

# 基于向量特征的车辆轨迹预测

徐鑫, 王孝兰\*

上海工程技术大学机械与汽车工程学院, 上海

收稿日期: 2023年2月21日; 录用日期: 2023年5月22日; 发布日期: 2023年5月29日

## 摘要

在复杂交通场景中轨迹预测是智能驾驶汽车中一个至关重要的问题, 这是因为道路结构、车辆间相互作用、智能体移动状态和环境信息的难以表示。本文提出一种多层图神经网络, 首先利用向量分别表示车道线、目标车辆等其他交通成员的空间局部特性, 然后建模所有成员之间的高阶相互作用。目前, 大多数方法将动态目标车辆的轨迹和道路结构环境信息的俯视图用卷积神经网络进行编码。而本文通过向量化表示高精度地图和智能体轨迹, 解决了计算密集的卷积网络编码步骤。为了进一步提高向量化学习上下文特征能力, 提出一种新的辅助任务根据上下文恢复随机掩码智能体特征。本文根据行为预测基准和ArgoVerse预测数据集对本文提出的算法进行评估。本文的方法表现了很好的性能, 同时节省了70%的模型参数。它在ArgoVerse数据集上的表现也超过了其他方法。

## 关键词

智能驾驶, 轨迹预测, 道路结构, 高精度地图, 特征工程

# Vehicle Trajectory Prediction Based on Vector Features

Xin Xu, Xiaolan Wang\*

School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai

Received: Feb. 21<sup>st</sup>, 2023; accepted: May 22<sup>nd</sup>, 2023; published: May 29<sup>th</sup>, 2023

## Abstract

Trajectory prediction in complex traffic scenarios is a crucial problem in smart driving vehicles due to the difficulty of representing road structure, inter-vehicle interactions, intelligent body movement states and environmental information. In this paper, we propose a multilayer graph

\*通讯作者。

neural network that first uses vectors to represent the spatial local characteristics of other traffic members such as lane lines and target vehicles separately, and then models the higher-order interactions among all members. Currently, most methods encode the top view of dynamic target vehicle trajectories and road structure environment information with convolutional neural networks. In contrast, this paper solves the computationally intensive convolutional network coding step by vectorizing the representation of high-definition (HD) maps and intelligent body trajectories. To further improve the vectorization learning contextual feature capability, a new auxiliary task is proposed to recover random masked Agent features according to the context. The algorithm proposed in this paper is evaluated against a behavioral prediction benchmark and the ArgoVerse prediction dataset. The approach in this paper achieves better performance on both benchmarks while saving 70% of the model parameters. It also outperforms other methods on the ArgoVerse dataset.

## Keywords

Intelligent Driving, Trajectory Prediction, Road Structure, High-Definition Maps, Feature Engineering

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着高精度地图被广泛用于提供上下文信息, 动态智能体的行为预测在自动驾驶应用中变得越来越重要[1] [2]。例如, IntentNet [3]提出联合检测车辆, 根据雷达点云和渲染的 HD 预测它们的轨迹。Hong 等[4]通过车辆目标检测编码实体与 CNN 的交互关注行为预测。卷积网络也被使用作多路径编码, [5]采用预处理的数据集来回归多模态可能未来轨迹。Precg [6]基于流的生成模型来捕捉未来的随机性。以上研究方法用 ConvNet 来编码渲染的道路地图, 而本文是直接编码向量化的场景上下文和代理动态。

在自动驾驶领域中, 理解动态智能体交互是至关重要的。例如行人[7] [8]、人类活动[9]和体育运动员[10] [11] [12]。Social LSTM 将单个智能体的轨迹建模为单独的 LSTM 网络, 并基于空间邻近度聚合 LSTM 隐藏状态以建模交互关系。Social GAN [13]简化了交互模块, 提出了一种相互博弈来预测不同的未来轨迹。Sun 等人[14]将图神经网络[15]与变分 RNN [16]相结合对不同的交互进行建模。Kipf 等人[17]将相互作用视为潜在变量; 图注意网络[18] [19]自注意机制对自定义图中的边进行加权。本文提出了一个统一的多层次图网络来联合建模多个代理之间的交互, 以及它们与道路地图中智能体的交互。

传统的机器感知算法一直专注于高维连续信号, 如图像、视频或音频。而 3D 感知它的输入通常是由深度传感器给出的无序点集的形式。齐等人提出了 PointNet 模型[20]和 PointNet++ [21]在嵌入层上应用排列不变运算(例如最大池化)。与点集不同, 高清地图上的道路特征和动态车辆迹形成封闭的形状或定向的轨迹, 它们也可能与属性信息相关联。因此, 本文通过向量化输入来保留这些信息, 并将属性编码为图中的节点特征。

最近, NLP 领域提出了一种自我监督的建模语言上下文[22] [23]。当转移到下游任务时, 他们学习的表示法可以显著提高性能。本文提出了一种图表示的辅助损失, 该方法学习从相邻节点预测丢失的节点特征。其目标是激励该模型更好地捕获节点之间的交互。

## 2. 模型框架

本文主要研究自动驾驶汽车等复杂多智能体中的轨迹预测问题。核心问题是找到一种统一的表示, 包括由感知系统提供的目标检测和跟踪得到的动态智能体与场景上下文, 场景上下文通常以高精度地图(HD)的形式作为先验知识提供。建立一个学习预测车辆意图的系统, 然后其意图被参数化为轨迹。

传统的行为预测方法是基于规则的, 其中多个行为假设是基于道路地图的约束而生成的。最近, 许多基于学习的方法被提出[24]; 他们提供了对不同行为假设进行概率解释, 但需要建立一个表示来编码地图和轨迹信息。虽然 HD 地图是高度结构化的, 包含车道和红绿灯等实体, 但这些方法中的大多数选择将 HD 地图呈现为颜色编码的属性, 这需要手动规范; 并且使用 CNN 来编码场景上下文信息, 其中 CNN 只有有限的感受野。

本文从多智能体动力学和结构化场景上下文的向量化形式中学习它们的统一表示(图 1)。道路结构可以由坐标中的点、多边形或曲线表示。例如, 车道边界包含样条线的多个控制点; 人行横道是由多个点定义的多边形; 停车标志由单个点表示。所有这些交通实体都可以看作由多个控制点及其属性定义的多段线。同样, 动态智能体的动力学也可以根据运动轨迹用多段线来表示, 这些多段线都可以表示为向量集。

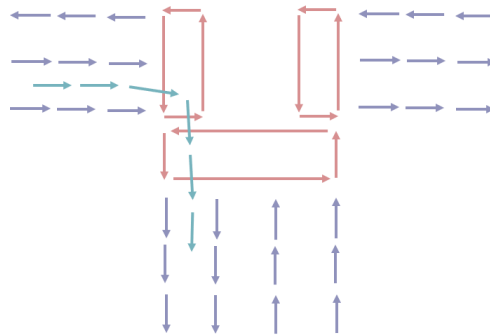


Figure 1. Vectorized representation  
图 1. 向量化表示

本文使用图神经网络(GNN)来拼接这些向量集合。将每个向量视为图中的一个节点, 并将节点特征设置为每个向量的起始位置和结束位置。来自 HD 的上下文信息和其他动态智能体的轨迹通过 GNN 传播到目标车辆节点。然后, 利用与目标车辆对应的输出节点特征来解码其未来的轨迹。

基于节点的空间和语义邻近度来约束图的连通性是重要的。因此, 本文提出了一种多层次图结构, 将相同折线的语义标签向量连接并嵌入到折线特征中, 然后所有折线完全相互连接以交换信息。用平均池化卷积实现局部图, 利用自注意力机制实现全局图。如图 2 所示。

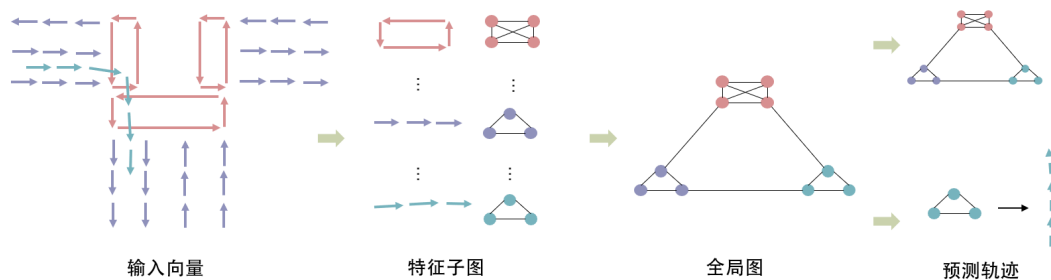


Figure 2. The overall structure of model  
图 2. 模型总体框架

最后, 受到自然语言[25]和机器视觉自我监督学习的启发, 本文还提出了一个辅助图补全目标。随机地屏蔽出属于场景上下文或目标车辆轨迹的输入节点特征, 模型预测被屏蔽的特征。使图神经网络更好地捕捉智能体动态和场景上下文之间的交互。在 ArgoVerse 数据集上对所提出的方法进行了评估, 结果表明本文方法与其他方法相比取得更好的性能, 模型大小节省了 70%。

### 3. 向量化方法

本节介绍向量化表示方法。首先, 描述如何向量化智能体轨迹和高精度地图。然后, 本文提出了层次图网络, 它从单个折线聚集局部信息。最后, 在所有轨迹和地图特征上全局聚集信息, 该图可用于轨迹预测。

#### 3.1. 高精度地图和动态智能体的向量化表示

来自 HD 地图的大多数注释是线条(例如车道)、闭合形状(例如十字路口区域)和点(例如交通信号灯)的形式, 还有注释的语义标签及其当前状态(例如交通信号灯的颜色、道路的速度限制)。对于智能体的轨迹是关于时间的有向折线的形式。所有这些元素都可以近似为向量序列: 对于地图要素, 从相同空间距离的线条中选取起点和方向, 对关键点进行均匀采样, 并将相邻的关键点依次连接成向量; 对于轨迹, 需以 0.1 s 的采样频率, 从  $t = 0$  开始, 并将其连接成向量。给定足够小的空间或时间间隔, 生成的折线用作原始地图和轨迹的近似值。

向量化是连续轨迹、地图注释的一对一映射, 尽管后者是无序的。这使得在向量集的顶部形成一个图表示, 该图可以通过图神经网络进行编码。本文将属于折线  $\mathcal{P}_j$  的每个向量  $v_i$  视为图中的节点, 其节点特征为

$$v_i = [d_i^s, d_i^e, a_i, j] \quad (1)$$

其中,  $d_i^s$  和  $d_i^e$  是向量的起点和终点坐标,  $d$  表示为 2D 坐标  $(x, y)$  或 3D 坐标  $(x, y, z)$ ;  $a_i$  是属性特征, 例如对象类型、轨迹时间戳、车道特征类型和速度限制;  $j$  是  $\mathcal{P}_j$  的全部 id,  $v_i \in \mathcal{P}_j$ 。

为了使输入节点特征对目标车辆的位置不变, 将所有向量的坐标归一化, 使其以目标车辆在最后观察到的时间步长的位置为中心。然后, 共享所有交互代理的坐标中心, 以便可以同时预测它们的轨迹。

#### 3.2. 构造多线段子图

为了利用节点的空间和语义局部性, 本文采取了一种分层的方法, 首先在向量级别构建子图, 其中属于同一折线的所有向量节点相互连接。考虑具有节点  $\{v_1, v_2, \dots, v_p\}$  的折线  $\mathcal{P}$ , 将单层次子图传播操作定义为

$$v_i^{(l+1)} = \varphi_{rel} \left( g_{enc} \left( v_i^{(l)} \right), \varphi_{agg} \left( \left\{ g_{enc} \left( v_j^{(l)} \right) \right\} \right) \right) \quad (2)$$

其中  $v_i^{(l)}$  是子图网络的第  $l$  层的节点特征,  $v_i^{(0)}$  是输入特征  $v_i$ , 函数  $g_{enc}$  是转换单个节点特征,  $\varphi_{agg}$  整合所有相邻节点的信息,  $\varphi_{rel}$  是节点  $v_i$  与它邻居节点关系运算符。

$g_{enc}$  是一个权重共享的多层感知器(MLP); MLP 包含一个完全连接的层, 随后是层归一化[3]和是 ReLU 非线性激活函数。  $\varphi_{agg}$  是最大池化操作, 而  $\varphi_{rel}$  是一个简单的拼接。如图 2 所示。堆叠了多层次子图网络, 其中  $g_{enc}$  的权重是不同的。最后, 子图特征计算为

$$P = \varphi_{agg} \left( \left\{ v_i^{(L_p)} \right\} \right) \quad (3)$$

其中  $\varphi_{agg}$  是最大池化计算。

本文的多段子图网络是 PointNet 改进: 当  $d_s = d_e$  且  $a, l = 0$ , 与 PointNet 有相同的输入和计算流程。然而, 通过将排序信息嵌入到向量中, 基于折线分组约束子图的连通性, 并将属性编码为节点特征, 该方法适用于结构化地图标注和车辆轨迹的编码。

### 3.3. 交通主体相互作用的全局图

本文使用全局交互图对节点和边特征  $\{P_1, P_2, \dots, P_p\}$  进行高阶交互作用建模:

$$\{p_i^{(l+1)}\} = \text{GNN}\left(\{p_i^{(l)}\}, A\right) \quad (4)$$

其中,  $p_i^{(l)}$  是线段节点特征集合, GNN 对应于图神经网络层, 而  $A$  是线段节点集合的邻接矩阵。假设  $A$  是一个完全连通的图, 子图网络的自注意力机制运算为:

$$\text{GNN}(P) = \text{softmax}\left(P_Q P_K^T P_V\right) \quad (5)$$

其中,  $P$  是节点特征,  $P_Q P_K^T P_V$  是  $P$  的线性映射。

对动态智能体的节点解码未来轨迹:

$$v_i^{\text{future}} = \varphi_{\text{traj}}\left(p_i^{(L_t)}\right) \quad (6)$$

其中  $L_t$  是 GNN 的总层数,  $\varphi_{\text{traj}}$  是轨迹解码器, MLP 为解码函数。

在本文的算法中, 使用一层 GNN, 计算与目标车辆对应的节点特征。然而, 还可以多层叠加 GNN 对高阶的交互进行建模。

为了使全局交互图更好地捕捉不同轨迹之间的交互并映射折线, 引入了一个辅助图任务。在训练期间, 随机地遮蔽出折线节点的子集的特征。然后, 尝试恢复其屏蔽特征, 如下所示:

$$\hat{P}_i = \varphi_{\text{node}}\left(p_i^{(L_t)}\right) \quad (7)$$

其中  $\varphi_{\text{node}}$  实现 LSTM 的节点特征解码器。

$p_i$  是全连接无序图中的一个节点。为了在子图的特征被遮盖时识别该节点, 计算其所有归属向量的起始点坐标的最小值, 以获得嵌入  $p_i$  的标识。然后, 输入结点要素变为

$$p_i^{(0)} = \left[ p_i; p_i^{\text{id}} \right] \quad (8)$$

本文图的补全特征算法与在处理自然语言的 BERT 方法是相似的, 该方法基于离散和顺序文本数据的双向上下文来预测丢失的标记。将预测目标特征推广无序图上, 节点特征在端到端框架中联合优化。

## 4. 实验

### 4.1. 数据集

本文使用两个关于车辆轨迹预测的数据集, ArgoVerse 数据集和 Apollo 数据集。ArgoVerse 数据集包含两大任务 3D 追踪和运动预测。数据集是由 324,557 个场景的精选集合, 每个场景 5 秒, 每个场景都包含以 10 赫兹采样的每个跟踪对象的 2d 鸟瞰质心。为了创建这个集合, 我们从自动驾驶测试车队中筛选了 1000 多个小时的驾驶数据, 以查找最具挑战性的细分市场-包括显示交叉路口的车辆, 向左转或向右转弯的车辆以及改变车道的车辆。除了车辆轨迹外, 每个序列还与高精度地图信息相关联。Apollo 数据集包含仿真数据集、演示数据集和标注数据集。由 53 分钟的训练序列和 50 分钟的测试序列组成, 以两帧/秒的速度捕获。该数据集是在中国北京收集的。数据的格式包括帧 ID、道路代理 ID、道路代理类型、3D 空间坐标、方向角和对象的高度、长度和宽度。基于相机的图像、激光雷达扫描的点云和手动注释的

轨迹组成。包含各种照明条件, 交通密集度收集的。

## 4.2. 指标

评估本文的算法, 在整个轨迹时间序列上计算平均位移误差 (ADE) 和最终位移误差(FDE)。其中,  $t = \{1.0, 2.0, 3.0\}$  秒。位移是以米为单位。

## 4.3. 向量化表示的向量网络

为了确保公平比较, 模型分别输入由矢量化表示与光栅化表示的信息。首先, 提取的地图元素集与渲染时完全相同。使目标车辆的道路特征向量和光栅化表示中的相同。然而, 向量化表示更复杂的道路特征, 这些特征对于渲染是没有的。

本文建立的向量化表示的轨迹预测模型的网络架构如表 1 所示, 其中包含编-解码器两部分。编码器基于多层 GNN 对道路结构、车辆轨迹和车道线编码, 利用卷积池化层提取道路成员交互特征。解码器基于结构化的 LSTM 输出未来轨迹。其中, 归一化和 ReLU 非线性激活函数, 对矢量坐标进行归一化, 使其以目标车辆在最后一个观测时间步长的位置为中心。与栅格化模型类似, 向量网络在 8 个 GPU 上与 ADAM 优化器同步训练。学习速率每 5 个周期衰减 0.3 倍, 训练 300 个周期, 初始学习率为 0.001, 如表 2 所示。

**Table 1.** Network design of trajectory prediction model based on vector representation

**表 1.** 基于向量表示的轨迹预测模型的网络设计

	参数名	备注
编码器	Fully connected	输入
	Batch normalization	提高计算效率
	激活函数	ReLU
	Maxpooling	提取特征
	Embedding	固定高维向量
	GNN_1	车辆轨迹子图
	GNN_2	道路结构子图
解码器	GNN_3	车道线子图
	LSTM	全局交互图
	激活函数	SeLU
	Fully connected	轨迹输出

**Table 2.** Model main parameters setting

**表 2.** 模型主要参数设置

参数	值	备注
input_embed_size	128	输入嵌入维度大小
hidden_size	128	隐藏层大小
batch_size	8	批量尺寸大小
num_epochs	300	训练周期
learning_rate	0.001	初始学习率
dropout	0.1	防止过拟合参数
seq_length	20	序列长度
obs_length	8	观测序列长度
pre_length	4/6/8/10/12	预测序列长度

#### 4.4. 向量网络实验

输入节点类型的影响。研究了在向量化表示中同时包含地图特征和车辆轨迹对预测未来轨迹是否有帮助。表 3 中的分别对应是只考虑目标车辆的历史轨迹、仅考虑高精度地图和同时考虑高精度地图和动态智能体状态。可以清楚地观察到, 添加地图信息显著提高了轨迹预测性能, 又结合其他车辆轨迹信息进一步提高了性能。表 3 比较了特征补全对目标车辆的影响, 增加辅助任务有助于减少预测误差, 特别是长时间范围内。

**Table 3.** Comparative error analysis

**表 3.** 误差对比分析

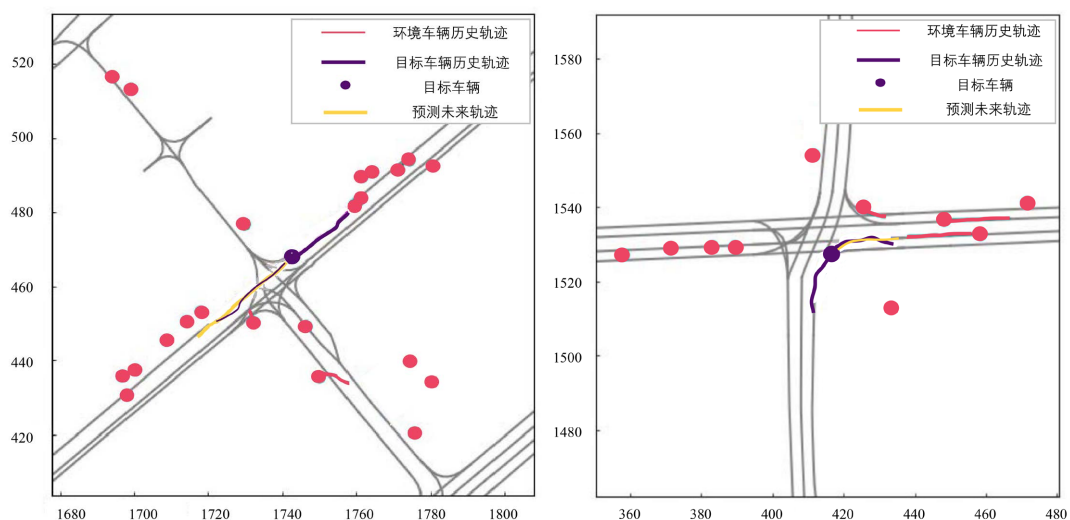
	补全特征	Apollo 数据集				Argoverse 数据集			
		1 s	2 s	3 s	ADE	1 s	2 s	3 s	ADE
-	-	0.79	0.89	1.23	0.93	1.29	2.97	5.23	2.38
HD	否	0.57	0.85	1.11	0.71	0.95	2.16	3.93	1.73
HD + 智能体	否	0.54	0.67	1.05	0.93	2.15	2.16	3.69	1.73
HD	是	0.57	0.79	1.08	0.69	0.95	2.43	3.67	1.80
HD + 智能体	是	0.53	0.73	0.99	0.73	1.23	2.04	3.68	1.65

在表 4 中, 研究了解码时 MLP 和 LSTM 对轨迹预测性能的影响。可以看到, LSTM 相对于 MLP 预测误差更小, 是因为 LSTM 有遗忘门、输入门和输出门。是因为 LSTM 有很强大的序列建模能力, 而 MLP 作为基础模型。LSTM 神经网络在处理长期语境或输入序列有优势。图 3 显示了有关预测轨迹可视化。

**Table 4.** Influence of MLP and LSTM on trajectory prediction performance

**表 4.** MLP 和 LSTM 对轨迹预测性能的影响

解码器网络	Apollo			Argovers		
	1 s	2 s	3 s	1 s	2 s	3 s
MLP	0.53	0.82	1.16	1.12	2.62	4.83
LSTM	0.50	0.77	1.09	1.09	2.27	4.81



**Figure 3.** Forecast trajectory visualization

**图 3.** 预测轨迹可视化图

车道显示为灰色, 环境车辆为红色, 目标车辆的真实轨迹为紫色, 预测轨迹为黄色。道路结构和动态智能体的可视化。可以看到真实轨迹和预测轨迹非常接近。

将向量表示与 ConvNet 模型进行了比较, 对于 apollo 数据集, 模型实现 ResNet 模型相同的性能。对于 ArgoVerse 数据集, 本文的方法优于最好的 ConvNet 模型, 位移误差减少了 12%。由于驾驶场景的自然分布, 内部数据集包含了大量静止的车辆, 向量化方法的性能超过最佳的 ConvNet 基线: 是因为它能够通过分层图网络捕获远程上下文信息。

## 5. 总结与展望

本文提出向量化表示 HD 和动态智能体运动状态, 预测车辆未来轨迹。为此设计多层次图网络, 其中第一层是折线内部向量之间的信息, 第二层模拟了子图间的高阶关系。在大规模 Apollo 数据集和公开可用的 ArgoVerse 数据集上的实验表明, 所提出的向量化表示在性能上优于 ConvNet, 同时大大降低了计算代价。本文方法还在 ArgoVerse 测试集上实现了最先进的性能。

在不久的将来, 希望在其他数据集上对所提出的模型进行评估, 其中的数据不仅包含高速公路场景, 还包含城道路场景, 提高模型的泛化性。通过以 RGB 摄像机等采集的视觉数据作为模型的输入, 以提高预测模型性能。

## 参考文献

- [1] Chang, M.F., Lambert, J., Sangkloy, P., *et al.* (2019) Argoverse: 3D Tracking and Forecasting with Rich Maps. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 8748-8757. <https://doi.org/10.1109/CVPR.2019.00895>
- [2] Krajewski, R., Bock, J., Kloeker, L., *et al.* (2018) The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Hawaii, 4-7 November 2018, 2118-2125. <https://doi.org/10.1109/ITSC.2018.8569552>
- [3] Casas, S., Luo, W. and Urtasun, R. (2018) IntentNet: Learning to Predict Intention from Raw Sensor Data. *Conference on Robot Learning*, PMLR, Tokyo, 17-20 September 2018, 947-956.
- [4] Hong, J., Sapp, B. and Philbin, J. (2019) Rules of the Road: Predicting Driving Behavior with a Convolutional Model of Semantic Interactions. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 8454-8462. <https://doi.org/10.1109/CVPR.2019.00865>
- [5] Chai, Y., Sapp, B., Bansal, M., *et al.* (2019) MultiPath: Multiple Probabilistic Anchor Trajectory Hypotheses for Behavior Prediction. *Proceedings of the Conference on Robot Learning*, PMLR, Vol. 100, 86-99.
- [6] Rhinehart, N., McAllister, R., Kitani, K., *et al.* (2019) Precog: Prediction Conditioned on Goals in Visual Multi-Agent Settings. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 20-26 October 2019, 2821-2830. <https://doi.org/10.1109/ICCV.2019.00291>
- [7] Alahi, A., Goel, K., Ramanathan, V., *et al.* (2016) Social LSTM: Human Trajectory Prediction in Crowded Spaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 26 June-1 July 2016, 961-971.
- [8] Zhao, J., Li, J., Cheng, Y., *et al.* (2018) Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and a New Benchmark for Multi-Human Parsing. *Proceedings of the 26th ACM International Conference on Multimedia*. Romania, Alba, 6-9 June 2018, 792-800. <https://doi.org/10.1145/3240508.3240509>
- [9] Sun, C., Shrivastava, A., Vondrick, C., *et al.* (2019) Relational Action Forecasting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 273-283. <https://doi.org/10.1109/CVPR.2019.00036>
- [10] Felsen, P., Agrawal, P. and Malik, J. (2017) What Will Happen Next? Forecasting Player Moves in Sports Videos. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 24-27 October 2017, 3342-3351. <https://doi.org/10.1109/ICCV.2017.362>
- [11] Yeh, R.A., Schwing, A.G., Huang, J., *et al.* (2019) Diverse Generation for Multi-Agent Sports Games. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 4610-4619. <https://doi.org/10.1109/CVPR.2019.00474>



- 
- [12] Zhan, E., Zheng, S., Yue, Y., *et al.* (2018) Generative Multi-Agent Behavioral Cloning.
- [13] Gupta, A., Johnson, J., *et al.* (2018) Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Alabama, 18-23 June 2018, 2255-2264. <https://doi.org/10.1109/CVPR.2018.00240>
- [14] Sun, C., Karlsson, P., Wu, J., *et al.* (2019) Stochastic Prediction of Multi-Agent Interactions from Partial Observations.
- [15] Battaglia, P.W., Hamrick, J.B., Bapst, V., *et al.* (2018) Relational Inductive Biases, Deep Learning, and Graph Networks.
- [16] Chung, J., Kastner, K., Dinh, L., *et al.* (2015) A Recurrent Latent Variable Model for Sequential Data. *Computer Science*, **35**, 1340-1353. <https://doi.org/10.1002/sim.6784>
- [17] Kipf, T., Fetaya, E., Wang, K.C., *et al.* (2018) Neural Relational Inference for Interacting Systems. *International Conference on Machine Learning*, PMLR, Chengdu, 16-18 July 2018, 2688-2697.
- [18] Hoshen, Y. (2017) Vain: Attentional Multi-Agent Predictive Modeling. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 2698-2708.
- [19] 范丽丽, 赵宏伟, 赵浩宇, 等. 基于深度卷积神经网络的目标检测研究综述[J]. 光学精密工程, 2020, 28(5): 1152-1164.
- [20] Qi, C.R., Su, H., Mo, K., *et al.* (2017) Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 21-26 July 2017, 652-660.
- [21] Qi, C.R., Yi, L., Su, H., *et al.* (2017) Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 5105-5114.
- [22] Devlin, J., Chag, M.W., Lee, K., *et al.* (2018) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.
- [23] Radford, A., Wu, J., Child, R., *et al.* (2019) Language Models Are Unsupervised Multitask Learners. *OpenAI Blog*, **1**, 9.
- [24] Cui, H., Radosavljevic, V., Chou, F.C., *et al.* (2019) Multimodal Trajectory Predictions for Autonomous Driving Using Deep Convolutional Networks. *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, 20-24 May 2019, 2090-2096. <https://doi.org/10.1109/ICRA.2019.8793868>
- [25] Sun, C., Myers, A., Vondrick, C., *et al.* (2019) Videobert: A Joint Model for Video and Language Representation Learning. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 20-26 October 2019, 7464-7473. <https://doi.org/10.1109/ICCV.2019.00756>