**轨迹预测方法综述**

**摘 要：**随着深度学习技术的突破和大型数据集的提出，智能体轨迹预测的准确度已经成为人工智能领域的研究热点之一。伴随着深度学习在轨迹预测领域的广泛应用，其预测效果得到了显著的提升，本文主要对交通场景中不同智能体轨迹预测的技术分类和研究现状进行详细综述。根据模型建模方式的不同，将现有方法分为基于浅层学习的轨迹预测算法和基于深度学习的轨迹预测算法，对每类方法中具有代表性的算法效果及优缺点进行了分析介绍，归纳了当前主流的轨迹预测公开数据集，并在数据集中对比了主流轨迹预测方法的性能，最后对轨迹预测技术面临的挑战与发展趋势进行了展望。

**关键词：**轨迹预测；深度学习；序列决策

**1 引言**

人类环境中出现越来越多的智能自主系统，这些系统感知、理解和预测智能体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。它涉及到表征、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：预见涉及多个智能体的场景将随时间的推移而开展，从而以主动的方式整合这些知识，即允许增强主动感知、预测性规划、模型预测性控制或人机交互。因此, 近年来智能体运动预测在多个领域中受到越来越多的关注，存在许多重要的应用领域，例如自动驾驶汽车、服务机器人、智能交通、智慧城市等诸多高级监视系统领域。

保证交通场景中道路使用者的自身安全是自动驾驶车辆普及应用的前提条件[2]。其中行人、机动车、非机动车等智能体作为交通场景中的主要参与者，其未来运动轨迹的合理推理和预测对自动驾驶和道路安全至关重要。轨迹预测的难点在于: 首先,各智能体运动特性复杂多样，精准预测未来的运动轨迹几乎是一个不可能完成的任务，但是通过观察其历史时刻的运动轨迹和所处的环境，根据一些算法可以大致推测出智能体未来的运动轨迹（例如最简单的是匀速直线运动）。在实际道路场景中，自行车、汽车等智能体建立运动学模型相对行人而言较为容易，而行人运动更加灵活多变，很难对行人建立合理的动力学模型（因为行人可以随时转弯、停止和运动等），有时甚至连人类驾驶员都很难预测出行人未来的运动，其次，在实际交通场景中，某特定智能体未来的运动不仅受自己意图的支配，同样也受周围其他智能体及其周围所处环境的影响（例如避障，交互等）, 根据 Moussaid等[3]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，类似行人这种智能体间的交互非常抽象，在算法中往往很难精确地建模出来。最后，常规算法模型的实现是找到一个输入到输出的函数映射，对于轨迹预测的模型来说，其对应着一个序列到序列的映射，常规模型或训练方式容易使模型预测结果陷入平均状态（预测结果倾向于预测一种折中的轨迹），显然使用常规的训练模型无法很好地对类似智能体轨迹进行有效地预测。最后，机动车以及非机动车需要在符合交通规则的情况下行驶，其对于交通规则和道路状况的敏感程度相较于行人更为敏感，因此相较于行人的高随机性和高动态性，机动车以及非机动车又具有一定的确定性。因此寻求一个在智能交通场景中对所有交通参与者进行精准的轨迹预测是困难的。

交通环境中智能体轨迹高动态性、随机性以及与交通环境智能体之间复杂的交互使得轨迹预测问题充满挑战，但是对智能体轨迹进行长时间的预测仍有必要，这对实现自动驾驶汽车的主动规划和决策[4]影响巨大。目前国内外有关轨迹预测的研究与日俱增，对该领域的相关技术和文献进行综述十分必要。本文对轨迹预测的问题描述、研究挑战以及问题处理流程进行了梳理，对轨迹预测的研究进展，尤其是近几年的研究成果进行了综述。文中第二节给出了轨迹预测问题的定义和流程描述。第三节对现有的轨迹预测算法进行了分类，并在此基础上简要地总结了基于浅层学习及机器学习的方法，同时以发现问题为导向，解决问题为目标，对基于深度学习的不同智能体的轨迹预测方法分别进行了详细地分析介绍。第四节介绍了目前业内主流的轨迹预测公开数据集，将一些轨迹预测算法在数据集中进行测试后，对取得优异测试结果的方法进行分析和比较。第五节对轨迹预测技术在今后发展面临的挑战和发展趋势进行了总结和展望。

**2 轨迹预测的问题描述**

预测问题的本质是根据得到的智能体历史特征信息以及所处的环境信息，推断出智能体未来时刻可能出现的状态。而轨迹预测问题作为预测问题的子问题，可以被视为序列决策问题，即通过引入智能体的位置信息，自我运动历史信息，所处的环境信息等，根据智能体在过去时间段内的状态来预测未来固定时刻的状态（位置坐标，速度，行为等）。

抽象来说，通常我们将轨迹预测问题分为三个方面，如图1所示。首先是来自外部的输入激励信息，运动意图或外部环境直接（间接）地影响着智能体的运动行为过程，诸多预测方法很大程度上都依赖智能体运动的历史信息和运动学特征（位置，速度，角速度及其他属性），同时还存在一些其他形式的环境信息输入（例如道路的几何场景、树木等静态物理信息以及人-车辆、人-人等移动交互信息），而对于端到端的方法而言，实际上其主要依赖于原始的传感器数据序列。此外轨迹预测的复杂性还来自于不同的社会行为，因此通俗来讲在进行预测过程中一般还需要考虑智能体与环境（如障碍物）以及不同智能体和相同智能体之间的复杂交互。其次对于轨迹预测的建模方法来说，有基于运动学和动力学方法的参数化模型，有基于浅层学习和深度学习的方法，而基于学习的方法不一，本文重点对该部分类别方法进行归类，对模型优缺点和相似性进行对比分析，并对当前轨迹预测领域存在的问题与未来发展趋势进行彻底的剖析和展望。最后，采用不同的建模方法，就会使用不同的损失函数，产生不一样的预测结果，例如预测单条或多条轨迹，预测网格的概率分布，预测代理状态上的双变量高斯分布或者基于图模型的运动模式等。



图1　轨迹预测问题的组成元素

具体而言，典型的轨迹预测问题的步骤包括，前期数据集的获取，数据集的预处理和输入，特征的编码、提取和汇集，预测结果的输出和轨迹预测的可视化，如图2所示。现有的轨迹预测方法一般默认相应场景的数据集已经完成，由数据集直接提供轨迹信息或者场景信息。具体的常见公开数据集见下文第4节。



图2　轨迹预测问题的处理流程

**3 轨迹预测的技术分类**

20世纪90年代以来，大量基于浅层学习的轨迹预测模型率先被提出，这些方法对计算设备的算力要求比较大，且缺乏统一的评价标准，用于测试该模型的数据集也鱼龙混杂，质量参差不齐，因此基于浅层学习的轨迹预测模型相关整理及综述很少被研究者提出。近十几年来，随着机器学习技术（尤其是深度学习技术）的兴起，由于递归神经网络在处理时序数据上优异的性能，基于递归神经网络的时序预测模型层出不穷。下面本文将对国内外主要的轨迹预测算法进行分类综述。

根据预测模型的建模方式，本文大致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模块化，考虑和利用的信息也逐渐完善，预测的精度和实时性也逐渐提高。综上所述，轨迹预测方法分类如图3所示。



图3　轨迹预测方法分类

实际上，一个好的轨迹预测模型往往需要综合利用不同的信息（轨迹信息，语义信息，社交信息）等，因此很难从其诸多信息中的单一角度去区分预测模型的好坏。为了对现有的轨迹预测方法进行更加细致的梳理和归纳。接下来本文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进行详细地介绍。

**3.1 基于浅层学习的轨迹预测方法**

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

Helbing等[10]提出了一种具有吸引力和排斥力的运动模型，称为社会力模型。该模型现广泛应用于机器人和活动理解等[11]领域。Alahi等[12]通过从人群里的人类轨迹中学习他们的相对位置来呈现社会亲和力特征，而Yi等[13]提出使用人类属性来改善人群中的预测。

随着机器学习的快速发展，基于运动学的方法在做预测时有时会使用一些基于机器学习的跟踪算法来改进跟踪和预测，例如卡尔曼滤波（Kalman Filter, KF）、马尔可夫模型（Markov Model, MM）和高斯过程[14-16]（Gaussian Process, GP）等。通过结合KF模型，其优势在于处理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，但其对轨迹的波动较为敏感，且不适用于中长期的轨迹预测，一阶MM模型仅考虑了当前运动轨迹点对未来轨迹点的影响，历史轨迹点的数据信息无法被尽可能地利用，而高阶MM模型大大增加了模型计算的复杂度。GP模型通过假设隐变量服从高斯分布，为概率预测提供非参数模型，其中预测轨迹是从历史轨迹数据中学习的[14]，通过在GP中指定适当的核函数（协方差函数）来明确进行预测轨迹建模所涉及到的不确定性，其中*μ，θ*为协方差函数的超参数，参数需要通过数据训练得到，其能够较为有效地预测具有噪声点的轨迹数据，同时可以很好地避免轨迹数据离散性质的不足，并在此基础上有效地表达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轨迹的预测，并随着数据的增多，其计算量大为增加。而且在单独使用GP模型时，在不同时间点，其预测误差差别也比较大。

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

总而言之，基于浅层学习的轨迹预测算法在早期取得了一些成果，为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运动模型及场景难以进行较为精确的预测。

**3.2 基于深度学习的轨迹预测方法**

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

基于深度学习的轨迹预测方法有两个关键要素。一是网络结构的选择。研究者不同，其对轨迹预测问题和场景辅助信息的使用则有着不同的理解，选择的网络结构也可能不相同，那么对轨迹预测特征进行提取、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能否得到充分有效地利用。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，用于序列预测的递归神经网络（RNN）及其变体，包括长短期记忆网络（LSTM）[17]和门控递归神经网络（GRU）[18]在序列预测任务中（如语音识别[19,20]，标题生成[21]，机器翻译[11]）取得了巨大的成功。鉴于递归神经网络在处理序列数据尤其是解决长时依赖时的优越性能，自然而然的可以将其引入轨迹预测领域。而序列模型Seq2Seq为序列生成任务提供了一个编码器-解码器（Encoder-Decoder）框架，该框架旨在学习以输入序列为条件的输出序列（即）的条件分布，框架中可以将LSTM单元同时用于编码器和解码器中。编码器LSTM首先遍历输入序列，将其编码为语义向量，后将向量传递给解码器LSTM，随后生成输出序列，近几年的模型大多采用了该预测框架[22]。编码器-解码器（Encoder-Decoder）框架虽然较为经典，但是也存在一定的不足之处，即在编码和解码两者之间建立联系的唯一桥梁就是上述中提到的固定长度的语义向量，整个序列所携带的信息都将被压缩到此向量中，这就会出现两个问题，其一，输入序列的信息不能充分的被语义向量表示，其二，后输入的序列信息会将之前输入的信息覆盖掉。此现象会随着输入序列长度的增加愈发严重，最终造成解码准确度的下降。为解决此问题，注意力模型诞生，此模型不再要求编码器把整个序列所携带的信息都压缩到此向量中，而是将其将其输入信息编码成一个向量序列，解码时选择性的对其子集进行处理，保证了对序列所携带信息的充分利用。在此基础上，Karatzoglou等[23]通过对人类轨迹采用Seq2Seq框架扩展了LSTM网络，并探讨了基于注意力的Seq2Seq的影响，验证了Seq2Seq框架在轨迹建模和运动模式预测中的有效性。

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

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

**3.2.1**  **基于生成式网络架构的预测方法**

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

为解决先前模型中的平均轨迹的问题并对池化汇集模型进行改进，Gupta等[30]将基于生成式对抗网络的方法引入轨迹预测领域，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交互有用的其他特征信息，并且采用了传统的GAN架构，网络训练不稳定，容易崩溃。Amirian等[31]通过引入Info-GAN架构来改进GAN的架构，以此来解决 Social-GAN模型训练崩溃和掉落的问题，通过舍弃L2代价函数，引入基于互信息的 Information Loss损失函数，增强了模型对多条合理轨迹的预测能力，并引入注意力机制[32]使模型自主分配对交互信息的关注，该论文表明最新提出的Info-GAN架构可以极大地改善多模式轨迹预测，避免类似训练崩溃的问题。在此基础上 Sadeghian等[33]通过融合环境中场景的上下文信息以及智能体的历史轨迹，使用GAN的网络架构生成多条物理条件下的可接受轨迹。Kosaraju等[34]采用了基于Cycle-GAN的网络架构和训练方法，保证GAN在生成轨迹时对于噪声的敏感性，从而有助于生成多样性的轨迹，文中使用VGG网络提取场景图像特征，使用LSTM提取智能体轨迹特征，根据提取到的特征差异，分别使用基于 Scale-Dot和GAT的多种注意力机制，以得到对于各种输入最为合理的注意力向量，该模型可以理解较为复杂的智能体运动社会本质，不仅能够为特定的智能体生成多个轨迹，而且可以通过多模式方式同时为多个智能体预测更真实的运动轨迹。

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

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

总而言之，基于生成式网络架构的预测方法在进行智能体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。

**3.2.2**  **基于图网络的预测方法**

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

Gupta等[30]和Sadeghian等[33]利用具有交互机制的GAN或变分编码器来考虑场景中所有个体。但是这两种模型都无法学习人类行为的真正多模态分布，而是学习具有高方差的单一行为模式。此外，两种模型都受到他们学习社交行为的建模方式的限制，尽管前者通过对场景中的所有智能体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，可以将智能体之间的交互表达为图形[45]，其中节点是指智能体，而边缘代表智能体之间的互动；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的信息，空间边捕捉智能体交互的信息，输出采用双变量高斯分布，该方法能较好地对时空信息进行有效建模，但该方式计算较为复杂。Haddad等[47]基于LSTM神经网络提出新型的Spatio-Temporal Graph（时空图），旨在实现在拥挤的环境下，通过将智能体-智能体，行人-静态物品两类交互纳入考虑，对智能体的轨迹做出预测。

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

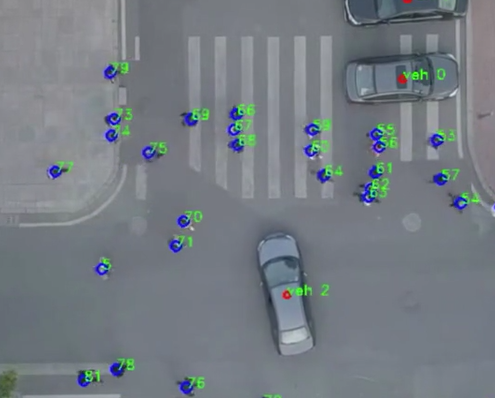
总而言之，随着深度学习的快速发展，基于深度学习的轨迹预测方法成为研究热点，理论上而言，基于深度学习的算法基本可以解决利用传统的浅层学习算法的轨迹预测问题，但当神经网络足够深，功能足够强大的时候，如果数据集数据量过小，就非常容易产生过拟合的问题，从而影响预测精度。数据集规模的增大，社会交互场景信息的丰富，数据集所涵盖的corner case 越来越完善，场景就会越来越接近于现实场景，并且随着数据量的不断增强以及神经网络结构及损失函数合理有效地设计，对于轨迹预测的效率及精度都会有很大的改善。近几年图卷积神经网络应用到轨迹预测问题，取得了亮眼的成绩，但是基于图网络算法的模型存在着模型建图效率低下，邻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

**4 数据集及性能比较**

**4.1 数据集介绍**

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，行人包含诸多复杂的行为，包括行人交互、非线性轨迹、避免碰撞、站立及群体行人的轨迹坐标等，同时包含从固定的俯视图记录的五个独特的室外环境信息。每个环境中单个场景的人群密度不同，所有视频的每秒帧数为25，行人轨迹以2.5fps的速度进行采样标记。其中ETH由ETH和Hotel两个数据集组成，UCY由Zara1、Zara2和Univ三个数据集组成。数据集包括从室外监控摄像头拍摄的五个视频，其中包含2 206条行人轨迹，表现出在直线运动和曲线运动样条之间变化的不同特征。随着视频捕捉到大学入口处人们的动作，ETH场景包含了更多笔直的轨迹，几乎没有社交互动，而UCY场景展示了更多与人空间互动有关的场景。例如，UCY-ZARA数据集中包括在商店入口处弯曲的行人轨迹，而UCY-UNIV则具有更多的社交互动。此外，除非考虑社会和空间环境，否则这些情况尤其会增加单个路径的不可预测性。目前绝大部分轨迹预测模型都以该数据集为训练集和测试集，并在此数据集上测试模型性能。

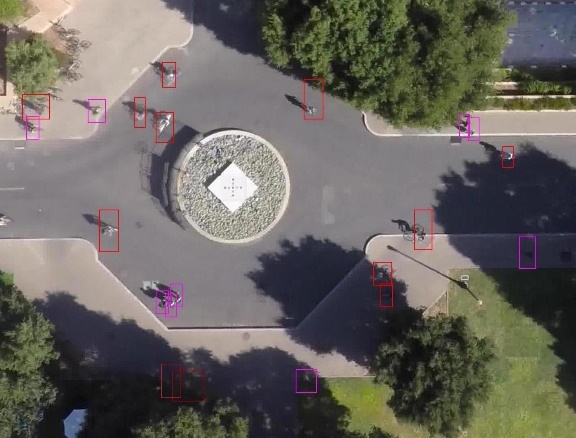
DUT数据集是在中国大连理工大学（DUT）主校区选择具有典型混杂交通特征的路段通过航拍采集的，旨在解决适用于人车交互运动的轨迹预测模型，如图4所示。图4（a）为无交通信号的交叉路口。由于没有交通信号灯，当行人与车辆交互时，可以体现人车交互的“社会性”。图4（b）为环岛，行人和车辆在较大的活动空间自由移动，而不受制于路面，可以得到多个方向上的预测结果。本课题组采用一架DJI-Mavic-Pro在行人和车辆很难察觉的高度拍摄，以达到不会因为无人机的出现影响行人运动的目的。视频分辨率为1 920×1 080，fps为25。数据集共有17个交叉路口场景片段和11个环岛场景片段，包含了1 793条轨迹。

（a）无交通信号交叉路口 （b）环岛

图4　DUT无人机数据集

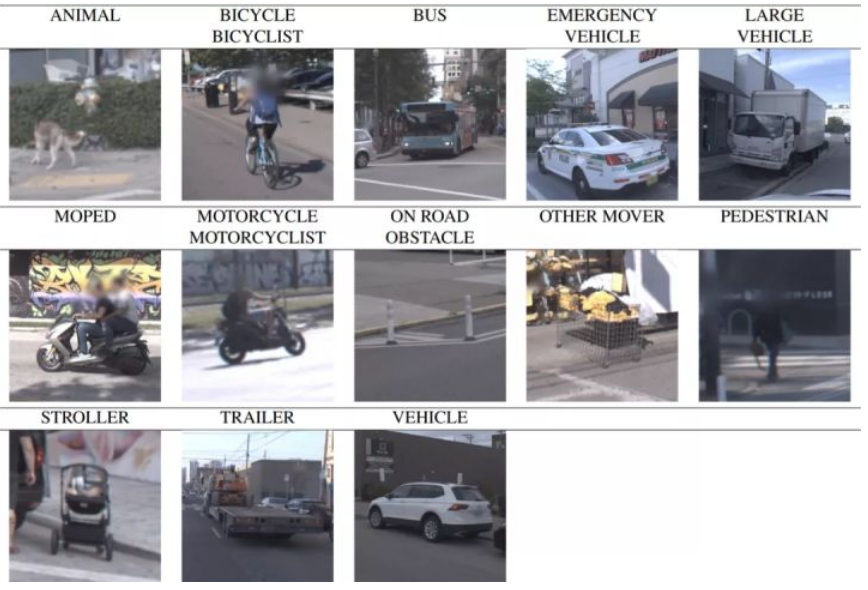
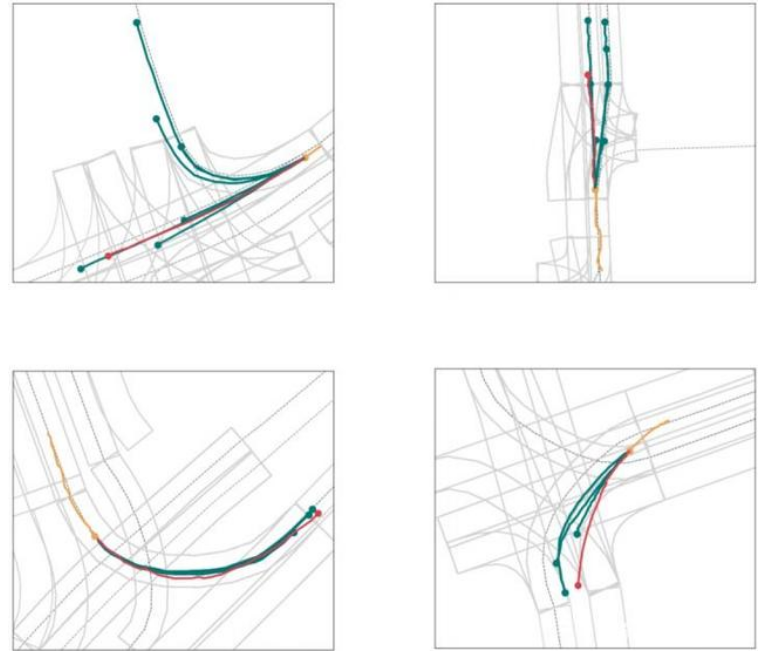
斯坦福无人机数据集（Stanford Drone Dataset）是由Robicquet等[51]提出的一个大型数据集，旨在解决目标跟踪或轨迹预测之类的任务，斯坦福无人机数据集是一个大型且先进的数据集，收集了各种类型的智能体（不仅是行人，还包括自行车，滑板，汽车，公共汽车和高尔夫球车）的图像和视频（如图5所示）在现实的户外环境（例如大学校园）中行驶。其中包含行人、骑自行车的人、滑板者、手推车、小汽车和在大学校园中行驶的公共汽车的视频以及轨迹信息。

（a）quad场景 （b）death-Circle场景

图5　斯坦福无人机数据集

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

（a）3D轨迹跟踪/预测分类 （b）运动预测

图6　Argoverse数据集

**4.2 指标评价**

轨迹预测问题作为预测问题的子问题，可以看作是一个序列生成问题，其中输入序列对应着观测到的所有行人的位置，之后生成一个输出序列，表示行人在未来不同时刻所处的位置。即假设在任意一个时刻，在场景中行人所处于的实际位置表示为,我们从时间到（表示观测序列长度）观察所有行人所处的位置序列，然后预测时间到（表示预测序列长度）所有行人所处的预测位置。那么在评估轨迹预测方法的性能时，通常主要使用平均位移误差（Average Displacement Error, ADE）以及最终位移误差（Final Displacement Error, FDE）两个评价指标对每个场景中智能体的轨迹进行预测评估。

（1）平均位移误差（ADE）

平均位移误差（ADE）是指每一个行人的真实轨迹与预测轨迹的位置序列在每一个时间步长内的平均欧式距离差值，用来评估预测过程的总体性能，如式（1）所示。

 （1）

（2）最终位移误差（FDE）

最终位移误差（FDE）是指每一个行人真实轨迹与预测轨迹位置序列在终点位置的平均欧式距离差值。如式（2）所示。

 （2）

**4.3 性能比较**

目前领域内最常用于性能测试的数据集是ETH和UCY数据集。虽然在后期有许多针对轨迹预测算法的大型数据集相继被提出，但是由于这些数据集出现的时间较为靠后，一些早期发布的轨迹预测算法在后期发布较晚的数据集上的验证缺乏充足的实验数据，难以相对准确地测试这些算法的效果，故本文只比较了基于ETH和UCY数据集的模型性能比较。表1列出了本文提到的一些轨迹预测方法在公开数据集上的性能比较结果。一些方法要用到场景或行人骨架等语义信息，故实验是在模型只使用轨迹信息时所得出的预测结果。

表1 各方法在ETH/UCY数据集中的性能表现

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 参考文献序号 | 评价指标(ADE/FDE) | | | | | |
| ETH | HOTEL | UNIV | ZARA1 | ZARA2 | AVG |
| Linear | [24] | 1.33/2.94 | 0.39/0.72 | 0.82/1.59 | 0.62/1.21 | 0.77/1.48 | 0.79/1.59 |
| S-LSTM | [24] | 1.09/2.35 | 0.79/1.76 | 0.67/1.40 | 0.47/1.00 | 0.56/1.17 | 0.72/1.54 |
| O-LSTM | [24] | 1.05/2.21 | 0.81/1.68 | 0.71/1.45 | 0.47/1.02 | 0.64/1.25 | 0.74/1.52 |
| S-RNN | [47] | 2.72/4.60 | 0.85/1.35 | 1.05/2.20 | 1.60/3.50 | 1.45/3.00 | 1.53/2.93 |
| Social-attention | [52] | 3.60/4.70 | 0.79/1.44 | 1.30/2.66 | 0.95/2.05 | 1.00/2.14 | 1.53/3.52 |
| S-GAN | [30] | 0.81/1.52 | 0.72/1.61 | 0.60/1.26 | 0.34/0.69 | 0.42/0.84 | 0.58/1.18 |
| S-GAN-P | [30] | 0.87/1.62 | 0.67/1.37 | 0.76/1.52 | 0.35/0.68 | 0.42/0.84 | 0.61/1.21 |
| Sophie | [33] | 0.70/1.43 | 0.76/1.67 | 0.54/1.24 | 0.30/0.63 | 0.38/0.78 | 0.54/1.15 |
| Social-ways | [31] | 0.39/0.64 | 0.39/0.66 | 0.55/1.31 | 0.44/0.64 | 0.51/0.92 | 0.46/0.82 |
| GAT | [34] | 0.68/1.29 | 0.68/1.40 | 0.57/1.29 | 0.29/0.60 | 0.37/0.75 | 0.52/1.07 |
| STGAT | [53] | 0.70/1.35 | 0.37/0.67 | 0.59/1.23 | 0.35/0.69 | 0.31/0.64 | 0.47/0.92 |
| Social-BiGAT | [34] | 0.69/1.29 | 0.49/1.01 | 0.55/1.32 | 0.30/0.62 | 0.36/0.75 | 0.48/1.00 |
| Next | [37] | 0.73/1.65 | 0.30/0.59 | 0.60/1.27 | 0.38/0.81 | 0.31/0.68 | 0.46/1.00 |
| TPNet-20 | [38] | 0.84/1.73 | 0.24/0.46 | 0.42/0.94 | 0.33/0.75 | 0.26/0.60 | 0.42/0.90 |
| Social-STGCNN | [48] | 0.64/1.11 | 0.79/0.85 | 0.44/0.79 | 0.34/0.53 | 0.30/0.48 | 0.44/0.75 |

从表1中可得，伴随着轨迹预测领域内相关研究人员不知疲倦地思考尝试、创新突破，各种轨迹预测方法层出不穷，其性能在主流公开数据集上的最佳效果不断被突破刷新。从表1中数据可知，基于深度学习的轨迹预测方法效果更佳，其识别精度远远高于基于传统的浅层学习方法，且在ETH,UCY两个主流公开数据集上的不同场景中性能表现都十分优异。随着基于生成式网络与图神经网络技术的发展以及与各种语义场景信息的结合，诸多轨迹预测方法在公开数据集上的测试性能愈发优异，其中以Social-Ways为代表的基于生成式对抗网络的学习方法性能显著，识别率最高，Social-Ways方法的ADE与FDE指标达到了0.46和0.82，在性能上较之前提出的方法有了大幅度的提高。而且可以发现基于图网络的轨迹预测方法，进展迅速，在S-RNN提出之时，其表现性能仅仅为基于LSTM方法的50%，但随着图模型的改进以及GCN和GAT模型的应用，基于该方法的预测准确率也获得了逐步的提升，尤其是今年提出的几个模型，其预测效果已经超过了基于生成式网络的预测方法，且较S-LSTM提高了50%以上。总而言之，通过合理地设计网络结构和损失函数，基于深度学习的方法可以相对精确地预测出行人接下来的轨迹，是可以较好地完成轨迹预测任务的基础框架。

**5 结束语**

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

·当前大部分智能体的轨迹预测数据集为俯瞰视角，该数据集距离轨迹预测的实际应用具有比较大的差距，由于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。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轨迹预测来说也是研究热点之一。

·目前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速度和角速度分布为条件定义了车辆舒适加速/减速率的规范等）。那么认为对于智能车基本紧急制动这一具体场景用例，其解决方案达到了工业化消费品的标准。对于其他具体实际场景用例，更加标准化及具体化的行业标准的提出对于加快领域内预测技术的发展也至关重要。

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

**参考文献：**

1. RUDENKO A, PALMIERI L, HERMAN M, et al. Human motion trajectory prediction: A Survey[J]. arXiv Preprint, 2019.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. HELBING D, MOLNAR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.
11. BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv Preprint, 2014.
12. 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.
13. 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.
14. 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.
15. 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.
16. RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning [M]. Cambridge: MIT Press, 2005.
17. HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
18. CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv Preprint, 2014.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. KIPF T, WELLING M. Variational graph auto-encoders[J]. arXiv preprint, 2016.
30. 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.
31. AMIRIAN J, HAYET J, PETTRE J. Social Ways: Learning multi-modal distributions of pedestrian trajectories with GANs[J]. arXiv preprint, 2019.
32. VARSHNEYA D, SRINIVASARAGHAVAN G. Human trajectory prediction using spatially aware deep attention models[J]. arXiv preprint, 2017.
33. 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.
34. 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.
35. 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.
36. YANG B, YAN G, WANG P, et al. TPPO: A novel trajectory predictor with pseudo oracle[J]. arXiv preprint, 2020.
37. 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.
38. SUN J, JIANG Q, LU C. Recursive social behavior graph for trajectory prediction[J]. arXiv preprint, 2004.
39. 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.
40. KIPF T N, WELLING M. Semi-supervised classiﬁcation with graph convolutional networks[J]. arXiv preprint, 2016.
41. VERT J, TSUDA K, SCHOLKOPF B, et al. Kernel methods in computational biology [M]. Cambridge: MIT Press, 2004: 35–70.
42. 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.
43. BERG R V D, KIPF T, WELLING M. Graph convolutional matrix completion[J]. arXiv preprint, 2017.
44. 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.
45. ZHANG L, SHE Q, GUO P. Stochastic trajectory prediction with social graph network[J]. arXiv preprint, 2019.
46. 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.
47. HADDAD S, WU M, WEI H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[J]. arXiv preprint, 2019.
48. 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.
49. 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.
50. LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J]. Computer Graphics Forum, 2007, 26(3): 655-664.
51. 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.
52. VEMULA A, MUELLING K, OH J. Social attention: Modeling attention in human crowds[C]// 2018 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 4601-4607.
53. HUANG Y, Bi H, Li Z, et al. STGAT: Modeling spatial-temporal interactions for human trajectory prediction[C]// 2019 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6272-6281.
54. MERCAT J, GILLES T, ZOGHBY NE, et al. Multi-Modal Simultaneous Forecasting of Vehicle Position Sequences using Social Attention[J]. arXiv preprint, 2019.
55. KHANDELWAL S, QI W, SINGH J, et al. What-If Motion Prediction for Autonomous Driving[J]. arXiv preprint, 2020.
56. LIANG M, YANG B, HU R, et al. Learning Lane Graph Representations for Motion Forecasting[J]. arXiv preprint, 2020.
57. MARCHETTI F, BECATTINI F, SEIDENARI L, et al. MANTRA: Memory Augmented Networks for Multiple Trajectory Prediction[J]. arXiv preprint, 2020.
58. PARK S, LEE G, SEO J, et al. Diverse and Admissible Trajectory Forecasting Through Multimodal Context Understanding[J]. arXiv preprint, 2020.
59. DEO N, TRIVEDI M. Multi-Modal Trajectory Prediction of Surrounding Vehicles With maneuver based LSTMs[J]. arXiv preprint, 2018.
60. LEE N, CHOI W, VERNAZA P, et al. DESIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents[J]. arXiv preprint, 2017.
61. ZHAO T, XU Y, MONFORT M, et al. Multi-Agent Tensor Fusion for Contextual Trajectory Prediction[J]. arXiv preprint, 2019.
62. HUANG X, MCGILL S, DECASTRO J, et al. Diversity GAN: Diversity-Aware Vehicle Motion Prediction via Latent Semantic Sampling[J]. arXiv preprint, 2020.
63. GAO J, SUN C, ZHAO H, et al. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation[J].arXiv preprint, 2020.
64. SRIKANTH S, ANSARI J, RAM R.K, et al. INFER: INtermediate representations for Future Prediction[J]. arXiv preprint, 2019.