用一些基于机器学习的跟踪算法来改进跟踪和预测，例如卡尔曼滤波（Kalman Filter, KF）、马尔可夫模型（Markov Model, MM）和高斯过程[14-16]（Gaussian Process, GP）等。通过结合KF模型，其优势在于处理行人轨迹预测问题时能够有效地处理无噪声点的轨迹数据，对于短时间(1步/2步)内的预测精确度较高。相反，对于长时间的预测（10秒/5步以上）误差较大，模型复杂度增高，严重影响预测精确性，且KF模型随着噪声的增大变得愈发敏感，预测精度也近似成线性降低。MM模型对于智能体行人运动过程的状态预测效果良好，但其对行人轨迹的波动较为敏感，且不适用于中长期的行人轨迹预测，一阶MM模型仅考虑了当前行人运动轨迹点对未来轨迹点的影响，历史轨迹点的数据信息无法被尽可能地利用，而高阶MM模型大大增加了模型计算的复杂度。GP模型通过假设隐变量服从高斯分布，为概率预测提供非参数模型，其中预测轨迹是从历史轨迹数据中学习的[14]，通过在GP中指定适当的核函数（协方差函数）来明确进行预测轨迹建模所涉及到的不确定性，其中*μ*，*θ*为协方差函数的超参数，参数需要通过数据训练得到，其能够较为有效地预测具有噪声点的轨迹数据，同时可以很好地避免轨迹数据离散性质的不足，并在此基础上有效地表达行人运动轨迹分布的统计特征，但是构造高斯过程十分复杂，需要付出较高的时间代价，且其为非稀疏模型，需要完整的样本或特征信息来进行行人轨迹的预测，并随着数据的增多，其计算量大为增加。而且在单独使用GP模型时，在不同时间点，其预测误差差别也比较大。

这类方法都有一个很明显的特点，需要根据历史时序的数据，建立时序递推数学公式：或者。因为这类方法通常都具有严格的数学证明和假设，也能处理一些常规的问题，但是对于一些复杂的问题就变得“束手无策”了。这是因为基于运动学的算法中都会引入一些先验假设，例如隐变量服从高斯分布，线性的状态转换方程及观测方程等，而最终这些假设同时也限制和约束了算法的整体性能。

总而言之，基于浅层学习的轨迹预测算法在早期取得了一定的成果，为轨迹预测领域（特别是行人轨迹预测领域）的发展做出了诸多贡献。但是由于基于运动学（恒定速度/加速度/转弯）方法的局限性，运动特征信息提取的不足，特定场景信息的缺失，模型构建的复杂性，以及当时大型数据集样本及行人智能体运动类型有限等问题，其与实际情况存在着一定差距，传统的方法对较为复杂的智能体运动模型及场景难以进行较为精确的预测。

### 1.2.2 基于深度学习的轨迹预测方法的研究现状

基于浅层学习的方法需要对模型进行复杂、严谨的数学建模和理论推导，而基于深度学习的方法一般不需要假设固定的数学模型，凭借大规模的数据集促使网络学习更加合理的映射关系。近几年，随着深度学习热潮的兴起，各种用于处理时序数据的模型如雨后春笋般被提出，使得基于神经网络的轨迹预测算法流行起来，且预测效果较传统算法有了较大的提升。

基于深度学习的轨迹预测方法有两个关键要素。一是网络结构的选择。研究者不同，其对轨迹预测问题和智能体所处场景的辅助信息的使用亦有着不同的理解，选择的网络结构也会有所不相同，那么对轨迹预测的特征进行提取、汇集及预测时，结果也会有所差异。二是损失函数的设计。在网络优化的过程中，最终模型预测效果的好坏取决于能否设计出合理有效的损失函数。损失函数不仅影响神经网络的优化过程，而且决定着大型公开数据集所提供的特征信息能否得到较为充分有效地利用。总的来说，如何选择设计出合理有效的网络结构和损失函数是目前基于深度学习的轨迹预测算法的研究重点。

是否添加对于RNN, LSTM, GRU以及编码解码 和 attention机制的阐述。

近年来，用于序列预测的递归神经网络（Recursive Neural Network，RNN）及其变体，包括长短期记忆网络（Long Short-Term Memory Network，LSTM）[17]和门控递归神经网络（Gated Recurrent Unit，GRU）[18]在序列预测任务中（如语音识别[19,20]，标题生成[21]，机器翻译[11]）取得了巨大的成功。鉴于递归神经网络在处理序列数据尤其是解决长时依赖时的优越性能，自然而然的可以将其引入轨迹预测领域。而序列模型Seq2Seq为序列生成任务提供了一个编码器-解码器（Encoder-Decoder）框架，该框架旨在学习以输入序列为条件的输出序列（即）的条件分布，框架中可以将LSTM单元同时用于编码器和解码器中。编码器LSTM首先遍历输入序列，将其编码为语义向量，后将向量传递给解码器LSTM，随后生成输出序列，近几年的模型大多采用了该预测框架[22]。编码器-解码器（Encoder-Decoder）框架虽然较为经典，但是也存在一定的不足之处，即在编码和解码两者之间建立联系的唯一桥梁就是上述中提到的固定长度的语义向量，整个序列所携带的信息都将被压缩到此向量中，这就会出现两个问题，其一，输入序列的信息不能充分的被语义向量表示，其二，后输入的序列信息会将之前输入的信息覆盖掉。此现象会随着输入序列长度的增加愈发严重，最终造成解码准确度的下降。为解决此问题，注意力模型诞生，此模型不再要求编码器把整个序列所携带的信息都压缩到此向量中，而是将其将其输入信息编码成一个向量序列，解码时选择性的对其子集进行处理，保证了对序列所携带信息的充分利用。在此基础上，Karatzoglou等[23]通过对人类轨迹采用Seq2Seq框架扩展了LSTM网络，并探讨了基于注意力的Seq2Seq的影响，验证了Seq2Seq框架在轨迹建模和运动模式预测中的有效性。

在行人轨迹预测方面：基于先前的研究，Alahi等[24]提出了用于轨迹预测的社会长短时记忆网络模型（Social-LSTM，S-LSTM），该模型对智能体之间可能发生相互冲突的社交互动进行了建模。每条轨迹被建模为一个LSTM层，并且不同的LSTM可以通过社交池化层（Social-Pooling）共享信息，从而生成无冲突的轨迹，具体的计算思路是将该智能体周围的区域划分成*N\*N*个网格，每个网络都是相同的大小，落入这些网格中的智能体将会参与交互的计算。该模型成功预测了不同社交互动引起的非线性行为（例如人群同时移动），但是此模型仅仅建立了单一的模型设置（智能体共享空间），在此基础上还可建立诸多对象（例如行人，自行车，滑板车，手推车等）的社交池化层来共享空间信息，另外也可通过加入场景中的图像信息建立人与空间的交互信息。除此之外，基于社交池化的模型在每次训练预测时都要对社交向量进行计算，使得模型预测的实时性不高，文献中同样提出了将O-LSTM（Occuapy Map-LSTM）结构作为社交池化向量的简化版，以此来提升预测速度。Kitani等[25]已经证明，将静态环境的语义特征信息（人行道的位置，草地区域的延伸等）输入模型有助于更准确地预测未来时刻的智能体轨迹。文献[26]中也通过使用语义场景信息对人-空交互进行建模，并以此来推断场景中的可穿越区域，从而预测智能体的未来轨迹。因此，在O-LSTM的基础上，Xue等[27]采用分层LSTM结构，提出了SS-LSTM模型，在O-LSTM考虑智能体交互的基础上，额外考虑该智能体所处的场景信息，做出轨迹预测，且相较于其他基于LSTM的模型，其在ETH、UCY数据集[ ]中有更好的表现。未来，增加智能体交互信息所占的比重（比如引入智能体之间的距离）或是增加新的注意力机制（空间-时间）也会对其预测效果产生新的影响。

在交通场景其他参与者（机动车与非机动车）轨迹预测方面：MERCAT等[54]利用多头注意模块，对道路场景中的所有车辆的交互情况进行关联，之后使记忆层（LSTM层）进行编码和预测。不同于大多数联合预测模型，需要空间网格来描述所处场景，该模型仅仅使用单纯的轨迹序列，通过由LSTM和两层自注意力构成的网络模型，便得到轨迹的高斯混合预测。PARK等[58]使用注意力模块来融合各个部分的交互，特别强调预测轨迹的多样性。该模型不依赖ground truth，而是通过可行驶区域来估计真实的轨迹分布，提高了预测的多样性。MARCHETTI等[57]首次将记忆增强网络应用于轨迹预测领域，实现多模轨迹预测。相比于RNN中，过去的记忆被视为一个整体，无法针对某一特定场景进行寻址，记忆增强网络则可以其存储知识进行单独的操纵和推理。DEO等[59]将多个LSTM模块用于预测高速公路场景下，车辆轨迹的预测。该模型同时考虑车辆的轨迹及相对位置，针对不同的意图做出轨迹预测。其中，轨迹编码模块将预测车辆及主车轨迹和相对位置上下文进行编码，然后通过解码器输出各种意图下的车辆轨迹，由意图分类分支负责对每条轨迹分配置信度，以此实现多样轨迹的预测。SRIKANTH等[64]利用处理后的语义图像，训练自回归模型来预测车辆的轨迹，且模型有较好的泛化能力，对于不同的数据集均有较好效果。从语义图中，能够获得有关车道、障碍物和相关车辆的信息，将包含上述信息的图像通过一个卷积LSTM（Conv LSTM）网络之后，在相应场景的栅格图中，得到预测轨迹。LEE等[60]提出了一种随机的递归编码器-解码器网络，用于预测动态环境中车辆的未来轨迹。模型中，由RNN和CVAE组成的样本生成模块根据输入的历史轨迹生成多个合理的预测样本，然后通过RNN-回归模块对产生样本进行排名细化，通过计算累计的未来奖励对样本进行评分，实现类似于IOC框架的长期战略决策。另外，RNN的场景上下文融合模块共同捕获过去的运动历史、语义的场景上下文以及多代理交互。采用反馈机制迭代地执行排序和细化，提高预测的准确性。ZHAO等[61]提出将多个代理的过去轨迹和场景上下文编码为多代理张量，然后应用卷积融合捕获多智能体交互，同时保持智能体与场景间的空间结构，该模型用对抗性损失学习随机预测，递归地解码出多代理未来轨迹。

#### 1.2.2.1 基于生成式网络架构的轨迹预测方法的研究现状

虽然基于简单的LSTM网络以及Seq2Seq架构的轨迹预测网络在精度方面能够取得较为理想的结果，但是其预测结果为单一的预测轨迹输出，预测结果是模型预测轨迹的平均状态，并且与数据集结果高度拟合，这与智能体行人的高动态性与随机性相冲突。为解决模型架构带来的问题，基于生成模型的一系列预测方法（如生成式对抗网络（Generative Adversarial Network, GAN）[28]，变分自编码器（Variational Auto- Encoder, VAE））[29]被相继提出。

为解决先前模型中的平均轨迹的问题并对池化汇集模型进行改进，Gupta等[30]将基于生成式对抗网络的方法引入行人轨迹预测领域，基于先前的Seq2Seq框架和池化汇集思想，通过引入噪声z、改进多样性损失函数，使得模型趋向于生成多样性的轨迹，并且使用最大池化方法对全局的智能体行人进行交互分析，提出了Social-GAN模型，其优势在于强调预测轨迹在社会规则上的规范性、合理性，即相对于其他预测模型，该模型生成路径更加合理。同时解决了预测结果与现实不符的单一预测轨迹输出问题，且相比于Vanilla LSTM、Social LSTM等模型速度有了较大提升，但是该模型在进行池化汇集时提取的特征是经过最大池化后的最大特征，模型忽略了对智能体行人交互有用的其他特征信息，并且采用了传统的GAN架构，网络训练不稳定，容易崩溃。Amirian等[31]通过引入Info-GAN架构来改进GAN的架构，以此来解决 Social-GAN模型训练崩溃和掉落的问题，通过舍弃L2代价函数，引入基于互信息的 Information Loss损失函数，增强了模型对多条合理轨迹的预测能力，并引入注意力机制[32]使模型自主分配对交互信息的关注，该论文表明最新提出的Info-GAN架构可以极大地改善多模式行人轨迹预测，避免类似训练崩溃的问题。在此基础上 Sadeghian等[33]通过融合环境中场景的上下文信息以及智能体行人的历史轨迹，使用GAN的网络架构生成多条物理条件下的可接受轨迹。Kosaraju等[34]采用了基于Cycle-GAN的网络架构和训练方法，保证GAN在生成轨迹时对于噪声的敏感性，从而有助于生成多样性的轨迹，文中使用VGG网络提取场景图像特征，使用LSTM提取智能体行人轨迹特征，根据提取到的特征差异，分别使用基于 Scale-Dot和GAT的多种注意力机制，以得到对于各种输入最为合理的注意力向量，该模型可以理解较为复杂的智能体行人运动社会本质，不仅能够为特定的智能体行人生成多个轨迹，而且可以通过多模式方式同时为多个智能体人预测更真实的行人运动轨迹。

Cheng等[35]提出了一种不同于上述GAN架构的轨迹预测生成模型，模型使用了带条件的变分自编码器（Conditional-VAE, CVAE）用于实现对轨迹多样性的预测，为了刻画噪声与已知轨迹和预测轨迹之间的分布情况P(z|X,Y)，模型通过引入变分估计，将噪声的分布情况简化成含有参数（均值，方差）的高斯分布，训练模型预测假定分布情况后的后验分布，并隐式地从轨迹X和Y中建立与假定分布之间的联系。由于基于GAN网络的方法中不可避免地涉及到采样操作，但其在进行反向传播时不可微，文献[35]中找到的新的解决方案，相比于在噪声分布z∼N(μ,σ)直接采样，文中的噪声分布为z=μ+σ⊙ϵ，其中ϵ满足高斯分布，但在传播中可以理解为常数，避免采样问题，预测结果为一个目标在某一时刻的多次预测，并按随机变量釆样的原理拟合高斯分布，按照似然进行排序。对于噪声的处理，Yang等[36]也采用了这种方法。

Liang等[37]同样使用LSTM来接收历史信息并预测智能体行人的未来轨迹。不同于其他算法的地方在于，这个模型不仅接收智能体行人的历史位置，轨迹信息，同时也提取智能体行人外观、人体骨架、周围场景布局以及周围智能体行人的位置关系，通过增加输入信息提升预测性能。除了预测具体的轨迹，算法还会做粗粒度预测（决策预测），输出智能体行人未来时刻可能所在的区域。自此基于深度学习方法的轨迹预测算法，开始往多任务以及模块化方向发展。而Sun等[38]将轨迹预测问题视为分类和回归问题的结合体，模型预测不同意图的多个终点并基于这些终点生成不同的候选轨迹，为减少模型预测轨迹的搜索空间，将轨迹建模为三次曲线，通过生成曲线簇来生成候选轨迹集合，并对候选轨迹进行分类和回归运算，分类模块对每个候选轨迹进行二分类，回归模块对候选轨迹进行修正得到更加精准的预测结果。HUANG等[62]将潜在语义层纳入轨迹生成，提出了一个能够生成既准确又多样化轨迹的模型。其中，轨迹生成器获取目标车辆过去的轨迹，车道中心线图和噪声样本，用来生成未来轨迹的样本。鉴别器识别所生成的轨迹是否真实。模型除了生成器和鉴别网络，还有轨迹语义的监督部分，以实现对轨迹更精确的预测。

总而言之，基于生成式网络架构的预测方法在进行智能体行人轨迹预测时能保证较高精度的轨迹，提升了模型的预测速度，并大大关注了预测轨迹在社会规则上的规范性，使复杂模型地建立更为合理有效。

#### 1.2.2.2 基于图网络架构的轨迹预测方法的研究现状

图卷积神经网络（Graph Convolutional Network, GCN）的综述文献[39, 40]介绍了GCN网络的研究现状，将卷积神经网络（Convolutional Neural Networks, CNN）的概念扩展到图中，图上定义的卷积运算将目标节点属性与其相邻节点属性的加权聚合[41]。GCN总体与CNN相似，但是在图上进行卷积操作需要对图的邻接矩阵进行相关的定义和计算。文献[29, 42, 43, 44]将GCN扩展到其他应用，例如矩阵完成和变分自动编码器。

Gupta等[30]和Sadeghian等[33]利用具有交互机制的GAN或变分编码器来考虑场景中所有个体。但是这两种模型都无法学习人类行为的真正多模态分布，而是学习具有高方差的单一行为模式。此外，两种模型都受到他们学习社交行为的建模方式的限制，尽管前者通过对场景中的所有智能体行人使用相同的社交矢量来提取信息，但后者需要手动定义排序操作，操作复杂且实时性较差。在轨迹预测问题中，可以将智能体行人之间的交互表达为图形[45]，其中节点是指智能体行人，而边缘代表智能体行人之间的互动；较高的边缘权重对应于更重要的交互。通过使图完全连接，以高效的方式对人类之间的局部和全局交互进行建模，而无需使用可能丢失重要特征的模型，如合并或排序等。在文献[46]中，智能体行人集合被建模为时空图，其中边（时间和空间）与RNN相连，时间边捕捉单个智能体行人的信息，空间边捕捉智能体行人交互的信息，输出采用双变量高斯分布，该方法能较好地对时空信息进行有效建模，但该方式计算较为复杂。Haddad等[47]基于LSTM神经网络提出新型的Spatio-Temporal Graph（时空图），旨在实现在拥挤的环境下，通过将智能体行人-智能体行人，行人-静态物品两类交互纳入考虑，对智能体行人的轨迹做出预测。

Vineet等[34]将图注意力（GAT, Graph Attention Network）网络[13]引入轨迹预测领域，图注意力网络引入了注意力机制来实现更好的邻居聚合，通过处理模型编码的轨迹信息，从而增强轨迹预测的推理能力，也赋予了模型一定的可解释性，图注意力网络允许在可以表示为图的任何类型的结构化数据上应用基于自我注意的架构。这些网络基于图卷积网络的先验而构建，允许模型隐式地为图中的节点分配不同的重要性。Mohamed等[48]将智能体行人的轨迹建模为时空图以替换聚集层，设计了一个特定的加权邻接矩阵，其中核函数定量地测量了智能体行人之间的影响，使用图卷积神经网终和时间卷积网络对时空图进行处理，解决递归神经网络训练时参数过多和效率低下的问题。KHANDELWAL等[55]提出了⼀种基于RNN的能够感知上下⽂的多模⾏为预测⽅法。通过将⻋辆轨迹和由路⽹转换⽽来的有向图输⼊模型，利⽤图注意⼒结合交互的上下⽂最终通过解码获得预测轨迹。LIANG等[56]提出⼀种同时融合智能体与智能体，智能体与⻋道，⻋道与⻋道，⻋道与智能体四种交互的模型。其中对⻋道与⻋道，不采⽤栅格化图像输⼊，而是根据路⽹⽮量图⽣成拓扑结构，并提出⼀种新的图卷积⽅法（LaneGCN），以体现⻋道的上下⽂关系。对输⼊的轨迹，则采⽤1D卷积来提取特征。之后通过注意⼒与LaneGCN结合的⽅法，实现将上述四种关系融合，实现最终的轨迹预测。GAO等[63]提出了 VectorNet架构，这是一个层次图神经网络，它首先利用矢量表示的单个道路组件的空间局部性，然后对所有组件之间的高阶交互进行建模。通过操作向量化的高清地图和智能体轨迹，避免了有损渲染的和髙密计算的卷积编码步骤。通过恢复随机掩盖的轨迹和路径上下文来增强 VectorNet能力。

总而言之，随着深度学习的快速发展，基于深度学习的轨迹预测方法成为研究热点，理论上而言，基于深度学习的算法基本可以解决利用传统的浅层学习算法的行人轨迹预测问题，但当神经网络足够深，功能足够强大的时候，如果数据集数据量过小，就非常容易产生过拟合的问题，从而影响预测精度。数据集规模的增大，社会交互场景信息的丰富，数据集所涵盖的corner case 越来越完善，场景就会越来越接近于现实场景，并且随着数据量的不断增强以及神经网络结构及损失函数合理有效地设计，对于行人轨迹预测的效率及精度都会有很大的改善。近几年图卷积神经网络应用到轨迹预测问题，取得了亮眼的成绩，但是基于图网络算法的模型存在着模型建图效率低下，邻接矩阵难以确定等问题。相信随着图网络技术的发展和成熟，基于图网络的轨迹预测方法发展前景一片大好。

## 1.3 本文主要研究内容

本文对近年来轨迹预测等方面工作领域进行研究，以模型的结构设计与优化为出发点，对目前轨迹预测方法进行了分类，并对不同算法的优缺点加以总结。结合轨迹预测的发展趋势可以发现，基于浅层学习方法的轨迹预测算法已成为历史，未来伴随着具有相当规模的大型数据集相继被提出，无论单独从精度还是效率上基于深度学习的方法都比浅层学习方法更为有效，目前基于深度学习的方法已有向多任务学习以及模型融合的趋势发展，对于模型结构的使用有显著的模块化特点，即先明确问题而后从备选模块中选用合适的模块进行拼接。未来伴随着神经网络结构递归构造为更为复杂的深度网络，数据集规模的发展以及特征信息的不断丰富，相对于浅层学习而言，基于深度学习方法的模型预测精度会不断提高，但是随之而来的是模型预测效率的降低。而在实际场景的应用过程中，需要同时保证轨迹预测算法运行的高效性及识别的高精度性，所以当前领域内轨迹预测技术在实际场景应用中还存在不足，需要结合实际考虑诸多问题，因此未来的研究工作还可从下面几个方面展开：

·当前大部分智能体的行人轨迹预测数据集为俯瞰视角，该数据集距离行人轨迹预测的实际应用具有比较大的差距，由于行人轨迹预测应用场景大部分为道路交通场景，创建一个各种信息完备（场景信息、轨迹信息以及人体骨架关键点信息等）的第一人称视角的大型数据集是必要的。

·实际场景中对行人轨迹预测算法的效率及识别精度有较高的要求，但当前存在两者不可兼顾的问题。如何设计出能够应用在嵌入式或移动设备上高效且轻量化的网络结构，并在此基础上保证高精度的识别效果，也是未来领域内的一大研究热点。

·目前的模型对计算到的交互信息缺乏可解释性。行人之间的交互是复杂而抽象的，在算法中很难精确建模，仍然依赖于数据驱动，即使使用基于图的算法，但随之而来的是图算法的各种缺点。如何通过对交互的可解释建模以及采用更为先进的图模型来提升模型预测的准确性和可解释性对于行人轨迹预测来说也是研究热点之一。

·目前行人轨迹预测框架强烈依赖于实际场景中的应用领域或者领域中的特定场景（例如，车辆汇集以及智能车领域中的行人过马路场景），因此达到任何形式的绝对要求几乎是一件不可能的事情。对于具体的场景用例，行业驱动的领域而言，诸如智能车领域内所提出的需求拟定，解决方案以及行业标准都是较为成熟的（例如国际标准化组织对城市驾驶场景下智能车的紧急制动的预测范围和度量精度的要求进行了描述，以最高车速，交通规则以及智能体行人速度和角速度分布为条件定义了车辆舒适加速/减速率的规范等）。那么认为对于智能车基本紧急制动这一具体场景用例，其解决方案达到了工业化消费品的标准。对于其他具体实际场景用例，更加标准化及具体化的行业标准的提出对于加快领域内预测技术的发展也至关重要。

轨迹预测技术在近五年的时间里已经取得了飞速的发展，随着思路的扩展，基于多任务学习和模块化思想的轨迹预测算法必将成为轨迹预测的一大趋势，随着技术的更新和进步，算法性能的提高，相信轨迹预测技术走进实际生活的距离不会很远

基于以上分析，本文针对智能车室外交通场景，解决智能车在无GPS信号的定位问题，并且构建三维语义地图为智能车提供充足的几何特征信息和语义信息，使智能车不仅仅对周围的几何环境有所理解，即知道障碍物所在位置，还能够为使智能车使用语义信息知道障碍物具体是什么以及障碍物的运动状态，为智能车的接下来进行决策以及复杂任务奠定基础。

本文创新性如下：

（1）在ORB-SLAM的基础上，提出了一种新型的双目语义SLAM系统，构建稠密的语义地图，弥补原ORB-SLAM中语义信息不足、受移动物体影响大的问题。

（2）针对移动物体中的状态估计问题，使用Yolo网络与光流跟踪融合的算法对车辆检测，同时使用直接法对移动车辆进行跟踪，估算车辆运行姿态。 针对以上研究内容，内容整体分为五章，分别为：

第一章，绪论。该章主要从无人驾驶平台的任务需求入手，对语义SLAM的意义进行阐述。并分别研究了传统的SLAM，环境语义理解，语义SLAM的研究现状、研究难点以及所存在问题，提出了本文的研究内容和论文框架。

第二章，基于全卷机网络的语义分割。该章主要详细研究本文语义理解的方法，使用端到端的方法对图像进行像素级别的语义分割。首先，详细介绍了课题组所提出的新的紧凑型卷积神经网络；然后，使用双目立体视觉获取图像的视差图，将视差图与彩色图融合成RGB-D四通道图作为输入。并且实验验证了使用RGB-D图作为输入能够提高语义分割的精度。

第三章，基于立体视觉的语义SLAM方法。该章详细研究语义SLAM方法原理。叙述了原ORB-SLAM的详细步骤，并对此进行相应改进，增加了语义地图构建线程，使用Octomap形式存储地图。并且分别以KITTI数据集以及车载平台取景构建数据集对所提出的方法进行实验验证和分析。

第四章，基于深度学习的移动车辆状态估计方法研究，该章主要研究YoloV3的物体检测方法，并且在其基础上，使用直接法对车辆进行跟踪，将跟踪与检测的结果相融提高物体检测的精度。并且使用ICP方法对跟踪的车辆进行状态估计，并且通过实车采集数据集验对算法进行了验证分析。

第五章，对全文研究内容进行总结，在总结的基础上分析本文不足，对未来工作进行规划。

2 基于State-Anchor 的多模概率化机动车与非机动车轨迹预测模型

## 2.1 引言（意图不确定性/控制不确定性）

机动车（汽车，摩托车等）和非机动车（自行车，电动单车等）作为复杂交通场景中最为重要的交通参与者，是交通场景中主要的交通智能体。无论从难度还是潜在的社会影响的角度，自动驾驶车辆都是机器人和人工智能（Artificial Intelligence，AI）目前面临的最大课题。无人驾驶汽车对于防止交通事故保障道路交通安全，减少道路拥堵优化城市道路结构，改善人民的生活水平都有着至关重要的意义。在复杂的交通场景中驾驶汽车是一个尤其危险的事情，因为这是一个各种智能体进行复杂交互的场景，虽然各大汽车供应商以及主机厂正在努力通过更优的设计以及ADAS系统来提高驾驶安全性，但是严峻的统计数字表明，我们还有许多工作要做。

为了使自动驾驶系统在现实世界中安全有效地运行，自动驾驶系统的关键部分是正确预测周围交通场景参与者的运动，一个成功的系统还需要考虑其固有的多模态特性。于主车而言，对周围车辆的轨迹进行精准的预测，有利于主车更为精准的进行速度控制，规避即将到来的危险，采取最优的局部路径，对于保障主车行驶的稳定性，安全性和经济性都有深远的意义。例如，当一辆车要插入到主车前方，对于主车来说，是否让路以及寻找最佳的并入车流时间都具有重要的意义。另外，对于未来状态和行为的预测本质上是随机的，因为我们无法得知场景中每一个智能体最为准确的动机。开车时，我们永远无法确定其他司机接下来会怎么做，因此在本文的轨迹预测模型中考虑多种可能性是非常有意义的。

在接下来的介绍中，本章节设计的机动车轨迹预测模型和非机动车轨迹预测模型基于一组固定的State-Anchor序列作为整个模型的建模基础。将复杂交通场景中机动车和非机动车的不确定性使用State-Anchor表示，并对他们的不确定性等级进行分类：首先是意图不确定性，通过一组Anchor轨迹的分布，来捕获智能体在未来时刻的意图；其次是控制不确定性，在给定机动车的意图之后，控制不确定性可以将机动车与非机动车的未来时刻的偏移表示出来。在本章节中，意图不确定性使用一组Anchor轨迹分布建模，而控制不确定性以假设智能体在未来时刻的轨迹点服从正态分布为建模基础。

## 2.2 轨迹预测问题建模

### 2.2.1 轨迹预测问题定义

在本文中我们将轨迹预测问题表示为基于过往机动车与非机动车的状态和信息来估计未来时刻智能体的状态。在下面的例子中，在任意时刻*t*，场景中的机动车和非机动车都由它们的*xy*坐标表示。使用索引来表示机动车的编号，其中*N*是场景包含的智能体的总数。模型中我们假设每个智能体的轨迹都受到先验运动、其他交通智能体的位置、物理场景约束以及交通规则的影响。因此，对于复杂道路交叉路口场景，模型的输入有两层，分别是交通场景信息和智能体轨迹信息，其中智能体的轨迹信息由四维向量编码，其中。在给定这些输入特征和每个行人的真实轨迹的情况下（其中），我们的模型将通过学习模型参数来预测未来每个行人的轨迹。

其中，，模型参数为模型中使用的所有深度神经网络的权值。模型使用反向传播算法训练模型的所有权值参数，通过随机梯度下降算法，最小化行人预测轨迹和真实轨迹的损失函数。在接下来的小节中，文章将分别介绍基于State-Anchor的机动车与非机动车的轨迹预测方法。首先文章将介绍如何捕获智能体的意图不确定性和控制不确定性，接下来将介绍轨迹预测领域的相关数据集并描述如何在选定的数据集中使用复合栅格化地图中描述交通场景和交通智能体的轨迹信息。接下来，将介绍如何使用这些属性信息并行地作为模型网络的输入(如位置、速度和转向角等)。然后通过独特的交互注意网络(Interaction - Attnet)将处理的隐藏特征结合起来。同时模型还设计了多模预测网络，提出了多模损失函数来预测多种概率轨迹。

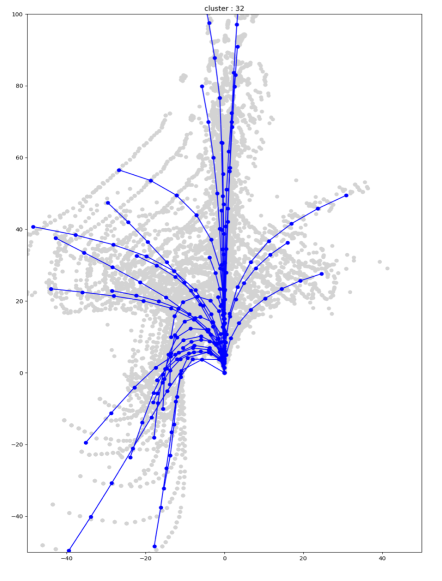
### 2.2.2 机动车与非机动车意图不确定性与控制不确定性建模

如上所述，在本文中我们将模型的不确定性分解为两个相对独立的量，智能体的意图不确定性建模智能体潜在的意图和行为目标。例如，在交通场景真实的驾驶环境中，智能体尝试行驶到哪条车道的不确定性。此外在意图不确定性的基础上，仍然存在着控制不确定性，它用来描述在满足智能体意图不确定性的前提下，智能体将遵循的未来轨迹预测点的状态序列的不确定性。例如，在已经选择好行驶的车道好，如何到达该车道，以怎样的轨迹序列完成这项任务，就体现的是智能体的控制不确定性，这表现在智能体在行驶时的速度，转角等。

在模型的建模时，我们设计并使用一组离散的State-Anchor轨迹序列来表示智能体的意图不确定性，其中*K*为Anchor轨迹序列的总数，}表示每一条Anchor轨迹的轨迹状态序列。模型使用Softmax分布来对智能体的不确定性进行离散性建模：

其中，x为模型的输入，是深度神经网络模型的输出。该分布是基于Anchor轨迹参数化的，如[][]在文章中所述，直接对轨迹进行学习会造成模式崩溃的问题，借鉴在目标检测领域和人体姿态与关键点检测领域的处理方法，模型在固定Anchor之前先对模型进行估计。在本文中我们使用K-means聚类算法对轨迹进行聚类，来简要的获得轨迹之间平方距离的近似，

其中，，为仿射变换矩阵，用来将智能体的轨迹放在一个以智能体为中心的标准的旋转不变和平移不变的坐标系中，使用K-means算法得到数量为*K*的Anchor轨迹序列，如图 所示，此为时聚类结果可视化分析图。



接下来，对智能体的控制不确定性进行建模，如上文所述，在已经得知智能体的意图不确定性后，控制不确定性依赖于每条Anchor轨迹所得出轨迹预测点的双变量高斯分布：

其中，为第*k*个Anchor轨迹的预测序列，为每个Anchor轨迹点的双变量高斯分布的参数，代表着智能体的控制不确定性。这使得模型在不同的道路交通场景的静态Anchor的基础上，可以考虑智能体动态的控制不确定性。例如在特定的道路中，考虑与其他智能体或者场景的交互。

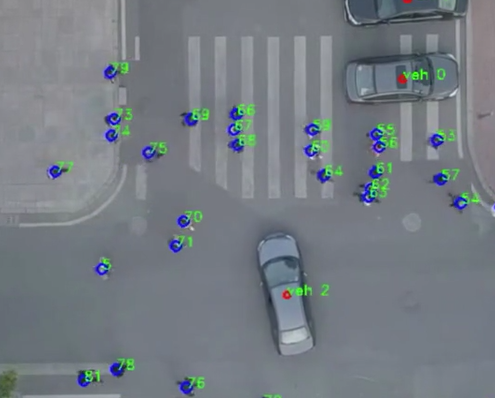
综上所述，为了获得整个状态空间的分布，考虑机动车与非机动的意图不确定性和控制不确定性，将模型建模为：

## 2.3 数据集

### 2.3.1 常用的轨迹预测数据集

在正式开始介绍我们的模型之前，我们将介绍关于在复杂交通场景中进行轨迹预测问题所要使用的数据集。ETH[49]和UCY[50]公开数据集包含在各种类型的社会交互场景下行人的全局轨迹坐标，在这些数据集中，行人作为数据的主体包含诸多复杂的行为，包括行人交互、非线性轨迹、避免碰撞、站立及群体行人的轨迹坐标等，同时包含从固定的俯视图记录的五个独特的室外环境信息。每个环境中单个场景的人群密度不同，所有视频的每秒帧数为25，行人轨迹以2.5fps的速度进行采样标记。其中ETH由ETH和Hotel两个数据集组成，UCY由Zara1、Zara2和Univ三个数据集组成。数据集包括从室外监控摄像头拍摄的五个视频，其中包含2206条行人轨迹，表现出在直线运动和曲线运动样条之间变化的不同特征。随着视频捕捉到大学入口处人们的动作，ETH场景包含了更多笔直的轨迹，几乎没有社交互动，而UCY场景展示了更多与人空间互动有关的场景。例如，UCY-ZARA数据集中包括在商店入口处弯曲的行人轨迹，而UCY-UNIV则具有更多的社交互动。此外，除非考虑社会和空间环境，否则这些情况尤其会增加单个路径的不可预测性。目前绝大部分轨迹预测模型都以该数据集为训练集和测试集，并在此数据集上测试模型性能。

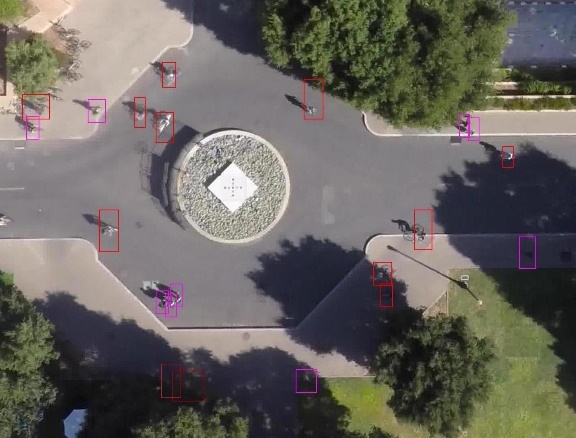
DUT数据集是在中国大连理工大学（DUT）主校区选择具有典型混杂交通特征的路段通过航拍采集的，旨在解决适用于人车交互运动的轨迹预测模型，如图4所示。图4（a）为无交通信号的交叉路口。由于没有交通信号灯，当行人与车辆交互时，可以体现人车交互的“社会性”。图4（b）为环岛，行人和车辆在较大的活动空间自由移动，而不受制于路面，可以得到多个方向上的预测结果。本课题组采用一架DJI-Mavic-Pro在行人和车辆很难察觉的高度拍摄，以达到不会因为无人机的出现影响行人运动的目的。视频分辨率为1 920×1 080，fps为25。数据集共有17个交叉路口场景片段和11个环岛场景片段，包含了1 793条轨迹。

（a）无交通信号交叉路口 （b）环岛

图4　DUT无人机数据集

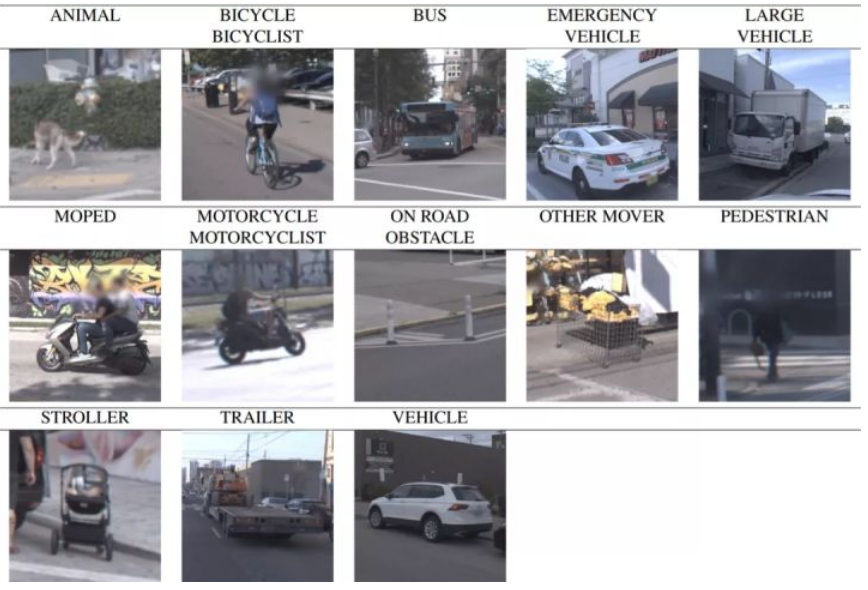
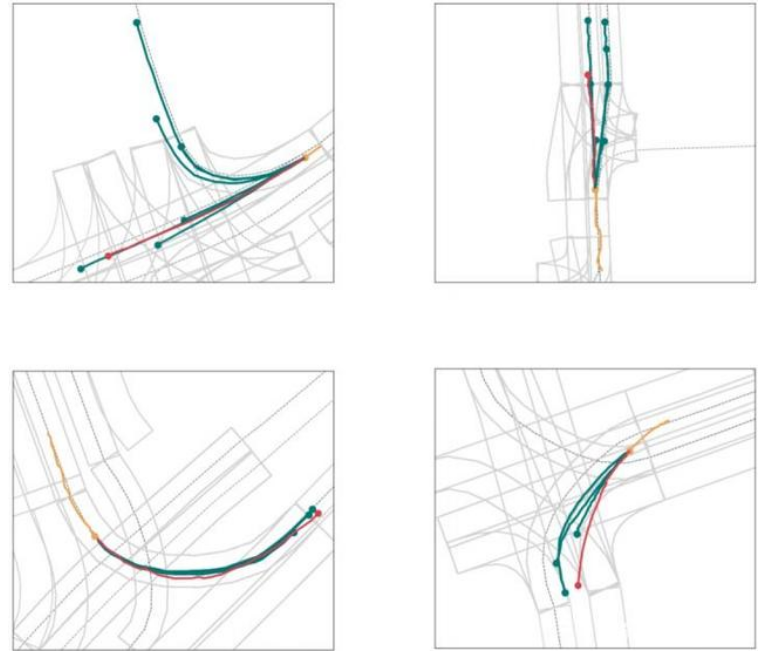
斯坦福无人机数据集（Stanford Drone Dataset）是由Robicquet等[51]提出的一个大型数据集，旨在解决目标跟踪或轨迹预测之类的任务，斯坦福无人机数据集是一个大型且先进的数据集，收集了各种类型的智能体（不仅是行人，还包括自行车，滑板，汽车，公共汽车和高尔夫球车）的图像和视频（如图5所示）在现实的户外环境（例如大学校园）中行驶。其中包含行人、骑自行车的人、滑板者、手推车、小汽车和在大学校园中行驶的公共汽车的视频以及轨迹信息。

（a）quad场景 （b）death-Circle场景

图5　斯坦福无人机数据集

Argoverse数据集是由Argo Al公司发布的第一个具备高精地图的大型无人驾驶数据集（如图6所示），旨在探究高精地图对于关键感知和预测任务的影响，该数据集主要包含3 D 轨迹跟踪和运动预测两部分，并将髙精度地图与3D轨迹跟踪和预测结合，用确定性的地图提高整体系统的确定性。3 D 轨迹跟踪数据集包含113个场景的三维跟踪注释。每个片段长度为15-30秒，共计包含11052个跟踪对象。训练集和测试集的每个片段场景中包含了五米内的所有物体的注释，可被理解为检测汽车可驾驶区域（5米）的所有物体，并以3D框架形式展现。超过70%的被跟踪对象是车辆，还观察到行人，自行车，轻便摩托车等。运动预测数据集包含324557个场景序列，主要包括（1）在十字路口，（2）左转或右转，（3）转向相邻车道（4）交通繁忙时等，每个序列时长5秒，且包含以10Hz采样的每个被跟踪对象的2D鸟瞰图，每个序列的“焦点”对象始终是车辆，但是其他跟踪的对象可以是车辆，行人或自行车，它们的轨迹可用作“社会（Social）”预测模型的上下文信息，该数据集由超过1000小时的街道驾驶所获取。

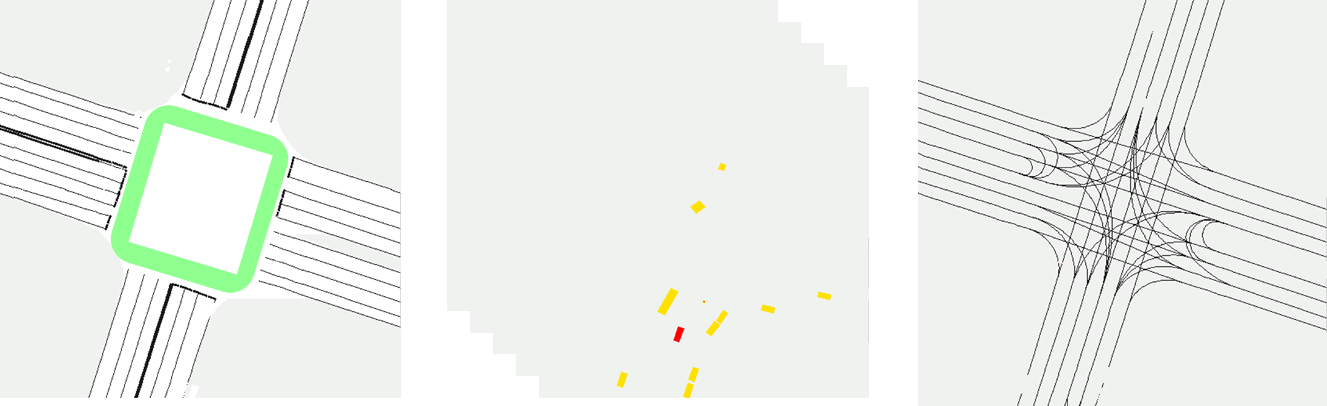
 

（a）3D轨迹跟踪/预测分类 （b）运动预测

图6 Argoverse数据集

### 2.3.2 构建数据集

复杂的交通场景中涉及各种智能体之间复杂的交互，例如，当车辆行驶在道路上时，司机总是受到周围物理环境(如车道线、红绿灯、障碍物信息)和交通环境中各类智能体(如行人、车辆、骑自行车的人)的影响。而从上文阐述中现有的大多数轨迹预测或行为预测类的数据集都无法提供如此全面且庞大的数据样本，基于此本文采用的数据样本为复合栅格地图，其由静态语义地图、动态智能体位置地图和实时交通灯地图组成。如图 所示，静态地图包括可通行区域、不可通行区域、车道线、人行道、停车线信息，用不同的RGB值表示。动态智能体位置地图是障碍物历史位置的映射序列，每一步都有一个通道。交通灯地图包含实时交通灯信息。这些元素组成了我们的复合光栅图。这种简单的地图格式不仅提供了基本的几何和语义特征，而且创新性地整合了时间序列信息，用于准确预测，这也符合交通场景中智能体做决定时的直观感受。



(a) HD static map (b) Historical location of agents (c) Traffic light illustration

Fig. 3. The composition of raster map. Illustration of the rasterized rendering to represent high-definition map and agent trajectories, where (a) is HD static map which contains the topology and semantic information in the junction scene, (b) is historical location of different agents which contains 5 channels represents 5 frames and (c) is the dynamic traffic light graph that represents traffic information of different lanes.

本文所采用的数据集采集自中国北京亦庄，其中包含了高清地图数据（HD-map）和障碍物历史数据。通过车载传感器（激光雷达、摄像头和雷达）可以准确的捕获周围包括行人、车辆和非机动车在内的所有智能体的位置和轨迹信息。在该数据集中，进行采集的感知车辆被视为交通场景中的附加障碍车辆，与其他车辆没有区别。所有智能体轨迹总数为800M，其中车辆轨迹655M，非机动车轨迹80M，行人轨迹65M。

## 2.4 机动车多模概率化轨迹预测模型



如图 所示，机动车多模概率化轨迹预测模型基于State-Anchor技术，模型通过Encoder-Interaction-Decoder架构搭建，采用MapNET和AgentNET对数据集中的栅格地图信息和历史轨迹特征信息并行的进行编码，通过交互网络对提取到的历史轨迹特征和地图场景特征进行整合并提取智能体之间，智能体与环境之间的交互信息，最后通过预测网络，对轨迹预测的结果进行预测。下面章节中我们将分别介绍这些网络模块。

### 2.4.1 MapNET：精确的时空表示

在本小节中我们提出了一种新的网络模型MapNET，用于从复合栅格图中学习道路交通场景中复杂的拓扑关系。如图 所示，MapNET由两个模块组成：首先，使用2.3节所介绍的方法搭建好复合栅格地图，之后将复杂栅格地图，使用Mobile\_net\_V2提取地图的语义特征表示，然后设计使用FCN模块进一步提取地图的拓扑信息,并返回一个与输入图像大小相同的特征图：

其中表示MapNET深度网络的所有权重，使用这种分支模型的一个重要好处是，网络可以完全提取时空表示。

1. 卷积神经网络

卷积神经网络（CNN）作为一种以数据驱动为基础的计算方法，通过将各个计算层排列起来，使用特殊的计算方法进行可微的函数计算，之后使用激活函数对计算结果进行非线性转化，实现对于非线性问题的拟合。其中现有的卷积神经网络大多由四种类型层组成：卷积层，池化层，非线性激活层以及全连接层。

卷积层是整个卷积神经网络的核心所在，若以图像为卷积的输入，卷积的过程其实为卷积核在整个图像上进行滑动，这也是卷积层区别与全连接层局部连接的特性，其中卷积核主要参数有卷积核大小F、步长S。卷积核大小其实为局部连接的大小也是卷积层感受区域的大小，卷积的操作如图2.2所示，绿色区域为处理数据大小5x5、卷积核大小为3x3，步长S为2。蓝色为输出的特征图，卷积后的特征图尺寸宽度、高度分别为：

 （2.1）

 （2.2）

其中、分别为输入数据的大小。P为在输入特征图零填充（zero-padding）的个数，卷积核在滑动过程中，若无法按照步长全部覆盖数据，则需要零填充来满足卷积核按照部长全部遍历图像。其卷积的数学表达式为：

 （2.3）

其中，为输出特征图行列结果，为输入特征图行列的数值，为卷积核的权重，表示卷积核的偏置项。这两项参数也是卷积神经网络训练的参数，为激活函数，激活函数在下文进行介绍。



图2.1 卷积操作示意图

Fig. 2.1 Convolution operation diagram

在研究中发现在传统的神经网络中，全连接层对于类似于图像大型数据的处理效果不佳，而卷积层因其局部链接、参数共享的特点减少了网络中的参数数量，同时局部链接的特性提取了图像的特征，提高了网络对于图像的处理能力。卷积层中局部链接主要区别与全连接层的全部链接，例如假设输入的图像大小为500X500X3，全连接层的权重参数为750000，若是使用卷积核的大小为5X5，卷积核的个数为3，参数数量则为75，参数数量大大减少。



图2.2 池化操作示意图

Fig. 2.2 Pooling operation diagram

通常在连续的卷积层之间会周期性插入一个下采样层，下采样层能够逐渐缩小数据体的尺寸，减少计算机计算资源消耗，还能够有效控制过拟合。其中下采样层主要有最大（Max）采样和均值采样。下采样层的主要参数有感兴趣大小和步长。最大采样主要选取感兴趣池化窗口中最大值，均值采样则为计算感兴趣池化窗口中的平均值。假设输入数据体的尺寸为4x4，感兴趣大小=2，=2，则下采样操作示意如图2.2：

其计算公式为：

 （2.4）

 （2.5）

其中，分别为是输出数据体的高和宽，，分别是输入数据体的高和宽。

目前，卷积神经网络中常用的下采样方式为最大下采样，并且从实验结果上看最大在一般情况下优于均值下采样[65]，常用的池化的感兴趣大小一般为2。就下采样发展趋势角度理解池化层，在文献[66]中不使用池化层，而是使用更大的卷积步长来缩小数据体的大小。并且在变化自编码器[67]（VAEs：variational autoencoders）和生成性对抗网络[68]（GANs：generative adversarial networks）中，弃用下采样层也很重要。在未来卷积神经网络中，可能很少或者不使用下采样层。

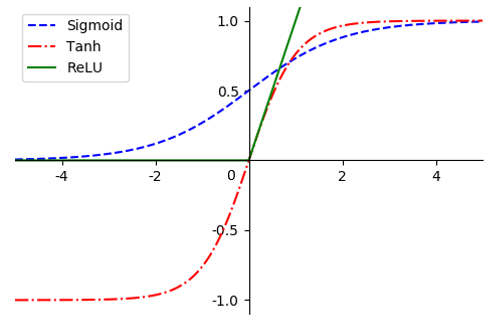


图2.3 常用激活函数曲线示意图

Fig. 2.3 Common used activation function curve diagram

激活函数一般作用在卷积之后，其作用之一在于过滤掉特征较弱的数据，增强网络的合理性。并且激活函数一般为非线性函数，从这个角度讲也增强卷积神经网络的非线性，这也符合大部分系统的属性。常用的激活函数有sigmoid函数、tanh函数、ReLU函数等，其函数的曲线如图2.3所示。

其中Sigmoid函数的计算公式和其导数分别为：

 （2.6）

 （2.7）

Sigmoid函数属于饱和函数，当输入数据较大或者较小的时候，其梯度均为零。这在网络训练，反向传播中会导致网络的梯度消失，导致权重无法更新学习，并且Sigmoid函数的值域为(0, 1)，经过激活函数计算的输出均值不为0，全部为正的结果会导致在反向传播过程中单向传播，导致训练收敛缓慢。

Tanh函数的计算公式和其导数分别为：

 （2.8）

 （2.9）

与Sigmoid函数相似，Tanh函数也是一种饱和函数，因此其同样存在梯度消失的问题，但是Tanh函数的值域范围为[-1，,1]，有效的解决了Sigmoid函数在训练中的单向传播问题。

相较于前两个函数，ReLU函数为非饱和函数，其公式为：

 （2.10）

与Sigmoid函数和Tanh函数相比较，ReLU函数没有指数运算，只需要一个阀值就能获得激活值，计算机源消耗少。而在文献[43]中证明，ReLU函数的不饱和性在训练过程中能够有效解决梯度消失问题，并且训练时收敛速度明显优于前两个函数。ReLU函数也是目前最常用的激活函数。

1. MobileNet-V2

ImageNet图像分类大赛是深度学习热潮的推动者之一，在ImageNet比赛上涌现出了许多经典的卷积神经网络VGG，AlexNet等。这些网络成为后续卷积神经网络开发的重要参考。2012年的比赛冠军AlexNet在图像处理方面具有重要的意义，他不仅证明了CNN对于复杂模型的有效性，并且GPU使得训练时间在可接受范围内。AlexNet网络的具体结构如图2.4所示[43]，AlexNet总体结构有8层网络，前五成的卷积层以及后三层的全连接层。AlexNet输入为,输出为1000维的向量，表示为1000中分类结果的得分率。

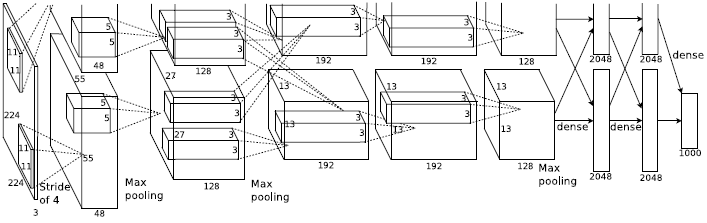


图2.4 AlexNet网络结构

Fig. 2.4 AlexNet network structure

相较于之前的CNN，AlexNet网络使用了一些以下经典方法优化网络性能。

1）激活函数选用ReLU，改善了使用其他激活函数梯度消失问题，在这之后激活函数基本选用ReLU函数以及相应的变种

2）在下采样过程中，下采样步长应小于下采样窗口大小，这种方法在实验中证明能够提高分类结果的精确率，减少过拟合。

3）在全连接层后加入Dropout层，Dropout层对全连接的输出以1/2的概率进行置0，并且在反向传播过程中也训练，这种方法防止了网络的过拟合。

4）在第一层以及第二层卷积层后进行了LRN操作（LRN，Local Response Normalization），对当前层的输出进行平滑处理，总体可以提高网络的精确率。

AlexNet应用以上方法防止了网络的过拟合。但是最后的三层全连接层导致网络具有60M以上的参数总量，计算资源消耗较大。

相比较AlexNet，VGG16网络增加了网络层数，并且展示了越深的网络能够取得更好的精确度。图2.5为VGG16网络的网络结构[44]，一共由13层卷积层和3层全连接层构成。VGG16一个突出的特点是全部网络使用的卷积核大小均为3×3，几个小卷积核的叠加要比一个大卷积核的参数数量少，却能达到大卷积核同样的感受野。3个3×3卷积核的叠加与一个7×7卷积核的效果一样，假设数据体的通道数为C，则3个3×3的参数数量为，而7×7的参数数量为。具有相同效果的感受野，小的卷积核的参数少量要少了将近50%。

对于智能车而言，语义分割实时性和精度都需要一定保证，VGG16网络较深，模型数量较大，实时性较差。并且AlexNet与VGG16中全连接层增加了模型的参数数量，计算机计算资源消耗较大。



图2.5 VGG16网络结构

Fig. 2.5 VGG16 network structure

课题组基于Alexnet提出了新的紧凑型全卷积网络，去掉了全连接层，仅仅使用卷积层，网络的整体框架如图2.6所示[69]，网络整体分为编码器和解码器两部分，编码器由5层卷积层构成，编码器由5层卷积层够成，解码器则为编码器的逆向操作，把下采样层改为上采样层，并且在最后加入Softmax层对像素进行分类。

图2.6 全卷积网络结构

Fig. 2.6 Full convolutional network structure

相比较Alexnet，网络的卷积层参照于VGG16采用小尺寸卷积核，卷积核的大小均为3，步长均为1。下采样的感兴趣区域设计为2×2，步长为2；减小卷积核以及池化区域的大小，可以增加特征图的分辨率，使训练过程中的特征图保留更多的图像信息。并且在卷积层的使用批规范化的（BN，Batch Normalization）方法对卷积的结果进行处理，这种方法主要解决卷积神经网络在训练中出现的internal covariate shift[70]问题，internal covariate shift问题主要是在神经网络小批量训练过程中，随着输入数据的变化，网络各层参数更新去满足新的数据分布，这会增加训练难度，使网络难以拟合。而BN将样本考虑在内，一定程度上解决了该问题，BN方法的公式如.1所示：

 （2.11）

为数据输入，为数据输出，和是与样本数据相关的参数，为样本的均值和方差，和为可训练的参数分别是缩放和平移。

解码器为编码器的逆操作，并且将下采样层修改为上采样，其中上采样的具体原理如图2.7所示，上采样将特征图中的值赋给最上角值，其他值补充为0。上采样将编码器缩小的特征图转变与原图大小一致的图像，在一定程度上还原了图像的特征，增加了SoftMax分类的精度。

解码器最终输出96个特征图，之后根据分类的个数使用卷积核为1的卷积进行降维，使用SoftMax计算属于每一类的精度。SoftMax公式为2.12：

 （2.12）

表示第类的输出值，表示属于该类别的概率。表示该像素点的第类的输入值，即像素对应的每一类的得分。



图2.7 最大池化和反池化原理

Fig. 2.7 Principle of Max-pooling and unpooling

解码器最终输出96个特征图，之后根据分类的个数使用卷积核为1的卷积进行降维，使用SoftMax计算属于每一类的精度。SoftMax公式为2.12：

 （2.12）

表示第类的输出值，表示属于该类别的概率。表示该像素点的第类的输入值，即像素对应的每一类的得分。

### 2.4.2 AgentNET:提取交通参与者的历史轨迹信息

AgentNET接收机动车的历史状态作为输入，利用LSTM提取其历史轨迹的序列特征。对于每个智能体，模型首先使用多层感知器(Muti-Layer Perception，MLP)将输入特征嵌入到更高维的行人中，然后使用LSTM对这些行人状态进行编码，编码生成行人的特征向量。

1. MLP
2. LSTM

### 2.4.3 Interaction：提取交互信息

在本小节中，我们提出了一种新的网络来结合从AgentNet和MapNet中提取的信息。如上文所述，机动车的行为取决于周围环境、其他障碍物以及整个地图的相互交互作用。在之前的工作[10]、[11]、[22]、[23]中，他们要么使用排列不变的对称函数，如max、mean或gumbel池化函数，要么使用排序函数（sort function），如基于欧几里德距离[13]的排序函数等。通过使用这些方法，模型可以很好地处理与其他智能体之间的交互。但这些方法要么需要舍弃一些可能对交互而言非常重要的特征信息，例如要么需要设置场景内智能体的最大容纳数量来保持数据维度的一致性，要么需要引入先验偏差，如使用欧氏距离来衡量交互强度[16]。

注意力机制类似于人类视觉注意机制。针对任务目标从大量信息中获得的最有效和最关键的信息。现在注意力机制已被广泛用于语音识别和机器翻译等自然语言处理领域。借助编码器-解码器模型[26]，在相关领域[27]中取得了良好的结果。本文中的注意力模块类似于人类的视觉注意力机制，它更加关注感兴趣的信息和特征。正如人类在观察图像时，不会一次性就了解所有的图像信息，而是首先关注图像的局部重要的特征。与观察图像类似，机动车驾驶司机在处理复杂的交互场景时，通常会结合当前状态和周围环境的影响，并专注于对他们影响较为重要的部分，以便快速做出相应的决策和改变机动车的行驶轨迹[2]。

在本文中，我们创新性地提出了一种处理智能体与环境之间相互作用的方法。如图2所示，Interaction-AttNet接收AgentNet的输出和MapNet的输出，整个模型由两个网络模块组成，第一个模块是通过接受MapNET的输出特征图进行仿射变换来提取以行人为中心的特征。然后使用多头注意力网络[27]来组合仿射特征和序列特征。

其中是由多尺度点积注意力模块（Multi-scale dot Attention Modul）组成的具有多头注意力机制[27]功能的模块，该模块已经在NLP领域中取得了不错的成绩和效果。对于模块，我们使用它来整合一定区域内智能体的特征信息。和主观印象一致，整张地图的信息对于单个智能体来说大多是无用的，我们只会关注在某一区域对自己有影响的信息。例如，行人不关心另一个人行道上的行人，车辆不会关注于远在另一个车道上的车辆，所以基于此我们设计使用仿射变换网络来提取一个固定大小的特征映射，通过一个以智能体为中心的，使用智能体转向角为旋转角度的仿射矩阵将整个特征映射到交互感兴趣区域内,并使用注意模块提取感兴趣区域的交互特征,使用特征作为包含时空信息的最终编码特征，然后使用提取到的特征输入预测网络内Multiple-PredictionNET进行解码。

### 2.4.4 Multiple-PredictionNET

如2.2.2所述，在本文中，我们将不确定性的概念分解为独立的两个独立的量，分别是意图不确定性和控制不确定性。以注意力机制处理后的机动车特征作为输入，Multiple-PredictionNet输出最终的行为预测，其中包括多条轨迹，其概率分别与意图不确定性和控制不确定性相对应。换句话说，对于每个智能体来说，模型将预测K个可能的未来轨迹及其置信度得分。

如图2所示，Multiple-PredictionNet具有两个分支，一个用于预测*K*轨迹的回归分支，该分支在每个时间步中生成描述双变量高斯分布的*K×T×5*参数（参数有）和分类分支来预测每条产生*K×T×1*个参数的轨迹，描述*K*个softmax来表示。

## 2.5 非机动车多模概率化轨迹预测模型

在2.4节中，本文对基于State-Anchor的机动车轨迹预测方法进行了建模，接下来，我们将介绍对于复杂交通场景中非机动车的轨迹预测模型的建模。相比于机动车，非机动车在道路交通场景中的机动性，随机性都比较大，且受个人主观意志影响较多。现有的对于非机动车轨迹预测的相关研究比较少，但是作为复杂道路交通场景中重要的交通智能体之一，准确的预测非机动车的轨迹对于道路交通安全，实现更高级别的自动驾驶研究都具有深远的意义。

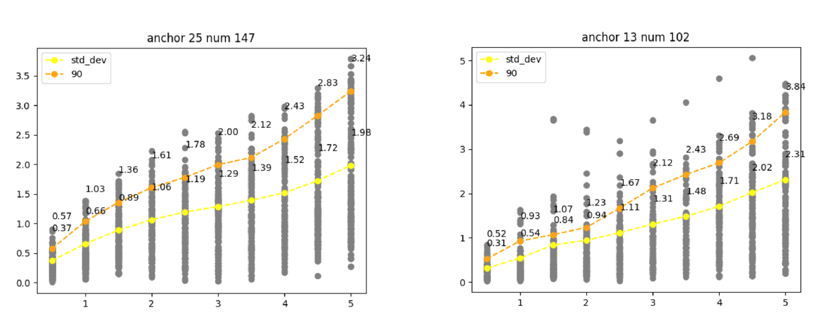
受启发于上文机动车的轨迹预测的相关研究技术，课题组研发了基于State-Anchor

的非机动车轨迹预测模型，在本章节中，首先我们对于非机动车的State-Anchor进行分析，确保其是否具有与机动车类似的Anchor类别，然后基于此开发非机动车的轨迹预测模型。

### 2.5.1 非机动车轨迹预测模型验证分析

在本章节中我们采用和上文一致的聚类方法，使用使用K-means聚类算法对轨迹进行聚类，获得轨迹之间平方距离的近似，

并对非机动车Anchor取值分别为25以及13时，对anchor的聚类效果进行数据分析，从图中可得当anchor取值25、13时，其均值和90分位非机动车的轨迹仍具有很好的聚类性质，且Anchor的数目为25时的优于Anchor为13时的效果。这是由于虽然非机动车的随机性要比机动车大，但是非机动车在复杂交通场景中仍然要受到交通规则和道路交通场景的限制，相比较于机动车使用38个anchor聚类结果，非机动车的Anchor数目要比机动车使用的Anchor数目少，推测是由于非机动车所能行驶的车道数目和采取的交通行为要比机动车少。基于此对于非机动车的轨迹预测，我们可以使用基于State-Anchor的方法对非机动车轨迹预测问题进行数学建模。



### 2.5.2 非机动车多模概率化轨迹预测模型



## 2.6 网络训练（训练参数，损失函数）

机动车与非机动车多模概率化轨迹预测模型的训练过程包括两个部分：分别是前向传播和模型参数的反向传播。前向传播是将模型的输入数据经过网络模型参数的计算，激活等操作，获得模型的输出，这一部分由评分函数构成（即前文所述的意图不确定性和控制不确定性函数，，），在本节中，我们设计用来训练我们的多模概率化轨迹预测模型的损失函数。由于整个模块是可微的，所以我们可以用端到端的方式训练机动车和非机动车模型。我们通过深度学习的方法训练机动车和非机动车模型，拟合网络模型的参数。通过使用负对数似然的方法去优化网络模型的输出，构成损失函数。损失函数通过使用最优化理论进行优化，使得损失函数的输出收敛到零，或接近于零。

现有的损失函数构建方法有很多，包括0-1损失函数，二次代价函数（平方差损失函数/L2损失函数），绝对值损失函数（L1损失函数），对数损失函数，指数损失函数，hinge损失函数（合页损失函数）等方法。在本文中我们基于对数损失函数的负对数似然函数进行损失函数的设计，对数损失函数的定义如下：

其中为网络模型的输出。

在2.2节中，我们将机动车与非机动车的轨迹预测问题，使用数学模型的方式表述了出来：

基于此我们使用对数似然函数，推导机动车与非机动车的损失函数：

其中式 中，符号为0-1函数，是与真实轨迹（Ground Truth，GT）最为接近的Anchor类别轨迹，使用L2范数测量二者之间的距离。

此外，经过模型训练结果分析，发现模型对于机动车低速段的拟合效果较差，进行相关数据分析发现，对于机动车而言，低俗段的数据量相比较于中高速段的数据量较少，于是对于机动车在现有的损失函数的基础上强化低速段的Loss，受启发于图像识别领域样本不均衡时的处理方法，在机动车的Loss函数中引入Focal Loss进行机动车损失函数的改进和优化：

在推导和改进完机动车与非机动车的损失函数后，我们使用Adam优化方法对模型参数进行反向传播优化，Adam算法结合了Momentum和RMSprop优化算法，通过引入一、二阶矩估计达到极佳的优化效果，是现如今最为常见的网络模型的学习优化算法。

## 2.5 网络模型的训练

本文模型架构基于TensorFlow搭建，网络结构的MLP层中的FC层之后都与L1正则化层和激活函数ReLU层相连。模型采用(2\*64)嵌入层进行轨迹编码。LSTM网络中隐含层单元数设为16，最小批量大小为64，优化器的学习率为0.000 5，训练次数150k次。其中计算机操作环境如表2.1所示。对于机动车多模概率化轨迹预测模型和非机动车多模概率化轨迹预测模型的可视化预测效果统一在接下来的章节中介绍。

表2.1 计算机硬件和软件配置

Tab. 2.1 Computer hardware and software configuration

|  |  |
| --- | --- |
| 项目 | 配置 |
| CPU | Intel Xeon E5-2620 |
| RAM | 32GB |
| GPU | NVIDIA TITAN X |
| 操作系统 | Ubuntu 14.04 LTS |
| Cuda | Cuda 8.0 with CuDNN v5 |
| 数据处理 | Python 2.7, OpenCV, etc. |

## 2.7 本章小结

本章首先对卷积神经网络的基本原理进行介绍，在此基础上对项目组提出的紧凑型全卷积神经网络的特点进行分析介绍。最后，在课题组提出网络的基础上，使用彩色图与深度图相融合作为卷积神经网络输入，并且实验验证了深度图有助于提高语义分割的精度。

4 基于Anchor-Free的多模概率化行人轨迹预测模型

## 4.1 引言

行人运动的数据分析对于道路安全、机器人导航[1]以及安全监控等领域具有重要的意义。研究行人轨迹需要收集行人的数据并进行离线分析，了解行人行为和周围环境并以此做出合理决策。在具有实时决策功能的系统中，对行人的行进路线进行预测，可以尽早发出警报或者采取相应的预防措施。

轨迹预测问题可视为序列决策问题，即根据行人过去时刻的位置预测未来时刻的轨迹。但该问题非常复杂。首先，行人的运动具有一定的随机性，在预测任务中生成一条确定性轨迹是不符合实际的。其次，每个行人并不是独立存在的，根据Mehdi等[2]的研究， 70％的行人倾向于成群行走，他们在同一时空下进行交互，这使处理该问题变得更加困难。

在上一章节中，我们对机动车和非机动车的轨迹预测模型进行了系统的建模，由于机动车以及非机动车受道路交通场景和交通规则的限制比较明显，在进行充分且有效的模型验证分析后我们设计使用State-Anchor技术，通过聚类分析的方法，对机动车和非机动车轨迹预测问题进行建模。但是与机动车和非机动车不同，行人作为交通场景中最为灵活，随机的智能体具有如下的特点：

1. 人际关系：每个行人虽然都是独立的个体，但是据研究表明[]，当人们在公共场所活动时，他们经常与其他行人进行交互，从而避免与其他行人碰撞或行人趋向于成群行走，且行人行走移动时，具有多种交互方式。
2. 物理场景：行人的行为不仅取决于周围的行人，而且还高度取决于周围的交通场景。这不仅包括无法躲避的障碍物（例如建筑物，栅栏等）和视觉提供的不同道路交通元素（例如人行道，红绿灯等），而且还包括与周围智能体的交互（与机动车或非机动车的交互）。这些因素可能都会影响和限制人类的活动。
3. 多模态：行人虽然没有像机动车和非机动车可以使用Anchor进行建模，但是行人在移动行走的过程中，仍然可能遵循多种合理的轨迹。例如行人之间并排行走时，行人可能会远离该行人，也可能会继续按原轨迹行走。

针对行人轨迹预测领域出现的这些特点，本章设计使用了一种新型模型建模方法，对于复杂道路交通场景的行人轨迹具有良好的预测效果，接下来，本章节将对该方法展开叙述。

## 4.2 行人轨迹预测模型建模

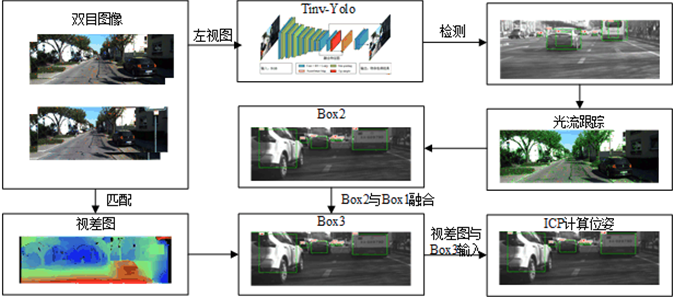


图4.1 系统框架

Fig. 4.1 System pipeline

如上文所述，在前文中我们将机动车和非机动车的模型不确定性分解为两个独立的量，智能体的意图不确定性和控制不确定性，进行联合建模，但是由于非机动车anchor聚类效果不理想，所以在进行模型建模时，我们设计并使用基于Anchor free的分类方法来表示智能体的意图不确定性，其中用*K*表示为要分类的轨迹序列类别的总数，使用表示轨迹状态序列的类别。该分布是基于Anchor-free分类轨迹参数化的，对模型的网络输出经过Softmax层得：

其中，*m*为模型的输入的第*m*个行人，*cls*表示行人多模轨迹预测网络模型的意图不确定性的分类分支。

接下来，对行人的控制不确定性进行建模，如上文所述，在已经得知智能体的意图不确定性后，控制不确定性依赖于每条分类轨迹所得出轨迹预测点的双变量高斯分布（GMM）：

其中，为第*k*个分类轨迹的预测序列，为每个Anchor轨迹点的双变量高斯分布的参数，代表着智能体的控制不确定性。这使得模型在不同的道路交通场景的静态

分类分支的基础上，可以考虑智能体动态的控制不确定性。例如在行人横穿的人行道时，考虑与其他行人或者场景的交互。

综上所述，为了获得整个状态空间的分布，在综合考虑行人的意图不确定性和控制不确定性后，将模型建模为：

## 4.3 行人多模轨迹预测模型网络结构设计



由网络结构模型图 可以得知，行人多模概率化轨迹预测模型的整体架构与机动车和非机动车的多模概率化轨迹预测模型基本一致，都是基于Encoder-Interaction-Decoder的架构对整个模型的网络进行设计，其中MapNET以上文所创建的复合栅格地图作为输入，经过MobileNET-V2网络提取输入信息特征图的特征信息，最后通过FCN网络实现特征图的尺度不变和特征信息的进一步提取，将特征从低维空间映射到高维空间；与此同时AgentNET网络输入行人的历史轨迹序列信息，使用行人的历史轨迹信息作为输入（位置，转向角，速度），利用LSTM提取时间序列信息的特点，经过LSTM网络提取历史序列轨迹信息；接着将这两部分的输出放入到交互注意力模块内，提取以行人为中心的小图特征信息以及对于该特征图的注意力信息；最后输入到预测模块，对行人轨迹进行预测。如上一章节介绍MapNET，AgentNET以及交互注意力模块的基本构成一致，唯一区别的模块是预测模块。在接下来的小节中，我们将着重介绍这一模块。

### 4.3.1 行人轨迹预测模型预测模块的设计

行人多模概率化轨迹预测模型的预测模块以交互注意力模块的输出作为输入，输出行人最终的运动预测。在该模块中，对于每一个行人，轨迹预测模块需要能够预测k条可能的轨迹以及每条可能轨迹的得分以实现行人轨迹的多模预测，为了实现多模概率化行人轨迹预测，我们设计了predictionNET，该模块有两个分支组成，一个回归分支用来预测每个模式的轨迹；一个分类分支，用来预测每个模式的置信度。



对于第m个行人来说，我们在回归分支中使用MLP对交互注意力模块的输出进行解码，其中交替使用残差块和线性层，对所要预测的k条轨迹进行回归计算，模型输出双变量高斯分布的五个参数。

其中，是回归网络所有结构的权重，为第m个行人第k条轨迹的第t时刻回归分支预测输出的五个参数：

对于分类分支，将回归分支的的均值以及交互注意力模块的输出作为分类分支的输入，经过MLP和残差块，softmax激活函数，输出每一个分支的置信度。

其中，是分类分支所有网络结构参数的权重。

### 4.3.2 行人多模概率化轨迹预测模型损失函数设计

与上文不同，机动车与非机动车模型使用的损失函数仍然是通过负对数似然函数进行构造，即：

为解决机动车低俗段loss的问题，引入了focal loss：

然而与基于anchor的机动车和非机动车轨迹预测问题不同，在进行行人轨迹预测建模时，我们使用了两个分支对行人的行为进行了预测，即分类分支和回归分支，与之类似，由于模型的每个模块都是可微的，对模型我们可以使用端到端的方式进行训练，为此我们设计了分类和回归联合损失函数去构建模型的损失函数：

其中，为权重参数。在进行分类损失计算时，针对于不同的分类类别寻找一个与GT的终点位移误差（FDE）最小的轨迹类别作为模型预测值与真值之间的差距，使用max-margin 损失构建损失函数：

其中，为一个很小的常数，用来保证损失函数数值的稳点性，M为场景中需要预测的行人总数。而对于回归损失，为了避免使用简单的L1损失函数，有不可导点的缺陷，我们使用Smooth-L1损失函数优化所有的轨迹预测序列：

其中，为第m个行人在第t时刻的真实轨迹点，

由smooth L1函数定义：

其中为L1距离。

## 4.4 网络训练



图4. 3 跟踪线程流程图

Fig. 4.3 Tracking flow chart

本文模型架构基于TensorFlow搭建，网络结构的MLP层中的FC层之后都与L1正则化层和激活函数ReLU层相连。模型采用(2\*64)嵌入层进行轨迹编码。LSTM网络中隐含层单元数设为16，最小批量大小为64，优化器的学习率为0.000 5，训练次数150k次。其中计算机操作环境如表2.1所示。对于机动车多模概率化轨迹预测模型和非机动车多模概率化轨迹预测模型的可视化预测效果统一在接下来的章节中介绍。本章所有的算法均在Jetson开发板上进行开发测试，在保证实时性的前提下尽可能提高算法的精度。其Jetson Tx2的软件硬件配置如表4.1所示。

表4.1 Jetson Tx2软件硬件配置表

Tab. 4.1 Jetson Tx2 Computer hardware and software configuration

|  |  |
| --- | --- |
| 项目 | 配置 |
| CPU | Dual-core NVIDIA Denver 2 64-bit |
| Memory | 4GB 128-bit LPDDR4 |
| GPU | NVIDIA PascalTM Architecture |
| 操作系统 | Ubuntu 14.04 LTS |
| Cuda | Cuda 8.0 with CuDNN v5 |
| 数据处理 | C++ 11, OpenCV, PCL 等. |

## 4.7 本章小结

本章在SLAM进行地图构建的基础上，为改善SLAM中容易受到运动物体的影响，提出了基于检测和跟踪融合的移动物体检测方法，并对其身份进行确认。同时基于双目立体视觉使用ICP方法对移动物物体的状态进行求解，为智能车进行自适应巡航等智能决策奠定基础。首先，研究了Tiny-Yolov3物体检测算法的原理，分析了Tiny-Yolov3在Tx2上工程化应用优缺点；然后，使用光流跟踪方法与Tiny-Yolov3算法融合消除Tiny-Yolov3漏检、重检问题；最后在光流跟踪的基础上，使用ICP算法对检测车辆进行位姿计算。

5. 复杂交通场景中多任务多模概率化轨迹预测模型

## 5.1 引言

目前为止我们已经对复杂道路交通场景中所有的交通智能体（机动车，非机动车和行人）都进行了轨迹预测模型的建模，并通过State-Anchor和Anchor-free的方法，完成了各自轨迹预测网络模型的搭建，但是在实际应用场景中，特别是在自动驾驶复杂的道路状况的场景中，将不同种类的智能体分开进行预测，这对于自动驾驶系统本就紧张的算力，无疑是雪上加霜的存在，所以在本章中，我们希望通过一个网络模型，对场景中所有类别的智能体实现多模态的轨迹预测，并能够有很好的精度和效率，进来对整个自动驾驶系统都产生深远的影响。为了将机动车，非机动车和行人的多模轨迹预测模型融合在一起，必须对模型的输入，交互注意力机制的设置以及模型的输出乃至模型的多任务损失函数进行全方位的设计，在接下来的内容中，我们将着重对这些内容进行详细的阐述，并对各智能体单独模型的预测结果进行定量和定性的分析，与多任务多模概率化轨迹预测模型的结果进行详尽的对比和可视化分析。

## 5.2 多任务网络模型结构设计



多任务多模概率化轨迹预测模型由四部分组成，首先是MapNET，用来将建模好的符合栅格地图进行处理，对复合栅格地图内的所有智能体（机动车，非机动车和行人）同时进行处理，通过MobileNET-V2提取栅格地图的初始特征图和并通过FCN模块进一步提取模型内交通物理场景的拓扑关系，用来表征地图上所有智能体的历史序列信息和交通环境（车道，人行道，红绿灯线等），最后输出一个和输入地图大小一致的特征图：

使用这种网络模型的一个优点就是所有种类的智能体的特征信息都经过同一个处理符合栅格地图的网络结构，并将特征集合在特征图中，通过共享特征图的信息，可以省去单独逐类别的进行复合栅格地图的特征提取和编码，提高模型预测多种类别智能体轨迹的速度与效率。在处理好复合栅格地图的输入后，接下来，在集中处理交通智能体的复合栅格地图的同时，并行的处理各智能体的历史序列轨迹信息，如上文所述，我们将每一个智能体的历史轨迹点信息、速度信息和各智能体的转向角信息拼接成一个时序向量Vector输入到AgentNET网络内，通过MLP将低维的序列数据特征映射到高维，最后通过LSTM模块提取各智能体的时序特征：

截止到现在，我们已经对复合栅格地图的特征信息和各个类别的智能体分别进行了特征信息的编码，完成了模型建模的Encoder阶段。

在拿到模型编码的特征信息和后，网络模型开始着手进行特征的融合和交互，实现对场景内智能体的交互建模，考虑到模型的效率和速度问题，我们采用共享特征图的建模方法，针对每一个智能体，使用仿射变换网络（STN），截取一定区域的特征图，作为该智能体的感兴趣区域，例如，我们在驾车行驶在一条道路上时，我们只需要注意周围环境和场景，而无需关注另一条车道的远处行驶而来的车辆，所以我们设计了放射变换模块，以实现感兴趣区域的提取，对于每一个不同种类型的智能体，我们在设计时，需要充分考虑每一种智能体的特点，实现每一种智能体对不同大小感兴趣区域的提取，在本文中，针对机动车，非机动车和行人，我们设计使用不同大小的截取区域，以满足不同种类型的智能体。

经过STN仿射变换网络截取后的感兴趣区域，与编码后智能体轨迹历史序列轨迹信息，使用缩放点积的多头注意力机制进行轨迹与特征图的注意力特征提取，得到经过注意力提取后的注意力特征。

通过使用注意力机制，可以获取加权过后的注意力特征向量，对于复杂道路交通场景，这表明着在各智能体感兴趣的区域内，智能体对何种信息更为敏感，哪些特征对于智能体做出精准预测有着不可或缺的作用。截止到现在，我们完成了Interaction模块的建模，接下里开始建模Decoder模块。

最后，将经过注意力网络处理过的特征向量，输入PredictionNET模块，实现行人多模概率轨迹的预测。如上文所述，预测模块由两部分组成，一部分是智能体的控制不确定性，一部分是智能体的意图不确定性。而对于机动车，非机动车以及行人来说，由于对于不确定性建模的方式不同，这也导致了Prediction模块的网络模型的搭建方式不同。首先是机动车和非机动车，由于这两类智能体的建模方式是采用State-Anchor的建模方式，对于不确定性而言，不同的Anchor代表着不同的行驶意图，而对于采取哪条，或者哪几条Anchor轨迹进行预测，需要对建模好的所有Anchor意图进行意图的分类预测，对于选择好的Anchor轨迹，需要在基于分类好的Anchor轨迹上进行控制不确定性的回归预测，拟合出预测目标轨迹所需的所有参数。其次是行人的预测，由于行人的高机动性和灵活性，无法使用Anchor表征行人的轨迹模式，如上文所述，我们设计使用两个分支来进行行人的轨迹预测，一是分类分支，分类行人的行为模式，一是回归分支，回归拟合行人轨迹所需的GMM的参数。

综上所述，多任务多模概率预测模型，由3部分组成，对于Encoder和Decoder模块，各智能体所需要的参数基本一致，但是对于Decoder模块，由于建模方式本质上的不同，需要对行人，机动车和非机动车分别进行建模，进而对各个不同种类的智能体进行预测，这样既兼顾了模型的准确性，又使得对于复杂多模概率预测模型预测的速度和效率显著提升，实现模型的多任务使用。

## 5.3 多任务网络模型训练

### 5.3.1 多任务模型损失函数设计

### 5.3.2 训练参数的设定

## 5.4 实验结果分析

## 5.5 本章小结

# 6 总结与展望

## 6.1 总结

本文基于立体视觉，使用深度学习的方法对智能车交通环境建模进行研究，提出了一种基于立体视觉的语义SLAM方法，对智能车准确定位的同时，对智能车交通环境的语义信息、几何信息、行为信息进行建模计算，分别对应本文的第二章、第三章、第四章。主要工作成果：

（1）基于新型卷积网络对图像进行像素级别的语义分割，其全卷机神经网络抛弃全卷机层，同时使用BN技术减使网络更易拟合。此外，为提高语义分割的精度，引入几何信息视差图作为卷积神经网络输入，将视差图与彩色图融合构成RGB-D图像，实验验证了相较于RGB图像，以RGB-D图像为输入能够提高语义分割的精度。

（2）在获取语义信息的基础上，将几何信息与语义信息相融合，提出了一种基于立体视觉的语义SLAM系统，在对智能车进行精确定位的同时构建稠密的三维语义地图。本文针对校园场景基于Fpgalena双目相机构建数据集验证了算法回环检测有效性以及普遍性；基于KITTI数据集展示了语义地图构建的效果，所构建语义地图验证了算法的有效性。整个系统平均速度接近10FPS,100m路程内角度误差平均值为0.25°，位移平均误差为1.3m。

（3）针对SLAM中容易受到运动物体的影响，针对时间序列的图像，提出了基于检测和跟踪融合的移动物体检测方法，并且使用ICP方法对检测目标进行状态估计。基于交通场景图像实验对比验证相较于Tiny-Yolo，本文提出的算法有效的改善了Tiny-Yolo的漏检、重检问题。

## 6.2 展望

本文基于立体视觉与深度学习对智能车环境建模方法进行研究，对环境几何信息与意义信息相融合进行了初步探索，并且在其基础上对移动物体的姿态进行估计。语义信息和几何信息本质是互利共赢的关系，语义信息能够帮助几何信息，几何信息同时也能提高语义信息获取的精度。未来将继续对物体的语义信息、几何信息、行为信息进行探索，期望工作如下：

（1）本文对于语义SLAM仅在地图模式上进行了深入研究，而语义信息能够帮助SLAM减少对于几何特征的依赖程度、增强数据关联。未来将语义信息应用于SLAM之间的帧间匹配，提高其匹配精度，并且将语义地图应用于定位之中。

（2）为满足实时性需求，以Tiny-Yolo为基础的移动物体检测方法在精度上存在缺陷，未来将在满足实时性的同时，结合视差图及K-means聚类提高目标检测的精度。

参 考 文 献

1. Waymo company, “On the road to fully self-driving,” <https://news.ycombinator.com/from>? Siteo- oglea apis.com,
2. Zhang J , Singh S . Low-drift and real-time lidar odometry and mapping[J]. Autonomous Robots, 2017, 41(2):401-416.
3. Tipaldi G D , Arras K O . FLIRT - Interest regions for 2D range data[C]// IEEE International Conference on Robotics & Automation. IEEE, 2010.
4. Alismail H, Baker L D, Browning B. Continuous trajectory estimation for 3D SLAM from actuated lidar[C]// IEEE International Conference on Robotics & Automation. 2014.
5. Triputen S, Schreve K, Tkachev V, et al. Closed-form Solution for IMU based LSD-SLAM Point Cloud Conversion into the Scaled 3D World Environment[J]. 2017.
6. Mur-Artal R, Tardós J D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras[J]. IEEE Transactions on Robotics, 2017, 33(5):1255-1262.
7. Whelan T , Kaess M , Johannsson H , et al. Real-time large-scale dense RGB-D SLAM with volumetric fusion[J]. International Journal of Robotics Research, 2015, 34(4-5):598-626.
8. Zhang J , Singh S . Visual-lidar Odometry and Mapping: Low-drift, Robust, and Fast[C]// IEEE Conf on Robotics and Automation (ICRA). IEEE, 2015.
9. Kasyanov A , Engelmann F , Stuckler J , et al. Keyframe-based visual-inertial online SLAM with relocalization[C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.
10. Surber J , Teixeira L , Chli M . Robust visual-inertial localization with weak GPS priors for repetitive UAV flights[J]. 2017 IEEE International Conference on Robotics and Automation (ICRA) .2017:6300-6306
11. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481-2495.
12. Yabuki N, Nishimura N, Fukuda T. Automatic Object Detection from Digital Images by Deep Learning with Transfer Learning[M]. 2018.
13. Zebin T, Scully P J, Ozanyan K B. Human activity recognition with inertial sensors using a deep learning approach[C]// Sensors. 2017.
14. GB/T 7714Balaban S. Deep learning and face recognition: the state of the art[C]// Spie Defense + Security. 2015.
15. 侯一民, 周慧琼, 王政一. 深度学习在语音识别中的研究进展综述[J]. 计算机应用研究, 2017(8).
16. Liu, Weiwei. Video Face Detection Based on Deep Learning[J]. Wireless Personal Communications, 2018.
17. Cadena C, Carlone L, Carrillo H, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age[J]. IEEE Transactions on Robotics, 2016, 32(6):1309-1332.
18. Jorge Othón EsparzaJiménez, Devy M , JoséLuis Gordillo. EKF-based SLAM fusing heterogeneous landmarks[C]// International Conference on Information Fusion. IEEE, 2014.
19. Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon,"Bundle adjustment: a modern synthesis," in Vison algorithms: theory and practice, pp. 298-372, Springer, 2000.
20. Galvez-López D, Tardos J D. Bags of Binary Words for Fast Place Recognition in Image Sequences[J]. IEEE Transactions on Robotics, 2012, 28(5):1188-1197.
21. E Rublee , V Rabaud , K Konolige, and G Bradski "SURP: An efficient alternative to SIFT or SURF." IEEE International Conference on Computer Vision IEEE, 2012:2564-2571.
22. Horn B K P, Schunck B G. Determining optical flow[M]// Determining Optical Flow. 1980.
23. Mur-Artal, Raúl, and J. D. Tardós. "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras." IEEE Transactions on Robotics 33.5(2016):1255-1262.
24. Mur-Artal, Raúl, J. M. M. Montiel, and J. D. Tardós. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System." IEEE Transactions on Robotics 31.5(2015):1147-1163.
25. Scherer S A, Zell A. Efficient onbard RGBD-SLAM for autonomous MAVs[C]// IEEE/RSJ International Conference on Intelligent Robots & Systems. 2013.
26. Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM.[J]. IEEE Trans Pattern Anal Mach Intell, 2007, 29(6):1052-1067.
27. Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C]// IEEE International Conference on Robotics & Automation. 2014.
28. Engel J, Schöps T, Cremers D. LSD-SLAM: Large-Scale Direct Monocular SLAM[M]// Computer Vision – ECCV 2014. 2014.
29. Engel J, Koltun V, Cremers D. Direct Sparse Odometry[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, PP(99):1-1.
30. Abdelrasoul Y, Saman A B S H, Sebastian P. A quantitative study of tuning ROS gmapping parameters and their effect on performing indoor 2D SLAM[C]// IEEE International Symposium on Robotics & Manufacturing Automation. 2017.
31. 梁明杰, 闵华清, 罗荣华. 基于图优化的同时定位与地图创建综述[J]. 机器人, 2013, 35(4):500-512.
32. 钱波. 基于深度学习的交通场景理解方法研究[D].辽宁:大连理工大学 2018.
33. Fritsch J, Kühnl T, Kummert F. Monocular Road Terrain Detection by Combining Visual and Spatial Information[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(4):1586-1596.
34. Pinz A. Exploiting Temporal and Spatial Constraints in Traffic Sign Detection from A Moving Vehicle[J]. Machine Vision & Applications, 2014, 25(3):649-665.
35. 郭春钊, 山部尚孝, 三田诚一. 基于立体视觉平面单应性的智能车辆可行驶道路边界检测[J]. 自动化学报, 2013, 39(4):371-380.
36. Jazayeri A, Cai H, Zheng J Y, et al. Vehicle Detection and Tracking in Car Video Based on Motion Model[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2): 583-595.
37. Keller C G, Enzweiler M, Rohrbach M, et al. The Benefits of Dense Stereo for Pedestrian Detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4):1096-1106.
38. Prioletti A, Grisleri P, Trivedi M M, et al. Design and Implementation of a High Performance Pedestrian Detection[C]// Intelligent Vehicles Symposium. IEEE, 2013:1398-1403.
39. Chu W, Liu Y, Shen C, et al. Multi-Task Vehicle Detection with Region-of-Interest Voting[J]. IEEE Transactions on Image Processing, 2018, 27(1): 432-441.
40. Zhang S, Cheng D, Gong Y, et al. Pedestrian Search in Surveillance Videos by Learning Discriminative Deep Features[J]. Neurocomputing, 2018, 283:120-128.
41. Boujemaa K S, Bouhoute A, Boubouh K, et al. Traffic Sign Recognition Using Convolutional Neural Networks[C]// International Conference on Wireless Networks and Mobile Communications. IEEE, 2017:1-6.
42. Saini S，Nikhil S，Konda K R, et al. An Efficient Vision-based Traffic Light Detection and State Recognition for Autonomous Vehicles[C]// Intelligent Vehicles Symposium (IV). IEEE, 2017: 606-611.
43. Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." Communications of the Acm 60.2(2012):2012.
44. Simonyan, Karen, and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." Computer Science (2014).
45. Szegedy, Christian, et al. "Going deeper with convolutions." Computer Vision and Pattern Recognition IEEE, 2015:1-9.
46. Jegou S, Drozdzal M, Vazquez D, et al. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation[J]. 2016.
47. He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. 2015.
48. Zhao H, Qi X, Shen X, et al. ICNet for Real-Time Semantic Segmentation on High-Resolution Images[J]. arXiv:1704.08545, 2017.
49. He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1-1.
50. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems. 2015.
51. Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. 2015.
52. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2017.
53. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. 2018.
54. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. 2016.
55. Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[J]. 2014.
56. Lin G, Milan A, Shen C, et al. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation[J]. arXiv:1611.06612, 2016.
57. Civera, J., Gálvez-López, D., Riazuelo, L., Tardós, J. D., and Montiel, J. M. M.. "Towards semantic SLAM using a monocular camera." Ieee/rsj International Conference on Intelligent Robots and Systems IEEE, 2011:1277-1284.
58. Salas-Moreno, R. F., Newcombe, R. A., Strasdat, H., Kelly, P. H. J., and Davison, A. J.. "SLAM++: Simultaneous Localisation and Mapping at the Level of Objects." IEEE Conference on Computer Vision and Pattern Recognition IEEE Computer Society, 2013:1352-1359.
59. A Kundu,Y Li , F Dellaert , F Li , and JM Rehg l. "Joint Semantic Segmentation and 3D Reconstruction from Monocular Video." European Conference on Computer Vision Springer, Cham, 2014:703-718.
60. Koenderink, J. J., and A. J. van Doorn. "Affine structure from motion. " Journal of the Optical Society of America A Optics & Image Science 8.2(1991):377-385.
61. Wang Y, Loe K F, Wu J K. A dynamic conditional random field model for foreground and shadow segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 28(2):279-289.
62. Mccormac, J., Handa, A., Davison, A., and Leutenegger, S. "SemanticFusion: Dense 3D Semantic Mapping with Convolutional Neural Networks." (2016).
63. Whelan, T., Leutenegger, S., Moreno, R. S., Glocker, B., and Davison "ElasticFusion: Dense SLAM Without A Pose Graph." Robotics: Science and Systems 2015.
64. Zhang X , Su Y , Zhu X . Loop closure detection for visual SLAM systems using convolutional neural network[C]// 2017 23rd International Conference on Automation and Computing (ICAC). IEEE, 2017.
65. Boureau Y L , Bach F , Lecun Y , et al. Learning mid-level features for recognition[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010.
66. Jost Tobias Springenberg∗, Alexey Dosovitskiy∗, Brox T , et al. STRIVING FOR SIMPLICITY: THE ALL CONVOLUTIONAL NET[J]. Eprint Arxiv, 2014.
67. Walker J, Doersch C, Gupta A, et al. An Uncertain Future: Forecasting from Static Images Using Variational Autoencoders[C]// 2016.
68. Quan T M, Nguyen-Duc T, Jeong, Won-Ki. Compressed Sensing MRI Reconstruction using a Generative Adversarial Network with a Cyclic Loss[J]. IEEE Transactions on Medical Imaging, 2017, PP(99).
69. IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [J]. arXiv preprint arXiv:1502.03167, 2015.
70. Li L , Qian B , Lian J , et al. Traffic Scene Segmentation Based on RGB-D Image and Deep Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, PP(99):1-6.
71. 黄海洋. 基于立体视觉的车辆检测与运动估计方法研究[D].辽宁：大连理工大学 2016.
72. Urtasun R, Lenz P, Geiger A. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2012.
73. Cordts M, Omran M, Ramos S, et al. The Cityscapes Dataset for Semantic Urban Scene Understanding[J]. 2016.
74. 钱波. 基于深度学习的交通场景理解方法研究[D].辽宁：大连理工大学， 2018.
75. Abdel-Hakim A E, Farag A A. CSIFT: A SIFT Descriptor with Color Invariant Characteristics[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. 2006.
76. Bay H, Tuytelaars T, Gool L V. SURF: Speeded Up Robust Features[C]// European Conference on Computer Vision. 2006.

攻读硕士学位期间发表学术论文情况

1 刘智杰;赵一兵;李琳辉;张溪桐;周雅夫;;基于卷积神经网络的语义同时定位以及地图构建方法[J];科学技术与工程;2019年09期。核心期刊。（已录用）

2 Linhui Li, Zhijie Liu, Ümit Özgüner, Jing Lian, Yafu Zhou, Yibing Zhao. Dense 3D Semantic SLAM of traffic environment based on stereo vision [C] //2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 965-970. （本硕士论文第三章）

致 谢

三年的硕士生活已临近尾声，非常幸运能够快乐而又充实的度过研究生三年。对此我充满了感激。

感谢的我的家人，他们让我感受到爱的意义，给予我信任的力量，教会我坚持的韧性。他们总是无条件的给予我支持，无论是经济上还是精神上，让我得以依靠，不畏失败。让我成长为一个独立的、健康、健全的人。

我一直认为自由、宽容、有尊严的学习氛围是科研工作基本土壤。三年来，非常幸运能够被赵一兵老师和李琳辉老师指导，老师对学生的尊重、信任以及宽容让我能够自由、快乐的进行科研工作。在权力不平等的师生关系中，这些品质弥足珍贵，如钻石一般照亮我科研之路。亦是老师帮助我从一名本科生完成了一名科研工作者的转换，是他带我走入了充满乐趣的科研世界。对此，我都深深感谢。此外，同样对连静老师对我生活上的帮助表示感谢。

在科研工作中，乐趣与困难并存。非常感谢帮助我解决困难的师兄师弟。感谢黄海洋、钱波两位师兄，他们丰富的学识，严谨的学术素养，对科研无尽的好奇心为我树立了榜样，亦是他们对我的帮助，让我能够高效的理解应用立体视觉与语义分割这两个工具。同时，也非常感谢清华高翔博士，是他以生动活泼的方式带我走入SLAM的世界。

三年来，我一直走在探索自我，探索世界的路上。感谢与我一起探索生活价值的挚友们，感谢Debby，你总是在我自我怀疑的时候肯定我的价值，总是在我困惑的时候与我一同寻找前进的方向，感谢龙易小熊，你们的支持和陪伴给予了我巨大的力量。下次我们再一起与海共舞，与山的共曲。

感谢我生活的这块社区，赐予我好吃的海蛎子、虾爬子、新鲜的蔬菜；感谢磨盘山，它为我带来了无数的心灵平和；感谢咪耨二、妹妹两只猫，你们总是陪伴；感谢桥牌，剑道。这一切，让我这三年充满了意义。

最后，感谢我的舍友、同教研室的同学、研究生同学，感谢你们宽容我的坏脾气，陪伴着我成长。

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目：

作 者 签 名 ： 日期： 年 月 日

导 师 签 名 ： 日期： 年 月 日