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

随着当前人工智能技术的飞速发展，自动驾驶技术已经成为汽车行业的研究热点，对于提高交通安全、增加社会和经济效益都具有重要意义。在多智能体交互的复杂道路交通场景中自动驾驶汽车能否准确的感知周围交通场景，做出合理的决策是保证自动驾驶技术安全性和有效性的重要前提，其中本文研究的预测问题就是连接感知和决策模块的重要环节。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分析，设计多任务网络，构建多任务多模概率轨迹预测模型，为自动驾驶汽车后续的行为决策和控制奠定基础。

首先，由于当前的行为预测数据集数据类别过于单一，无法对场景间的物理拓扑关系和历史序列信息进行有效的表达，并且多数数据集数据样本太少，无法使模型学习到尽可能多的行为模式。针对上述问题，本文开发了包含道路高清地图和障碍物历史序列的大型行为预测数据集，可以将场景中的环境及各类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和历史序列信息。

之后，基于 State-Anchor 与 Anchor-Free 的建模思想搭建机动车、非机动车与行人多模概率轨迹预测模型。首先通过将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设计和参数选用。最后，基于 Anchor-Free 的建模方法搭建行人多模概率轨迹预测模型。

最后，为提高模型预测效率，基于上述模型，提出多任务多模概率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并行预测。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，对交通场景中的各智能体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轨迹预测模型进行了定量和定性的分析比较，实证本文所提出模型的性能要优于当前主流模型。通过对模型预测结果进行可视化分析表明，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，为下游做出更合理，平顺的决策提供多条概率化可选轨迹。

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 current artificial intelligence technology, autonomous driving technology has become a research hotspot in the automotive industry. It is of great significance for improving traffic safety and increasing social and economic benefits. In the complex road traffic scenes with multi-agent interaction, whether autonomous vehicles can accurately perceive the surrounding traffic scenes and make reasonable decisions is an important prerequisite for ensuring the safety and effectiveness of autonomous driving technology. Among them, the prediction problem is an important link between the perception and decision-making modules in autonomous driving. Based on the modeling ideas of State-Anchor and Anchor-Free, this paper aims at complex road traffic scenes to model and analyze vehicles, cyclists and pedestrians respectively, build a multi-task multi-model probability trajectory prediction model, it will lay the foundation for the subsequent behavioral decision-making and control of autonomous vehicles.

First of all, because the current behavior prediction dataset' data category is too single, it cannot effectively indicate the physical topological relationship between the scenes and historical sequence information, and most data sets have too few data samples to support the model to learn as many behavior patterns as possible. In response to the above problems, this paper developed a large-scale behavior prediction data set which containing road High-Definition maps and obstacle historical sequences, which can rasterize the environment and various agents in the scene, and provide rich semantic topology information and historical sequence information for subsequent model modeling.

After that, based on the modeling ideas of State-Anchor and Anchor-Free, a multi-mode probabilistic trajectory prediction model for vehicles, cyclists and pedestrians is built. First, by clustering the trajectories of vehicles and cyclists, the intent uncertainty and control uncertainty of vehicles and cyclists are modeled. After that, the network architecture of Encoder-Interaction-Decoder is used to design the network models of vehicles and cyclists, and the design of the model loss function and the selection of parameters are completed. Finally, build a pedestrian multi-mode probability trajectory prediction model based on the Anchor-Free modeling method.

Finally, in order to improve the efficiency of model prediction, based on the above three models, a multi-task multi-mode probabilistic trajectory prediction model is proposed to

integrate vehicle, cyclist and pedestrian models to achieve real-time prediction of all agents in the scene. By using interest regions of different sizes in the Spatial Transformation network, combined with the use of a multi-head attention mechanism, the complex interaction of each agent in the traffic scene is modeled, and effective prediction of various agents is realized. At the end of the article, a quantitative and qualitative analysis and comparison between this model and the current mainstream single-category trajectory prediction models are carried out. The empirical performance of this model is better than that of the current mainstream model. Visual analysis of the model prediction results shows that the multi-task model can predict multiple socially acceptable trajectory sequences for each agent, and provide multiple probabilistic trajectories for downstream to make more reasonable and smooth decision.

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

目　　录

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[1]。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，在未来几年仍将获得更加突飞猛进的发展[2]。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[3]）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，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[4]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[5][6]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技术的极大兴趣，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。自动驾驶平台作为自动驾驶汽车的终端，无论是从国家战略意义上，还是从经济和社会效益上都具有举足轻重的作用。在国际上，谷歌（Google）子公司 Waymo 率先推出了自己的无人驾驶汽车（图 1.1（a））， 美国的自动驾驶初创公司 Uber（现已被收购）在亚利桑那州推出了自动驾驶出租车的服务（图 1.1（b）），

日本的丰田公司推出了无人巴士（图 1.1（c））；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，在进行 L4 级别自动驾驶汽车研发同时，并陆续推出各自 L3 级别的量产车型。在刚刚结束的中国上海车展中，华为，小米，百度，滴滴等一系列互联网公司都公开了自己的造车计划，一个属于自动驾驶汽车的时代俨然已经来临。

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”，

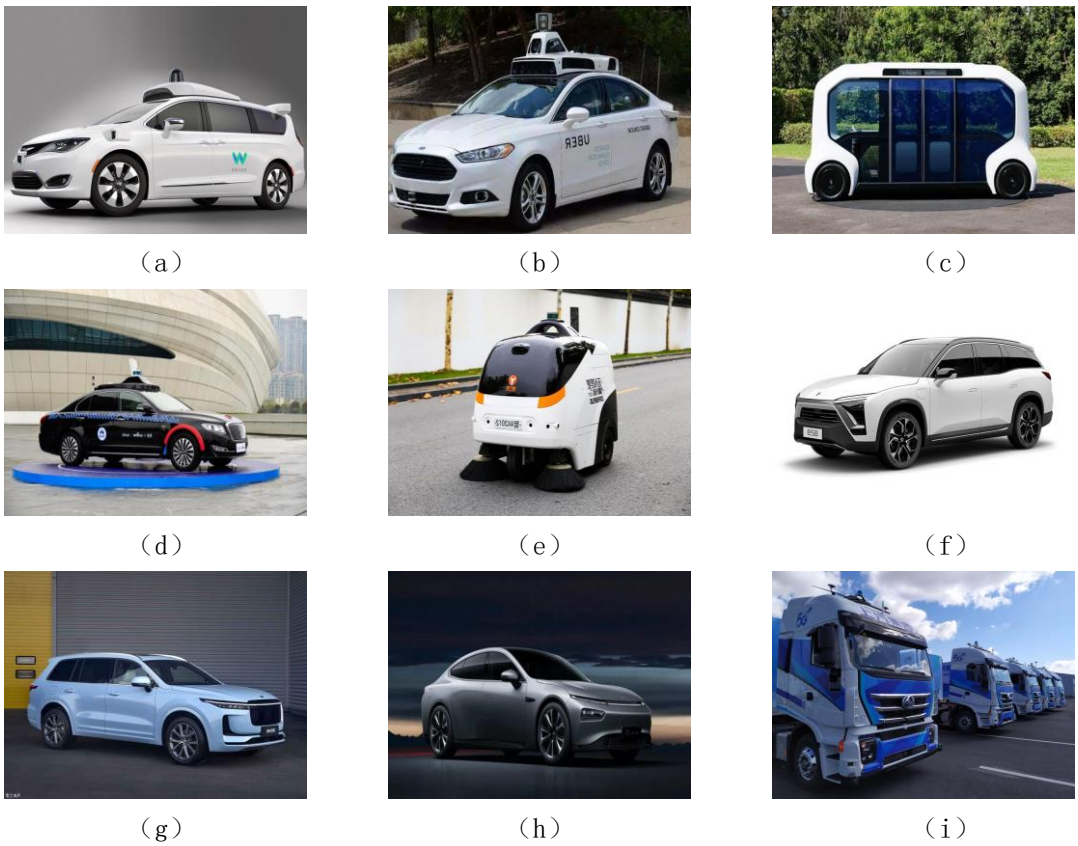


图 1.1 各类型无人驾驶车辆

Fig. 1.1 Different kinds of Self-driving cars

“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，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。

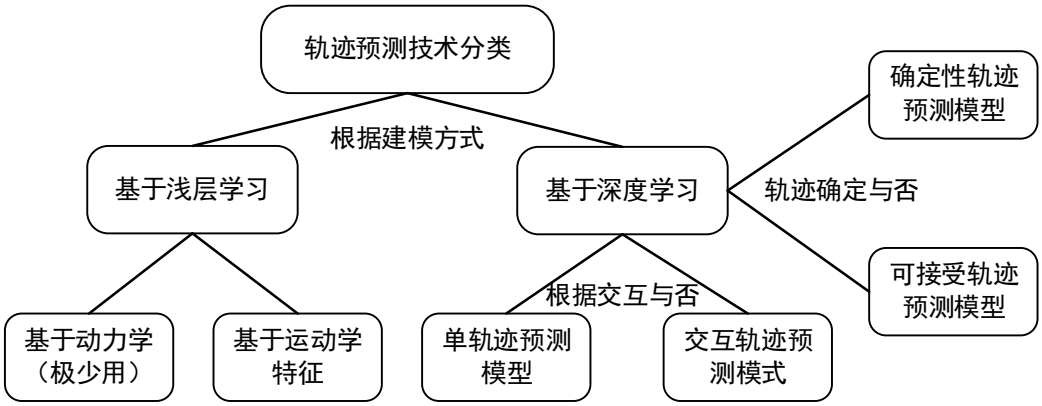


图 1.2 轨迹预测方法分类

Fig. 1.2 Classification of trajectory prediction methods

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。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[7]。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[8]。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[9]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[10]。Schneider 等[11]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横向位置预测提高 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[12]采用切换线性动力学系统（Switched linear dynamical system，SLDS）模型[13]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[14]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[15]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[16]领域。Alahi 等人[17]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[18]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[19-21]（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[19]，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：$X^t = f(X^{t-1})$或者$P(X^t|X^{t-1})$。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）[22]和门控递归神经网络（Gated Recurrent Unit，GRU）[23]在序列预测任务中（如语音识别[24,25]，标题生成[26]，机器翻译[16]）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$P(X^t|X^{t-1}$）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[27]。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[28]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[29]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[30]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[31]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[32]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[33]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[34]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[35]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[36]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[37]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[38]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[39]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）[40]，变分自编码器（Variational Auto- Encoder, VAE））[41]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[42]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[43]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[44]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[45]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[46]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[47]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$z \sim N(\mu,\sigma)$直接采样，文中的噪声分布为$z = \mu + \sigma \odot \epsilon$，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，并按随机变量采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[48]也采用了这种方法。

Liang 等人[49]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[50]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[51]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[52, 53]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[54]。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, 55, 56, 57]将 GCN 扩展到其他应用，例如矩阵计算和变分自动编码器。

Gupta 等人和 Sadeghian 等人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[58]，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[59]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[60]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[46]将图注意力（GAT, Graph Attention Network）网络[18]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[61]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[62]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[63]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[64]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高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主要工作进行研究，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，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的可预测性为出发点，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，将交通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有效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。

（4）本文通过引入多任务网络，设计多任务多模概率轨迹预测模型，实现对场景中所有智能体的实时预测，并在精度以及社会可接受性上取得了优异的性能。

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，为模型学习提供丰富的语义拓扑信息和轨迹序列信息。

第三章，基于 State-Anchor 的多模概率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意图不确定性和控制不确定性建模；之后设计基于 Encoder-Interaction-Decoder 的网络结构，分别对网络各个模块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不确定性建模，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轨迹预测模型。该章首先基于上述三章设计的单一任务的轨迹预测模型，从提高模型预测的精度和效率的角度，寻求模型融合预测的可能性；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与当前主流模型的性能进行了系统的定量和定性分析比较，并将单一模型与多任务模型进行定量和定性以及可视化分析，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，分析本文的不足，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 引言

第一章节中，本文对轨迹预测领域的国内外研究现状进行了详细的介绍了综述，并对当前轨迹预测问题的不足和缺点进行了总结，最后介绍整个论文主要研究内容，分别对论文的创新点和论文架构进行了介绍。在本章中，首先将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，为模型学习提供丰富的语义拓扑信息和轨迹序列信息。最后介绍在本文进行相关网络模型搭建所需要的一些深度学习基础知识。

2.2 轨迹预测问题定义

本文将轨迹预测问题表示为基于过往智能体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、非机动车和行人的位置都由它们的 xy 坐标(x_i^t, y_i^t)表示。使用索引$i, j \in \{1,2 \dots, N\}$来表示智能体的编号，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，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，模型的输入有两块，分别是交通场景信息和智能体轨迹序列信息，其中智能体的轨迹信息可以由四维向量X_i编码，其中$X_i = \{(x_i^t, y_i^t, v_i^t, \varphi_i^t) \in R^2 | t = 1, \dots, t_{obs}\}$。在给定这些输入特征和每个行人的真实轨迹$Y_i$的情况下（其中$Y_i = \{(x_i^t, y_i^t) \in R^2 | t = t_{obs} + 1, \dots, t_{pred}\}$），本文所提出的系列模型将通过学习模型的权重参数矩阵W'来预测未来每个行人的轨迹$\hat{Y}_i$。

$$\hat{Y}_i = f(X_i; W')$$

其中，$\hat{Y}_i = \{(\hat{X}_i^t, \hat{y}_i^t) \in R^2 | t = t_{obs} + 1, \dots, t_{pred}\}$，模型参数$W'$为模型中使用的所有深度神经网络的权值，$t_{pred}$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$\mathcal{L}$。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、非机动车和行人的多任务多模概率轨迹预测模型。首先在本章中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语义拓扑信息和交通智能体的轨迹信息。再接下来，本章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，本章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。

2.3 智能体不确定性建模

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，真实的司机都无法通过主观思想去判断每个智能体的意图和接下来的行为，更何况是由 AI 控制的自动驾驶汽车。本文将智能体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。

在模型的意图不确定性建模时，针对于交通场景中的机动车和非机动车，设计并使用一组离散的 State-Anchor 轨迹序列$\mathcal{A} = \{a^k\}_{k=1}^K$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，$a^k = \{a_1^k, a_2^k, \ldots, a_T^k\}$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：

$$\pi(a^k|x) = \frac{exp f_k(x)}{\sum_i exp f_i(x)} \tag{2.1}$$

其中，x 为模型的输入，$f_k(x)$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，研究[35,39,45]所得，直接对轨迹进行学习容易造成模式崩溃的问题，借鉴先去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的平方距离，

$$d(u, v) = \sum_t^T \left|\left|M_u u_t - M_v v_t\right|\right|_2 \tag{2.2}$$

其中，M_u，M_v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$K = 32$时机动车聚类结果的可视化分析图。

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：

$$\varphi\left(s_t^k \middle| a^k, x\right) = \phi(s_t^k|\mu_x^k, \mu_y^k, \sigma_x^k, \sigma_y^k, \rho^k) \tag{2.3}$$

其中，s_t^k为第 k 个 Anchor 轨迹的预测序列，$\mu_x^k, \mu_y^k, \sigma_x^k, \sigma_y^k, \rho^k$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：

$$p(s|x) = \sum_{k=1}^{K} \pi(a^k|x) \prod_{t=1}^{T} \varphi\left(s_t^k \middle| a^k, x\right) \quad (2.4)$$

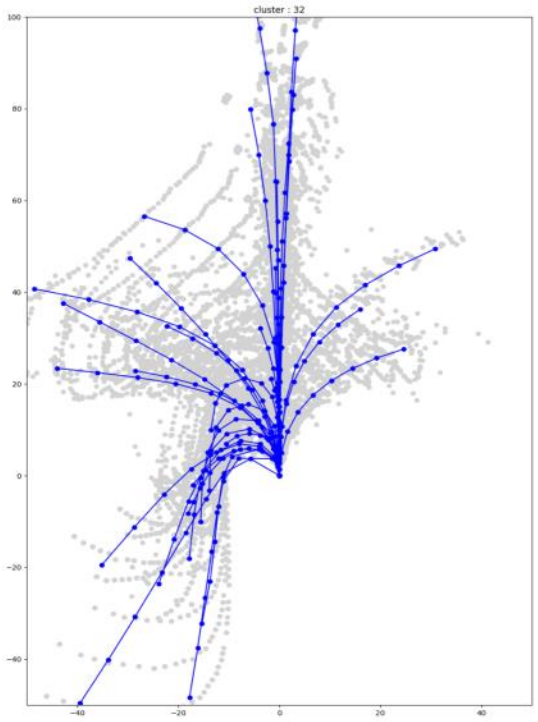


图 2.1　K = 32 时车辆聚类结果可视化图

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

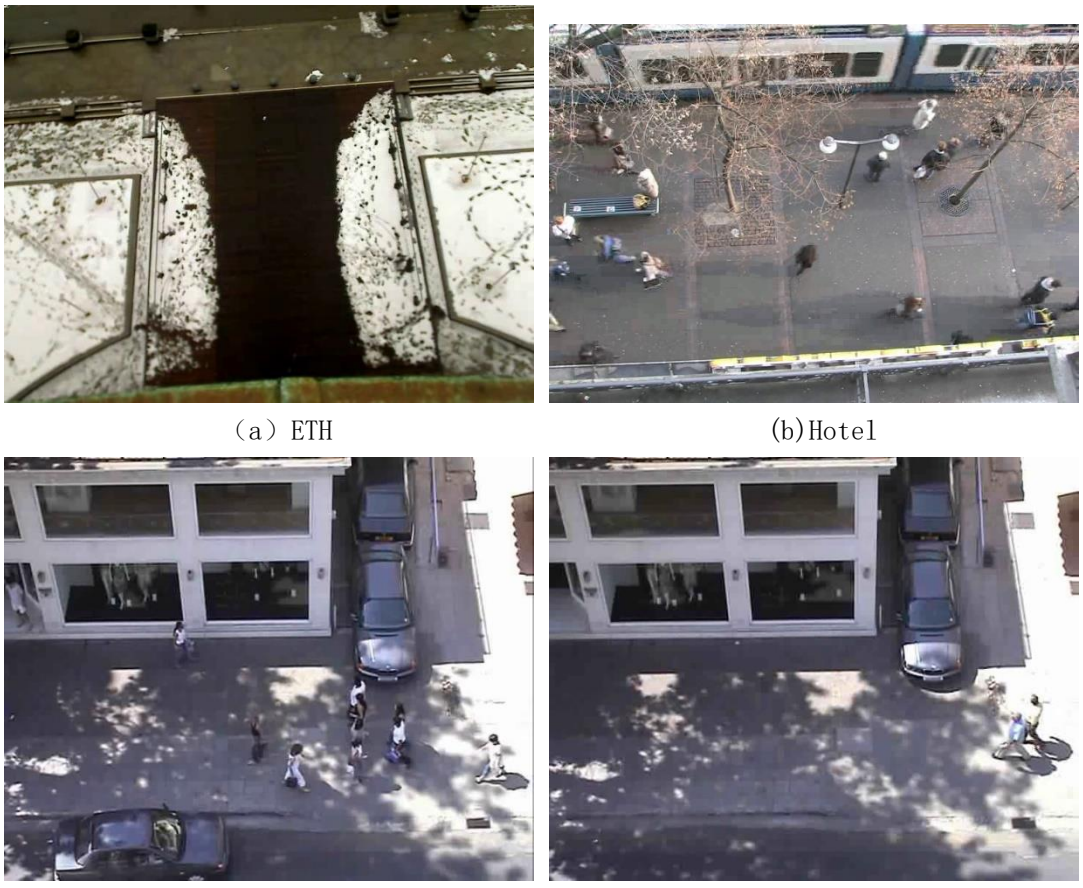
2.4　数据集

在正式开始介绍本文的模型之前，将介绍关于在复杂交通场景中进行轨迹预测问题的数据集。

2.4.1　常用的轨迹预测数据集

（1）ETH、UCY 数据集

ETH[65]和 UCY[66]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



（a）ETH　　(b)Hotel

（c）Zara-1　　(b)Zara-2

图 2.2　ETH、UCY 数据集

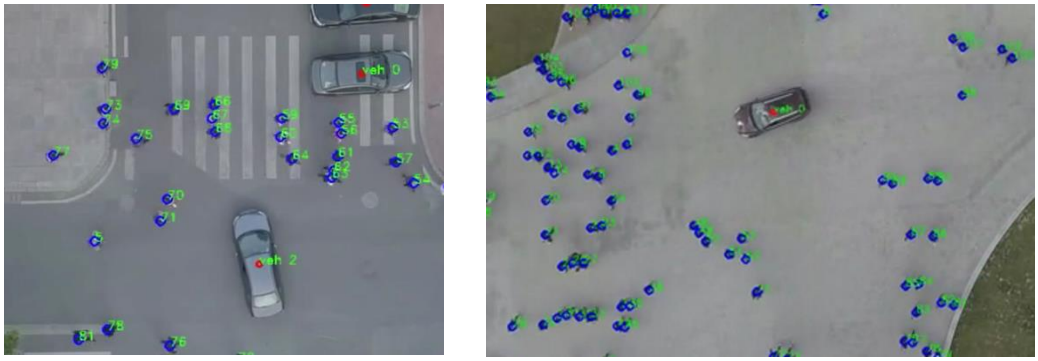
Fig. 2.2　DtaSet of ETH and UCY

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-Uuiv 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。

（2）DUT 数据集

DUT 数据集是具有复杂道路交通场景的路段通过航拍在中国大连理工大学校区内部采集的，旨在开发适用于人-人，人-车交互运动的轨迹预测模型，探索无交通规则或弱交通规则下轨迹预测模型，如图 2.3 所示。图 2.3-（a）为无交通信号的交叉路口。在该交通场景中由于没有交通信号灯，当行人与车辆进行交互时，可以体现人车交互具有显然的“社会性”。图 2.3-（b）为环岛场景，行人和车辆在不受交规的场景中自由移动。数据集的采集通过采用一架 DJI-Mavic-Pro 在行人和车辆难以察觉的高度拍摄。视频分辨率为 1920×1080，fps 为 25。数据集中共包括有 17 个交叉十字路口场景和 11 个环岛场景，共包含 1793 条轨迹序列。



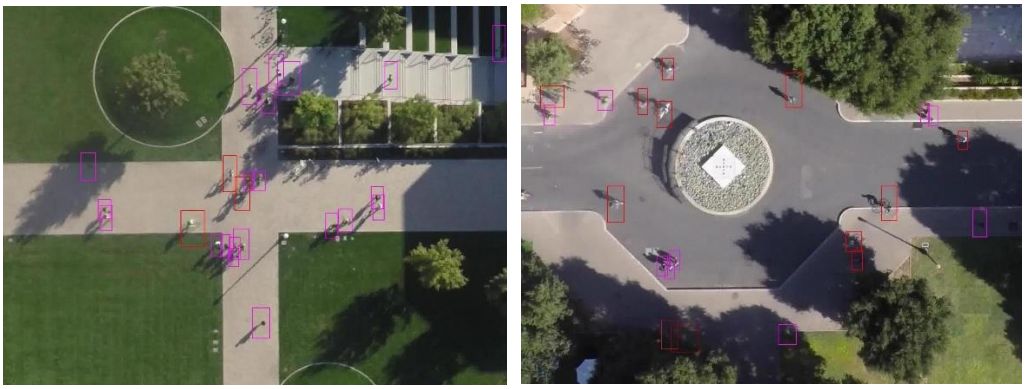
（a）无交通信号交叉路口　　（b）环岛

图 2.3　DUT 无人机数据集[67]

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[68]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 2.4 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，这些智能体在真实世界的户外环境中出现，在图片中，行人用粉红色的标签表示，自行车使用红色标签表示，滑板者用橙色标签表示，汽车使用绿色标签表示。



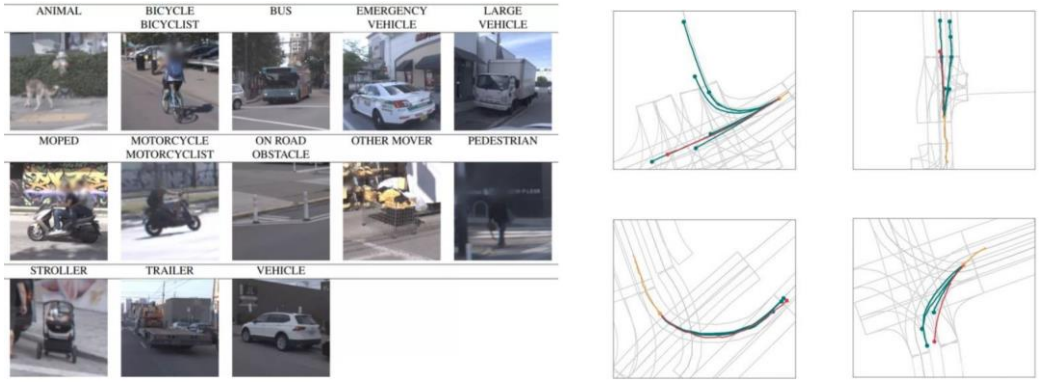
（a）quad 场景　　（b）death-Circle 场景

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，并将高精度地图与 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。



（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

该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。

2.4.2 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类型为复合栅格地图，由原始高清地图转化而来（图 2.6），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。

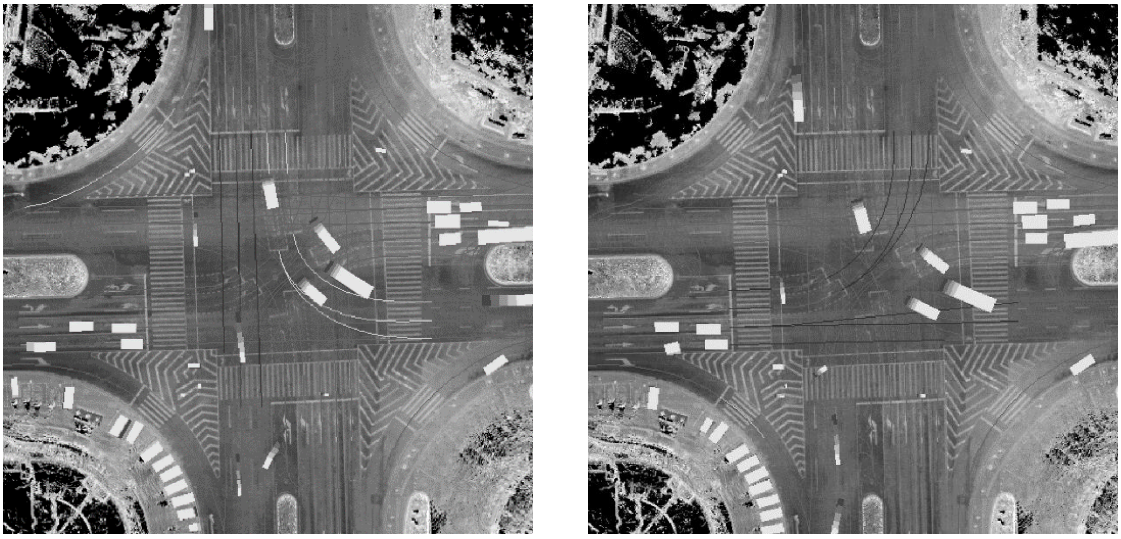


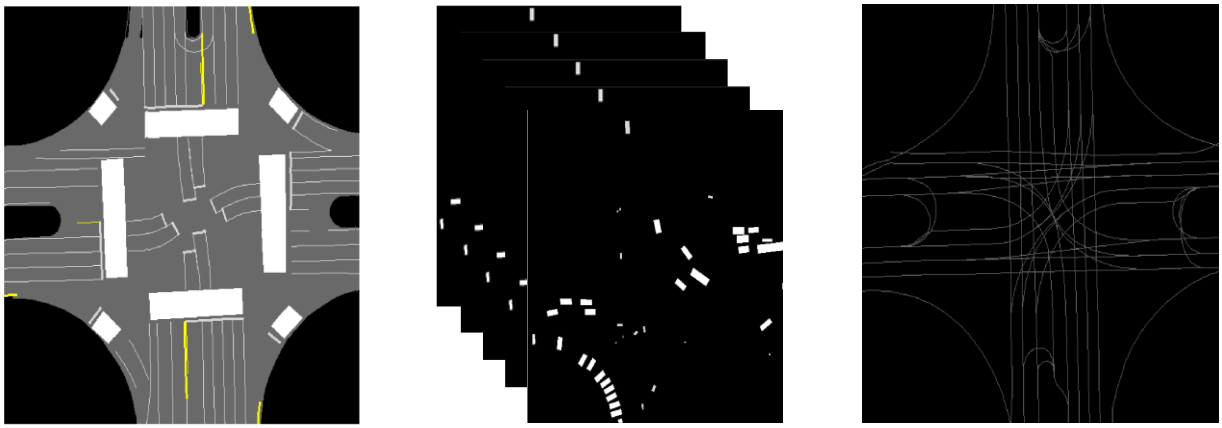
图 2.6 原始高清地图

Fig. 2.6 Original HD map

如图 2.7 所示，本文展示了两种地图表示形式，其中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。这些元素组成了复合栅格地图。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。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，通道数为 3；（b）是不同智能体的历史位置，包含 5 个通道，表示为 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均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以及障碍物的位置和轨迹信息。在该数据集中，进行采集的感知车辆（图中红车）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。

(a) 高清静态地图（白底）　（b）智能体历史位置图（白底）　（c）动态红绿灯示意图（白底）



(a) 高清静态地图（黑底）　（b）智能体历史位置图（黑底）　（c）动态红绿灯示意图（黑底）

图 2.7　复合栅格地图的组成

Fig. 2.7　The composition of a composite raster map

2.5 网络搭建相关基础知识

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

2.5.1 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）通过将各个神经网络层排列起来，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 CNN 及其变体大多由四种类型的神经网络层组成：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。

首先卷积神经网络的核心是卷积层，其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，以标准卷积为例进行讲解。假设使用图像作为卷积层的输入，卷积计算的流程是通过使用卷积核在要处理的输入图像上进行有规律的移动，来提取图像不同区域的特征信息，其中卷积核作为卷积层的基础计算组件，主要参数有卷积核大小 F（Filter）、步长 S（Stride）。其中卷积层感受野的大小就是卷积核的大小，如图 2.8 所示，蓝色的方格区域为输入的原始图像大小为 5x5、卷积核大小为 3x3，步长 S 为 2。紫色方格为卷积操作所输出的特征图，输出特征图尺寸为$W_{out} * H_{out}$为：

$$W_{out} = (W_{in} - F + 2P) / S + 1 \tag{2.5}$$

$$H_{out} = (H_{in} - F + 2P) / S + 1 \tag{2.6}$$

其中W_{in}、H_{in}为输入图像的长和宽。P 为是否采用全零填充，卷积层的计算公式为：

$$a_{i,j} = f(\sum_{m}\sum_{n} w_{m,n} x_{i+m,j+n} + w_b) \tag{2.7}$$

其中，$a_{i,j}$是输出特征图第i行j列的特征值，$x_{i,j}$是输入特征图第i行j列的特征值，w_b为卷积层的偏置项。$f(\bullet)$为激活函数。

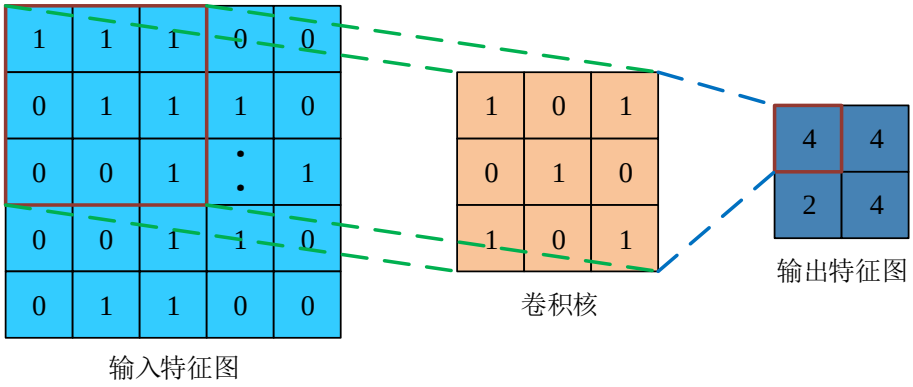


图 2.8　卷积操作示意图

Fig. 2.8　Convolution operation diagram

卷积层在处理图像等二维数据时，通过采用局部连接和参数共享的方法，可以极大的减少网络中的参数量，同时通过采样局部连接的技术可以更加有效的提取图像边缘特征。

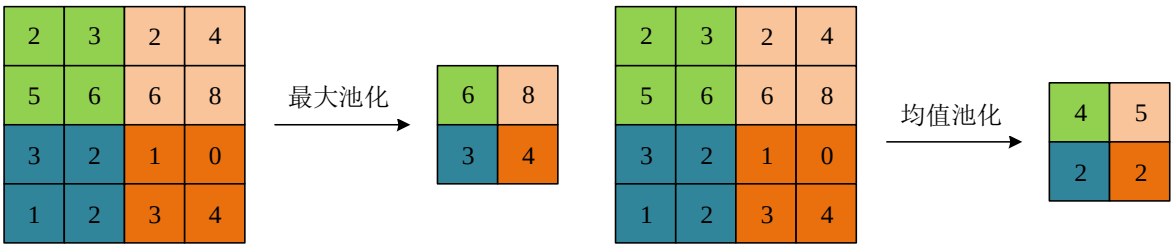


图 2.9　池化操作示意图

Fig. 2.9　Pooling operation diagram

为减少计算机计算资源的消耗，进一步有效的减少训练网络模型的参数量，减少在网络训练时出现过拟合现象，设计了池化层。现有的池化层主流方法有最大（Max）池化和平均（Mean）池化。最大池化是通过使用池化窗口$F \times F$中的最大值来代表窗口内的特征信息，平均池化则使用平均值来表征特征信息。上图 2.9 中假设特征图的输入尺寸为 4x4，池化核大小F为 2，步长S为 2，计算公式为：

$$H_{out} = (H_{in} - F) / S + 1 \tag{2.8}$$

$$W_{out} = (W_{in} - F) / S + 1 \tag{2.9}$$

其中H_{out}，W_{out}分别为是输出特征图的高和宽，H_{in}，W_{in}分别是输入特征图的高和宽。

ImageNet 图像分类大赛是深度学习领域的主要推动者之一，在比赛中涌现了许多经典的用于图像处理的神经网络如 GoogLeNet，AlexNet 和 VGG 等。

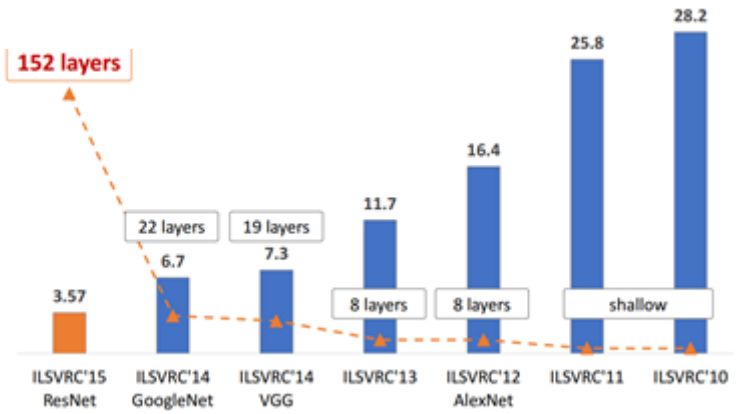


图 2.10　ImageNet 分类 Top-5 的误差率[69]

Fig. 2.10　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 2.11 所示。深度卷积与标准卷积不同，对于标准卷积，其卷积核作用在输入特征图的所有通道上，而深度卷积则是针对特征图的每个通道采用不同大小的卷积核进行卷积计算，换而言之就是一个卷积核只对应一个特征通道，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 2.12 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，是首先使用深度卷积网络对输入特征图的不同通道分别进行卷积操作，然后采用点式卷积将处理后的特征图输出进行结合，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，但与此同时会大大减少模型的计算量和参数量。

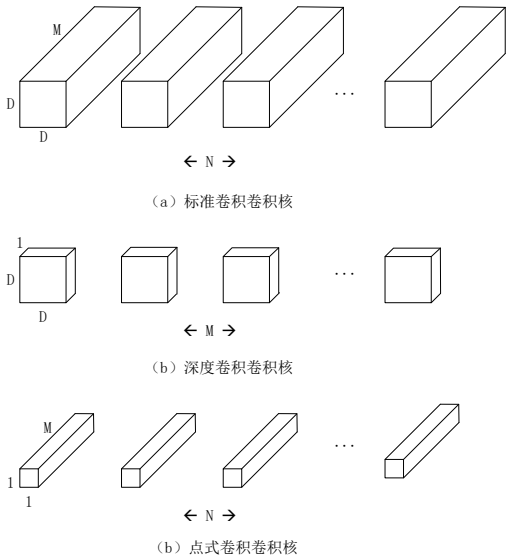


图 2.11　深度可分离卷积

Fig. 2.11　Depthwise Separable Convolution

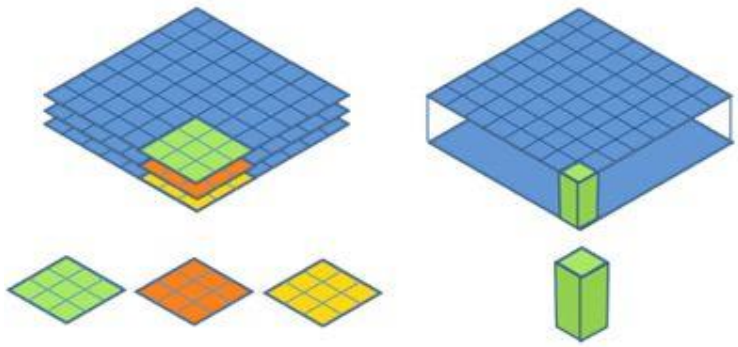


图 2.12　深度卷积和点式卷积

Fig. 2.12　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 2.13 所示：

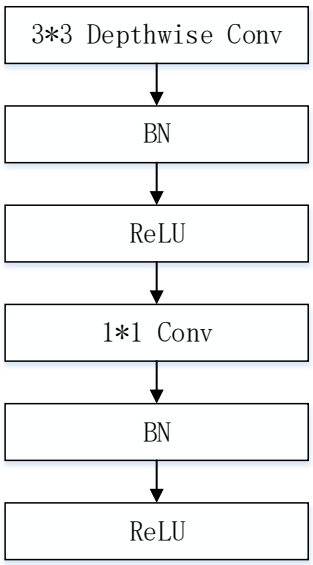


图 2.13　MobileNet 的基本组件

Fig. 2.13　Basic components of MobileNet

在论文里，作者将 MobileNet 网络与 GoogleNet 和 VGG16 做了对比，如表 2.1 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。

表 2.1　MobileNet 与 GoogleNet 和 VGG16 性能对比

Tab. 2.1　Performance comparison of MobileNet, GoogLenet and VGG16

模型	ImageNet Accuracy	Million Mult-Adds	Million Parameters
MobileNet	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

2.5.2　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 2.14 所示：

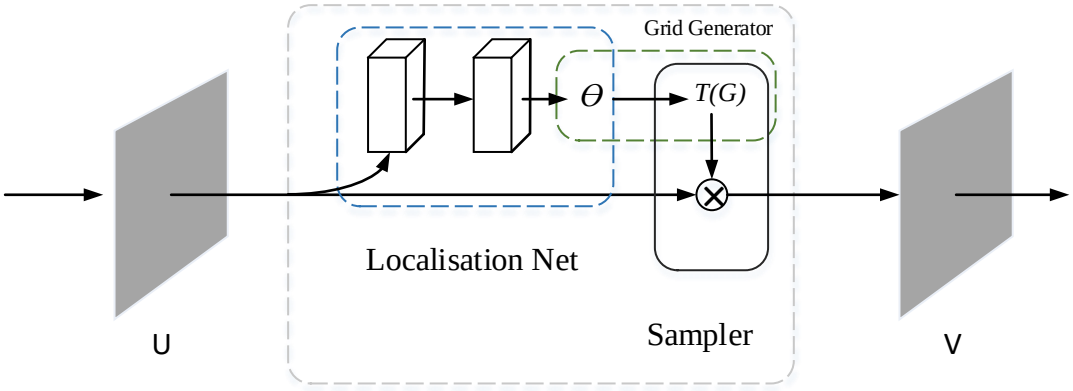


图 2.14　空间变换网络

Fig. 2.14　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)，网络根据输入特征图 U，计算输出参数θ，参数θ用来表示逆变换和 U 和 V 映射的坐标关系。Grid Generator 根据 V 中的坐标点和变化参数θ，计算出将要填充到 V 中的像素值在 U 中的坐标点。采样模块（Sampler）根据 Grid Generator 得到的一系列坐标和原始特征图 U 填充输出特征图 V。在应用 STN 时，首先先根据变换参数θ，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(i,j)的位置上，要对应输入图片的哪个位置。

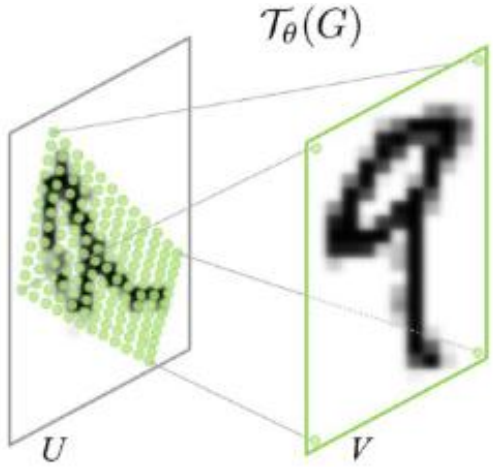


图 2.15　旋转缩放操作

Fig. 2.15　Rotate zoom operation

如图 2.15 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，经过变换$T_\theta(G)$，(θ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$G = G_i$表示，像素点的位置可以表示为$G_i = \{x_i^t, y_i^t\}$，这就是要确定的坐标。其次是应用仿射变换矩阵处理。

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = T_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \tag{2.10}$$

其中，$(\mathrm{x_i^t}, \mathrm{y_i^t})$是输出的目标图片的坐标，$(x_i^s, y_i^s)$是原图片的坐标，$A_\theta$表示仿射关系，即仿射矩阵。

2.6 本章小结

本章首先对轨迹预测问题的定义开始入手，对轨迹预测的不确定性进行意图和控制不确定性的分解，进而对整个交通场景中各智能体的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，并针对当前预测领域主流数据集数据类型单一，数据样本较少的问题，推出了自己的大型行为预测数据集。最后对模型搭建所需要的深度学习相关技术基础理论进行了简单的介绍。在接下来的章节中，本文将对机动车和非机动车的多模概率轨迹预测网络的搭建细节，进行详细地阐述。

3 基于 State-Anchor 的多模概率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。

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，对周围车辆的轨迹和行为进行精准的预测，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，对于主车而言，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，并通过上文的交通场景智能体的不确定性模型建模，对机动车和非机动车的不确定性等级进行分类：首先是智能体意图不确定性，通过一组固定的 Anchor 轨迹分布，用来捕获各智能体在未来时刻的意图；其次是智能体的控制不确定性，在给定好机动车和非机动车的行驶意图之后，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，进而生成未来时刻的一系列轨迹序列。在本章节中，机动车和非机动车的意图不确定性使用一组固定的 Anchor 轨迹分布建模，而控制不确定性则以假设智能体在未来时刻的轨迹点服从正态分布的先验条渐作为模型的建模基础。

3.2 机动车多模概率轨迹预测模型

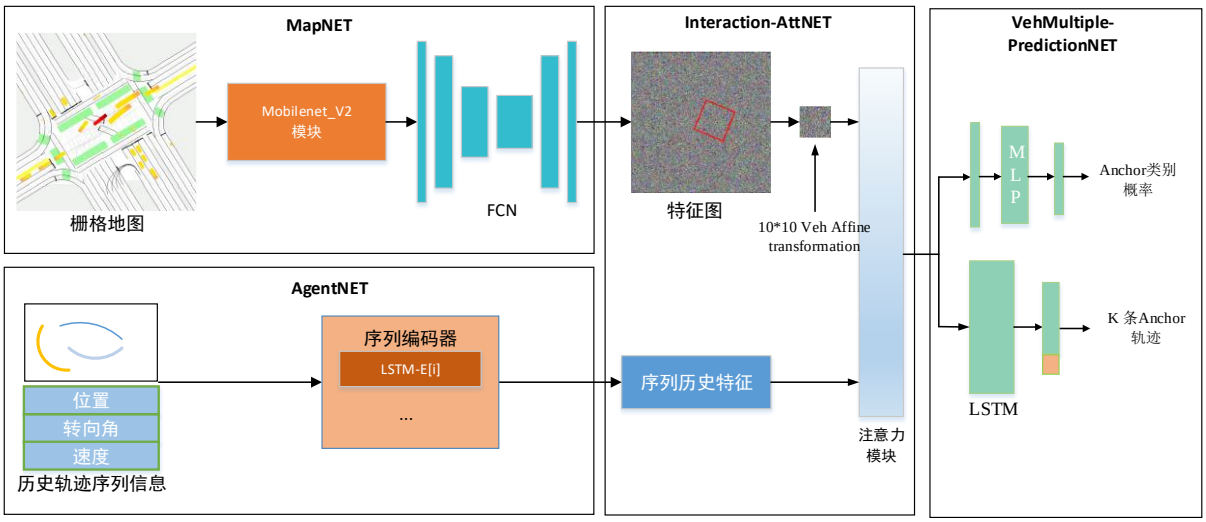


图 3.1　机动车多模概率化轨迹预测模型

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 Interaction-AttNET 交互注意力网络对提取到的历史轨迹序列特征和地图场景特征进行整合并提取智能体之间和智能体与环境之间的交互信息，最后通过 VehMultiple-PredictionNET 预测网络，对网络模型的输入进行预测。下面章节中将展开分别介绍这些网络模块。

3.2.1　MapNET：精确的时空表示

本小节提出了一种新的网络模型 MapNET，用于从复合栅格地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I，使用 MobileNet_V2 轻量特征提取网络，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V_p：

$$V_p = MapNET(I; W_p) \tag{3.1}$$

其中W_p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V_p，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 是用来接收机动车的历史状态X（四维向量：位置、速度和转向角，$X_i = \{(x_i^t, y_i^t, v_i^t, \varphi_i^t) \in R^2 | t = 1, \dots, t_{obs}\}$）作为输入，之后利用 LSTM 网络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X嵌入到更高维的特征向量中，然后使用 LSTM 对这些智能体的状态进行编码，编码生成特征向量$V_s(i)$。

$$V_s(i) = LSTM(MLP(X_i; W_{MLP}), h_{LSTM}(i); W_{LSTM}) \quad (3.2)$$

其中，智能体的编码特征向量$V_s(i)$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，提出了一种新交互注意力的网络来融合从 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[28]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[45]，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 Interaction-AttNET，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的视觉注意机制[70]，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应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，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。

在本章节中，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V_i和 MapNET 的特征图输出

V_p，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V_p进行空间变换网络（STN）来提取以预测智能体为中心的小特征图V_p'。然后另一个模块是使用多头注意力网络[54]来组合变换后的小特征图V_p'和序列特征$V_i(s)$。

$$V_p' = Affine(V_p) \tag{3.3}$$

$$C_p(i) = ATT\left(V_p', V_s(i); W_{ATT}\right) = \frac{1}{M}\sum_m c_p(i) \tag{3.4}$$

$$c_p(i) = softmax\left(\frac{V_p' \cdot MLP(V_s(i))}{\sqrt{d_{V_p'}}}\right) \cdot V_p' \tag{3.5}$$

其中$ATT(\cdot)$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$Affine(\cdot)$模块，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，所以基于此设计使用仿射变换网络$Affine(\cdot)$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$ATT(\cdot)$提取感兴趣区域的交互特征,使用特征$C_p(i)$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到机动车预测网络内 VehMultiple-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 10*10。

3.2.4 VehMultiple-PredictionNET

如 2.2 节所述，在本文中，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$C_p(i)$作为输入，VehMultiple-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 Anchor 的置信度得分。

如图 3.1 所示，VehMultiple-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 $K \times T \times (4+1)$ 参数（参数有$(\mu_x, \mu_y, \sigma_x, \sigma_y), \rho$）和分类分支来预测每条产生 $K \times T \times 1$ 个参数的 Anchor 类别的置信度，描述 *K* 个 Softmax 概率置信度来表示$\pi(s_t|X_t)$。

3.3 非机动车多模概率轨迹预测模型

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，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，本文研发基于 State-Anchor 的非机动车轨迹预测模型，在本章节中，首先对非机动车的 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 类别性质，然后基于此开发非机动车的轨迹预测模型。

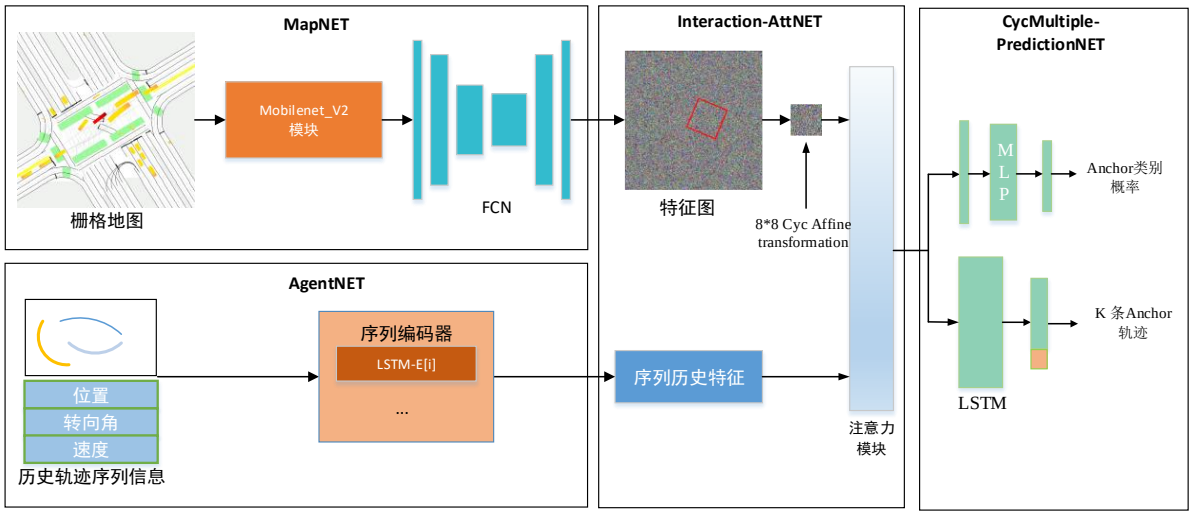


图 3.2　非机动车多模概率化轨迹预测模型

Fig. 3.2　Multi mode probabilistic trajectory prediction model for cyclists

3.3.1　非机动车轨迹预测模型验证分析

在本章节中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，

$$d(u,v)=\sum_t^T\left|\left|M_u u_t-M_v v_t\right|\right|_2 \tag{3.6}$$

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 3.3 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，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 CycMutiple-PredictionNET。模型的输入输出与上文介绍的基本一致，唯一有区别的点是，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，在设计非机动车的感兴趣区域时，设置其大小为 8*8。

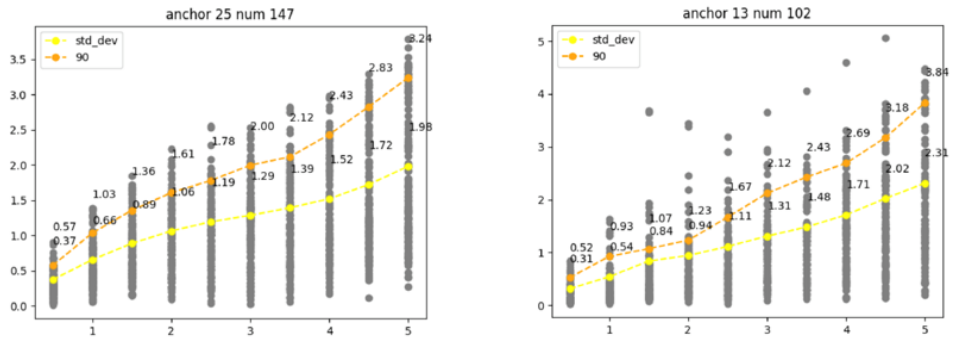


图 3.3　非机动车轨迹聚类分析

Fig. 3.3　Cluster analysis of cyclist trajectory

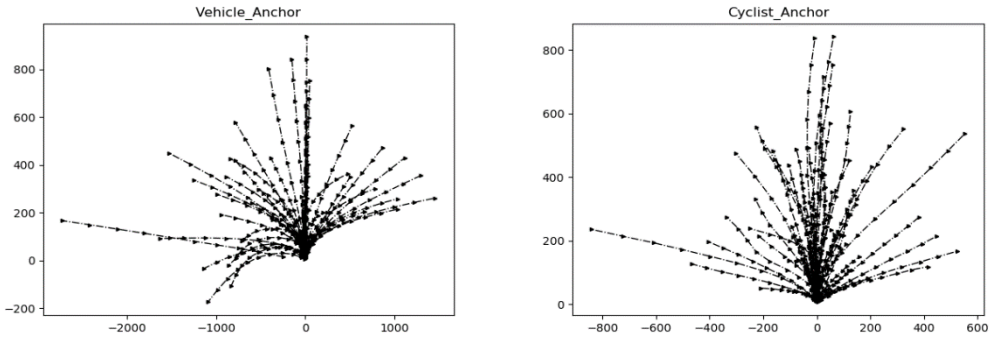


图 3.4　机动车与非机动车轨迹聚类对比

Fig. 3.4　Clustering comparison of vehicle and cyclist trajectories

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，$\pi(a^k|x)$，$\mu_x^k,\mu_y^k,\sigma_x^k,\sigma_y^k,\rho^k$），在本节中，将设计用来训练本文所提出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，所以模型可以使用端到端的方式分别训练机动车和非机动车模型。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：

$$L\big(Y,P(Y|X)\big)=-\log\big(P(Y|X)\big) \tag{3.7}$$

其中$P(Y|X)$为网络模型的输出。

在 2.2 节中，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：

$$p(s|x)=\sum_{k=1}^{K}\pi(\boldsymbol{a}^k|x)\prod_{t=1}^{T}\varphi\big(s_t^k\big|\boldsymbol{a}^k,\boldsymbol{x}\big) \tag{3.8}$$

基于此使用对数似然函数，推导机动车与非机动车的损失函数：

$$\mathcal{L}_{v,c}(\theta)=-log p(s|x) \tag{3.9}$$

$$\mathcal{L}_{v,c}(\theta)=-\sum_{m=1}^{M}\sum_{k=1}^{K}\begin{matrix}\mathbb{I}\big(k=\hat{k}^m\big)[log\pi(a^k|x^m;\theta)\\+\sum_{t=1}^{T}\log\ \varphi\big(s_t^k\big|a^k+\mu^k,x^m;\theta\big)]\end{matrix} \tag{3.10}$$

其中式 3.10 中，$\mathbb{I}(\cdot)$符号为 0-1 函数，$\hat{k}^m$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。

$$\hat{k}=k\ iff\ \min\big(\big|\big|\hat{s}_t-a_t^k\big|\big|_2\big) \tag{3.11}$$

此外，经过模型训练结果分析（图 3.5），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：

$$\mathcal{L}_v(\theta) = \begin{cases} \alpha \mathcal{L}_{v,c}(\theta)^{\lambda}, & v < 5m/s \\ \mathcal{L}_{v,c}(\theta), & v \geq 5m/s \end{cases} \tag{3.12}$$

即使用 Focal Loss 对速度低于 5m/s 的损失函数，引入α和λ变量。

在推导和改进完机动车与非机动车的损失函数后，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。

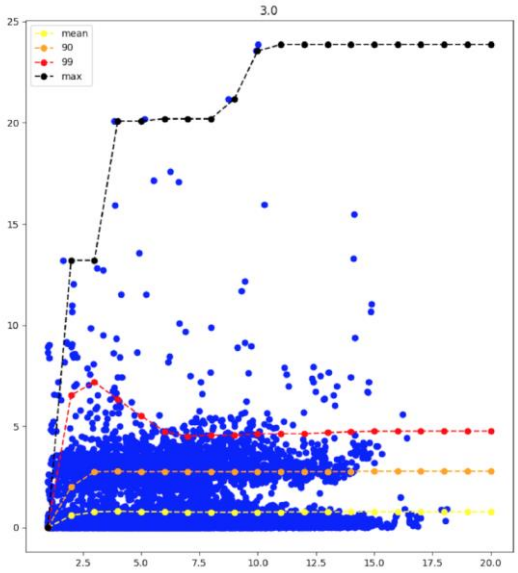


图 3.5　机动车模型预测速度-误差分布图

Fig. 3.5　Speed-Distance error distribution of vehicle prediction model

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 10.0 with CuDNN v8
数据处理	Python 2.7, Pandas, TFRecord etc.

3.5　网络模型的训练参数

本章节所设计的单一任务预测模型架构基于 TensorFlow 开源框架搭建，网络结构中的嵌入层 MLP 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 0.0005，训练次数 950k 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。

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，并分别介绍了设计网络模型的各个组成部分的技术细节。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选用。在下一章节中，本文将对道路交通场景中另一类复杂的交通智能体（行人）的轨迹预测模型进行介绍。

4 基于 Anchor-Free 的多模概率行人轨迹预测模型

4.1 引言

第三章中，文章对交通场景中的机动车、非机动车模型分别进行了建模，接下来将讨论交通场景中另一类主要的智能体，即行人。行人运动的数据分析对于道路安全、机器人导航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 Moussaid 等[71]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，文章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，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[71]，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，本章节将对该方法具体展开叙述。

4.2 行人轨迹预测模型建模

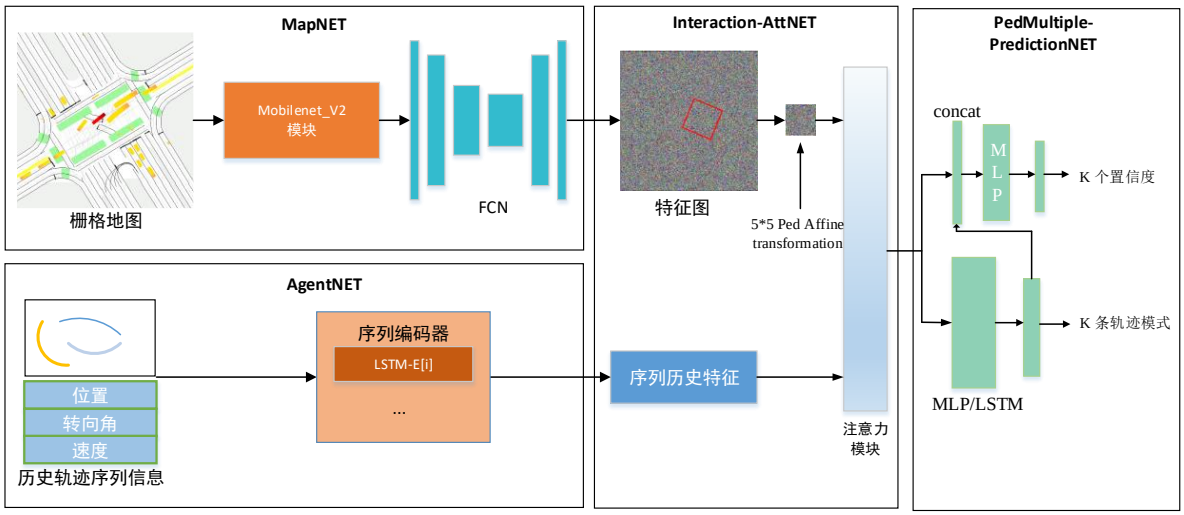


图 4.1 行人多模概率化轨迹预测模型

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，本章设计并使用基于 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$O_{m,cls}$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：

$$O_{m,cls} = \left(c_{m,0}, c_{m,1}, \dots, c_{m,K-1}\right) \tag{4.1}$$

$$\pi\left(O_{m,cls}^{k}\middle|x\right) = \frac{expc_{m,k}(x)}{\sum_i expc_{m,i}(x)} \tag{4.2}$$

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）：

$$\varphi\left(s_t^k\middle|O_{m,cls}, x\right) = \phi(s_t^k|\mu_x^k, \mu_y^k, \sigma_x^k, \sigma_y^k, \rho^k) \tag{4.3}$$

其中，s_t^k为第 k 个分类轨迹的预测序列，$\mu_x^k, \mu_y^k, \sigma_x^k, \sigma_y^k, \rho^k$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：

$$p_p(s|x) = \sum_{k=1}^{K} \pi\left(O_{m,cls}^{k}|x\right) \prod_{t=1}^{T} \varphi\left(s_t^k|O_{m,cls}, x\right) \quad (4.4)$$

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，本章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$C_p(i)$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，本章设计了 PedPredictionNET，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，在回归分支中使用 MLP 对交互注意力模块的输出$C_p(i)$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。

$$O_{m,reg} = RegNet(C_p(i); W_{reg}) \quad (4.5)$$

$$O_{m,reg} = \{(\boldsymbol{p}_{m,1}^{k}, \boldsymbol{p}_{m,2}^{k}, \boldsymbol{p}_{m,3}^{k}, \dots, \boldsymbol{p}_{m,t_{pred}}^{k}) | k \in [0, K-1]\} \quad (4.6)$$

其中，W_{reg}是回归网络所有结构的权重，$\boldsymbol{p}_{m,t}^{k}$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：

$$\boldsymbol{p}_{m,t}^{k}=\left(\mu_{x}^{k},\mu_{y}^{k},\sigma_{x}^{k},\sigma_{y}^{k},\rho^{k}\right)_{m,t} \tag{4.7}$$

对于分类分支，将回归分支的$\left(\mu_{x}^{k},\mu_{y}^{k}\right)_{m,t}$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。

$$O_{m,cls}=ClsNet(C_{p}(i),\left(\mu_{x}^{k},\mu_{y}^{k}\right)_{m,t};W_{cls}) \tag{4.8}$$

$$O_{m,cls}=\left(c_{m,0},c_{m,1},\ldots,c_{m,K-1}\right) \tag{4.9}$$

其中，W_{cls}是分类分支所有网络结构参数的权重。

4.3.2 行人多模概率轨迹预测模型损失函数设计

接下来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：

$$L\left(Y,P(Y|X)\right)=-\log\left(P(Y|X)\right) \tag{4.10}$$

$$\mathcal{L}_{v,c}(\theta)=-\sum_{m=1}^{M}\sum_{k=1}^{K}\begin{matrix}\mathbb{I}\left(k=\hat{k}^{m}\right)[log\pi(a^{k}|x^{m};\theta)+\\ \sum_{t=1}^{T}\log\ \varphi\left(s_{t}^{k}\middle|a^{k}+\mu^{k},x^{m};\theta\right)]\end{matrix} \tag{4.11}$$

为解决机动车低俗段 loss 的问题，引入了 focal loss：

$$\mathcal{L}_{v}(\theta)=\begin{cases}\alpha\mathcal{L}_{v,c}(\theta)^{\lambda},\ v<5m/s\\ \mathcal{L}_{v,c}(\theta),\qquad v\geq 5m/s\end{cases} \tag{4.12}$$

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，使用两个不同的模型分支对行人的行为进行预测，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，对模型可以使用端到端的方式进行训练。由于没有具体的 Anchor 类别的限制，模型训练过程中很容易出现模式崩溃的问题，即训练出的多模轨迹仍是平均化的单一轨迹，所谓的多模预测也由于模式崩溃变得毫无意义，为了解决这一问题，为此本文设计了分类和回归联合损失函数去构建模型的损失函数$\mathcal{L}_{p}$：

$$\mathcal{L}_{p}=\mathcal{L}_{cls}+\alpha\mathcal{L}_{reg} \tag{4.13}$$

其中，$\alpha\in(0,1]$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：

$$\mathcal{L}_{cls}=\frac{1}{M(K-1)}\sum_{m=1}^{M}\sum_{k\neq\hat{k}}\max\left(0,c_{m,k}+\varepsilon-c_{m,\hat{k}}\right) \tag{4.14}$$

其中，ε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。使用最大间隔损失函数的一大优点就是能够很好的避免模式崩溃的问题。

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，设计使用 Smooth-L1 损失函数优化所有的轨迹预测序列：

$$\mathcal{L}_{reg} = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} reg\left(\boldsymbol{p}_{m,t}^{k} - \boldsymbol{p}_{m,t}^{*}\right) \tag{4.15}$$

其中，$\boldsymbol{p}_{m,t}^{*}$为第 m 个行人在第 t 时刻的真实轨迹点，

$$reg(\cdot) = \sum_i d(x_i) \tag{4.16}$$

$d(x_i)$由 Smooth L1 函数定义：

$$d(x_i) = \begin{cases} 0.5x_i^2 & if\left||x_i|\right| < 1, \\ \left||x_i|\right| - 0.5 & otherwise \end{cases} \tag{4.17}$$

其中$\left||x_i|\right|$为 L1 距离。

4.4 网络训练

本章节行人轨迹预测模型的网络架构基于 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 0.0005，训练次数 950k 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 10.0 with CuDNN v8
数据处理	Python 2.7, Pandas, TFRecord，etc.

4.7 本章小结

本章在完成对机动车和非机动车的建模基础上，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，并对行人模型进行了建模；之后对行人多模概率预测模型的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。在下一

章节中，将对上述三个模型进行融合，以实现一个模型完成多个预测任务，并对上述几种模型进行相关性能指标的定量和定性分析。

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

目前为止本文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单一任务模型的建模，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，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多模概率轨迹预测，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，本章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

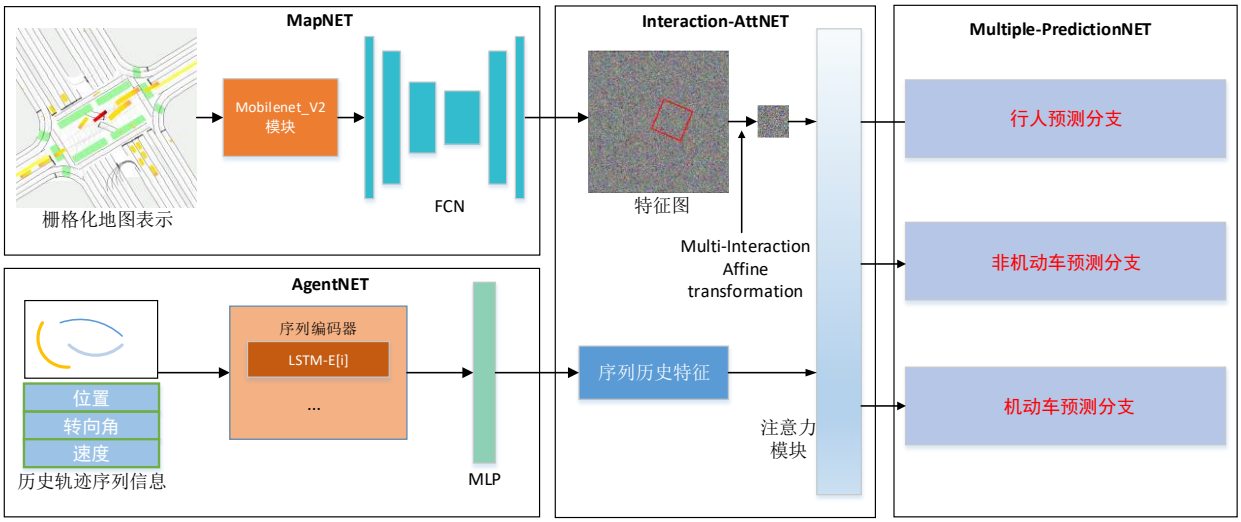


图 5.1 多任务多模概率化轨迹预测模型

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I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V_{p-all}：

$$V_{p-all} = MultiMapNet(I; W_{p'}) \tag{5.1}$$

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V_{p-all}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，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 VectorX_i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：

$$V_{s-all}(i) = LSTM(MLP(X_i; W_{MLP}), h_{LSTM}(i); W'_{LSTM}) \tag{5.2}$$

目前为止，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，完成了多任务模型建模的 Encoder 阶段。

在拿到模型编码的特征信息$V_{s-all}(i)$和V_{p-all}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，采用共享特征图V_{p-all}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，在驾车行驶在一条道路上时，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。所以本文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，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，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。

$$V'_{p-all} = Affine(V_{p-all}) \tag{5.3}$$

经过 STN 空间变换网络截取后的感兴趣区域V'_{p-all}，与编码后智能体轨迹历史序列轨迹信息$V_{s-all}(i)$，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$C_{p-all}(i)$。

$$C_{p-all}(i) = ATT\left(V'_{p-all}, V_{s-all}(i); W_{ATT}\right) = \frac{1}{M}\sum_m c_{p-all}(i) \tag{5.4}$$

$$c_{p-all}(i) = softmax\left(\frac{V'_{p-all} \cdot MLP(V_{s-all}(i))}{\sqrt{d_{V'_{p-all}}}}\right) \cdot V'_{p-all} \tag{5.5}$$

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，本章完成了 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$C_{p-all}(i)$，输入 Mulitiple-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。

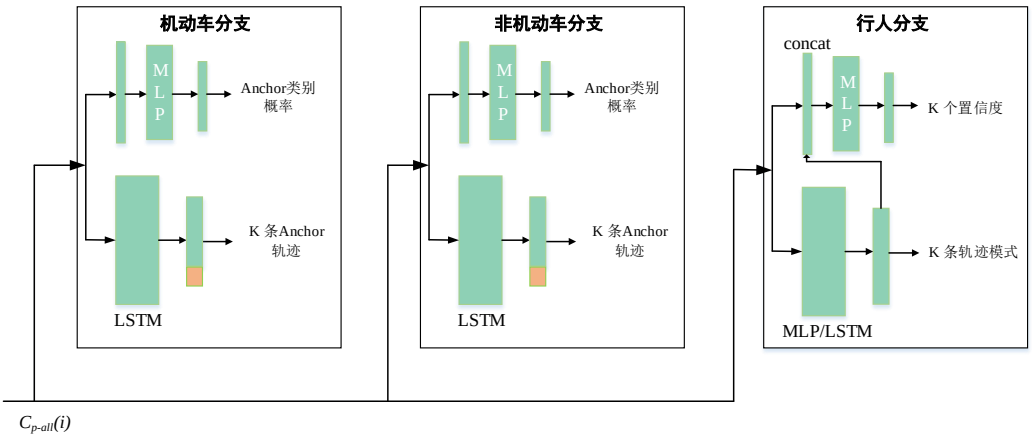


图 5.2　Mulitiple-PredictionNET 模型分支结构

Fig. 5.2　Branch structure of Mulitiple-PredictionNET model

如图（5.2）所示，首先搭建机动车和非机动车预测分支，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，本章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$\mathcal{L}$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$\mathcal{L}_v(\theta)$，非机动车多模概率化预测模型的损失函数$\mathcal{L}_c(\theta)$以及行人多模概率化轨迹预测模型的损失函数$\mathcal{L}_p(\theta)$。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。

$$\mathcal{L} = \alpha\mathcal{L}_v(\theta) + \beta\mathcal{L}_c(\theta) + \gamma\mathcal{L}_p(\theta) \tag{5.6}$$

$$\mathcal{L}_v(\theta) = \begin{cases} \alpha\ell_v(\theta)^{\lambda}, & v < 5m/s \\ \ell_v(\theta), & v \geq 5m/s \end{cases} \tag{5.7}$$

$$\ell_v(\theta) = -\sum_{m=1}^{M}\sum_{k=1}^{K}\begin{bmatrix}\mathbb{I}\left(k=\hat{k}^m\right)[log\pi(a^k|x^m;\theta)+\\ \sum_{t=1}^{T} log\ \varphi\left(s_t^k\middle|a^k+\mu^k,x^m;\theta\right)]\end{bmatrix} \tag{5.8}$$

$$\mathcal{L}_c(\theta) = -\sum_{m=1}^{M}\sum_{k=1}^{K}\begin{bmatrix}\mathbb{I}\left(k=\hat{k}^m\right)[log\pi(a^k|x^m;\theta)+\\ \sum_{t=1}^{T} log\ \varphi\left(s_t^k\middle|a^k+\mu^k,x^m;\theta\right)]\end{bmatrix} \tag{5.9}$$

$$\mathcal{L}_p = \mathcal{L}_{cls} + \delta\mathcal{L}_{reg} \tag{5.10}$$

$$\mathcal{L}_{cls} = \frac{1}{M(K-1)}\sum_{m=1}^{M}\sum_{k\neq\hat{k}}\max\left(0, c_{m,k} + \varepsilon - c_{m,\hat{k}}\right) \tag{5.11}$$

$$\mathcal{L}_{reg} = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} reg\left(\boldsymbol{p}_{m,t}^{k} - \boldsymbol{p}_{m,t}^{*}\right) \tag{5.12}$$

其中，$\alpha \in (0,1]$，$\beta \in (0,10]$，$\gamma \in (0,26]$，$\delta \in (0,1]$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

本章多任务模型架构基于 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 0.0005，训练次数 950k 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分析。

5.4 实验结果分析

在本小节中，将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，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 和 UCY。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（Zara-1、Zara-2 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。与之前的研究方法部分相似，使用以下指标来预测和评价每个智能体的轨迹：

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)；

$$minADE = min\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\|s_{i,t}^{k} - s_{i,t}^{gt}\|_2\right)_{k\in[1,K]} \tag{5.13}$$

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。

$$minFDE = min\left(\frac{1}{n}\sum_{i=1}^{n}\|s_{i,T}^{k} - s_{i,T}^{gt}\|_2\right)_{k\in[1,K]} \tag{5.14}$$

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）。

$$Recall = \begin{cases} (\sum_{k=0}^{K-1}\mathbb{1}_v(diff(s_{i,3}^{k}, s_{i,3}^{gt})))/K \\ (\sum_{k=0}^{K-1}\mathbb{1}_c(diff(s_{i,3}^{k}, s_{i,3}^{gt})))/K \\ (\sum_{k=0}^{K-1}\mathbb{1}_p(diff(s_{i,3}^{k}, s_{i,3}^{gt})))/K \end{cases} \tag{5.15}$$

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。接下来在实验过程中将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　测试轨迹数量及类型分布

Tab. 5.1　Number and type distribution of test trajectories

轨迹类型	未识别类型	直行	左转	右转	总计
机动车	8628	17740	3837	2282	32487
非机动车	14983	11146	2042	1418	29589
行人	10149	255	59	30	10493

由机动车和非机动车测试集的数据分布图 5.3 所示，可以清楚直观的看到直行的数据量占数据集的一大部分，是复杂交通场景中机动车和非机动车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%，这也进一步证明了对行人无法使用 Anchor 去区别各个模式的重要原因。

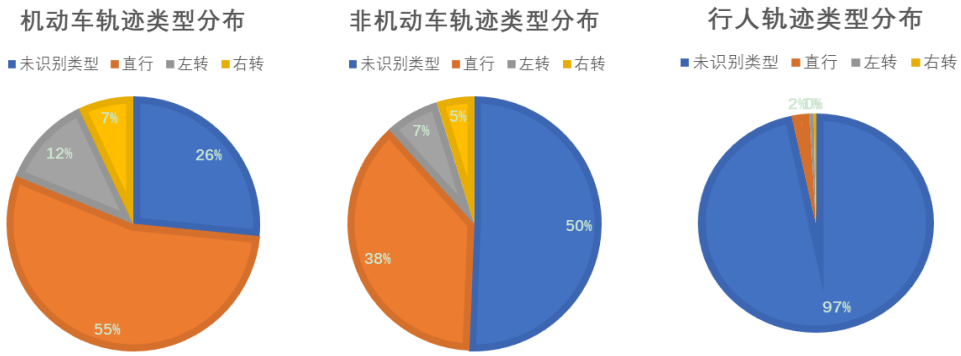


图 5.3　机动车、非机动车和行人轨迹类型分布

Fig. 5.3　Distribution of vehicle, non vehicle and pedestrian trajectory types

（2）机动车与非机动车多模概率预测模型定量分析

接下来，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（Unscented Kalman Filter，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，轨迹空间上的高斯混合模型。

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										
机动车	MinADE			MinFDE			Recall			Accuary（5s）
	1s	3s	5s	1s	3s	5s	1s	3s	5s	
UKF	1.03	2.16	3.98	1.19	3.83	6.99	0.71	0.24	0.09	0.22
STP	0.52	1.54	2.87	0.58	3.28	6.34	0.86	0.31	0.11	0.34
MDN	0.49	1.37	2.56	0.54	2.76	5.21	0.91	0.35	0.15	0.46
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**
非机动车	MinADE			MinFDE			Recall			Accuary（5s）
	1s	3s	5s	1s	3s	5s	1s	3s	5s	
UKF	1.15	2.07	4.38	1.23	3.47	7.28	0.69	0.27	0.12	0.22
STP	0.62	1.42	3.57	0.83	3.21	6.32	0.80	0.28	0.14	0.23
MDN	0.53	1.17	3.02	0.64	2.05	5.15	0.86	0.49	0.23	0.48
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										
机动车	MinADE			MinFDE			Recall			Accuary（5s）
	1s	3s	5s	1s	3s	5s	1s	3s	5s	
Multi-Ours1	0.25	0.81	1.62	0.34	1.63	3.65	0.98	0.58	0.26	0.64
Multi-Ours2	0.19	0.58	1.15	0.25	1.12	2.50	0.99	0.80	0.39	0.88
Multi-Ours3	**0.17**	**0.50**	**0.99**	**0.21**	**0.95**	**2.08**	**0.99**	**0.87**	**0.43**	**0.95**
非机动车	MinADE			MinFDE			Recall			Accuary（5s）
	1s	3s	5s	1s	3s	5s	1s	3s	5s	
Multi-Ours1	0.41	1.12	2.06	0.55	2.13	4.31	0.97	0.46	0.19	0.45
Multi-Ours2	0.35	0.91	1.61	0.45	1.65	3.25	0.98	0.58	0.31	0.67
Multi-Ours3	**0.32**	**0.80**	**1.41**	**0.41**	**1.43**	**2.81**	**0.98**	**0.64**	**0.38**	**0.77**

在表 5.2 中，本文将机动车和非机动车的轨迹预测模型与上述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，Recall 以及 Accuracy。

由表 5.2 可得，无论是本文提出的单一任务模型还是多任务模型的机动车和非机动车分支，相较于当前主流的机动车和非机动车预测模型有了很大的提升。其中对于机动车分支当 Anchor 取值为 1 时，机动车的 3s 点的 ADE、FDE 达到了 0.82、1.66，3s 点的召回率 0.59，较之前机动车的轨迹模型的最好效果召回率提高了 68.6%，且随着预测 Anchor 的数量的提高机动车的预测性能指标稳步上升，其中当预测 Anchor 的数目为 3 时，机动车 3s 点的 ADE、FDE 达到了 0.54 和 1.00，3s 点的召回率达到了 0.85，较当前主流模型提高了 142.9%，较 Anchor 值为 1 的 3s 点召回提高了 44.1%；对于非机动车分支的整体预测效果趋势与机动车一致，但是指标较机动车偏大，这是由于非机动车本身固有的灵活性和随机性，间接的限制了模型的性能。同样从表中可以得知，非机动车分支的预测效果要明显的优于当前主流模型，并在预测 Anchor 数目达到 3 时，模型的性能达到了最高。在附表中，展示了多任务模型的机动车和非机动车分支的预测效果的数据。进行相应的数据分析后，发现机动车的多任务模型要比单一任务模型的性能要好，3s 点的召回率要提高了 2%，而非机动车分支的性能略有下降，这是因为进行多任务模型的训练时，难免会使得模型在处理一个任务时，发挥的作用大，而另一个任务的性能会稍微减弱，且同一张复合栅格地图中，机动车的数量也要比非机动车的数量多，从表 5.1 中也可以发现这一规律，训练数据量的相对较少也是造成这一现象的主要原因。但由于交通场景中，机动车作为交通智能体中的主体，精度的提升要比非机动车来说更具有性价比。

综上所述，无论是单一任务模型还是多任务模型，多模概率轨迹预测模型的性能要比当前主流的轨迹预测模型优越，且随着多模预测模型的模式增多，模型的精度等指标也会越来越好。对比多任务模型和单一任务模型，发现机动车模型的性能较单一任务模型有略微的提高，而非机动车模型则稍有下降，这也符合机动车是交通场景中最为主要智能体的事实。

（3）不同 Anchor 数值的选择

本小节通过对数据集中智能体轨迹使用仿射变换，将智能体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 5.4 可以发现，轨迹点的有序程度依次递减，且随着聚类类别的增加，轨迹模式越来越清晰，大致可以分为左转，右转，直行，加速，

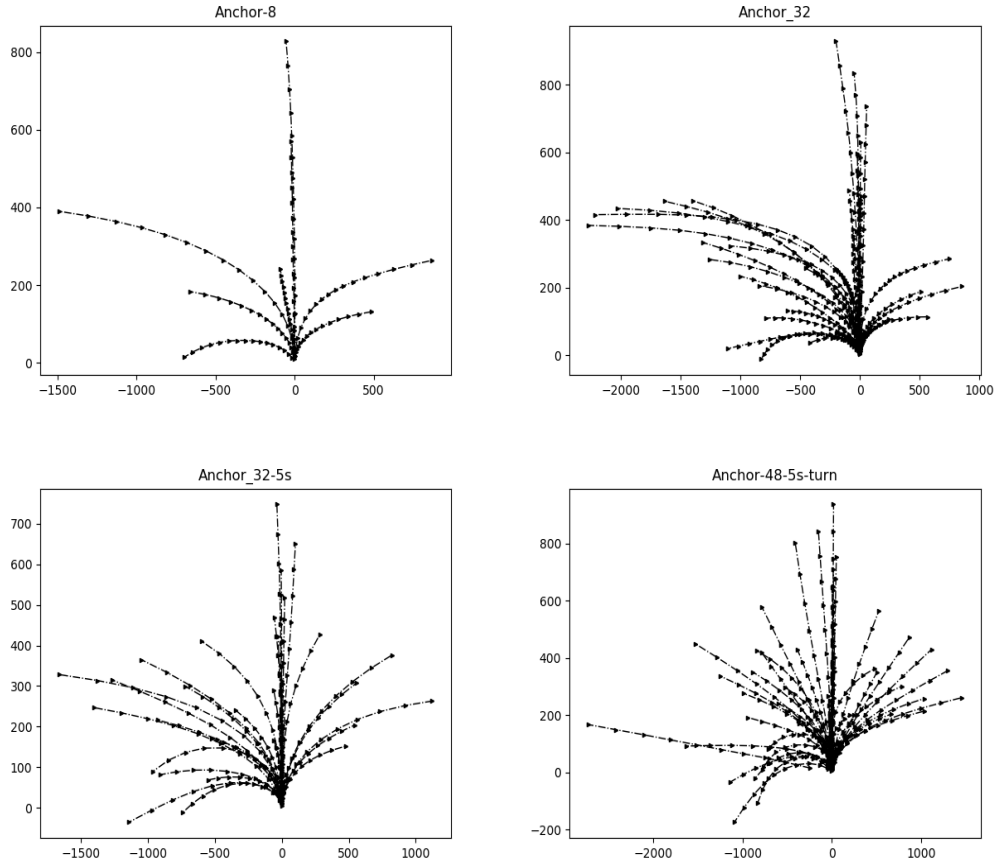
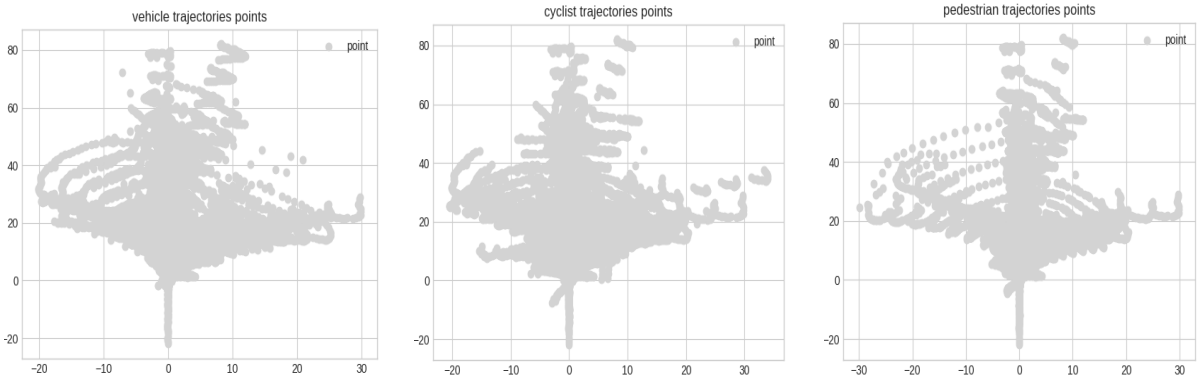


图 5.4　机动车 Anchor 不同取值的可视化图

Fig. 5.4　Visualization of different values of vehicle anchor



（a）机动车轨迹点　　（b）非机动车轨迹点　　（c）行人轨迹点

图 5.5　机动车、非机动车和行人轨迹点图

Fig. 5.5　Point chart of vehicle, cyclist and pedestrian trajectories

减速和匀速状态。如上文所述，本文分别对机动车、非机动车和行人的轨迹进行 Anchor 的可视化分析，可视化图在 5.5 中，结合图 5.4 和图 5.5 可以直观的看到随着 Anchor 聚类类别的增加，Anchor 所能代表的模式就会越具有代表性，最终应用到本文提出的机动车和非机动车模型中，采用 Anchor 类别为 48 时，作为模型预测时最终的 Anchor 划分依据。

（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 5.3 中给出了多任务概率化轨迹预测模型机动车分支在本文推出内部评价数据集上的可视化分析，其中图 5.6-（a）、（b）是为同一个交通路口场景，图 5.6-（c）、（d）则是另一个路口场景，图 5.6-（e）、（f）为另外的交通路口场景。在本节中，在进行模型可视化分析时，从多任务模型机动车分支预测的 48 条 Anchor 轨迹中，挑选了 top3 的轨迹进行可视化分析，在图中，红色的轨迹线为 Anchor 标签轨迹线，蓝色的轨迹线则为模型预测输出的轨迹预测线，绿色的轨迹线为 GT 轨迹，在蓝色的预测轨迹线上分别在 3s 点和 5s 点，将该预测点的双变量高斯分布绘制出来，其中双变量高斯由预测所得到的四个参数（$\mu_x^k + a_x^k, \mu_y^k + a_y^k, \sigma_x^k, \sigma_y^k, \rho^k$）表示的分布椭圆的大小，代表着预测轨迹点的不确定性区间的大小，不确定性区间越小，表明该点的方差越小，说明该点的预测置信度越高。分析图 5.6 中几个机动车轨迹的预测效果可得，图 5.6（a）中对于该机动车的右转预测较为准确，top3 的三条 anchor 轨迹都是右转轨迹，且对于每条轨迹的 3s 点和 5s 点，轨迹都能够准确的预测到 GT 轨迹的车道，能够达到很好右转预测效果。图 5.6（b）中，可以清楚的发现在进行轨迹预测时该机动车处于进入左转向待转区的状态，明确该车辆要进行左转，但是由于对向车道是开启的，复合栅格地图中的红绿灯线是没有的，即栅格地图所表示的拓扑信息是该向车道无法通行，所以经过本文所提出的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，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%，在实际应用中影响较小。图 5.6（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，该场景属于车辆与行人进行交互的交通场景，经过模型预测的轨迹线可得，机动车在与行人交互时倾向于避让行人，并有轻微右转趋势，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 5.6（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。图 d 的预测线位于车道 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，表明模型在预测时具有很强的鲁棒性。图 5.6-（e）、（f）两图对比发现，两个机动车的轨迹都是直线，但是不同的是，3s 点和 5s 点的置信

椭圆一个大，一个很小。再进一步分析可得，造成这种现象原因的是，图 f 所处的交通场景是复杂的交通场景，周围的机动车数量较图 e 要多的多，所以在图 f 中的预测线比图 e 中预测线的置信度要低，表现在可视化图中，就是 3s 点和 5s 点的轨迹置信椭圆的大小。

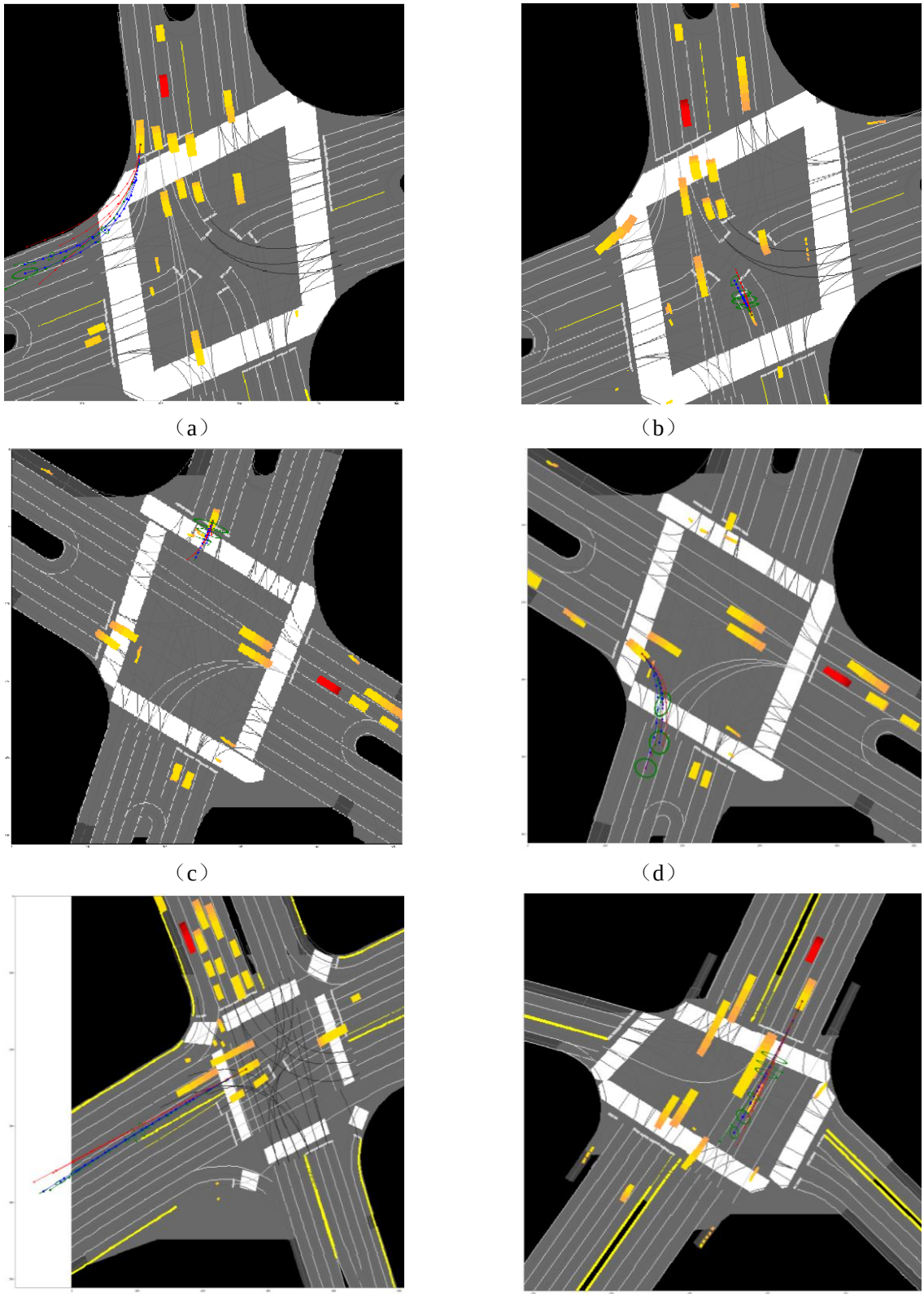
经过上述对多任务模型机动车分支的轨迹预测线的可视化分析，证明了模型输入以及模型网络结构的有效性，模型预测效果能够达到很好的精度和良好的社会可接受性。

（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，本文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 5.7 所示，图 5.7 的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。对比图 5.7-（a），（b）两图发现，单一模型趋向于对一侧集中进行预测，而多任务模型的预测 Anchor 却可以预测两个模式的 Anchor，这一点更加符合模型可以进行多模态预测的特性。对比图 5.7-（c），（d）两图，（c）图中的智能体在前方车辆已经停止的情况中，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，因为机动车在道路行驶过程中，发现前面是红灯，且前车已经减速或者停止，机动车是不会采取上述行为的，机动车大概率会采取停车或减速的措施。反之对于（d）图则多任务轨迹预测模型则能够很好的预测出，机动车下一步的动作，即该车辆会向前移动并逐渐靠近前方车辆进而停止前进，而这更符合驾驶员的驾驶习惯，更具有社会可接受性。对比图 5.7-（e），（f）两图，可以直观的分析出，对于转弯场景，单一任务激动车预测模型预测的更为发散，模型预测线分布于该向的所有车道上，而多任务模型则倾向于预测与 GT 轨迹相近或相邻的车道位置。

经过上述分析，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，即无论是在模型预测精度，还是在预测结果的社会可接受性上，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。这也体现了多任务模型在处理机动车分支上的优异性能。



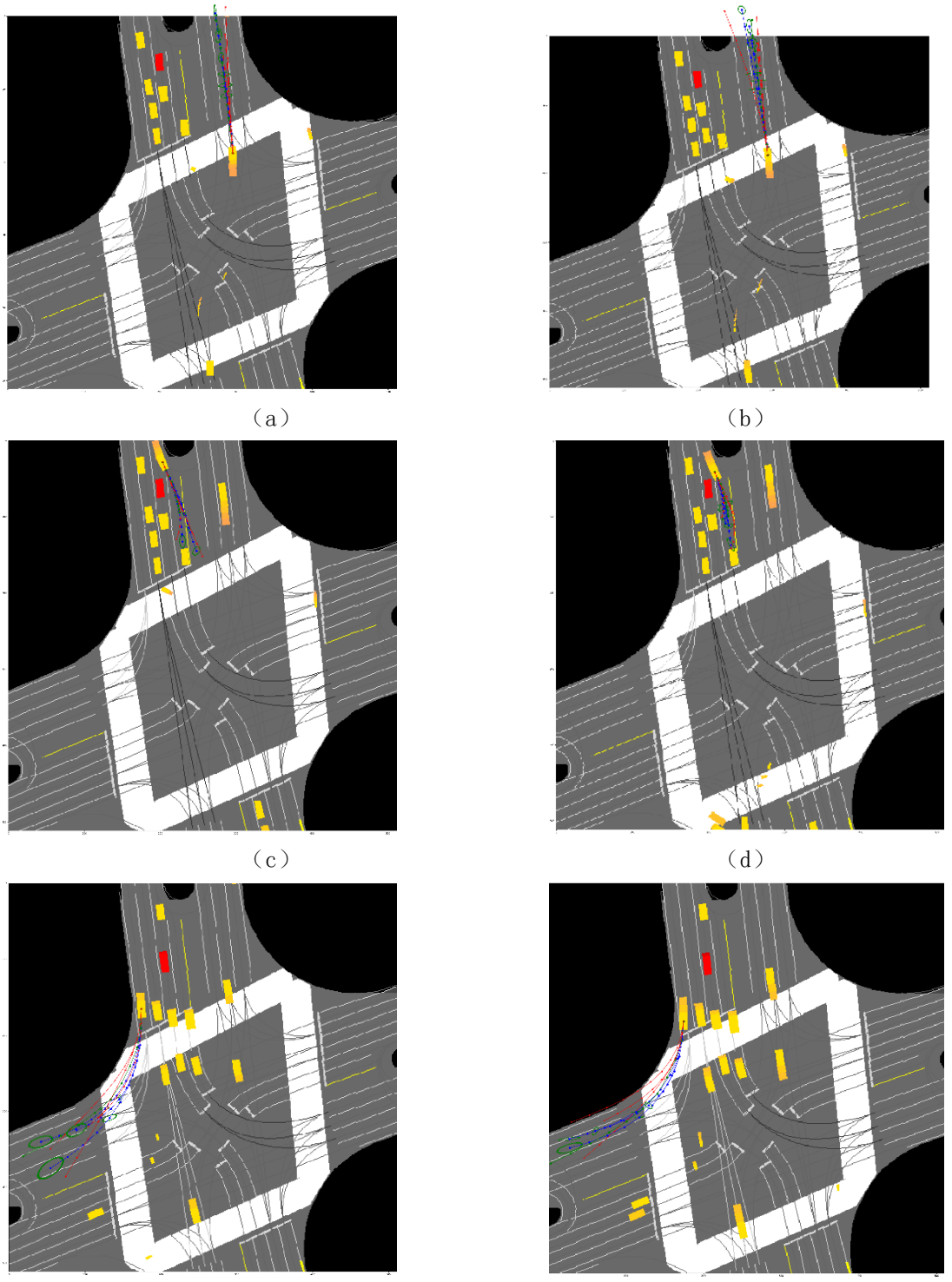
（a）　（b）

（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



（a）　（b）

（c）　（d）

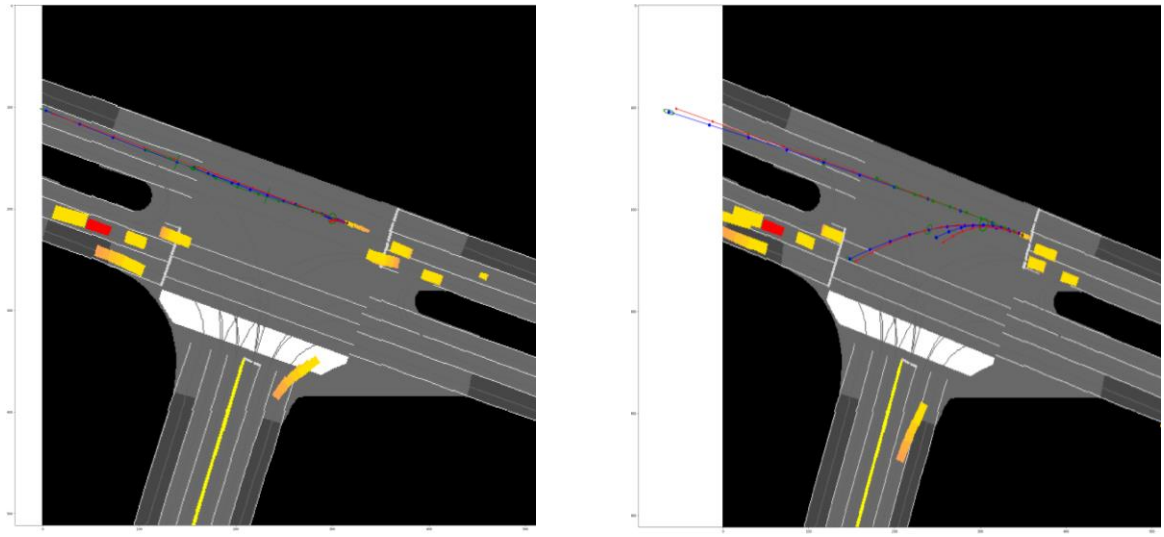
（e）　（f）

图 5.7　多任务模型机动车分支与机动车模型预测可视化对比图

Fig. 5.7　Visual comparison chart of multi task model vehicle branch and vehicle model prediction

（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 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 5.8　多任务模型非机动车分支预测可视化图

Fig. 5.8　Visualization of cyclist 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，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。

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。

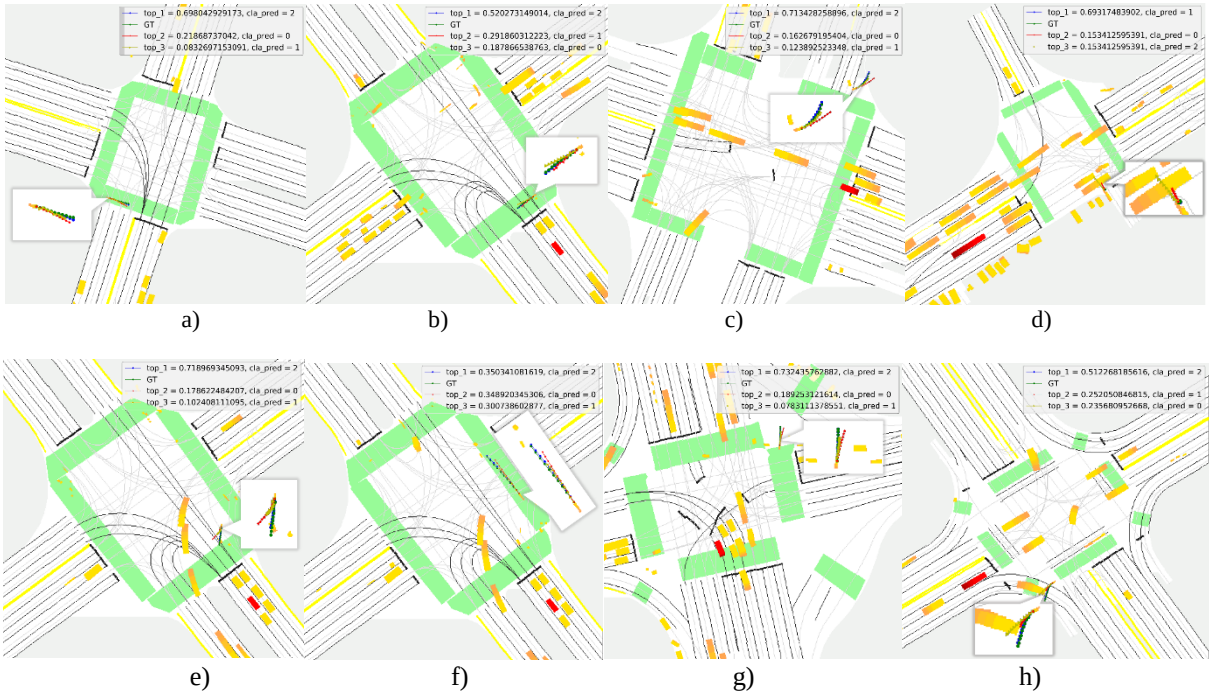
Social-STGCNN，一种使用图网络建模和时间卷积预测的轨迹预测模型。

将本文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，Gupta 等人提出生成预测 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，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，可以看到本文所提出的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。与本文的工作类似，Social-BiGAT 也使用语义场景信息和序列信息作为输入。但是本文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。这是因为多任务模型通过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，模型可以有效地学习地图的拓扑结构关系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，模型能够很好地建模语义特征。图 5.9 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，将展示多任务模型行人分支在处理不同场景、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 5.9-(a, b, c)所示，模型预测了不同路口轨迹的可靠性。模型预测了每个行人的 3 条轨迹，包括 top1, top2, top3,

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-model 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
Our Model (K=1)	1s	0.22	0.24	0.99
	3s	0.45	0.78	0.88
	5s	0.75	1.48	0.68
Our Model (K=3)	1s	**0.17**	**0.17**	**0.99**
	3s	**0.32**	**0.49**	**0.96**
	5s	**0.52**	**0.94**	**0.85**



a)　b)　c)　d)

e)　f)　g)　h)

图 5.9　多任务模型行人分支预测可视化图

Fig. 5.9　Visualization of Pedestrian Branch Prediction by Multi-task Model

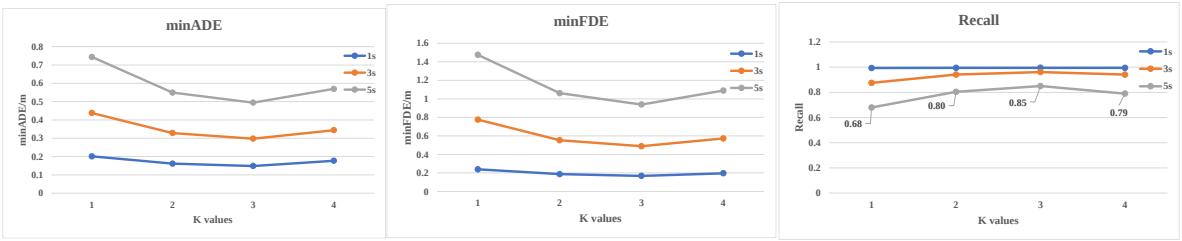
top1 的轨迹是概率值最高的预测轨迹，与 Ground Truth (GT)最为一致。在图 5.9-(d)中，行人处于交通密集的 t 形路口，模型预测的 top1 轨迹以 69%的概率与 GT 一致，此时行人状态处于等待状态，这种情况在自动驾驶任务中尤其常见。对于自动驾驶汽车来说，

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 5.9-(e, f)可以看出，本文所提出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 5.9-(f)中要求行人在人行道上行走，并快速穿过人行道。在最后一个图中，可以看到行人模型可以很好地处理与车辆的交互。

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 5.10　不同 K 值的模型指标的比较分析的可视化图

Fig. 5.10　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，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 5.10 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，本文所提出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，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，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提高轨迹预测精度和模型预测效率，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

本文针对复杂道路交通场景中智能体的行为预测的问题，提出了一种新型的轨迹预测模型，可以实现场景中所有类型智能体的实时预测，并相较于当前主流的轨迹预测模型获得了最优的精度和性能。主要工作成果：

（1）针对当前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的问题，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，可以实现场景中环境及交通智能体的栅格化表示，为后续模型的建模提供丰富的语义拓扑信息和历史序列信息。

（2）对机动车、非机动车和行人的轨迹预测问题进行建模。将智能体的不确定性分解建模为意图不确定性和控制不确定性，并根据交通场景中智能体的特点，多模式预测对应智能体意图不确定性，概率预测智能体的控制不确定性。

（3）针对机动车和非机动车，本文以 Encoder-Interaction-Decoder 为网络架构，提出基于 State-Anchor 的多模概率化轨迹预测模型，设计了机动车和非机动车模型损失函数，模型通过 MapNET 和 AgentNET 提取场景中的语义拓扑信息和历史序列信息，利用空间变换网络和注意力机制设计 Interaction-AttNET 模块以融合交互提取到的信息，最后使用 PredictionNET 对多模轨迹及其置信度进行预测；对行人，本文提出基于 Anchor-Free 的多模概率化轨迹预测模型，设计行人模型损失函数，搭建行人预测模型网络结构。

（4）在对机动车，非机动车和行人模型进行分别模型搭建后，进行多任务融合训练，设计多任务损失函数，搭建复杂道路交通场景的多任务多模概率化轨迹预测模型。

（5）本文将多任务模型，单一任务模型分别与当前业内主流的轨迹预测模型进行了定量和定性比较，并进行相应模型的可视化分析。模型实验结果表明，本文提出的模型不仅能对交通场景中所有的智能体进行轨迹预测，而且模型性能优于单一任务模型和当前主流的轨迹预测模型。

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领域的一大趋势，随着技术的更新和进步以及算法性能的提高，相信轨迹预测技术走进实际生活的距离不会很远。

本文通过上述几章的一系列模型的设计和研究，对于复杂道路交通场景的轨迹预测问题的预测结果已经能够达到很好的效果，并且在模型预测精度和轨迹的可接受性方面达到了非常优异的性能。但部署在移动端（尤其是，对实时性和算力要求很高的自动驾驶平台中），需要更进一步提高模型效率和实现模型压缩，这仍然是当前实际工作中重要的一环。因此未来的研究工作中还可从下面几个方面展开：

（1）首先模型对于高清地图中丰富的拓扑关系的提取，仍然需要依赖深度卷积网络（MobileNet-V2）提取，且目前的模型对计算、推理得到的交互信息缺乏可解释性。智能体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。如何通过对交互进行可解释性建模以及采用更为先进的地图拓扑关系表达来提升模型预测的准确性和可解释性对于轨迹预测来说将是研究热点之一。

（2）其次是当前大部分智能体的轨迹预测数据集为俯瞰视角，该数据集极度依赖于高清地图信息，对于平台的计算能力和算力都有很大的要求，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，对于处理没有实时高精地图的场景仍然是必要的[72]。

（3）实际场景中对轨迹预测算法的效率及识别精度有较高的要求，通过设计嵌入式或移动设备上高效且轻量化的网络结构、使用 TensorRT 框架进行推理提速以及进行相应网络结构的模型压缩或者障碍物筛选，可以在保证高精度的基础上，进一步提高模型预测的效率，这也将是未来领域内的一大研究热点。

参 考 文 献

[1] Waymo company, “On the road to fully self-driving,” https://news.ycombinator.com/from? Siteo-oglea apis.com.

[2] HARARI Y N, Reboot for the ai revolution[J]. Nature News, vol. 550, no. 7676, p. 324, 2017.

[3] LINDGREN A and CHEN F, State of the art analysis: An overview of advanced driver assistance systems (adas) and possible human factors issues[J]. Human factors and economics aspects on safety, pp. 38–50, 2006.

[4] POMERLEAU D A, ALVINN: An autonomous land vehicle in a neural network[J]. in Advances in neural information processing systems, 1989, pp. 305–313

[5] MONTEMERLO M, BECKER J, et al., Junior:The stanford entry in the urban challenge[J] Journal of field Robotics, vol. 25, no. 9, pp. 569–597, 2008.

[6] URMSON C et al., Self-driving cars and the urban challenge[C]. IEEE Intelligent Systems, vol. 23, no. 2, 2008.

[7] RUDENKO A, PALMIERI L, Herman M, et al. Human motion trajectory prediction: A Survey[J]. arXiv Preprint, 2019.

[8] DAI M, WANG J, YIN G, et al. Dynamic output-feedback robust control for vehicle path tracking considering different human drivers' characteristics[C]// The 36th Chinese Control Conference. Piscataway: IEEE Press, 2017: 9407-9412.

[9] LEFEVRE S, VASQUEZ D, LAUGIER C. A survey on motion prediction and risk assessment for intelligent vehicles[J]. ROBOMECH Journal, 2014, 1(1): 1-14.

[10] KELLER C G, GAVRILA D M. Will the pedestrian cross? A study on pedestrian path prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2): 494-506.

[11] SCHNEIDER N, GAVRILA D M. Pedestrian path prediction with recursive Bayesian filters: a comparative study[C]// The 35th German Conference on Pattern Recognition. Berlin: Springer Press, 2013: 174-183.

[12] PAVLOVIC V, REHG J M, MACCORMICK J. Learning switching linear models of human motion[C]// Conference and Workshop on Neural Information Processing Systems. New York: Curran Associates Press, 2000: 981-987.

[13] FOX E B, SUDDERTH E B, JORDAN M, et al. Bayesian nonparametric inference of switching dynamic linear models[J]. IEEE Transactions on Signal Processing, 2011, 59(4): 1569-1585.

[14] KOOIJ J F P, SCHNEIDER N, FLOHR F, et al. Context-based pedestrian path prediction[C]// The 13th European Conference on Computer Vision. Berlin: Springer Press, 2014: 618-633.

[15] HELBING D, MOLNAR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.

[16] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv Preprint, 2014.

[17] ALAHI A, RAMANATHAN V, LI F. Socially-aware large-scale crowd forecasting[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 2211-2218.

[18] YI S, LI H, WANG X. Understanding pedestrian behaviors from stationary crowd groups[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3488-3496.
[19] GOLI S A, FAR B H, FAPOJUWO A. Vehicle trajectory prediction with Gaussian process regression in connected vehicle environment[C]// 2018 IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2018: 550-555.
[20] ELLIS D, SOMMERLADE E, REID I. Modelling pedestrian trajectory patterns with Gaussian processes[C]// 2009 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2009: 1229-1234.
[21] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning [M]. Cambridge: MIT Press, 2005.
[22] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
[23] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv Preprint, 2014.
[24] GRAVES A, JAITLY N. Towards end-to-end speech recognition with recurrent neural networks[C]// The 31st International Conference on Machine Learning. New York: ACM Press, 2014: 1764-1772.
[25] CHOROWSKI J, BAHDANAU D, CHO K, et al. End-to-end continuous speech recognition using attention-based Recurrent NN: First results[J]. arXiv preprint, 2014.
[26] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 2625-2634.
[27] WU H, CHEN Z, SUN W, et al. Modeling trajectories with recurrent neural networks[C]// The 26th International Joint Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2017: 3083-3090.
[28] KARATZOLOU A, JABLONSKI A, BEIGL M. A Seq2Seq learning approach for modeling semantic trajectories and predicting the next location[C]// The 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2018: 528-531.
[29] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: Human trajectory prediction in crowded spaces[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 961-971.
[30] KITANI K M, ZIEBART B D, BAGNELL J A, et al. Activity forecasting[C]// The 12th European Conference on Computer Vision. Berlin: Springer Press, 2012: 201-214.
[31] KITANI K, ZIEBART B D, BAGNELL J A, et al. Activity forecasting[C]// The 12th European Conference on Computer Vision. Berlin: Springer Press, 2012: 201-214.
[32] XUE H, HUYNH D Q, REYNOLDS M. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction[C]// 2018 IEEE Workshop on Applications of Computer Vision. Piscataway: IEEE Press, 2018: 1186-1194.

[33] MERCAT J, GILLES T, ZOGHBY NE, et al. Multi-Modal Simultaneous Forecasting of Vehicle Position Sequences using Social Attention[J]. arXiv preprint, 2019.
[34] PARK S, LEE G, SEO J, et al. Diverse and Admissible Trajectory Forecasting Through Multimodal Context Understanding[J]. arXiv preprint, 2020.
[35] MARCHETTI F, BECATTINI F, SEIDENARI L, et al. MANTRA: Memory Augmented Networks for Multiple Trajectory Prediction[J]. arXiv preprint, 2020.
[36] DEO N, TRIVEDI M. Multi-Modal Trajectory Prediction of Surrounding Vehicles With maneuver based LSTMs[J]. arXiv preprint, 2018.
[37] SRIKANTH S, ANSARI J, RAM R.K, et al. INFER: INtermediate representations for Future Prediction[J]. arXiv preprint, 2019.
[38] LEE N, CHOI W, VERNAZA P, et al. DESIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents[J]. arXiv preprint, 2017.
[39] ZHAO T, XU Y, MONFORT M, et al. Multi-Agent Tensor Fusion for Contextual Trajectory Prediction[J]. arXiv preprint, 2019.
[40] GOODFELLOW I, POUGETABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Conference and Workshop on Neural Information Processing Systems. New York: Curran Associates Press, 2014: 2672-2680.
[41] KIPF T, WELLING M. Variational graph auto-encoders[J]. arXiv preprint, 2016.
[42] GUPTA A, JOHNSON J, LI F, et al. Social GAN: Socially acceptable trajectories with generative adversarial networks[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2255-2264.
[43] AMIRIAN J, HAYET J, PETTRE J. Social Ways: Learning multi-modal distributions of pedestrian trajectories with GANs[J]. arXiv preprint, 2019.
[44] VARSHNEYA D, SRINIVASARAGHAVAN G. Human trajectory prediction using spatially aware deep attention models[J]. arXiv preprint, 2017.
[45] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints[C]// 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 1349-1358.
[46] KOSARAJU V, SADEGHIAN A, MARTINMARTIN R, et al. Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks[C]// Conference and Workshop on Neural Information Processing Systems. New York: Curran Associates Press, 2019: 137-146.
[47] CHENG H, Yang W L M Y, SESTER M, et al. Context conditional variational autoencoder for predicting multi-path trajectories in mixed traffic[J]. arXiv preprint, 2020.
[48] YANG B, YAN G, WANG P, et al. TPPO: A novel trajectory predictor with pseudo oracle[J]. arXiv preprint, 2020.
[49] LIANG J, JIANG L, NIEBLES J C, et al. Peeking Into the Future: Predicting future person activities and locations in videos[C]// 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5725-5734.
[50] SUN J, JIANG Q, LU C. Recursive social behavior graph for trajectory prediction[J]. arXiv preprint, 2004.

[51] HUANG X, MCGILL S, DECASTRO J, et al. Diversity GAN: Diversity-Aware Vehicle Motion Prediction via Latent Semantic Sampling[J]. arXiv preprint, 2020.
[52] LI G, MULLER M, THABET A, et al. DeepGCNs: Can GCNs go as deep as CNNs?[C]// 2019 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9267-9276.
[53] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, 2016.
[54] VERT J, TSUDA K, SCHOLKOPF B, et al. Kernel methods in computational biology [M]. Cambridge: MIT Press, 2004: 35–70.
[55] SCHLICHTKRULL M S, KIPF T, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// European Semantic Web Conference. Berlin: Springer Press, 2018: 593-607.
[56] BERG R V D, KIPF T, WELLING M. Graph convolutional matrix completion[J]. arXiv preprint, 2017.
[57] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// European Semantic Web Conference. Berlin: Springer Press, 2018: 593-607.
[58] ZHANG L, SHE Q, GUO P. Stochastic trajectory prediction with social graph network[J]. arXiv preprint, 2019.
[59] YAN S, XIONG Y, LIN D, et al. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// National Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2018: 7444-7452.
[60] HADDAD S, WU M, WEI H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[J]. arXiv preprint, 2019.
[61] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction[J]. arXiv preprint, 2020.
[62] KHANDELWAL S, QI W, SINGH J, et al. What-If Motion Prediction for Autonomous Driving[J]. arXiv preprint, 2020.
[63] LIANG M, YANG B, HU R, et al. Learning Lane Graph Representations for Motion Forecasting[J]. arXiv preprint, 2020.
[64] GAO J, SUN C, ZHAO H, et al. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation[J].arXiv preprint, 2020.
[65] PELLEGRINI S, ESS A, GOOL V L. Improving data association by joint modeling of pedestrian trajectories and groupings[C]// The 11th European Conference on Computer Vision. Berlin: Springer Press, 2010: 452-465.
[66] 王欣然. 基于人-车交互的行人轨迹预测方法[D]. 辽宁:大连理工大学,2020.
[67] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J]. Computer Graphics Forum, 2007, 26(3): 655-664.
[68] ROBICQUET A, SADEGHIAN A, ALAHI A, et al, Learning Social Etiquette: Human trajectory prediction in crowded scenes[C]// The 14th European Conference on Computer Vision. Berlin: Springer Press, 2016: 549-565.

[69] HOWARD, ANDREW G., et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861 (2017).
[70] 李琳辉,周彬,连静,周雅夫.基于社会注意力机制的行人轨迹预测方法研究[J].通信学报,2020,41(06):175-183.
[71] MOUSSAID M, PEROZO N, GARNIER S, et al. The walking behaviour of pedestrian social groups and its impact on crowd dynamics[J]. PLoS ONE, 2010, 5(4):e10047.
[72] 李琳辉,周彬,连静,任威威.行人轨迹预测方法综述[J].智能科学与技术学报,2020,2020-061.

攻读硕士学位期间发表学术论文情况

1　李琳辉,周彬,连静,周雅夫.基于社会注意力机制的行人轨迹预测方法研究[J].通信学报,2020,41(06):175-183.核心期刊。（已录用）

2　Linhui Li, Bin Zhou, Jing Lian, et al. Multi-PPTP: Multiple Probabilistic Pedestrian Trajectory Prediction in the complex junction scene [J]. IEEE Transactions on Intelligent Transportation Systems (ITS). 2020. （Under review 本硕士论文第四章）

3　李琳辉,周彬,连静,任威威.行人轨迹预测方法综述[J].智能科学与技术学报,2020,2020-061.（已录用，本硕士论文第一章）

致　　谢

行文至此，落笔为终，三年的研究生生活打马而过，十数年的求学生涯也要画一个句号。时光荏苒，如白驹过隙，时间就是这么让人猝不及防的东西。目之所及，皆是回忆；心之所向，皆是过往。一路走来也曾迷茫过、失落过，但也曾被信任、被托付、被坚定，纵有万般不舍，但仍心存感激。

桃李不言，下自成蹊。首先，非常有幸能够被李琳辉老师和连静老师指导，得遇恩师，何其有幸。感谢老师对我的宽容、信任和教导，不仅提供自由，宽容和积极向上的科研氛围，让我能够自由，快乐的进行科研工作，更教导我许多为人处事的道理，让我可以受益终生。对此我都深深的表示感谢。

父字开头，母字结尾。借此机会感谢含辛茹苦养育我的父母周先生、张女士和我的奶奶刘女士，在我 23 年的清浅岁月里，十数年的求学之路中，感谢他们对我无条件的支持，无论是经济上还是精神上，护我周全，给我依靠，让我成长为一个独立、健康、健全的人。养育之恩，无以为报，唯有永无止境的探索和奋斗，成为他们的骄傲，祝愿父母身体健康、和睦美满，奶奶天国安好。

凡事过往，皆为序章。何德何能，所遇之人皆不偏不倚，传道授业，亦师亦友，感慨万千，无处安放。感谢杨曰凯师兄和王欣然师姐，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，教我如何科研，带我走入预测的大门。感谢吕星晨和张永康，总是在我最难的时候给予我最为中肯有效的建议。感谢挚友张琳、庄璐洁、张娇，感谢你们包容我，倾听我，陪伴我，鼓励我。感谢跆拳道让我遇到了他们。感谢我的舍友、同教研室的同学、研究生同学，感谢你们宽容我的脾气，陪伴着我成长。最后感谢一路走来，遇到的所有人所有事，好的坏的，谢谢你们成就了现在的我。

山水一程，三生有幸。初见乍惊欢，久处亦怦然，感谢我的女朋友王越陪我度过研究生最后的时光，包容我，爱护我，收拾我凌乱的心情。感谢相遇，感谢相识，感谢相爱，愿未来一切顺遂，喜乐有分享，共度日月长。

以梦为马，不负韶华，感谢自己，感谢自己一直以来有始有终的坚持和努力，让我可以遇到更好的自己。

祝万事胜意，平安喜乐，所得皆所盼。

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：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

导 师 签 名 ：_____ 日期：_____年____月日