硕 士 学 位 论 文

交通场景中多任务多模概率轨迹预测方法研究

Study on Multi-task and Multi-model Probabilistic Trajectory Prediction Method in Traffic Scenes

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： 交通场景中多任务多模概率轨迹预测方法研究

作 者 签 名 ： 日期： 年 月 日

摘 要

现阶段随着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。其中自动驾驶车辆对于提高交通安全、社会效益、经济效益具有重要意义，而多智能体的复杂道路交通场景中的轨迹预测是自动驾驶汽车中的一个重要问题。本文针对复杂的道路交通场景，基于State-Anchor和Anchor-free的建模思想，分别对机动车、非机动车和行人进行建模，并搭建多任务网络将三者融合，构建多任务多模概率轨迹预测模型，为自动驾驶车辆后续的行为决策和控制奠定基础。

首先，由于现有的行为预测数据集数据类别过于单一，无法有效表明交通场景间的物理拓扑关系和历史序列信息，且数据样本量太少，无法学习到交通场景中尽可能多的行为模式，针对上述问题。我们开发了一个包含道路高清地图和障碍物历史序列信息的大型行为预测数据集，可以对场景中的环境及交通智能体进行栅格化表示，为后续模型建模提供丰富的语义拓扑信息和历史序列信息。

之后，基于State-Anchor的建模方法搭建机动车和非机动车的多模概率化轨迹预测模型。首先通过对机动车和非机动车的轨迹进行聚类分析，对机动车和非机动车的意图不确定性和控制不确定性进行建模。之后采用Encoder-Interaction-Decoder的网络架构设计机动车和非机动车的网络模型，并完成模型损失函数的设计和参数的选用。最后，基于Anchor-free的建模方法搭建行人的多模概率化轨迹预测模型。

最后，在构建好上述三个模型的基础上，为提高模型在实车上预测的效率，提出多任务多模概率化轨迹预测模型以融合机动车、非机动车和行人模型，实现场景中所有智能体的实时预测。通过在空间变换网络中使用大小不同的感兴趣区域，结合使用多头注意力机制，有效的对交通场景中的复杂交互进行建模，实现对各类智能体的有效预测。文章最后将本文模型与当前主流的各单一类别的轨迹预测模型进行了定量和定性的分析比较，实证我们模型的性能要优于当前主流模型。并对模型预测结果进行了可视化分析，实验结果表明，多任务模型可以预测各智能体多条社会可接受的轨迹序列，为下游做出更为合理，平顺的决策打下基础。

关键词：不确定性建模；轨迹预测；多任务；多智能体；多模态

Study on Multi-task and Multi-model Probabilistic Trajectory Prediction Method in Traffic Scenes

Abstract

With the rapid development of artificial intelligence at this stage, it has had a huge impact on the automotive industry. The automotive industry is developing in a smarter, cleaner, and safer direction. And autonomous vehicles are of great significance for improving traffic safety, social benefits, and economic benefits, and trajectory prediction in complex road traffic scenes with multi-agents is an important issue in autonomous vehicles. Aiming at complex road traffic scene, this paper models vehicles, cyclists and pedestrians based on the State-Anchor and Anchor-free modeling ideas, and builds a multi-task network to integrate the three models to build a Multi-task Multi-modal probabilistic trajectory prediction model, and it will lay the foundation for the subsequent behavioral decision-making and control of autonomous vehicles.

First of all, because the existing behavior prediction data sets are too single in data categories, they cannot effectively indicate the physical topological relationship and historical sequence information between the traffic scenes, and the data sample size is too small to learn as many behavior patterns in the traffic scene as possible. Regarding the issue above. We have developed a large-scale behavior prediction data set containing road High-Definition maps and obstacle historical sequence information, which can rasterize the environment and traffic agents in the scene, and provide rich semantic topology information and history for subsequent model modeling sequence information.

After that, based on the State-Anchor modeling method, a multi-mode probabilistic trajectory prediction model for vehicles and cyclists is built. Firstly, through cluster analysis of the trajectories of vehicles and cyclists, the intent uncertainty and control uncertainty of vehicles and cyclists are modeled. Afterwards, the Encoder-Interaction-Decoder network architecture is used to design the network models of vehicles and cyclists, and then we will design the model loss function and select the model parameters. Finally, a multi-mode probabilistic trajectory prediction model for pedestrians is built based on the Anchor-free modeling method.

Finally, based on the construction of the above three models, in order to improve the efficiency and accuracy of the model's prediction on the actual vehicle, a multi-task and multi-mode probabilistic trajectory prediction model is proposed to integrate the vehicle, cyclist and pedestrian models to realize the real-time prediction of the agent in the scene. Through the use of different sizes of interest regions in STN, combined with the use of multi-head attention mechanism, we can effectively model the complex interaction in the traffic scene, and achieve effective prediction of various agents. At the end of the article, quantitative and qualitative analysis is carried out between the model in this paper and the current mainstream single-category trajectory prediction models. It is demonstrated that the performance of our model is better than the current mainstream model, and we also visualize the model prediction results. The experimental results show that the model can predict multiple socially acceptable trajectory sequences for each agent, laying the foundation for the downstream to make more reasonable and smooth decisions.

Key Words：Uncertainty modeling; trajectory prediction; multi-task; multi-agent；multi-modal

目 录

[摘 要 I](#_Toc69498414)

[Abstract II](#_Toc69498415)

[1 绪论 1](#_Toc69498416)

[1.1 研究背景和意义 1](#_Toc69498417)

[1.2 轨迹预测研究动态与现状 3](#_Toc69498418)

[1.2.1 基于浅层学习的轨迹预测方法的研究现状 4](#_Toc69498419)

[1.2.2 基于深度学习的轨迹预测方法的研究现状 6](#_Toc69498420)

[1.3 本文主要研究内容 12](#_Toc69498421)

[2 交通场景轨迹预测问题建模 14](#_Toc69498422)

[2.1 轨迹预测问题定义 14](#_Toc69498423)

[2.2 意图不确定性与控制不确定性建模 14](#_Toc69498424)

[2.3 数据集 17](#_Toc69498425)

[2.3.1 常用的轨迹预测数据集 17](#_Toc69498426)

[2.3.2 构建数据集 20](#_Toc69498427)

[2.4 深度学习相关基础知识 21](#_Toc69498428)

[2.4.1 卷积神经网络与MobileNet 21](#_Toc69498429)

[2.4.2 空间变换网络 26](#_Toc69498430)

[2.5 本章小结 28](#_Toc69498431)

[3 基于State-Anchor 的多模概率化机动车与非机动车轨迹预测模型 29](#_Toc69498432)

[3.1 引言 29](#_Toc69498433)

[3.2 机动车多模概率化轨迹预测模型 30](#_Toc69498434)

[3.2.1 MapNET：精确的时空表示 30](#_Toc69498435)

[3.2.2 AgentNET:提取交通参与者的历史轨迹信息 31](#_Toc69498436)

[3.2.3 Interaction-AttNET：提取交互信息 31](#_Toc69498437)

[3.2.4 Multiple-PredictionNET 32](#_Toc69498438)

[3.3 非机动车多模概率化轨迹预测模型 33](#_Toc69498439)

[3.3.1 非机动车轨迹预测模型验证分析 33](#_Toc69498440)

[3.4 网络模型的损失函数设计 35](#_Toc69498441)

[3.5 网络模型的训练参数 37](#_Toc69498442)

[3.6 本章小结 37](#_Toc69498443)

[4 基于Anchor-Free的多模概率化行人轨迹预测模型 38](#_Toc69498444)

[4.1 引言 38](#_Toc69498445)

[4.2 行人轨迹预测模型建模 39](#_Toc69498446)

[4.3 行人多模轨迹预测模型网络结构设计 40](#_Toc69498447)

[4.3.1 行人轨迹预测模型预测模块的设计 40](#_Toc69498448)

[4.3.2 行人多模概率化轨迹预测模型损失函数设计 41](#_Toc69498449)

[4.4 网络训练 42](#_Toc69498450)

[4.7 本章小结 42](#_Toc69498451)

[5. 复杂交通场景中多任务多模概率化轨迹预测模型 43](#_Toc69498452)

[5.1 引言 43](#_Toc69498453)

[5.2 多任务网络模型结构设计 43](#_Toc69498454)

[5.3 多任务网络模型训练 45](#_Toc69498455)

[5.3.1 多任务模型损失函数设计 45](#_Toc69498456)

[5.3.2 训练参数的设定 46](#_Toc69498457)

[5.4 实验结果分析 46](#_Toc69498458)

[5.4.1 实验细节 46](#_Toc69498459)

[5.4.2 多模概率化轨迹模型机动车与非机动车预测结果分析 48](#_Toc69498460)

[5.4.4 多模概率化行人轨迹预测模型的结果分析 57](#_Toc69498461)

[5.5 本章小结 60](#_Toc69498462)

[6 总结与展望 61](#_Toc69498463)

[6.1 总结 61](#_Toc69498464)

[6.2 展望 61](#_Toc69498465)

[参 考 文 献 63](#_Toc69498466)

[攻读硕士学位期间发表学术论文情况 68](#_Toc69498467)

[致 谢 69](#_Toc69498468)

[大连理工大学学位论文版权使用授权书 70](#_Toc69498469)

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，智能算法日渐成为目前人民日常生活中不可或缺的一部分。医院使用AI技术来对疾病进行诊断；自媒体短视频APP通过AI推荐算法获得巨大的流量，也在日益改变人们的日常生活，闲暇时间刷会儿视频成为现在社会生活中的常态。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，而且在未来几年仍将获得更加突飞猛进的发展[ ]。值得推敲的是，在广大AI技术的应用领域中，有一个并未完全受到人工智能影响的领域，即汽车领域。大型整车汽车制造商，如大众，通用等，虽然这些主机厂都在通过高级驾驶辅助系统（Advanced Driving Assistance System，ADAS）逐步提高人工智能技术在汽车领域的应用。但是根据目前由美国汽车工程师学会（Society of Automotive Engineers，SAE）车辆无人驾驶水平的等级划分，目前应用到量产车上的技术大多属于L2-L3级别，其中真正L3级别的量产车则少之又少，人工智能的全部技术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。众所周知，在交通中驾驶汽车是一项非常危险的任务，因为交通场景作为一个许多人的共同活动，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日益完善的ADAS系统来提高汽车的安全性，但严峻的统计数字表明，我们今后还有很多工作要做。

虽然无人驾驶技术是一项新兴技术，但是也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]，[15]中，各个参赛队伍被要求在复杂的城市交通环境中行驶，处理公共交通道路上遇到的常见情况，并且与人类或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研发团队和法律框架，使得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）子公司Waymo推出了无人驾驶汽车（图1.1（a）），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（c）），日本丰田推出了无人巴士（图1.1（b））；而在国内，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（图1.1（d））的服务，现在已经在北京市市区和湖南长沙运营；智行者科技推出蜗小白系列（图1.1（e）），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（图1.1（i）），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。蔚来汽车（图1.1（f）），小鹏汽车（图1.1（h））和理想汽车（图1.1（g））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。智能汽车正进入了蓬勃发展的时代。

（a） （b） （c）

（d） （e） （f）

（g） （h） （i）

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要解决“我现在在哪里？”、“我周围都有什么”、“我周围将来会发生什么”，“我接下来要怎么去做”四个问题，其中“我周围将来会发生什么”，意味着自动驾驶汽车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，在本文中，我们将围绕这一主题对道路交通复杂场景中多智能体的行为预测问题进行详尽的分析和研究。

## 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，这些方法对于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，并且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。

根据预测模型的不同建模方式，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，现阶段轨迹预测方法的分类如图1.2所示。



图1.2 轨迹预测方法分类

Fig. 1.2 Classification of trajectory prediction methods

实际上，一个好的轨迹预测模型往往需要综合利用不同的特征信息（包括轨迹信息，语义信息和社交信息等），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。

人类环境中出现了越来越多的智能自主系统，这些系统需要感知、理解和预测人类行为的能力变得越来越重要，具体来说，预测智能体（Agent）的未来位置并根据这些预测进行规划是自动驾驶车辆、服务机器人和高级监控系统（包括智能交通或者智能城市）的关键任务[1]。理解人体运动是智能系统与人类共存和互动的一项关键技能，其涉及表征、感知和运动分析等方面。而预测在人体运动分析中起着重要的作用：随着时间的推移，模型可对涉及多个智能体的场景进行预测，并以主动的方式对这些场景信息进行整合，即增强主动感知、预测性规划、模型预测性控制或人机交互的效果。因此，近年来行人轨迹预测在多个领域中受到越来越多的关注，例如自动驾驶汽车、服务机器人、智能交通、智慧城市等领域。保证交通场景中道路使用者的自身安全是自动驾驶车辆被普及应用的前提条件[2]。其中行人作为交通场景中的主要参与者，对其未来运动轨迹的合理推理和预测对于自动驾驶和道路安全至关重要。交通场景中各智能体轨迹的高动态性、随机性以及与交通环境智能体之间复杂的交互使轨迹预测问题充满挑战，但是对他们的轨迹进行长时间的预测仍是非常有必要，这对实现自动驾驶汽车的主动规划和决策[4]影响巨大。

### 1.2.1 基于浅层学习的轨迹预测方法的研究现状

最基本、也是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）和贝叶斯滤波器将其扩展组合起来，将当前状态传播到未来状态[5]。Schneider等[6]人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，这显然不符合真实的道路实际情况，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等人[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模型，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等人[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等人[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等人[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等人[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模型对于智能体运动过程的状态预测效果良好，但其对于轨迹的波动较为敏感，且不适用于中长期的智能体轨迹预测，一般一阶MM模型仅考虑了当前智能体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，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各智能体轨迹的预测，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。

上述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轨迹预测领域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

上一小节中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数据集和模型之间的映射关系。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。

近年来，用于历史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到同为预测领域的轨迹预测领域中。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，近几年的时序预测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，注意力模型随之诞生了，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和解码。在此基础上，Karatzoglou等人[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。

在交通场景的行人轨迹预测方面：基于先前的研究，Alahi等人[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，具体的建模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，该文章中同样提出了将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等人[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。文章[26]中也通过使用语义场景信息对行人-时空交互进行建模，并以此来推断场景中的可穿越区域和可通行区域，从而预测智能体的未来轨迹。因此，在O-LSTM的基础上，Xue等人[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等人[54]利用多头注意（Multi-head Attention）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（GMM）预测。PARK等人[58]使用注意力模块来融合各个部分输出的交互，并着重特别强调预测轨迹的多样性。该模型不依赖真实轨迹（Ground Truth，GT）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等人[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等人[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等人[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等人[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等人[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）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，但是其预测结果大多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，并且与数据集的数据分析结果高度拟合，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等人[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等人[31]通过引入Info-GAN架构来改进GAN模型的网络架构，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等人[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等人[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，分别使用基于缩放点积（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等人[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，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，相比于在噪声分布直接采样，文中的噪声分布为，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等人[36]也采用了这种方法。

Liang等人[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等人[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等人[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。

（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，例如矩阵计算和变分自动编码器。

Gupta等人[30]和Sadeghian等人[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智能体。但是这两种模型都无法学习人类驾驶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等人[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等人[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等人[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等人[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。通过将车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等人[56]提出一种同时融合智能体与智能体，智能体与车道，车道与车道，车道与智能体四种交互的模型。其中对车道与车道，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），以体现车道的上下文关系。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等人[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，数据集所涵盖的边界条件（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

上文对近年来轨迹预测领域的工作进行研究，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的提出，无论单独从精度还是效率上，基于深度学习的方法都比基于浅层学习浅层方法更为有效，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。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实际问题。

基于以上分析，本文以提高复杂道路交通场景中各智能体的行为可预测性为出发点，通过建立大型的行为预测数据集，对智能体的道路交通行为进行数据特征表示，为模型学习提供丰富的语义地图拓扑信息和历史轨迹序列信息，使模型不仅能够基于历史特征进行预测，还能够结合周围的环境进行合理的多模预测，为自动驾驶汽车在复杂道路交通场景中的决策和控制奠定基础。

本文创新性如下：

（1）针对现有的行为预测数据集数据类别过于单一和数据样本量过少的问题。本文开发了一个包含道路高清地图和障碍物历史序列信息的大型行为预测数据集，对场环境及交通智能体进行栅格化表示，为后续模型建模提供丰富的语义拓扑信息和历史序列信息。

（2）针对当前机动车和非机动车轨迹预测精度低且无法进行多模预测的问题，本文基于State-Anchor技术建模机动车和非机动车的多模概率化轨迹预测模型，实现对机动车和非机动车的多模概率预测。

（3）针对当前主流的行人预测模型需要依赖生成架构进行多模预测，且对于交互处理需要引入先验的问题，本文设计了基于Anchor-free的行人多模概率化预测模型，实现对行人的多模概率预测。

针对以上研究内容，本文文章整体分为六章，分别为：

第一章，绪论。该章主要从自动驾驶车辆的需要解决的问题入手，对轨迹预测任务的意义和重要性进行阐述。并分别研究了基于浅层学习的轨迹预测方法，和基于深度学习的轨迹预测方法的研究现状、研究难点以及所存在一系列问题，提出了本文的研究内容和论文框架。

第二章，交通场景轨迹预测问题建模。该章首先从轨迹预测问题的定义入手，对轨迹预测问题进行详细的阐述，并对轨迹预测问题的不确定性进行不确定建模；然后介绍了当前主流的行为和轨迹预测数据集，并针对数据集类别和数量单一性的问题，提出了大型行为预测数据集，为模型提供丰富的语义拓扑和轨迹序列特征信息。

第三章，基于State-Anchor的多模概率化机动车与非机动车轨迹预测模型。该章首先基于上文的不确定性建模思路进行机动车和非机动车的建模；之后设计搭建基于Encoder-Interaction-Decoder的网络结构，分别对网络的具体设计细节进行介绍；最后设计机动车和非机动车的损失函数和网络参数，并对其进行改进和优化。

第四章，基于Anchor free的多模概率化行人轨迹预测模型。该章首先将机动车与非机动车的建模思路和行人模型的建模思路进行了对比，验证模型使用Anchor建模的可行性；然后选用Anchor-free的技术对行人模型进行建模，完成行人模型的网络结构搭建；最后进行模型损失函数的设计和网络参数的选用。

第五章，复杂交通场景中多任务多模概率化轨迹预测模型。该章首先基于上述三章设计的单一智能体的轨迹预测模型，寻求模型融合预测的可能性，进而提高模型预测的精度和效率；然后设计多任务多模概率预测模型的整体架构，设计多任务模型的损失函数，完成对模型参数的选用；最后对上述三章模型的效果进行了系统的定量和定性分析，并将单一模型与多任务模型进行定量和定性比较，验证模型的有效性。

第六章，对全文的研究内容进行总结分析，在总结的基础上分析本文不足，以及对未来工作进行规划和展望。

# 2 交通场景轨迹预测问题建模

## 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的位置都由它们的*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。在本文接下来要提出的模型中，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，对于复杂道路交通路口（junction）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可以由四维向量编码，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），本文所提出的系列模型将通过学习模型的权重参数矩阵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，为模型预测时间步长。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。在接下来的章节中，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和基于Anchor-free的行人轨迹预测模型以及机动车、非机动车和行人多任务多模概率轨迹预测模型。首先在该章中我们将介绍如何捕获智能体的意图不确定性和控制不确定性以此作为多模概率预测的建模基础，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。再接下来，我们将系统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)。最后，我们将介绍模型建模搭建时所需要的深度学习的一些技术知识。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 意图不确定性与控制不确定性建模

如上所述，交通场景中的各类智能体都有着很大的不确定性和随机性，我们都无法通过主观判断各个智能体的意图和接下来的行文，何况是AI。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和控制不确定性。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，在交通场景的真实驾驶环境中，智能体尝试行驶到哪条车道上的不确定性。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，智能体将如何遵循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，转角以及与周围环境之间的交互等。

在模型的建模时，针对于交通场景中的机动车和非机动车，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，来简要的获得近似的轨迹之间平方距离，

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2.1所示，此为时机动车聚类结果的可视化分析图。

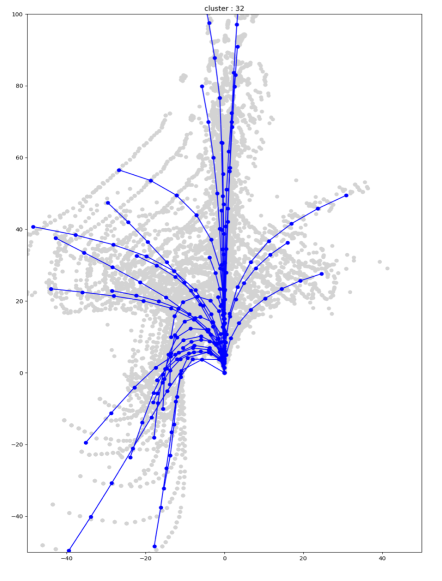


图2.1 轨迹预测方法分类

Fig. 2.1 Visual graph of vehicle clustering results when k = 32

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：

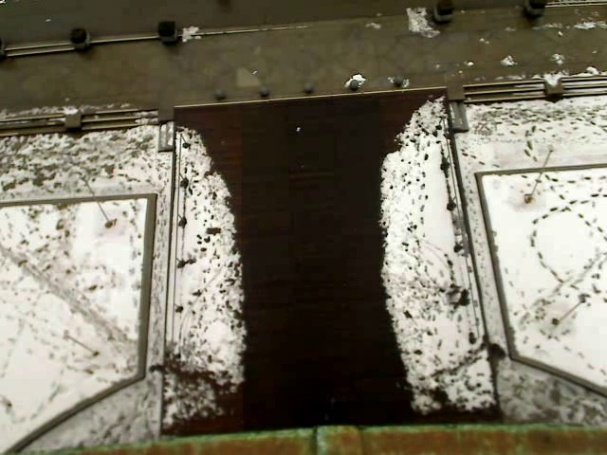
## 2.3 数据集

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。

### 2.3.1 常用的轨迹预测数据集

（1）ETH、UCY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两个子数据集组成，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。目前绝大部分行人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。

（a）ETH (b)HOTEL

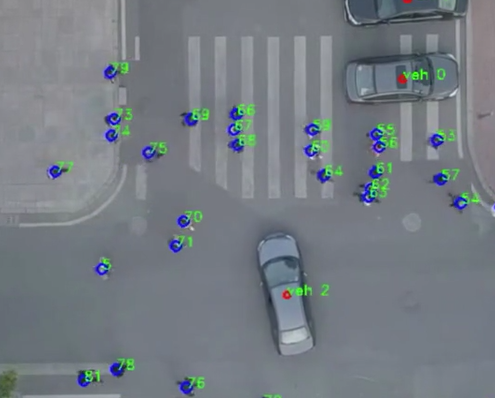
（c）ZARA-1 (b)ZARA-2

图2.2 ETH、UCY数据集

Fig. 2.2 DtaSet of ETH and UCY

（2）DUT数据集

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1920×1080，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1793条轨迹。

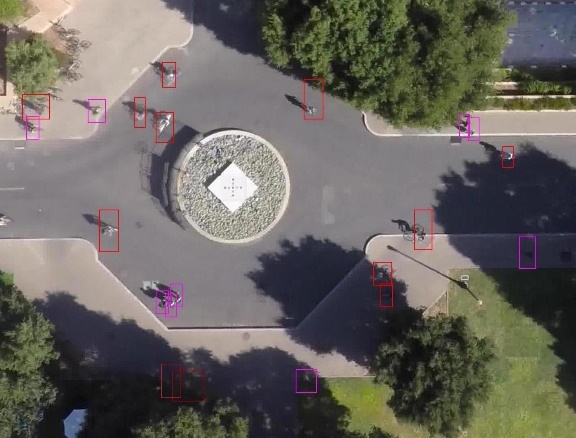
（a）无交通信号交叉路口 （b）环岛

图2.3 DUT无人机数据集

Fig. 2.3 DUT UAV Dataset

（3）斯坦福无人机数据集

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。

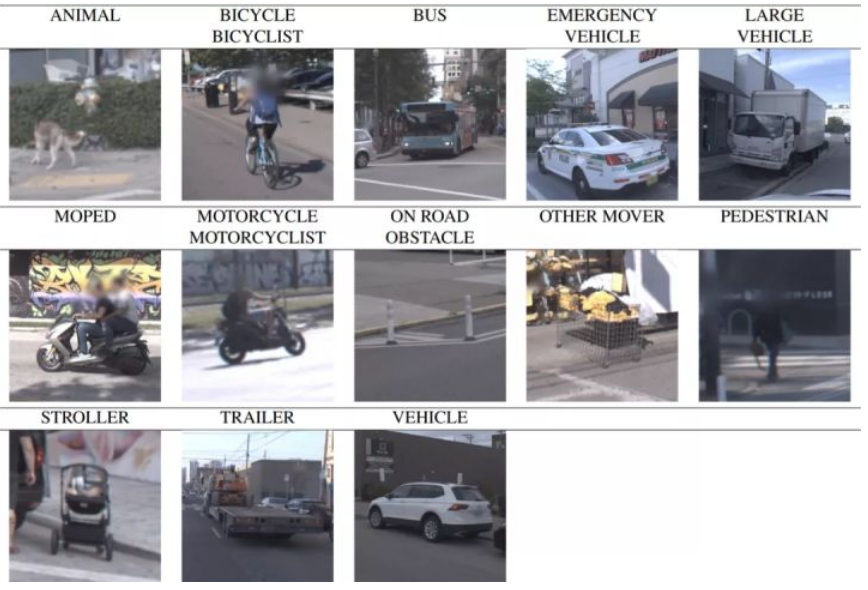
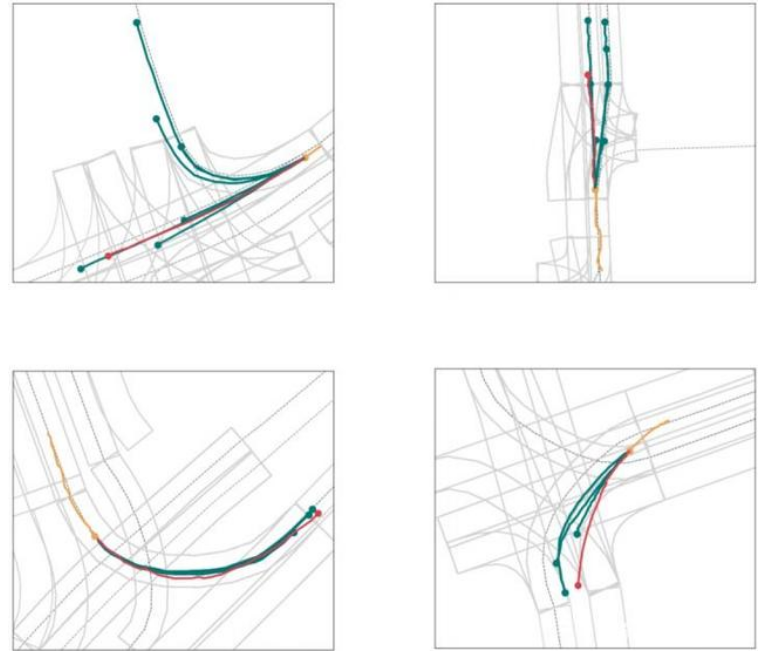
（a）quad场景 （b）death-Circle场景

图2.4 斯坦福无人机数据集

Fig. 2.4 Stanford UAV Dataset

（4）Argoverse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。

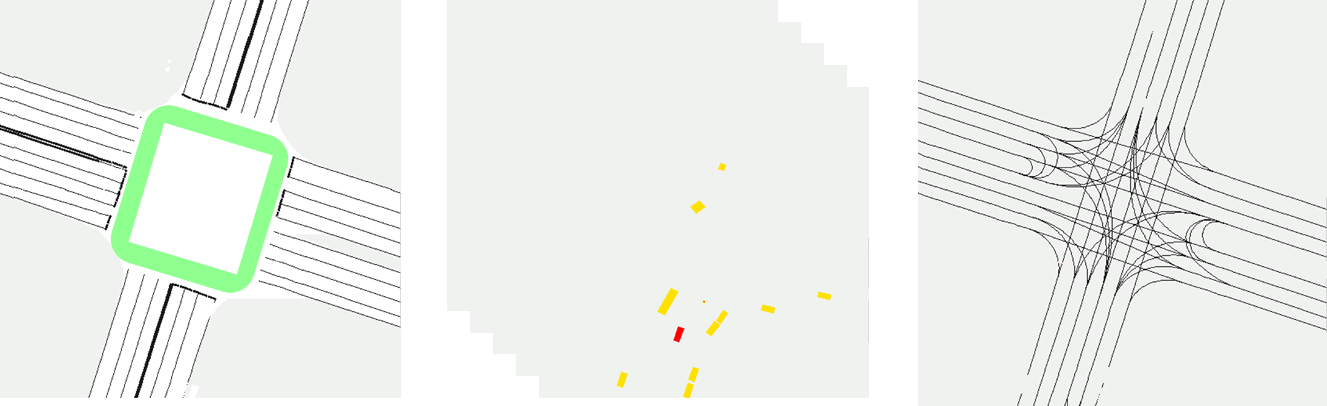
（a）3D轨迹跟踪/预测分类 （b）运动预测

图2.5 Argoverse数据集

Fig. 2.5 Dataset of Argoverse

### 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（Composed Raster Map），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以及序列历史轨迹特征组成。如图2.6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)高清静态地图 （b）智能体历史位置图 （c）动态红绿灯示意图

图2.6 复合栅格地图的组成

Fig. 2.6 Dataset of Argoverse

图2.6是光栅化渲染的高清地图和智能体轨迹的图示，其中（a）是高清静态地图，其中包含连接交通场景中的拓扑和语义信息，（b）是不同智能体的历史位置，包含5个通道，表示5帧，（c）是表示不同车道交通信息的动态交通灯图。本文上述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 深度学习相关基础知识

在正式开始介绍本文所提出的系列模型之前，我们将对模型建模设计所需要的系列技术进行简要的阐述。

### 2.4.1 卷积神经网络与MobileNet

卷积神经网络作为一种提取局部和整体之间相关性的计算单元，现在已经被广泛应用在图像识别，自然语言处理等领域，并且由于图像局部和整体的相关关系，卷积神经网络在图像识别领域（目标检测，语义分割等）获得了巨大的成功。卷积神经网络（CNN）通过将各个计算层排列起来，使用卷积的计算方法进行可微可导的函数计算，之后使用激活函数对输出的计算结果进行非线性激活，实现对于非线性问题的拟合。其中当前主流的卷积神经网络及其变体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，除此之外在不同的网络里，还相继提出了归一化等方法提高网络在不同任务上的性能。

卷积层作为整个卷积神经网络的核心，有着诸多的卷积计算方式，包括全卷积，局部卷积，转置卷积等，在下文的介绍中，我们以标准卷积为例进行讲解。若以图像为卷积的输入，卷积的过程其实通过卷积核在整个图像上进行滑动，来提取图像不同区域的特征信息，同时这也是卷积层区别与线性层局部连接的特性，其中卷积核的主要参数有卷积核（Filter）大小F、步长（Stride）S。卷积核大小是卷积层感受区域的大小，如图2.7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为在输入的特征图进行全零填充（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。



图2.7 卷积操作示意图

Fig. 2.7 Convolution operation diagram

卷积层在处理图像等二维数据时，因为卷积核的存在，采用局部连接和参数共享的技术，极大的减少了网络中的参数数量，同时局部连接可以更加有效的提取图像边缘特征，这对于图像而言尤其重要，提高了网络对于图像的处理能力。例如输入的图像大小为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75，参数数量以的速度极大的减少。



图2.8 池化操作示意图

Fig. 2.8 Pooling operation diagram

为减少计算机计算资源的消耗，进一步有效的减少网络模型的参数量，并且有效控制过拟合，设计了池化层。其中池化层现有的主流方法为最大（Max）池化和平均池化。池化操作的主要参数有池化窗口的大小和步长。最大池化主要通过池化窗口中最大值来表示这一个窗口内的特征信息，平均池化则使用平均值来表示特征信息。上图2.8中假设输入数据体的尺寸为4x4，感兴趣大小=2，=2，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。

ImageNet图像分类大赛是当前深度学习热潮的最为主要的推动者之一，如图2.9中在ImageNet比赛上涌现出了许多经典的卷积神经网络如VGG，AlexNet， GoogLeNet等。这些网络中使用的一系列技术都成为了后续卷积神经网络开发的重要技术参考。

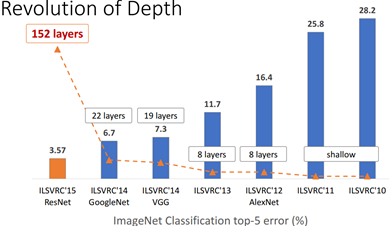


图2.9 ImageNet分类Top-5的误差率

Fig.2.9 ImageNet Classification top-5 error rate

由上图可以看到为了追求图片分类的准确率，模型的纵向深度越来越深，模型的横向宽度也越来越宽，模型的复杂程度也越来越高，如深度残差网络（ResNet）的层数已经高达152层。在图像识别领域的一些，某些真实的应用场景中，如移动或者嵌入式设备，应用如此之大并且相当复杂的模型是基本不可能实现的。首先是由于网络模型过于的庞大，随时面临着计算平台的内存不足的问题，其次是在这些实际的应用场景中，网络输出要求高效率，或者说高的响应速度，例如在自动驾驶系统这样的应用场景中，未能实时的完成图像识别任务，或者模型的响应速度稍有延迟，都会带来非常恐怖的后果。目前图像识别的发展方向日益向着小模型和高效率模型进展。目前主流的方法分为两种，一是对训练好的复杂模型进行模型压缩；另一个是直接设计小的网络模型架构并进行训练。而接下来要说的MobileNet就是Google公司开发的小的高实时性的网络模型。

MobileNet网络模型的基础架构是深度可分离卷积（Depthwise Separable Convolution，DSC），作为MobileNet网络的基础架构，其最早被应用在Inception模型中，深度可分离卷积作为一种可分解卷积，分为深度卷积（Depthwise Convolution，DC）和点式卷积（Pointwise Convolution，PC），如图2.10所示。深度卷积和标准卷积不同，对于标准的卷积如上文所述，其卷积核作用在所有的输入通道上，而深度卷积则是针对每个输入的通道采用不同大小的卷积核，换而言之就是一个卷积核对应一个输入通道，这也是深度卷积操作，深度一词的由来。而点式卷积其实就是普通的卷积操作，只不过点式卷积则是采用1x1的卷积核进行相应的卷积操作。图2.11中清晰的展示了这两种不同的操作。对于深度可分离卷积的卷积操作，是首先采用深度卷积对不同的输入通道分别进行卷积，然后采用点式卷积将上面的输出再进行结合，上述计算过程保证了深度可分离卷积的输出结果和标准卷积输出结果的一致性，与此同时会大大减少模型的计算量和参数量。

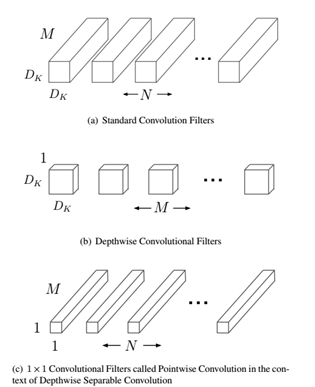


图2.10 深度可分离卷积

Fig.2.10 Depthwise Separable Convolution

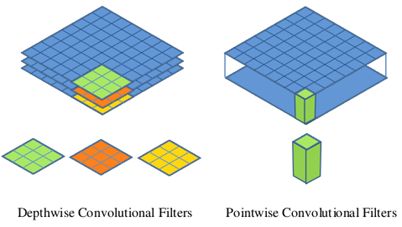


图2.11 深度卷积和点式卷积

Fig.2.11 Depthwise Convolution and Pointwise Convolution

深度可分离卷积是MobileNet的基本构件，如上文所述在实际应用时通常会加入池化层和激活层，而MobileNet在实际应用时会加入BN层（BatchNormalization，BN），并使用ReLU激活函数，所以深度可分离卷积的基本组件如图2.12所示：

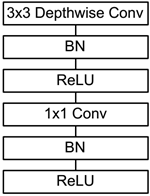


图2.12 MobileNet的基本组件

Fig.2.11 Basic components of MobileNet

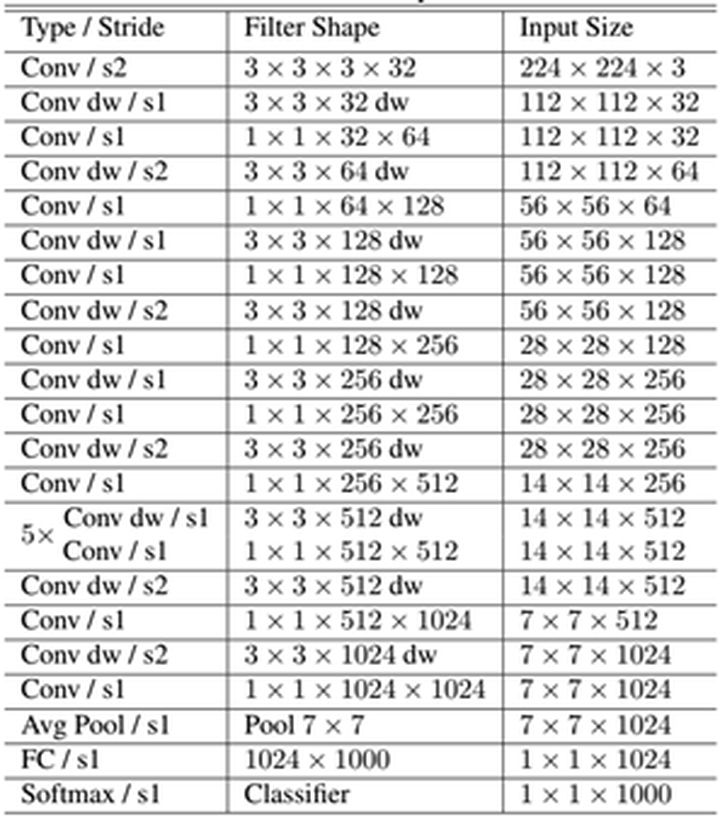


图2.13 MobileNet的网络架构

Fig.2.13 MobileNet's network architecture

在论文[ ]里，作者将MobileNet网络与GoogleNet和VGG16做了对比，如图2.14所示。相比于VGG16，MobileNet的准确度稍微下降，但是优于GoogleNet。然而，从计算量和参数量上MobileNet具有绝对的优势。

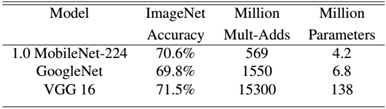


图2.14 MobileNet与GoogleNet和VGG16性能对比

Fig.2.14 Performance comparison of MobileNet, GoogLenet and VGG16

### 2.4.2 空间变换网络

仿射变换（Affine Transformation）涵盖了平移、旋转和缩放三种变换操作，而空间变换网络（Spatial Transforms Networks，STN） 能够实现对图像的平移、缩放、旋转以及裁剪，它主要有三部分结构组成：

Localisation net，用来进行参数计算、Grid generator用来对目标进行坐标映射以及 Sampler用来进行像素采集，如下图2.15所示：

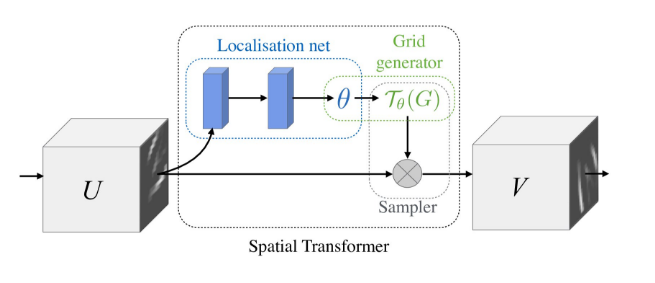


图2.15 空间变换网络

Fig.2.15 Spatial Transforms Networks

其中，Localisation net：是一个自定义的网络(可以使用全连接或者卷积神经网络外加一个回归层)，它根据输入U，输出变化参数，这个参数用来描述逆变换和映射U和V的坐标关系。Grid generator：根据V中的坐标点和变化参数，计算出将要填充到V中的像素值在U中的坐标点。Sampler：要做的是填充V，根据Grid generator得到的U中一系列坐标和原图U（因为像素值要从U中取）来填充。如果计算出来的坐标值为小数，可以用另外的方法来填充，比如双线性插值。

在应用STN时，首先先根据变换参数，在原样本上采样，拿到对应的像素点。通俗点说，就是输出的图片的位置上，要对应输入图片的哪个位置。

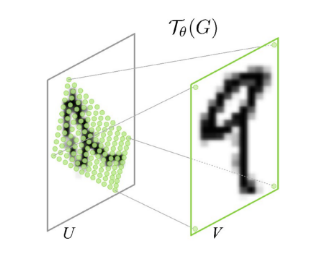


图2.16 旋转缩放操作

Fig.2.16 Rotate zoom operation

如图2.16所示旋转缩放操作，把像素点看成是坐标中的一个小方格，输入的图片U可以是一张图片或网络输出的特征图（feature map），经过变换，(是Localisation net生成的参数)，生成图片V，它的像素相当于被贴在了图片的固定位置上，用表示，像素点的位置可以表示为，这就是我们要确定的坐标。其次是应用仿射变换矩阵处理

其中，是输出的目标图片的坐标,，

## 2.5 本章小结

本章首先对轨迹预测问题的定义开始入手，分别对轨迹预测的不确定性进行建模，进而对整个轨迹预测问题进行建模。之后对轨迹预测领域的相关数据集进行了介绍，并推出了自己的数据集。最后对模型搭建所需要的深度学习相关技术进行了详细的介绍。

3 基于State-Anchor 的多模概率化机动车与非机动车轨迹预测模型

## 3.1 引言

机动车（汽车，摩托车等）和非机动车（自行车，电动单车等）作为复杂道路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。无论从模型算法难度还是潜在的社会影响的角度，自动驾驶都是机器人和人工智能（Artificial Intelligence，AI）领域目前所面临的最大课题。无人驾驶汽车对于避免交通事故，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，因为这是一个多种智能体进行复杂交互的场景，虽然各大汽车供应商以及汽车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，我们还有许多的工作要做。

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，解决“我周围将来发生什么”的问题，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，有利于主车做出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，是否让路以及何时寻找最佳的并入车流的时间点都具有重要的意义。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的可能性是非常有意义的。

在本章节接下来的介绍中，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序列以此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，并通过上文的模型建模，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，来捕获各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量表示出来。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作为建模基础。

## 3.2 机动车多模概率化轨迹预测模型



图3.1 机动车多模概率化轨迹预测模型

Fig.3.1 Multi-mode probabilistic trajectory prediction model for vehicles

如图3.1所示，机动车多模概率化轨迹预测模型基于State-Anchor技术，模型的整个架构是通过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。下面章节中我们将展开分别介绍这些网络模块。

### 3.2.1 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3.1所示，MapNET由两个模块组成：首先，使用2.3.2小节中所介绍的方法搭建好复合栅格地图，以此表示周围复杂的交通环境和历史信息，之后将复杂栅格地图，使用MobileNet\_V2特征提取网络，提取地图的语义特征表示，输出特征提取后的特征图，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：

（3.1）

其中表示MapNET深度网络的所有权重，通过使用上述分支模型可以完全提取复杂交通场景中的时空表示，并输出经过特征提取过后的特征图，其中包含着该场景下做出预测所需要的环境信息、交互信息和历史序列信息。

### 3.2.2 AgentNET:提取交通参与者的历史轨迹信息

由图3.1可以直观的看到，AgentNET用来接收机动车的历史状态（四维向量：）作为输入，之后利用LSTM处理时序序列信息的特点，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特征向量中，然后使用LSTM对这些智能体状态进行编码，编码生成特征向量。

（3.2）

其中，智能体的编码特征向量中包含着输入的历史序列特征信息，即智能体的速度、位置变化、转向角等特征。

### 3.2.3 Interaction-AttNET：提取交互信息

在本小节中，我们提出了一种新的网络来融合从AgentNET和MapNET中提取的特征图和特征向量信息。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智能体之间相互交互的作用。在之前的工作[10]、[11]、[22]、[23]中，针对这个问题，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。但是使用这些方法，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]，这些都会对预测结果造成负面的影响。基于此本文提出了一种新型的交互网络，用来提取智能体与周围智能体和周围环境的交互信息。

注意力机制类似于人类视觉注意机制，即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。当前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（Encoder-Decoder）[26]的架构，在相关领域[27]中取得了良好的表现。在本文中，我们提出的注意力模块类似于人类的视觉注意力机制，它会更倾向于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，进而再去考虑整体特征。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中较为重要的部分，以便快速做出相应的决策和改变机动车的行驶轨迹。

在章节中，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3.1所示，Interaction-AttNET接收AgentNET的特征向量输出和MapNET的特征图输出，整个模型一共由两个主要的网络模块组成，第一个模块是通过接受MapNET的输出特征图进行空间变换网络来提取以预测智能体为中心的小特征图。然后另一个模块是使用多头注意力网络[27]来组合变换后的小特征图和序列特征。

（3.3）

（3.4）

（3.5）

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和语音识别等领域中取得了不错的成绩和效果。对于空间变换模块，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。在实际应用中，机动车的仿射变换的感兴趣区域初步设为8\*8。

### 3.2.4 Multiple-PredictionNET

如2.2节所述，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3.1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概率置信度来表示。

## 3.3 非机动车多模概率化轨迹预测模型

在3.2节中，文章对基于State-Anchor的机动车轨迹预测方法分四部分进行了网络模型的搭建，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网络模型的搭建。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。

受启发于上文种机动车的轨迹预测的相关技术研究，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模型建模的可行性进行验证分析，确保其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。

### 3.3.1 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K-means聚类算法对非机动车的轨迹进行聚类，获得轨迹之间平方距离的近似，

（3.6）

并在对非机动车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3.2中可得，当anchor取值分别为25、13时，其均值和90分位的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。

因此非机动车多模概率化轨迹预测模型的整体网络架构和上文的机动车多模概率化轨迹预测模型一致，由四部分模块组成：MapNET，AgentNET，Interaction-AttNET以及Mutiple-PredictionNET。模型的输入输出与上文介绍的基本一致，唯一一个有区别的点是，在进行交互特征提取时，空间变换网络的仿射变换的感兴趣区域的提取，对于机动车感兴趣区域的大小为10\*10，而由于非机动车自身的特点，我们在设计非机动车的感兴趣区域时，设置其大小为8\*8。

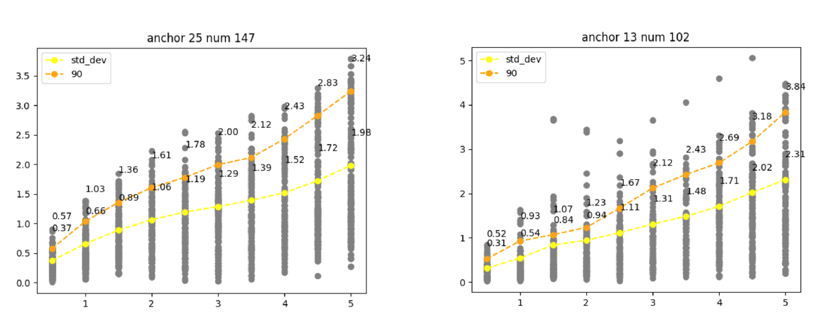


图3.2 非机动车轨迹聚类分析

Fig. 3.2 Cluster analysis of cyclist trajectory

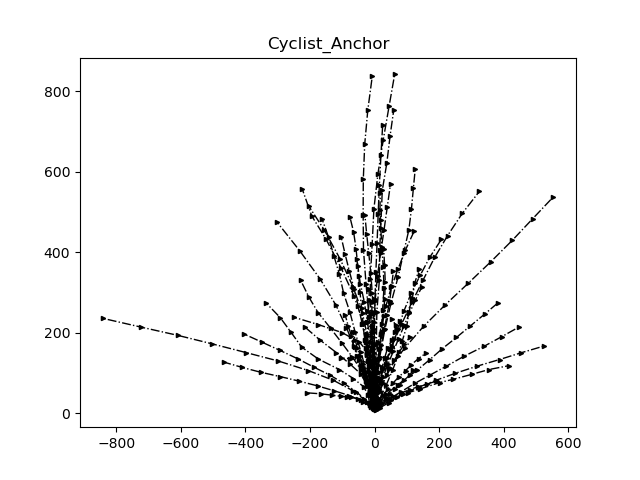
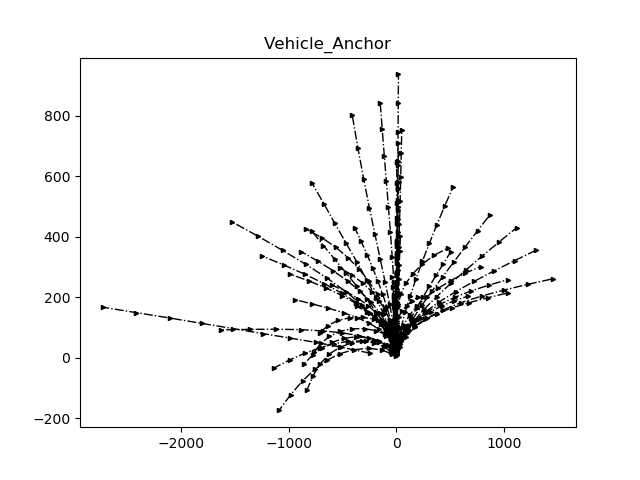


图3.3 机动车与非机动车轨迹聚类对比

Fig. 3.3 Clustering comparison of vehicle and cyclist trajectories



图3.5 非机动车多模概率化轨迹预测模型

Fig. 3.5 Multi mode probabilistic trajectory prediction model for cyclists

## 3.4 网络模型的损失函数设计

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：分别是前向传播计算和模型参数的反向传播修正。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，我们将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，所以我们可以使用端到端的方式分别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：

（3.7）

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：

（3.8）

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：

（3.9）

（3.10）

其中式3.10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。

（3.11）

此外，经过模型训练结果分析（图3.6），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：

（3.12）

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。

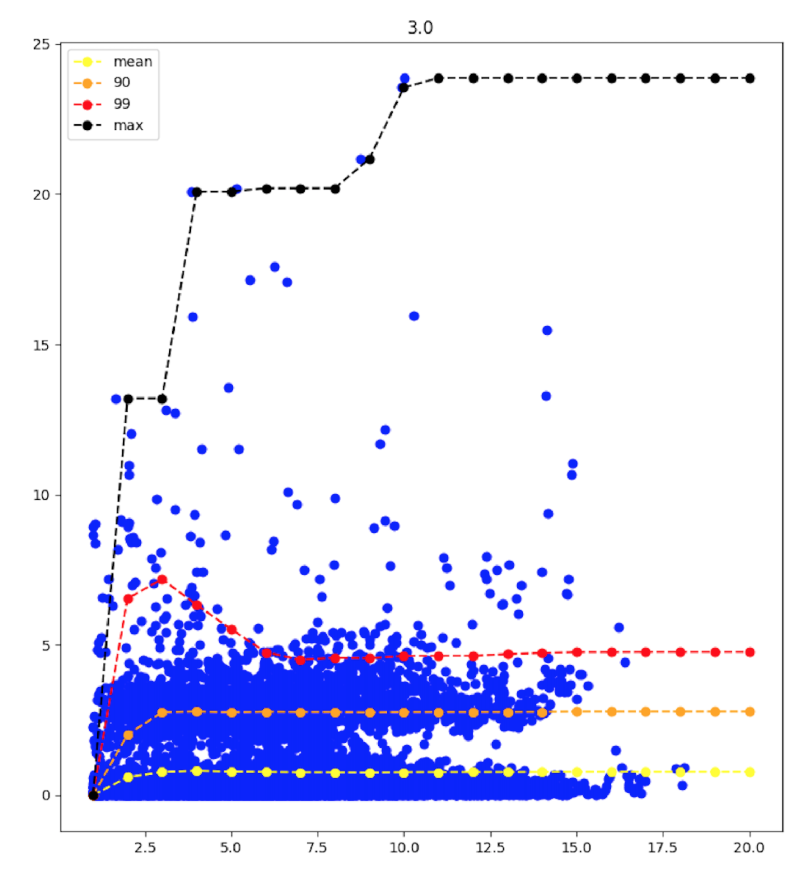


图3.6 机动车模型预测速度-误差分布图

Fig. 3.6 Speed-Distance error distribution of vehicle prediction model

## 3.5 网络模型的训练参数

本章节模型架构基于TensorFlow开源框架搭建，网络结构种的MLP嵌入层中的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3.1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3.1 计算机硬件和软件配置

Tab. 3.1 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 18.04 LTS |
| Cuda | Cuda 8.0 with CuDNN v5 |
| 数据处理 | Python 2.7, OpenCV, etc. |

## 3.6 本章小结

本章首先对机动车的多模概率化轨迹预测模型进行了搭建，并分别介绍了网络模型的各个组成部分。之后再在此基础上对非机动车的多模概率化轨迹预测模型的建模方法进行了相关验证分析，确定了最后的非机动车网络模型的结构。最后，分别介绍了机动车与非机动车网络模型的损失函数的设计和网络参数的确定。

4 基于Anchor-Free的多模概率化行人轨迹预测模型

## 4.1 引言

第三章中，本文对交通场景中的机动车与非机动车分别进行了建模，接下来我们将讨论交通场景中另一类主要的智能体，行人。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。

行人的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行为轨迹。但该问题非常复杂。首先，行人的运动具有很高的随机性，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

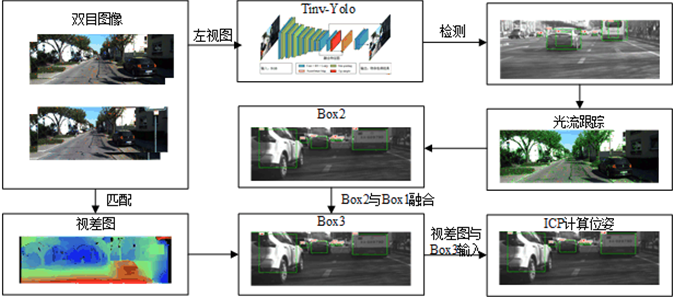


图4.1 行人多模概率化轨迹预测模型

Fig. 4.1 Multi mode probabilistic trajectory prediction model for pedestrians

如上文所述，在第三章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，将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的多模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：

（4.1）

（4.2）

其中，*m*为模型输入的第*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，*k*为分类分支的数量。

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）：

（4.3）

其中，为第*k*个分类轨迹的预测序列，为每个预测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：

（4.4）

## 4.3 行人多模轨迹预测模型网络结构设计

由行人多模概率化轨迹预测网络结构模型图4.1所示，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特征作为输入，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。

（4.5）

（4.6）

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：

（4.7）

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。

（4.8）

（4.9）

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

接下来我们对行人多模概率化模型的损失函数进行设计，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：

（4.10）

（4.11）

为解决机动车低俗段loss的问题，引入了focal loss：

（4.12）

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，我们使用了两个不同的模型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：

(4.13)

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，使用最大间隔（Max-Margin）损失构建损失函数：

(4.14)

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：

（4.15）

其中，为第m个行人在第t时刻的真实轨迹点，

（4.16）

由smooth L1函数定义：

（4.17）

其中为L1距离。

## 4.4 网络训练

本章节模型架构基于TensorFlow开源框架搭建，网络结构中的MLP嵌入层中的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。行人预测分类分支k = 6，进行数据分析时，取Top3的轨迹指标进行评价。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3.1所示。对于行人多模概率化轨迹模型的预测结果，在第五章中将和机动车与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介绍。

表4.1 计算机硬件和软件配置

Tab. 4.1 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 18.04 LTS |
| Cuda | Cuda 8.0 with CuDNN v5 |
| 数据处理 | Python 2.7, OpenCV, etc. |

## 4.7 本章小结

本章在对机动车和非机动车建模完成的基础上，讨论了行人这种高机动性，高灵活性智能体的相关模型的建模方法的研究，并对行人模型进行了建模，之后对网络结构进行了设计。最后完成行人模型损失函数的设计和参数的取值。

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，并分别通过State-Anchor和Anchor-free的方法，完成了各智能体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交通场景中，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



图5.1 多任务多模概率化轨迹预测模型

Fig. 5.1 Multi-task Multi-mode probabilistic trajectory prediction model

为了融合上述模型，多任务多模概率化轨迹预测模型同样应由四部分组成，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）同时进行图像级别的特征提取，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：

(5.1)

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复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，再处理交通智能体的复合栅格地图的同时，并行的提取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：

（5.2）

目前为止，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，实现对场景内各智能体之间的交互建模，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，使用空间变换网络（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。所以我们设计了空间变换模块，通过仿射变换，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的感兴趣区域的提取，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。其中机动车的感兴趣区域为10\*10，非机动车8\*8，行人4\*4。

（5.3）

经过STN空间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。

(5.4)

(5.5)

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，实现多智能体的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，机动车和非机动车各分支分别进行建模，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

复杂场景下多任务多模概率化轨迹预测模型的损失函数由三部分组成，分别是机动车多模概率化预测模型的损失函数，非机动车多模概率化预测模型的损失函数以及行人多模概率化轨迹预测模型的损失函数。由于多任务模型三部分的分支需要在数据集中一起进行训练，需要对这三个分支模型的权重进行分析设计。

（5.6）

（5.7）

（5.8）

（5.9）

（5.10）

（5.11）

（5.12）

其中，，，，分别为机动车分支损失函数的权重参数，非机动车分支损失函数的权重参数，行人分支损失函数的权重参数以及行人损失函数的回归分支的权重参数，公式其余参数详见上文，在此就不过多赘述。

### 5.3.2 训练参数的设定

本章模型架构基于TensorFlow 1.0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4.1所示。对于机动车多模概率化轨迹预测模型、非机动车多模概率化轨迹预测模型和行人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小节中进行介绍，并完成多任务网络的定量和定性对比。

## 5.4 实验结果分析

在本小节中，我们首先介绍实验的一些细节，其中包括数据集的具体设置和各模型的评价指标。然后，分别对上面两个章节中所提出的单一任务的智能体轨迹预测模型的实验结果进行定量和定性分析。最后，对本章节提出的多任务多模概率化轨迹预测模型进行定性，定量分析，并将模型的各个分支的性能与当前最为先进的模型性能指标进行比较，展示出模型所有指标的巨大改进，以此说明本文所提出的多任务多模概率化轨迹预测模型的优势。

### 5.4.1 实验细节

（1）数据集的设置

为了验证本文所提出的各类模型的有效性，我们在课题组内部建立的行为预测数据集和公开数据集上分别评测相关结果指标。内部数据集是来自中国北京的真实驾驶场景。它包含由百度Apollo自动驾驶汽车采集的高清地图数据和障碍物数据，通过使用车载的激光雷达、摄像头和毫米波雷达等传感器，可以为包括行人、机动车和非机动车在内的所有附近的智能体提供足够准确的位置、轨迹和高精地图信息。在该数据集中，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。所有种类的智能体轨迹数量总数为800M，其中机动车轨迹655M，非机动车轨迹80M，行人轨迹65M。在文中的实验任务中，需要使用智能体的轨迹来验证本文所提出的各类模型的性能。其中机动车，非机动车和行人的轨迹序列分别被分为训练集、验证集和测试集，机动车集合有555M，50M，50M；非机动车集合有60M、10M、10M；行人集合有45M、10M和10M。每条轨迹的长度为11秒，其中(0-3)秒是观测到的历史记录，(4-9)秒是用于预测的真值轨迹。行人轨迹是由真实世界捕捉得到的，包括静止、快速通过、转弯、匀速前进、群体结伴行走等。机动车轨迹包括停车、超车、加减速等状态行为。非机动车轨迹包括转弯、匀速、加减速、停车等。对于高清地图，该数据集则包括车道线，停车线，交通灯信息和人行道信息。而对于公开数据集上的评测，本文实验使用两个公开数据集，分别为ETH[21]和UCY[22]。数据集包含各种类型的社会交互场景下行人的轨迹坐标，其中包含行人交互、避免碰撞和行人结伴的轨迹坐标，以2.5帧/s的速度进行手动采样标记。其中ETH包含2个数据集（ETH和Hotel），UCY包含3个数据集（Zara1、Zara2和Univ）。为评估算法的性能，在上述5个数据集中进行验证，采用交叉验证的方法，分别在其中4个数据集中进行训练，在另外一个数据集中测试验证。使用该数据集纵向比较、评价单一行人模型和多任务模型行人分支的性能。

（2）评价指标

在本文中我们设计使用多种评价指标对上述两种数据集上的各类模型的训练结果进行比较和分析。实验运行在Ubuntu 18.04 LTS操作系统上，GPU为NVIDIA Titan X，使用Tensorflow 1.1.0、CUDA 10.1和Cudnn v7.5.0深度学习框架。与之前的研究方法[11]部分相似，我们使用以下指标来预测和评价每个行人的轨迹：

1. 平均位移误差(Average Displacement Error，ADE)

平均位移误差定义为每个时间步长中实际轨迹与预测轨迹序列之间的欧氏距离(Euclidean Distance)；

（5.13）

2）最终位移误差(Final Displacement Error，FDE)

最终位移误差定义为实际轨迹与最终位置预测轨迹序列之间的欧氏距离。

（5.14）

3）召回率（Recall）

召回率用来衡量预测轨迹的稳定性，即在3s时刻预测点与真实点的差值分别小于（1.5m，1m，0.85m）。

（5.15）

在本文中，对于基于公开数据集的评测，由于公开数据集中没有高清地图的栅格地图信息，模型在训练比较时，只走行人历史轨迹编码分支，进而对模型的性能进行预测。由于各类别轨迹预测本质上是多模态的，本文使用前K个预测的最小ADE (minADE)和最小FDE (minFDE)作为度量。当K等于1时，minADE和minFDE等于ADE和FDE。

### 5.4.2 多模概率化轨迹模型机动车与非机动车预测结果分析

（1）轨迹数据类型分析

对于机动车多模轨迹概率化预测模型的模型搭建和参数设置，在上文第三章节，本文已经做出了详细的阐述，在此就不过多赘述。接下来我们使用单独的测试集对模型进行相关测试分析，其中在该小节的部分实验中，机动车的轨迹数量为32487条，非机动车轨迹29589条，行人轨迹10493条，各条轨迹的类型识别如下：

表5.1 轨迹数量及类型分布

Tab. 5.1 Conception and scope of Logistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 轨迹类型 | 未识别类型 | 直行 | 左转 | 右转 | 总计 |
| 机动车 | 8628 | 17740 | 3149+688 | 2282 | 32487 |
| 非机动车 | 14983 | 11146 | 1832+210 | 1418 | 29589 |
| 行人 | 10149 | 255 | 59 | 30 | 10493 |

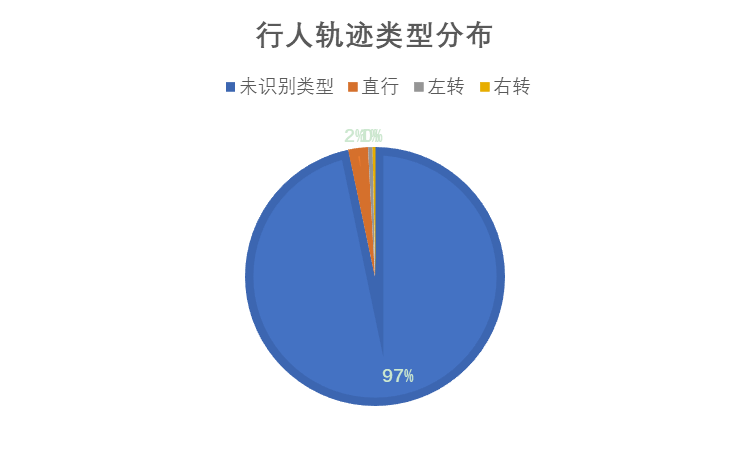
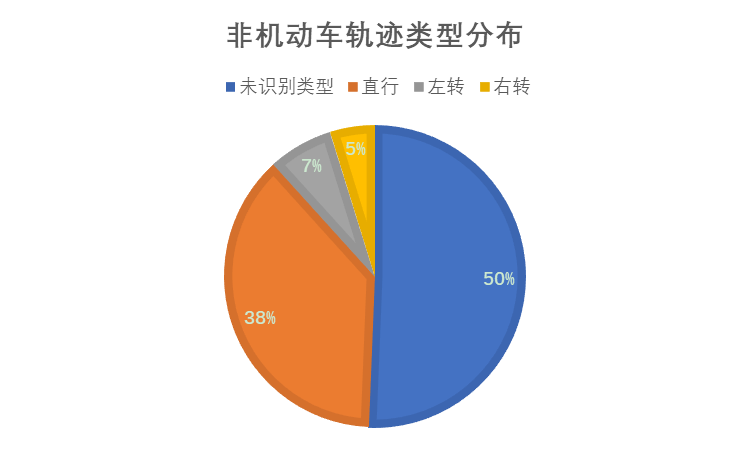
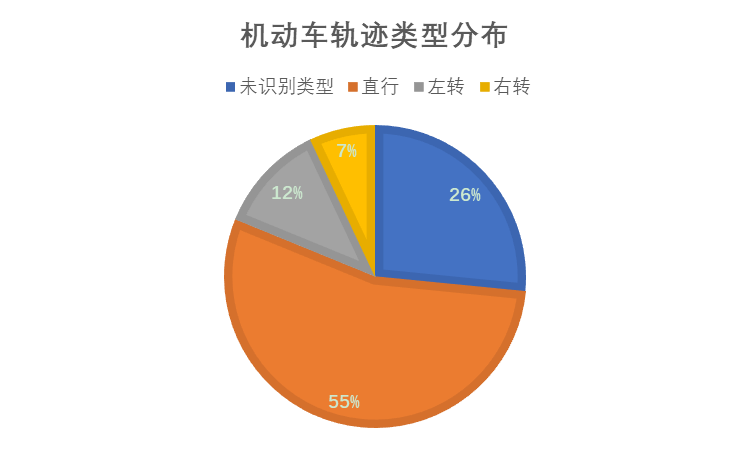


图5.2 机动车、非机动车和行人轨迹类型分布

Fig. 5.2 Distribution of vehicle, non vehicle and pedestrian trajectory types

由机动车和非机动车测试集的数据分布图，可以清楚直观的看到直行的数据量占数据集的一大部分，是复杂交通场景中最为主要的交通行为，而且轨迹类型中未识别类型也占有一定的比例，表明机动车轨迹有很大一部分是不确定的，进一步表明了机动车轨迹的随机性和不确定性。而对于行人来说，由于其最高的不确定性和随机性，未识别类型达到了97%，这也是对于行人无法使用Anchor去区别各个模式的重要原因。

（2）机动车与非机动车定量分析

接下来，我们将就第二章到第五章所提出的机动车与非机动车的轨迹预测方法与几种基线模型进行了比较：

•无迹卡尔曼滤波（UKF），按照时间的先后顺序向前传播估计状态;

•单一轨迹预测模型（Single Trajectory Prediction，STP）;

•MDN[35]，轨迹空间上的高斯混合模型。

在表5.2中，我们将机动车和非机动车的轨迹预测模型与几种业内主流的机动车轨迹预测模型分别进行比较。表中的比较指标包括模型在不同轨迹类型中预测1s，3s，5s轨迹时的minADE，minFDE以及Recall，其中当将我们的注意力转移到MTP方法上，我们观察到相对于竞争方法有显著的改进。平均误差和6s-horizon误差都全面下降，这表明这些方法捕捉到了交通问题的多模态特性。短期1s和长期6s视界的预测误差均低于其他方法，但在较长预测视界的收益更大。有趣的是，结果显示，在所有指标上都获得了最佳性能。

表5.2 机动车与非机动车单一模型和多任务模型指标定量分析

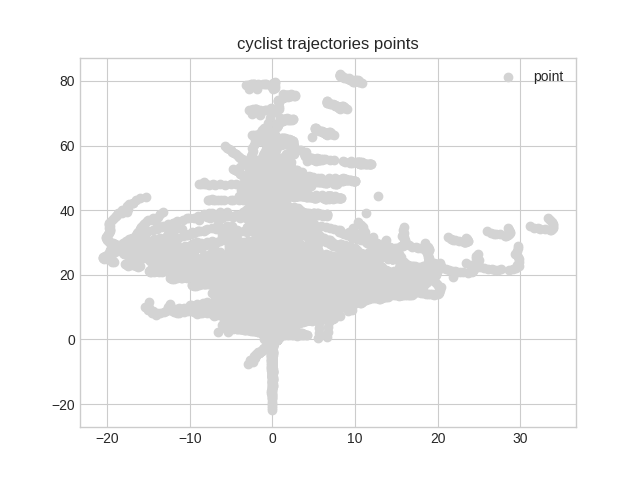
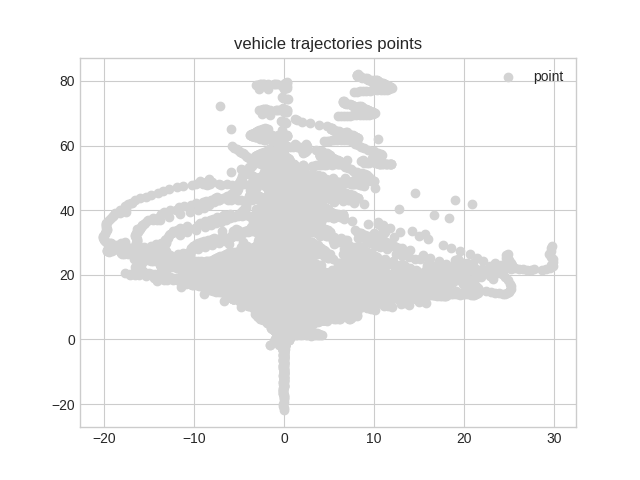
Tab. 5.2 Single model and multi-task model metrics quantitative analysis of vehicle and cyclist

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 单一任务的轨迹预测模型 | | | | | | | | | |
| V | MinADE | | | MinFDE | | | Recall | | | Accuary（5s） |
| 1s | 3s | 5s | 1s | 3s | 5s | 1s | 3s | 5s |
| UKF |  |  |  |  |  |  |  |  |  |  |
| STP |  |  |  |  |  |  |  |  |  |  |
| MDN |  |  |  |  |  |  |  |  |  |  |
| Ours1 | 0.28 | 0.84 | 1.65 | 0.38 | 1.66 | 3.66 | 0.98 | 0.59 | 0.27 | 0.65 |
| Ours2 | 0.23 | 0.62 | 1.20 | 0.29 | 1.18 | 2.57 | 0.98 | 0.78 | 0.39 | 0.87 |
| Ours3 | **0.20** | **0.54** | **1.04** | **0.25** | **1.00** | **2.16** | **0.99** | **0.85** | **0.43** | **0.94** |
| C | MinADE | | | MinFDE | | | Recall | | | Accuary（5s） |
| 1s | 3s | 5s | 1s | 3s | 5s | 1s | 3s | 5s |
| UKF |  |  |  |  |  |  |  |  |  |  |
| STP |  |  |  |  |  |  |  |  |  |  |
| MDN |  |  |  |  |  |  |  |  |  |  |
| Ours1 | 0.36 | 1.06 | 2.04 | 0.48 | 2.08 | 4.46 | 0.98 | 0.53 | 0.23 | 0.47 |
| Ours2 | 0.30 | 0.85 | 1.61 | 0.40 | 1.62 | 3.43 | 0.99 | 0.65 | 0.35 | 0.67 |
| Ours3 | **0.27** | **0.75** | **1.40** | **0.35** | **1.39** | **2.92** | **0.99** | **0.72** | **0.41** | **0.78** |

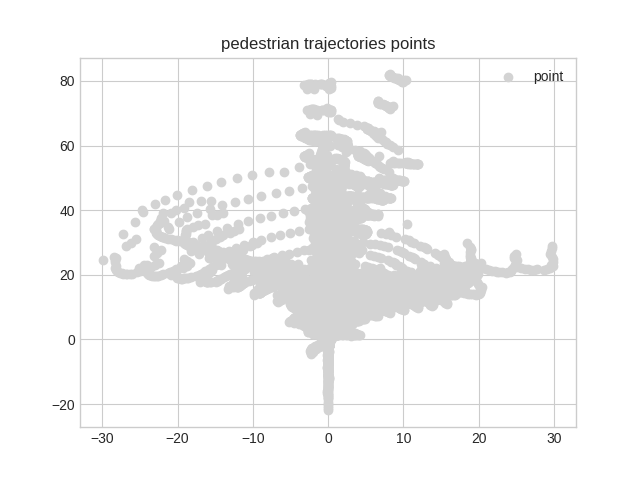
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 多任务轨迹预测模型 | | | | | | | | | |
| V | MinADE | | | MinFDE | | | Recall | | | Accuary（5s） |
| 1s | 3s | 5s | 1s | 3s | 5s | 1s | 3s | 5s |
| UKF |  |  |  |  |  |  |  |  |  |  |
| STP |  |  |  |  |  |  |  |  |  |  |
| MDN |  |  |  |  |  |  |  |  |  |  |
| Ours1 | 0.25 | 0.81 | 1.62 | 0.34 | 1.63 | 3.65 | 0.98 | 0.58 | 0.26 | 0.64 |
| Ours2 | 0.19 | 0.58 | 1.15 | 0.25 | 1.12 | 2.50 | 0.99 | 0.80 | 0.39 | 0.88 |
| Ours3 | **0.17** | **0.50** | **0.99** | 0.21 | 0.95 | 2.08 | 0.99 | 0.87 | 0.43 | **0.95** |
| C | MinADE | | | MinFDE | | | Recall | | | Accuary（5s） |
| 1s | 3s | 5s | 1s | 3s | 5s |  |  |  |
| UKF |  |  |  |  |  |  |  |  |  |  |
| STP |  |  |  |  |  |  |  |  |  |  |
| MDN |  |  |  |  |  |  |  |  |  |  |
| Ours1 | 0.41 | 1.12 | 2.06 | 0.55 | 2.13 | 4.31 | 0.97 | 0.46 | 0.19 | 0.45 |
| Ours2 | 0.35 | 0.91 | 1.61 | 0.45 | 1.65 | 3.25 | 0.98 | 0.58 | 0.31 | 0.67 |
| Ours3 | **0.32** | **0.80** | **1.41** | **0.41** | **1.43** | **2.81** | **0.98** | **0.64** | **0.38** | **0.77** |

（3）不同Anchor数值的选择

我们对数据集中的智能体轨迹使用仿射变换置于起点为原点的坐标系中，对轨迹点进行可视化分析，由图4.2可以发现，轨迹点的有序程度依次递减。由上文所述，我们对机动车的轨迹进行Anchor聚类，聚类可视化图在4.3中，由图4.3可以直观的看到随着Anchor聚类类别的增加，Anchor所能代表的模式更具有代表性，最终应用到模型中，我们采用Anchor类别为48时，作为模型预测时最终Anchor划分。



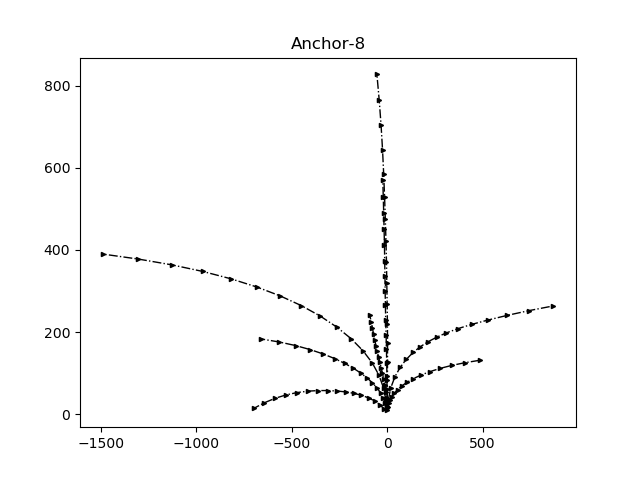
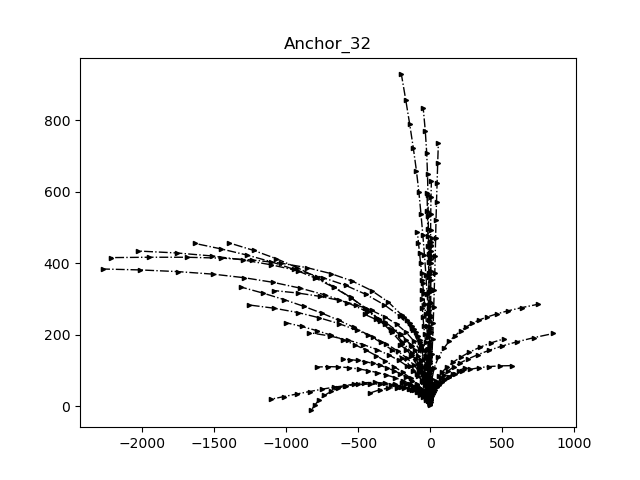
（a）机动车轨迹点 （b）非机动车轨迹点



（c）行人轨迹点

图5.3 机动车、非机动车和行人轨迹点图

Fig. 5.3 Point chart of vehicle, cyclist and pedestrian trajectory

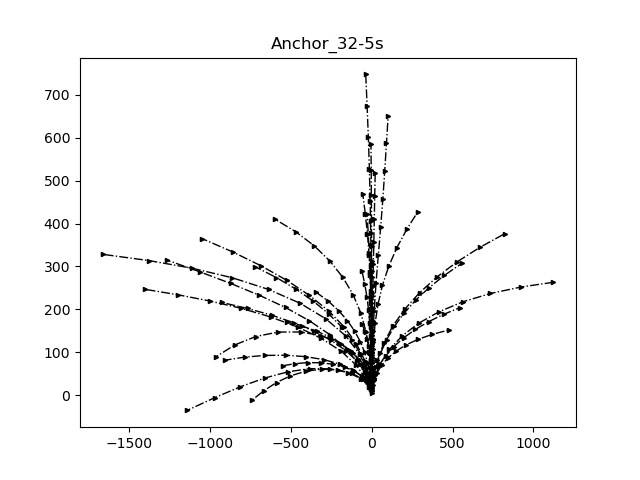
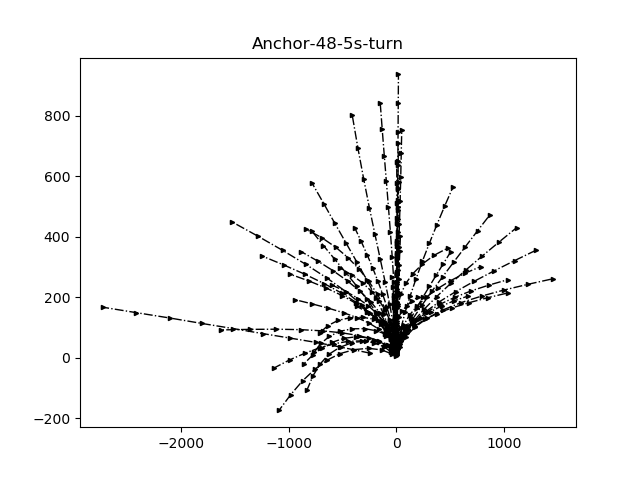
 

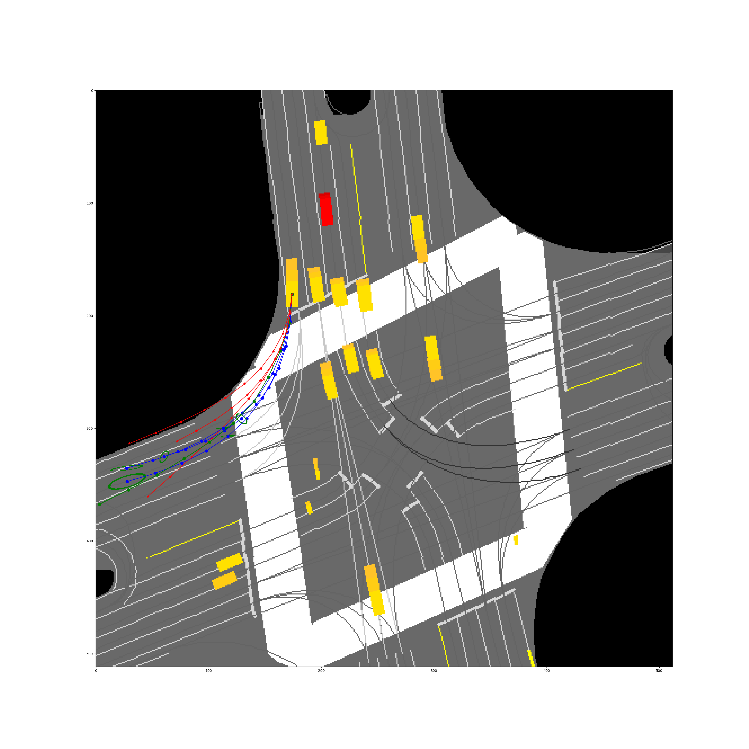
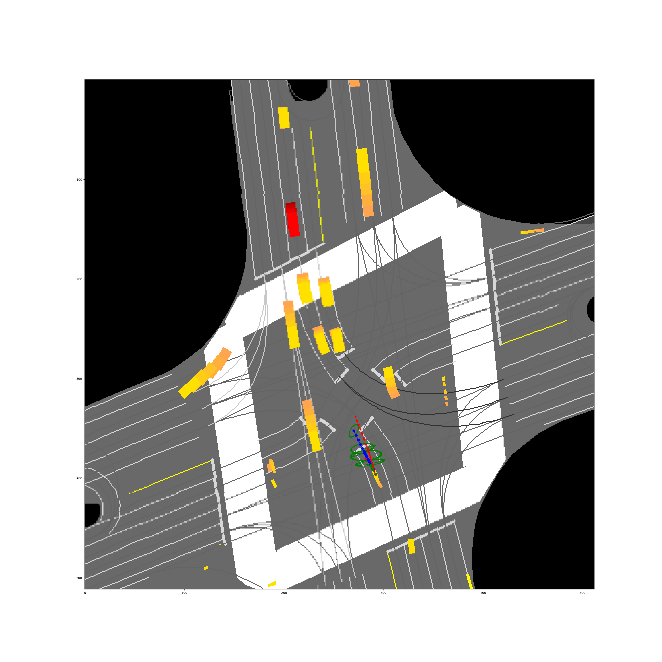
图5.4 机动车Anchor不同取值的可视化图

Fig. 5.4 Visualization of different values of vehicle anchor

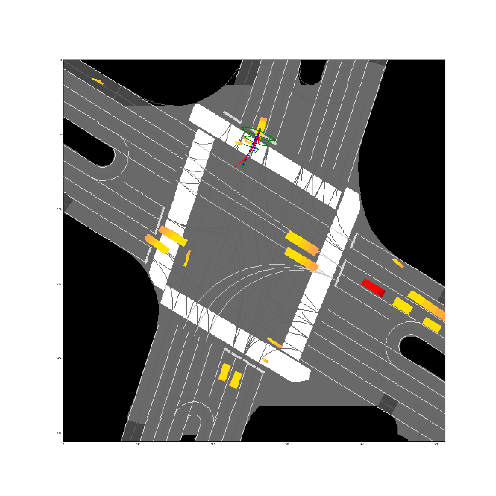
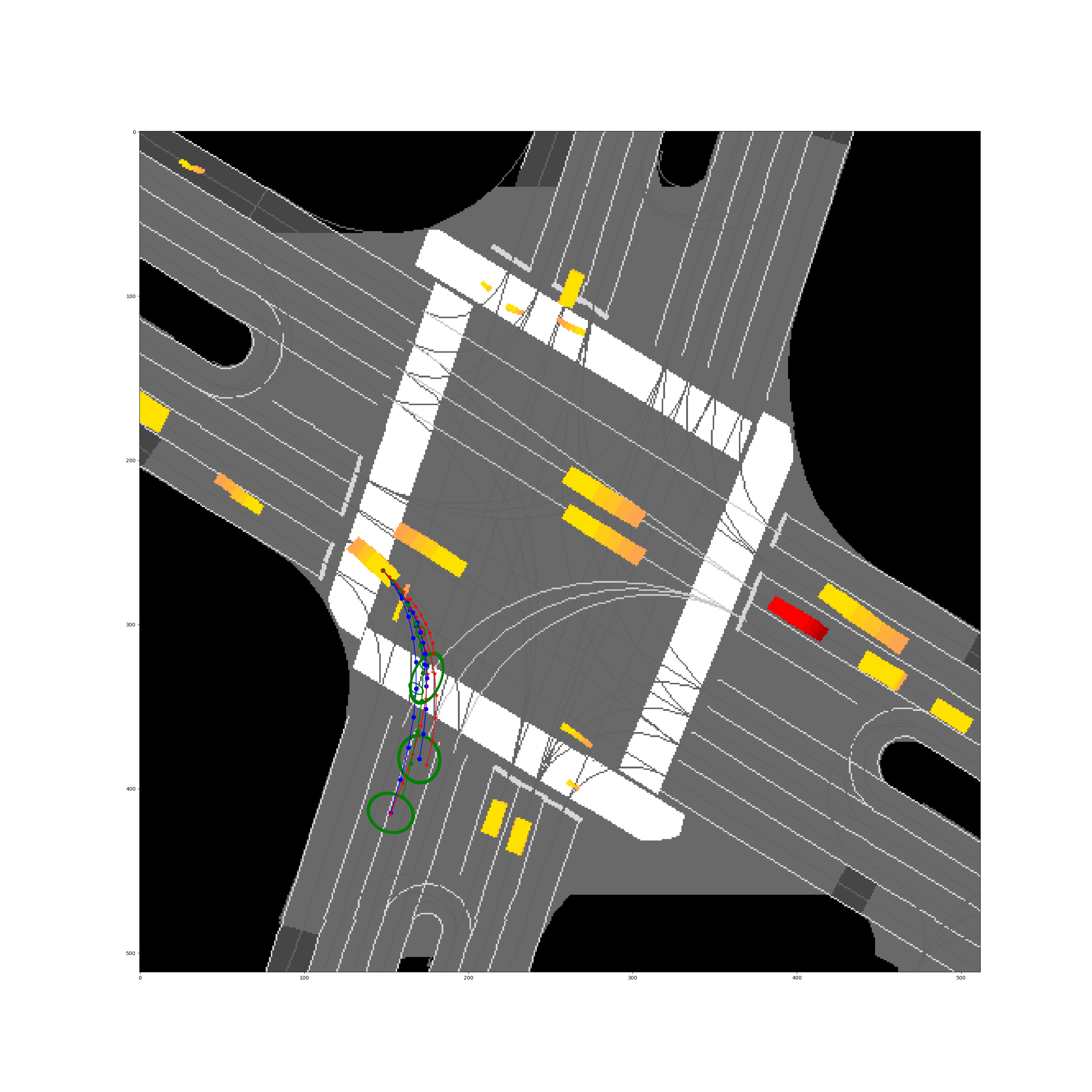
（4）机动车多任务多模概率化轨迹预测模型的可视化分析

图4.4中给出了多任务概率化轨迹预测模型机动车分支在我们内部评价数据集上的可视化分析，其中图4.4-（a）、（b）是同一个路口场景，而图4.4-（c）、（d）则是另一个路口场景，我们从预测的48条Anchor轨迹中，我们挑选了top3的轨迹进行可视化，其中，红色的为Anchor标签轨迹，而蓝色的则为预测线，绿色轨迹为GT轨迹，在蓝色的预测轨迹线上我们分别在3s点和5s点，将该预测点的双变量高斯分布绘制出来，其中双变量高斯分布的椭圆的大小，代表着预测轨迹点的不确定性区间，区间越小也就是说方差越小，说明该点的预测置信度越高。分析图4.4中几个机动车轨迹的预测效果可得，图4.4（a）中对于该车的右转预测较为准确，top3三条anchor轨迹都是右转轨迹，对于每条轨迹的3s点和5s点，轨迹都准确的预测到GT轨迹的车道，能够达到很好的预测效果。图4.4（b）中，表明在轨迹预测时该车辆处于进入左向待转区状态，该车辆要进行左转，但是对向车道是开启的，经过多模模型预测出来的两条预测线都是减速到停止线前，其中一条预测线，稍微预测出来，属于预测出车辆将要启动，但通过预测置信度可以发现，其置信度较低仅为9%。图4.4（c）场景中，预测车辆处于红灯要结束的状态，而前方人行道中还有行人通过，经过模型预测的轨迹线，能够很好的预测出来，与行人交互时避让行人的轻微右转趋势，并能在红绿灯红灯结束后，就能预测出该行为，表明模型输入实时红绿灯线以及使用仿射变换进行交互分析的建模有效性。接下来是图4.4（d）右转进入另一条道路的场景中，由于GT轨迹转向后压到了，车道2和车道3之间的实线上，而在真实道路交通场景中，这种行为是违反交通规则的，所以在模型预测时，将其作为一种特例。故预测线位于车道2和车道3中，而不是单纯为了拟合轨迹而去违返交通规则，这也不具有代表性。

经过上述对多任务模型的机动车分支的轨迹预测线可视化分析，证明了模型输入以及模型网络结构的有效性，模型预测效果能够达到很好的精度和良好的社会可接受性。

（a） （b）

（c） （d）

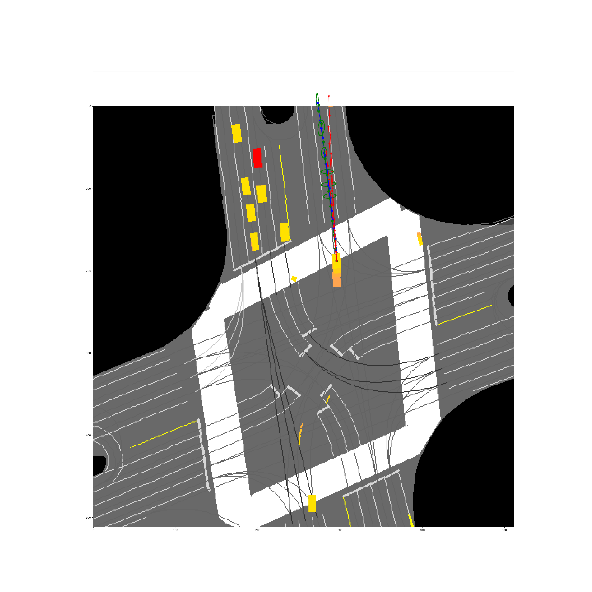
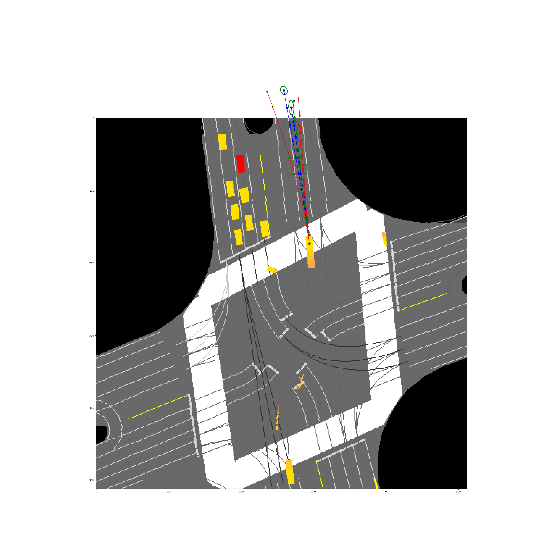
图5.5 多任务模型机动车分支预测可视化图

Fig. 5.5 Visualization of vehicle branch prediction in multi task model

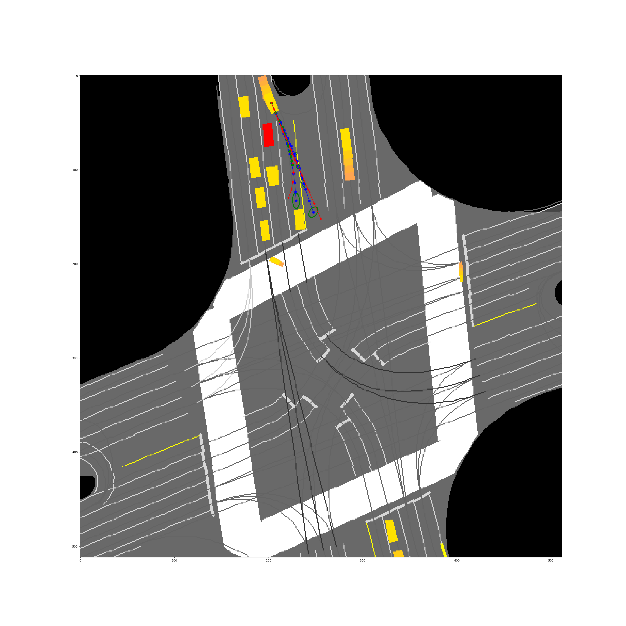
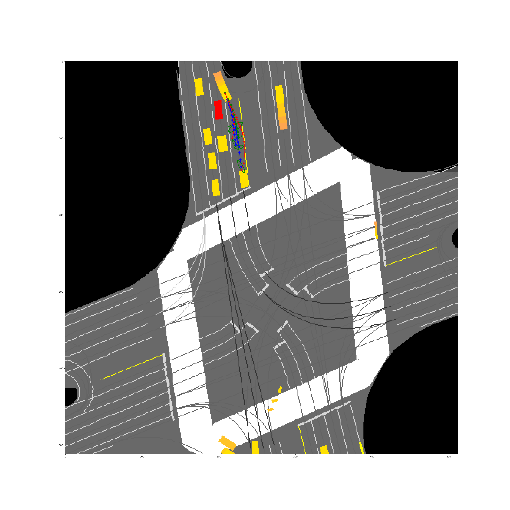
（5）机动车单一任务模型和多任务多模概率化轨迹预测模型对比分析

在证明了多任务模型机动车分支的有效性后，我们开始着手将上述模型与单一任务模型进行对比，在表5.2中，通过数据我们可以直观的发现，在采用多任务模型训练后，机动车分支的评价指标都有所提升，接下来我们将进一步分析他们的可视化图，进而对他们的性能进行分析比较。

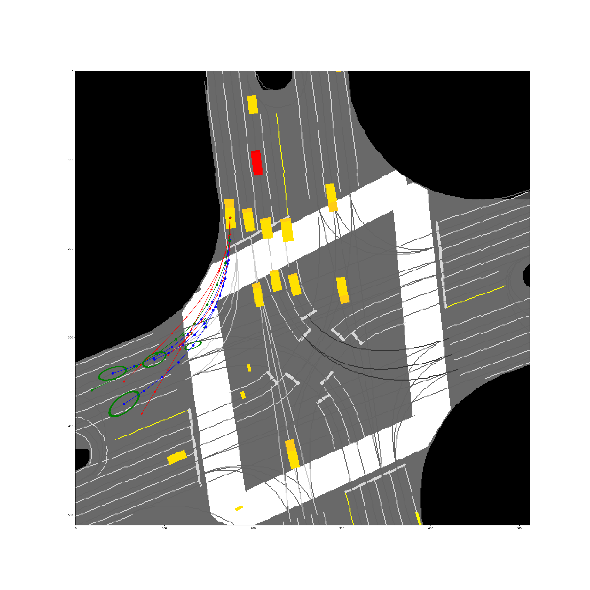
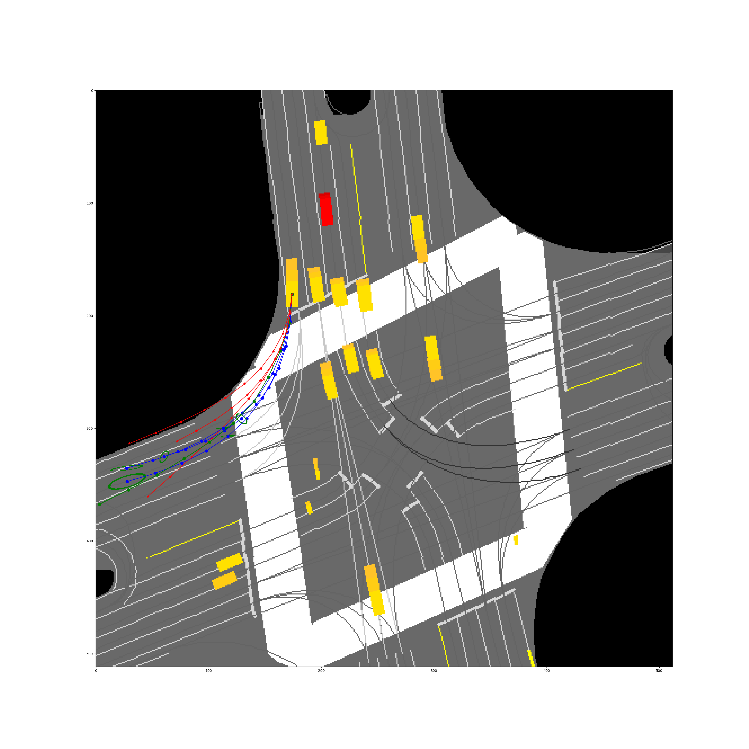
如图5.5所示，左侧图为单一任务模型的轨迹预测效果，右侧图为多任务模型的轨迹预测效果图，截取同一时刻同一个智能体的轨迹预测效果图进行可视化对比分析。对比（a），（b）两图发现，单一模型趋向于一侧集中进行预测，而多任务模型的预测Anchor却是两个模式的Anchor，这一点更符合多模预测的直观体现。对比（c），（d）两图，（c）图中智能体在前方车辆已经停止的场景中，预测出一条，绕过该车辆的轨迹预测线，而这条轨迹线，往往是不符合预期的，反之（d）图则能很好的预测出来，该车辆会向前移动并逐渐靠近前方车辆，停止。对比（e）（f）两图，发现对于转弯场景，单一模型预测的更为发散，模型预测线分布于改向的所有车道上，而多任务模型则倾向于预测与GT轨迹相近的车道位置。

（a） （b）

（c） （d）

（e） （f）

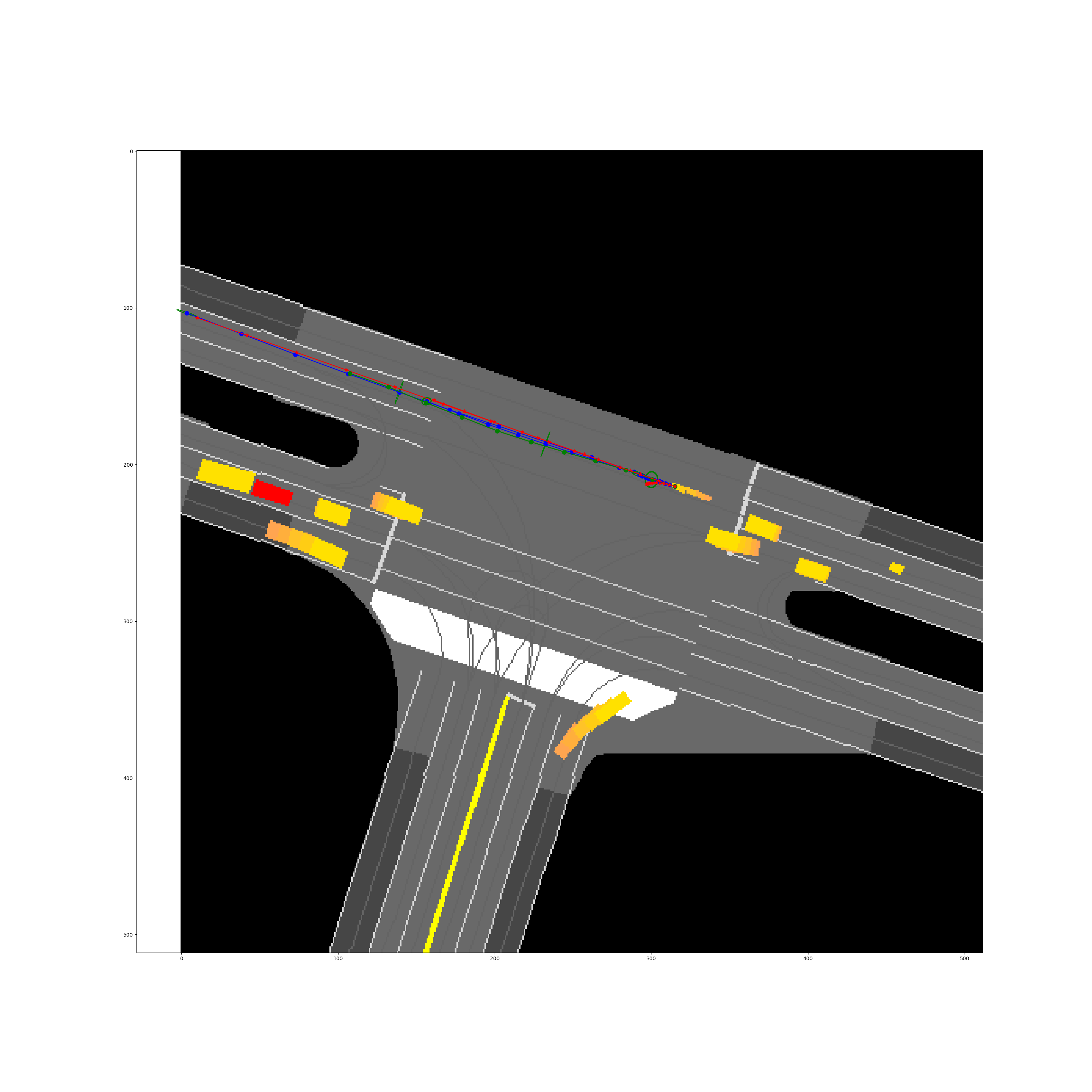
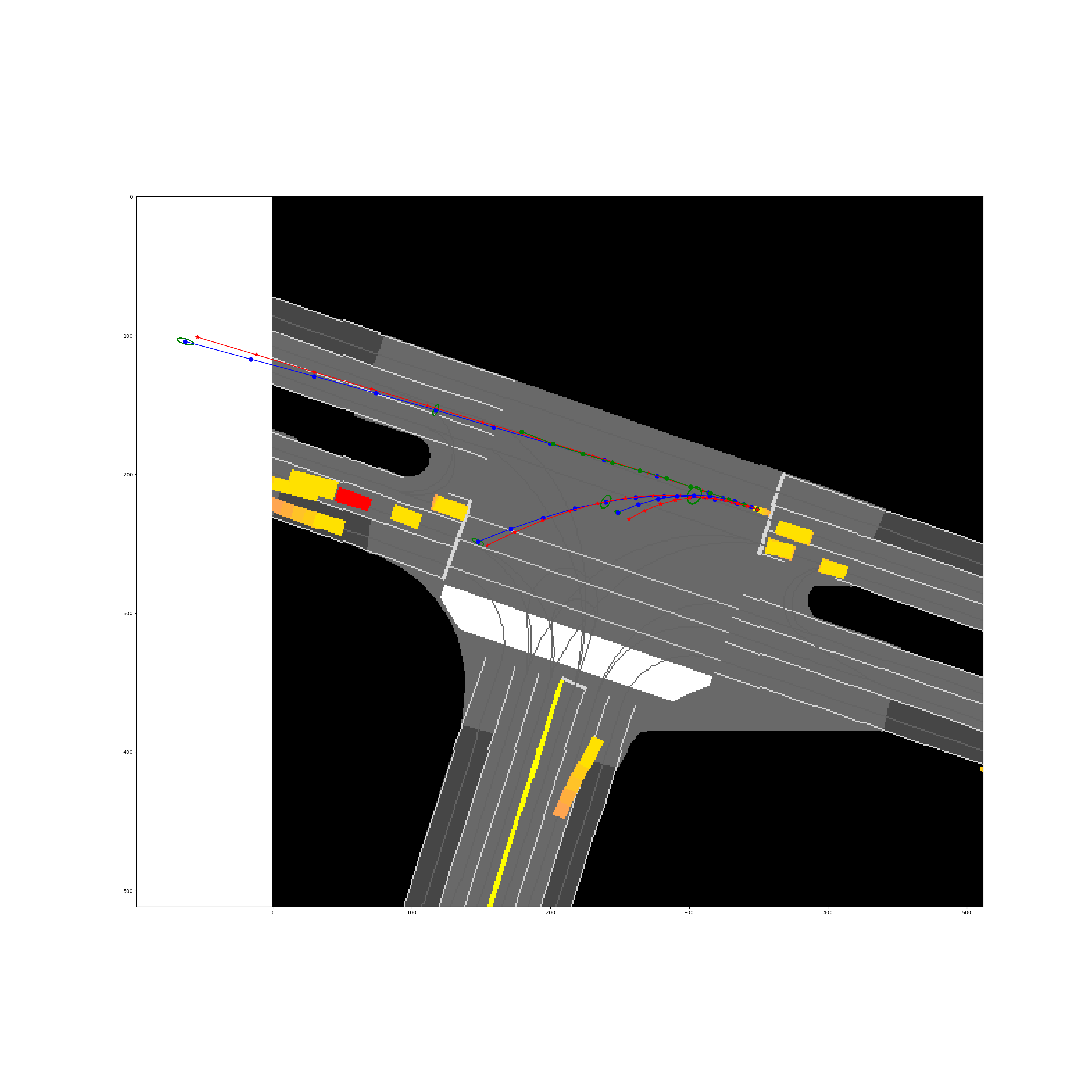
图5.6 多任务模型机动车分支预测可视化图

Fig. 5.6 Visualization of vehicle branch prediction in multi task model

通过上述分析，在已经得到单一模型的机动车分支的指标比多任务模型的指标要稍有逊色的基础上，分别通过对他们的可视化分析，得到更为准确的结论，即采用多任务模型预测出来的机动车轨迹的性能要优于单一模型预测出的机动车轨迹。

（6）非机动车单一任务模型和多任务多模概率化轨迹预测模型对比分析

由表5.2的分析可得，单一任务模型的非机动车分支的精度比多任务模型的非机动车分支的精度要高，有下图5.5我们可以更直观的发现，多任务模型的非机动车分支，倾向于多模预测，在不仅仅预测与GT最为接近轨迹的同时，还预测另外的社会可接受的轨迹，这也体现了多任务模型在多模预测上的优点。

（a）单一任务模型 （b）多任务模型

图5.7 多任务模型机动车分支预测可视化图

Fig. 5.7 Visualization of vehicle branch prediction in multi task model

### 5.4.4 多模概率化行人轨迹预测模型的结果分析

（1）行人轨迹预测模型定量分析

与上文介绍机动车与非机动车的预测结果分析不同，由于行人轨迹预测领域是轨迹预测中最为发展起来的，其模型算法与机动车和非机动车算法不同。在本节中，我们采用最具代表性的Linear, LSTM, Social LSTM, Social GAN, Sophie, Social- bigat, Social- stgcnn作为比较基准，使用最小化MSE对回归参数进行优化。

LSTM，一种数据驱动的轨迹预测方法。

Social -LSTM，将每个行人与一个LSTM单元关联起来，利用栅格占有图模型收集并预测邻近行人的隐藏状态。

Social GAN，基于GAN网络方法，对行人轨迹进行多模预测。

Sophie，基于物理场景和社会注意的轨迹预测生成模型。

Social-BiGAT，基于图注意力网络的轨迹预测生成模型。

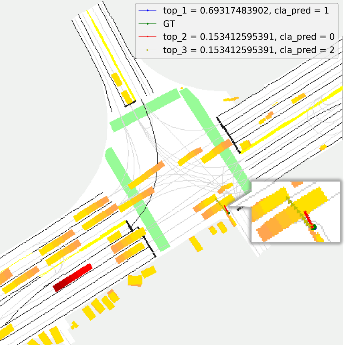
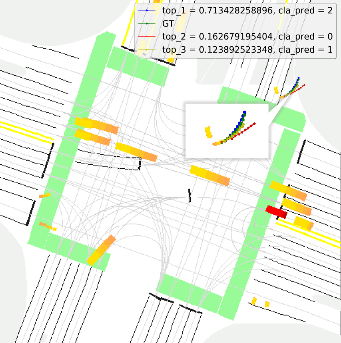
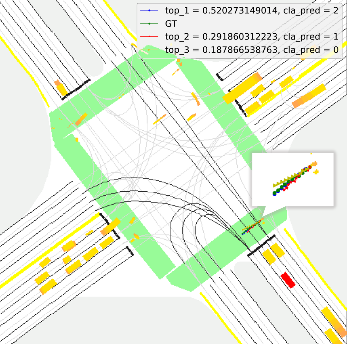
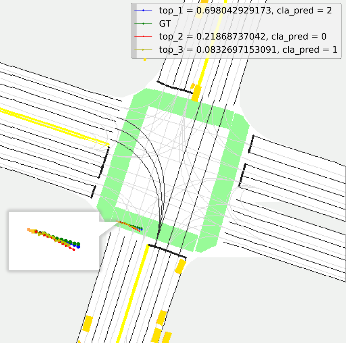
Social-STGCNN，一种使用图网络建模和时间卷积预测的轨迹预测模型。

将我们的模型与上述一系列的基准模型进行比较。为了维持基准的一致性，在评估模型时，对于基于GAN网络的模型预测结构，古普塔等人[11]提出生成预测N条轨迹，使用最接近于GT的轨迹进行性能和指标评估，在实验中，我们采集的数据样本数量N = 20。由于有些模型只能预测单个轨迹输出，如Linear、S-LSTM等，在这种情况下，默认模型预测轨迹的ADE等于minADE，类似地，FDE等于minFDE。在表5.3中，我们可以看到我们多任务多模概率预测模型在各个指标上都明显优于其他模型。对比分析每种方法，模型在预测步长下的预测轨迹指标随着预测距离的增加而增加，所有网络的准确率和召回率指标都显著降低。事实证明，轨迹预测的时间越长，难度越大。在整个评价实验中，每个模型在预测步骤内的预测误差都有相似的趋势。正如预期的那样，由于线性模型Linear不能有效地模拟不同行人之间复杂的社会相互作用和理解交通路口的场景，因此在所有的轨迹预测任务中表现最差。随着LSTM在序列预测领域的应用，模型预测的精度逐步提高，也证明了历史序列数据对预测问题的有效性。与基于网格划分的S-LSTM交互池化模型相比，S-GAN模型倾向于生成多个随机轨迹，具有更高的精度和社会可接受性。但GAN网络很难训练，尤其是当发生轨迹发生器和轨迹判别器不平衡的状况，就很容易导致梯度消失或模式崩溃(采样合成数据没有多样性)的问题，特别是当预测行人轨迹，避免模式崩溃对于自动驾驶决策和行车安全至关重要。Social-STGCNN采用了完全不同的方法，使用无向图对人群交互行为进行建模，并使用时域卷积网络代替递归循环结构，提高模型的精度和效率。但上述模型都没有考虑场景中的空间拓扑关系，这在很大程度上都限制了模型的性能。与我们的工作类似，Social-BiGAT也使用语义场景信息和序列信息作为输入。但是我们提出的多任务多模概率化轨迹预测模型的所有指标都比这个模型要好，并且能够预测场景中所有种类的智能体。这是因为我们使用了复合栅格地图，而不是直接将传感器的图像输入到模型中，我们的模型可以有效地学习地图拓扑结构和交通智能体之间复杂的相互作用。在上述的工作中，行人的状态是独立编码的，因此不能准确捕捉全局地图拓扑与单个行人之间的联系。本文的模型提出了一种通过多头注意力机制，将场景特征图和行人序列特征连接起来的交互注意模块，因此，我模型能够很好地建模语义特征。图5.6中给出了交通路口场景中多个预测轨迹的概率可视化图。在这一部分中，我们将展示我们的模型在处理不同场景、交通状况和紧急情况时准确预测行人未来轨迹的能力。如图5.6-(a, b, c)所示，模型预测了不同路口轨迹的可靠性。我们的模型预测了每个行人的3条轨迹，包括top1, top2, top3, top1的轨迹是概率值最高的预测轨迹，与Ground Truth (GT)最一致。在图5.6-(d)中，行人处于交通密集的t形路口，模型预测的top1轨迹以69%的概率与GT一致，此时行人状态处于等待状态，这种情况在自动驾驶任务中尤其常见。对于自动驾驶汽车来说，准确预测行人在道路行驶时的位置和意图对于道路交通安全和缓解交通压力非常重要，尤其是在交通密集的场景中。从图5.6-(e, f)可以看出，我们的模型可以准确预测出行人违反人行道交通规则的特殊情况，图5.6-(f)中要求行人在人行道上行走，并快速穿过人行道。最后，在最后一个图中，我们可以看到我们的模型可以很好地处理与车辆的交互。

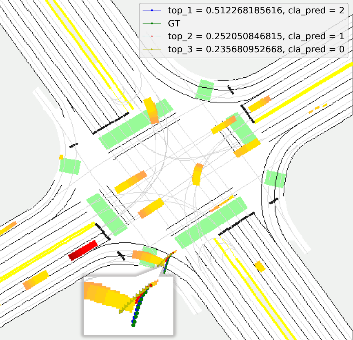
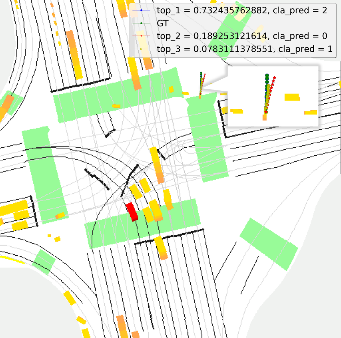
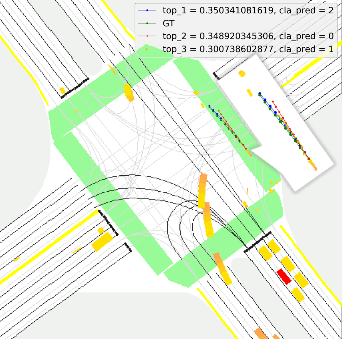
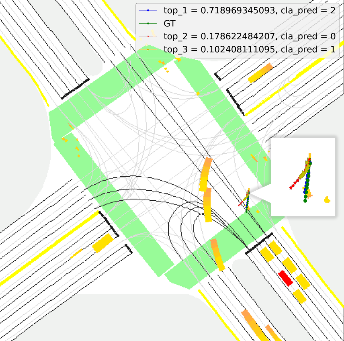
表5.3 多任务多模概率化轨迹预测模型行人分支定量分析

Tab. 5.3 Quantitative Analysis of Pedestrian Branch of Multi-task Multi-mode Probabilistic Trajectory Prediction Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 时间 | minADE | minFDE | Recall |
| Linear | 5s | 1.43 | 2.98 | 0.15 |
| LSTM | 5s | 1.01 | 1.98 | 0.43 |
| S-LSTM | 5s | 0.89 | 1.72 | 0.51 |
| S-GAN-20P | 5s | 0.61 | 1.21 | 0.75 |
| Sophie | 5s | 0.70 | 1.43 | 0.59 |
| Social-BiGAT | 5s | 0.69 | 1.30 | 0.71 |
| Social-STGCNN | 5s | 0.71 | 1.35 | 0.82 |
|  | 1s | 0.22 | 0.24 | 0.99 |
| Our Model (K=1) | 3s | 0.45 | 0.78 | 0.88 |
|  | 5s | 0.75 | 1.48 | 0.68 |
|  | 1s | **0.17** | **0.17** | **0.99** |
| Our Model (K=3) | 3s | **0.32** | **0.49** | **0.96** |
|  | 5s | **0.52** | **0.94** | **0.85** |



a) b) c) d)



e) f) g) h)

图5.8 多任务模型行人分支预测可视化图

Fig. 5.8 Visualization of Pedestrian Branch Prediction by Multi-task Model

表5.4 不同K值的模型指标比较分析

Tab. 5.4 Comparative analysis of model metrics with different K values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| K | Forecast duration | minADE | minFDE | Recall |
| 1 | 1s | 0.22 | 0.24 | 0.99 |
| 3s | 0.45 | 0.78 | 0.88 |
| 5s | 0.75 | 1.48 | 0.68 |
| 2 | 1s | 0.18 | 0.19 | 0.99 |
| 3s | 0.35 | 0.55 | 0.94 |
| 5s | 0.56 | 1.06 | 0.80 |
| 3 | 1s | **0.17** | **0.17** | **0.99** |
| 3s | **0.32** | **0.49** | **0.96** |
| 5s | **0.52** | **0.94** | **0.85** |
| 4 | 1s | 0.18 | 0.20 | 0.99 |
| 3s | 0.34 | 0.57 | 0.94 |
| 5s | 0.57 | 1.09 | 0.79 |



a) b) c)

图5.9 不同K值的模型指标的比较分析的可视化图

Fig. 5.9 Visualization diagram of comparative analysis of model metrics with different K values.

（2）轨迹预测分类分支K值的选择

在本节中，我们分析了不同预测K值的选择对模型预测精度的影响。如表5.4和图5.6所示，当模型选择不同的K值时，模型有着不同的预测精度，其中当K值为3时对应于最高的模型预测效果。从结果中可以进一步得出一个更大的结论。K表示行人未来可采取的行为模式的数量。不同的K对应不同的行为模式。直观地说，当K = 3时，我们的模型可以很好地拟合行人轨迹预测问题。在本文的其他实验中，我们选取使用K = 3作为多任务多模概率轨迹预测模型的预测参数。

## 5.5 本章小结

在本章中，我们对复杂道路交通场景中多任务多模概率化轨迹预测模型进行了网络模型的结构设计，设计多任务模型的损失函数和训练参数的确定。最后，对机动车、非机动车的单一模型的结果进行了定量和定性分析，并将其与多任务模型进行了对比分析，分析得到对于机动车和非机动车，使用多任务网络模型不仅能够取得很高的轨迹预测精度，提高模型预测效率，而且其能够产生更为鲁棒性、更符合社会可接受性的轨迹序列；然后通过行人轨迹预测模型与一系列基线模型的对比分析，得到多任务模型的行人分支具有最优的精度和性能。

# 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

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。

，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

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