

# 驾驶环境中的行人穿越检测方法进展研究

(1. 长安大学交通视觉安全实验室, 西安 710064)

**摘要:** 随着自动驾驶和计算机视觉领域的飞速发展, 人们对交通安全的重视程度越来越大, 其中驾驶环境中行人穿越检测问题就是备受瞩目的难点之一。本文对行人穿越检测这一领域进行了综合评述, 并对未来研究进行了展望。按照行人试图穿越、正在穿越及穿越预测三个阶段进行分析, 首先分析了基于行人姿态穿越意图的研究进展(试图穿越); 随后讨论了基于轨迹跟踪的穿越行为检测方法(正在穿越); 接着概括了基于轨迹预测的穿越行人检测方法(穿越预测); 最后总结一部分现阶段比较新颖的研究角度并且对未来该领域的需重点发展的方向提出意见。

**关键词:** 穿越行人; 穿越意图; 轨迹跟踪; 轨迹预测; 注意力机制; 车路协同;

**中图分类号:** TP391.4 **文献标识码:** A

## Research on Pedestrian Crossing Detection Method in Driving Scenarios

(1. Lab Of Traffic Vision Safety Research (LOTVS), Chang an University, Xi'an, 710064, China)

**Abstract:** With the rapid development of autonomous driving and computer vision, more and more efforts are being made for traffic safety research, among which pedestrian crossing detection in driving environment is one of the most prominent problems. This paper reviews the field of pedestrian crossing detection, and prospects its future research. According to the three stages of pedestrian crossing process of crossing intention, crossing behavior and crossing prediction, this paper firstly reviews the crossing intention. Secondly, the tracking frameworks in crossing are discussed. Then the pedestrian crossing prediction works based on trajectory prediction is summarized. Finally, some new research perspectives are summarized and some suggestions are put forward for the future development of this field.

**Keywords:** Crossing Pedestrians; Crossing Intention; Trajectory Tracking; Trajectory Prediction; Attention Mechanism; Vehicle-to-Infrastructure cooperation.

### 1 引言

当前智能驾驶技术更加注重安全性的提升。据统计, 由于“车辆未按规定避让行人”这一因素造成的交通事故伤亡占据中国交通事故发生的最大比重<sup>[1]</sup>, 因此, 行人穿越检测是智能驾驶车辆面临的重要感知任务之一。同时, 由于长期以来人类在驾驶中积攒的大量经验, 类人驾驶是目前智能驾驶技术研究的主要途径之一。所以, 面对行人穿越检测这一问题, 如果智能车辆系统能够像人一样理解行人的穿越意图、感知行人穿越的运动状态、预测其运动趋向, 则可以有效地减少交通事故的发生, 从而提高驾驶的安全性, 促进智能交通领域发展。

近些年伴随深度神经网络理论研究的深入和计算力的不断提升, 深度学习得到了前所未有的关注与发展, 在计算机视觉领域催生并推动了许多应用场景, 在一些诸如物体检测, 图像分类等任务中都取得了超过人类检测的水平。因此, 在需要感知系统的多种无人系统平台中, 例如机器人、智能驾车、

无人机等, 被广泛使用<sup>[1][3][4][5]</sup>。在驾驶环境中行人是最主要的交通要素, 所以, 许多学者及研究机构着力于研究如何检测行人穿越<sup>[6][7][8][9][10]</sup>。发展到现在, 由于行人目标的高机动性、主观意图不明确性、场景交互的多样性, 这个问题现在仍然面临许多挑战<sup>[11]</sup>。本文主要综述当前驾驶环境行人穿越检测的发展现状及主要技术线路, 从试图穿越、正在穿越、穿越预测三个阶段分别以穿越意图理解角度、穿越轨迹跟踪角度及穿越行为预测角度进行分析。最后对行人穿越检测问题可能的新思路及挑战因素进行探讨, 亟期为无人驾驶领域的发展提供一些新思路。

### 2 基于行人穿越意图的方法: 试图穿越

#### 2.1 基于人体姿态意图的行人穿越检测研究

近年来, 一些关于穿越行人检测的方法利用行人的轮廓、姿态<sup>[12][13]</sup>等信息来进行穿越意图检测, 这些方法的目的是在人体姿态发生细微变化时进行穿越与否的判断。比如, 当某个行人要发生穿越行为时, 首先会左右观望交通状况, 此时尚未发生穿

越动作，只是头部进行了细微的转向。一般而言，如果路侧行人与车辆存在视觉注意的交互时，其穿越意图与未发生眼神交互的行人而言更加明显。因此路侧的行人头部朝向和眼神的交互是行人穿越可能发生的重要线索。此外，当一个行人站在道路边缘时，其比站在距离路边较远的行人更有意发生穿越。在行人的穿越意图理解中，头部和躯干的特征是最主要的特征。

在计算机视觉领域内，行人的特征通常是从视觉影像中获取能够充分表征行人区分性的高纬度表征。基于此，采用若干机器学习算法进行识别与分类，比如：支持向量机<sup>[6][14][15]</sup>，概率决策树<sup>[16]</sup>，k-NN<sup>[17]</sup>，人工神经网络<sup>[15]</sup>和卷积神经网络<sup>[18]</sup>等。在这些工作中有一类方法结合头部和腿部的特征来预测行人是否具有穿越意图的方法，比如<sup>[19]</sup>中的工作，首先利用头部图像进行头部朝向的估计，而腿部（躯干部分）用于估计行人的运动。图 1 是这类方法的一种典型示例，描述了这种特征提取的体系结构。

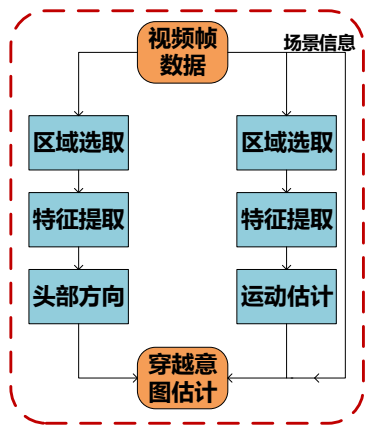


图 1. 基于人体特征识别穿越意图的框架

## 2.2 基于高斯动力学模型的穿越意图研究

高斯动力学模型在计算机视觉、模式分析及多媒体处理领域已经得到了大量的应用，并在很多具体应用中取得了突出进展。究其根本，高斯动力学模型的研究重点集中于如何将时空高维信息转换到低维空间来求解。在行人穿越意图估计中，高斯过程动力学模型将行人身体的关键点或肢体关节提取的三维时空信息约减到一个由位置和位移表征的低维高斯空间。例如 Raúl 等人<sup>[20]</sup>提出的平衡化的高斯动力学模型（Balanced Gaussian Process Dynamical Models，简称 B-GPDM）模型可以提前 1 秒预测未来行人的意图，该方法将行人骨架或者关节点提取

的三维位置和低维度隐空间上分布的位移关系融合进模型中，使其具有推断未来的可能位置和预测位置的特性。在此工作中，专门提出了四种基础模型：行走(walking)、停止(stopping)、开始(beginning)、站立(standing)。

这种基于高斯运动学过程模型进行意图检测的核心过程是匹配一个训练好的行人识别模型来估计每个目标的意图，此类方法的示意如图 2 所示，将行人执行不同活动的运动序列数据集进行训练。首先将几种类型的行人穿越序列分成若干子集，然后对视频内包含某一个活动的短时序建模一个高斯过程动力学模型。对于持续到来的新视频输入不断选择最佳行为模式匹配的模型，这种匹配的方法精度取决于模型的训练准确程度。最后所选模型用于预测行人意图是否具有四种意图属性中的一种从而完成意图检测，为自动驾驶车辆提供信息。本质上来讲，此工作是一种行人运动意图的分类方法。

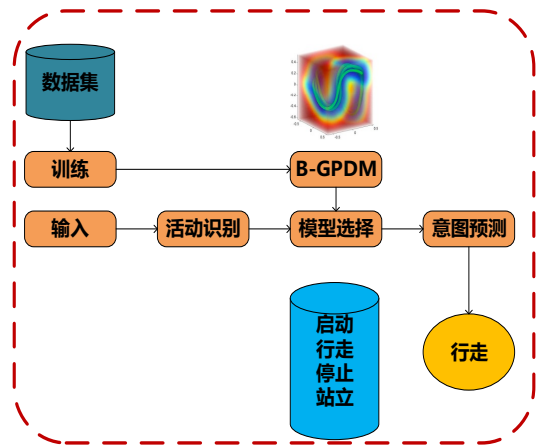


图 2. 训练及匹配示意图（该图来源于文献<sup>[20]</sup>）

## 2.3 基于数据驱动的穿越意图研究

随着大规模自动驾驶数据的不断涌现，深度学习逐渐在行人穿越任务中发挥作用。基于此，产生了一类数据驱动的行人穿越意图估计方法。当然，根据自动驾驶车辆配置传感器数据的模态差异，将有若干工作将不同传感数据引入进穿越意图估计，比如，采用三维激光点云数据。

在数据驱动下，多种深度学习网络结构被提出，其中，文献<sup>[22]</sup>将深度神经网络的时间序列分类体系结构进行比较，专门创建了预测路旁行人的穿越意图的多尺度密集记忆网络（multi-scale dense neural networks）和长短时记忆网络（Long-Short-Term-

Memory, 简称 LSTM), 并验证了其效果。多尺度密集记忆网络包含两个全联接层、两个 dropout 层以及一个 sigmoid 激活层, 用于直接将卷积神经网络提取的行人特征进行分类。而 LSTM 结构进一步在时序上进行关联, 捕捉行人运动特征。基于此, 可以推断出行人的运动意图, 比如行人未来姿势和移动方向。

由于图像坐标系是一种透视变换, 在行人穿越意图估计上存在偏差, 因此文献[22]采用三维点云数据进行意图估计, 其将三维点云数据的映射成一个二维图像, 图像分辨率为三维点云采样点的范围, 图像数值为三维点云的远近。这些图像可以收集行人姿态信息。对于每个独立的行人, 记录下点云集合组成的点阵编号, 以及点阵本身的旋转角度。这样, 卷积神经网络卓越的图像处理分类能力就可以得到利用。

#### 2.4 小结

如表 1 所示, 显示出文中提到的几种典型方法的特征。这些方法是在行人还未发生穿越行为时, 或者说行人将要穿越时进行意图的检测, 从而避免可能发生的事故。

表1 意图检测方法对比

方法	文献出处	模型	数据集	数据类型
头部特征	2018 ITSC	DKN	JAAD <sup>[25]</sup>	RGB
B-GPDM	2019 ITSC	GPDM	CMU-UAH <sup>[20]</sup>	RGB
CNN-LSTM	2016 ITSC	LSTM	STC <sup>[22]</sup>	激光点云
Multi-directional intention	2020 (Arxiv.com)	LSTM	BIT-Interaction <sup>[56]</sup>	RGB

### 3 基于轨迹跟踪的方法: 正在穿越

#### 3.1 行人穿越检测中跟踪方法的特性

在驾驶环境中, 由于交通参与者种类多样, 行人的轨迹跟踪问题需要克服动态场景的变化的干扰。同时, 驾驶环境行人目标数目众多, 一般采用多目标跟踪的方法进行研究。针对行人穿越而言, 其主要呈现两个大的挑战因素。

**频繁遮挡:** 行人穿越过程中, 由于其所处的复杂的交通环境, 难免存在被其他车辆或者行人目标遮挡的情况, 且经常存在持续遮挡的问题。

**强尺度变化:** 由于车辆的运动特性, 穿越的行人目标存在明显的尺度变化问题。比如远处的行人尺度较小, 随着行人的不断接近, 尺度极速增大。

**高机动性:** 行人穿越过程中, 由于其主观判断的变化, 其运动方向、运动快慢存在频繁的变化, 因此行人穿越过程具有高度的机动性, 所以要将行人在运动中的机动性考虑进去。

轨迹跟踪中为了克服这些问题, 目前工作大致可以分成离线跟在线训练的方法, 这是从数据的训练方面叙述几类典型的方法。其次, 对最新的基于图卷积的轨迹跟踪方法进行描述。

#### 3.2 基于两阶段训练和端到端训练的跟踪方法研究

##### 3.2.1 两阶段训练的跟踪方法研究

行人跟踪的分步训练方法一般来说需要预先收集整个序列后进行所有视频帧中行人目标的检测。考虑到相邻的视频帧之间的目标时空连接关系, 一般会用到图模型来求解多目标跟踪问题。这些图结构可以通过最小损失流<sup>[30][31]</sup>、最短路径<sup>[32]</sup>、多次剪枝<sup>[33]</sup>或多区域<sup>[34]</sup>的子图分解来求解。可以向图结构中的点或者权边施加概率约束来描述运动和遮挡问题<sup>[35]</sup>。现如今, 随着深度学习网络的发展, 行人特征描述已经全面采用深度卷积特征。

两阶段训练过程中, 在基于卷积神经网络进行行人目标特征提取基础上<sup>[36][37]</sup>, 关键问题是如何学习行人目标两两之间的相似度和短时轨迹之间的相似度。为了获得更精确的相似度度量, 在文献[38]中考虑了额外的时间顺序信息。这些连续的特征在轨迹分裂重联结模块中 (Generation Cleaving Re-connection Association, 简称 GCRA)<sup>[39]</sup>中得到了强化使用。此外, 联合多图切割方法 (Joint Multi-cut, 简称 JointMC)<sup>[40]</sup>提取行人目标的深度光流特征提升图结构边切割优化性能。DeepNetWork<sup>[41]</sup>采用端到端学习的方式对网络流进行优化从而改善跟踪结果, 从而更好地实现行人穿越跟踪。

在此类方法中, 我们研究组之前也设计了基于 RGB 图像数据和三维点云融合的多目标跟踪方法。方法流程图如图 4 所示。通过预先检测行人目标后,

将行人的视觉特征和三维点云特征进行融合，并输入到匈牙利矩阵中计算相邻视频帧之间的目标关联。

### 3.2.2 端到端的跟踪方法研究

与两阶段训练方式不同，端到端训练框架将行人目标检测和轨迹跟踪集成为一个端到端的网络结构。其主要难点在于如何同时学习目标检测器及轨迹关联模型，并进行新数据进入后的时序更新。

一些经过设计的深度学习方法为在线训练提供了有效的处理方法。比如工作[42]从行人再识别任务中学习外观特征，与检测相关联。而另一类方法<sup>[43][44]</sup>则通过使用长短时记忆（LSTM）网络进行自回归和匹配分类来预测运动和外观特征，比如基于时空注意机制的多目标跟踪（Spatial-Temporal Attention Mechanism MOT，简称 STAM-MOT）<sup>[44]</sup>应用空间和时间注意力映射来处理跟踪中的局部遮挡问题，并设计端到端的递归神经网络（RNN、LSTM）来学习轨迹与检测、语句更新、初始化和终止轨迹之间的关联。为了寻找目标的最优位置，文献[46]采用了用于单个目标跟踪的密集搜索策略。采用多假设跟踪（Multiple Hypothesis Tracking，简称 MHT）框架的两种基于深度学习的方法是 MHT-DAM 和 MHT-bLSTM<sup>[47]</sup>，其中采用 CNN 和双线性 LSTM 网络学习外观特征。

### 3.3 基于图卷积（GNN）的跟踪方法研究

随着图卷积网络（Graph neural network，简称 GNN）的发展，由于其对于关系的建模特性非常适合于多目标跟踪领域，本小节通过介绍近期基于 GNN 的跟踪算法的来分析其特点。

最早将 GNN 用于行人预测的是在 Jiang 等人<sup>[23]</sup>的工作中，可以看出这种方法通过一个孪生网络，得到观测的目标和当前的目标轨迹的表征相似度，然后将目标的历史轨迹位置作为输入，输入一个 LSTM 网络得到相应的位置，计算这个点与观测到的目标的运动相似度，由两个相似度构成相似度矩阵。轨迹与观测目标的相似度之间构成了关系图，以目标和观测信息作为节点，相似度作为边权，表现特征和位置信息拼接作为节点属性特征。

再比如另外一类基于 GNN 的多目标跟踪方法，深度关联网络（Deep Association Network，简称 DAN）<sup>[24]</sup>先将行人表征特征和实际运动特征进行提取，并构建了两种特征的网络联系图，通过 GNN 后得到

关联矩阵，其中目标重合度信息被用于相似度矩阵计算中。另外的一种方法，即 GNMOT（Graph Networks Multiple Object Tracking）<sup>[25]</sup>将 GNN 网络用于表征行人整体和躯干的运动，二者融合在一起得到一种相似度矩阵。

我们可以发现，基于 GNN 的多目标跟踪方法关键问题是怎样构建一个好的图关联结构。为此，文献[26]设计了一种基于信息传递网络（Message Passing Networks，简称 MPN）的跟踪器，其中，MPN 的图关联结构的构建方法是比较新颖的，图的节点是由所有的视频帧中的全部目标组成，直接将观测的信息作为一个总的节点。节点的属性特征是由训练得到的行人外观表征特征与人体姿态几何特征组成，并且特征由欧氏距离度量，然后通过创建的 MLP（多层感知机）网络得到最终的特征表达。由于图结构边之间的连接的是跨越不同帧的不同行人目标之间的关系，所以这可以看做一种端到端的多目标跟踪方法。

此外，在一些利用三维激光点云进行多目标跟踪的工作上，比如在 GNN3DMOT<sup>[27]</sup>，运动特征是通过 LSTM 网络得到，表现特征是通过卷积神经网络得到的。这个方法的改进版本 GNNTrk Forecast<sup>[28][29]</sup>中将多目标跟踪与轨迹预测结合，取得了更好的跟踪效果，其示意图如图 5 所示。

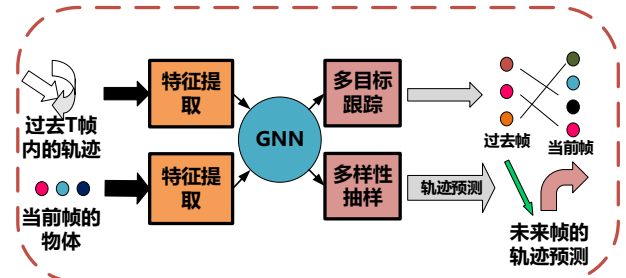


图 5. GNN 用于跟踪示意图（此图来源于文献 [28]）。

### 3.4 小结

表 2 为大家详细展示每种方法的具体网络模型。行人跟踪的方法着重于行人发生穿越的开始阶段，当行人发生了穿越行为的一小段时间内，通过跟踪的方法发现并检测出来穿越行人为智能车提供信息。表 2 跟踪方法详情



训练方式	时间	方法	特征提取模型+关联模型
两阶段	2016	Siamese CNN <sup>[36]</sup>	CNN
	2016	CNNTCM <sup>[37]</sup>	CNN
	2017	QuadMOT <sup>[38]</sup>	CNN
	2018	GCRA <sup>[39]</sup>	GRU
端到端	2015	MHT-DAM <sup>[41]</sup>	CNN
	2017	DeepSort <sup>[42]</sup>	CNN
	2017	STAM-MOT <sup>[46]</sup>	CNN
	2017	RNN-LSTM-MOT <sup>[45]</sup>	LSTM+RNN
	2018	RAN <sup>[43]</sup>	LSTM
	2018	MHT-bLSTM <sup>[47]</sup>	LSTM
	2020	MPN-Tracker <sup>[26]</sup>	GNN
2020	GNN3DMOT <sup>[27]</sup>	GNN	

#### 4 基于轨迹预测的方法研究：穿越预测

在基于穿越跟踪的基础上许多学者一直在研究穿越行人的轨迹来进行穿越预测，近些年的发展来看大致可以分为几类方法，例如基于概率匹配模型和物理匹配模型的方法，基于光流的方法，近些年还有基于深度学习的方法。

##### 4.1 基于马尔可夫决策过程的穿越预测方法

在行人穿越预测中有短期时序依赖的穿越行为预测和长期时序依赖的穿越行为预测。**Vasilyi【】**提出了一种针对行人长期行为的预测方法，将他们的行为建模为跳跃马尔可夫过程，目标是一个隐藏变量。假设近似理性的行为，并结合环境约束和偏差，包括环境的约束和偏差，将穿越行为建模为马尔可夫决策过程框架中的策略学习。使用时序滤波器推断行人状态，并通过随机策略进行规划来预测穿越动作。

将穿越行为预测定义为将目标映射到行动的策略函数  $\pi$ : 给定自然观测状态  $g$  和当前状态  $X=(X, \theta, S)$ , 其中  $X$  为行人位置,  $\theta$  为行人移动方向,  $S$  为行人移动距离。策略  $\pi$  产生从当前时间  $t$  到实现目标的未来某个时间的未来物理状态轨迹。因为具有相同目标的不同个体会有不同的行为, 所以策略  $\pi$  是随机的。预测结果示意图如图 6 所示。

实际上, 在轨迹预测过程中, 结合道路的几何结构和行人目标的运动朝向, 利用强化学习或者深度

强化学习进行轨迹预测具有很大的实用价值。只不过, 难点在于奖励空间的设计和与环境交互中轨迹预测动作集合的设计。

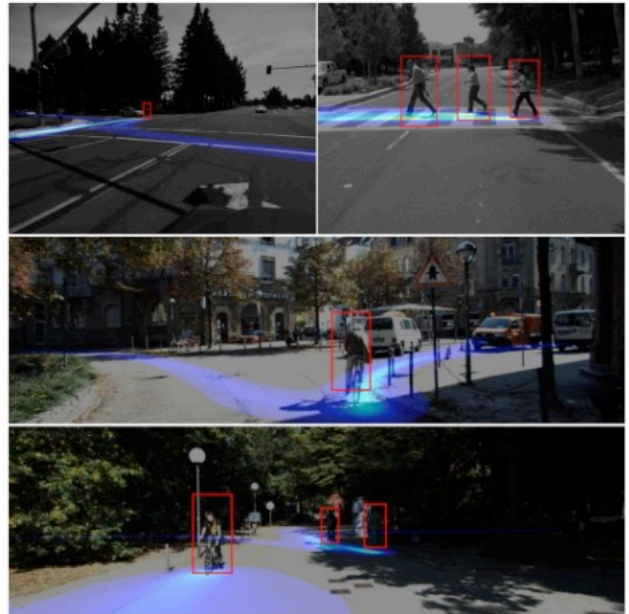


图 6. 蓝色表示可能的潜在运动区域 (该图来源于文献<sup>[59]</sup>)

##### 4.2 基于树状模型搜索的穿越预测方法

基于树状模型搜索的穿越预测方法, 本质上需要预先初始化若干包含行人运动未来状态的固定长度轨迹片段。在轨迹片段特征提取基础上, 并将这些特征嵌入到由历史轨迹片段构建的树状图结构中。然后给定目标轨迹的历史轨迹片段, 遍历搜索整个树状结构, 得到最佳的未来匹配轨迹片段。比如, **文献【】**通过对片段中每个时刻的状态进行特征提取, 并聚合到固定维度的特征向量中, 并用主成分分析方法对这些向量组进行特征值分解。然后可以根据最大特征值对它们的主维数进行排序。经过特征分解后的不同成分的特征向量用于构建搜索树, 通过遍历搜索的方法检索整个历史轨迹库来获得最佳匹配轨迹, 从而预测行人的穿越行为。图 7 表示出了树状模型搜索框架示意图。其中 (a) 将包含每个条目的位置  $X$ 、 $Z$  和特征信息  $F$  的历史长度为  $H$  的测试轨迹与训练数据库进行匹配。所得到的与训练数据库中轨迹的匹配位置和相似距离描述了可能的轨迹和类别标签。(b) 轨迹训练数据库的树表示。叶节点  $C_n$  表示固定长度的轨迹片段。通过遍历整个搜索树来搜索类似的轨迹。

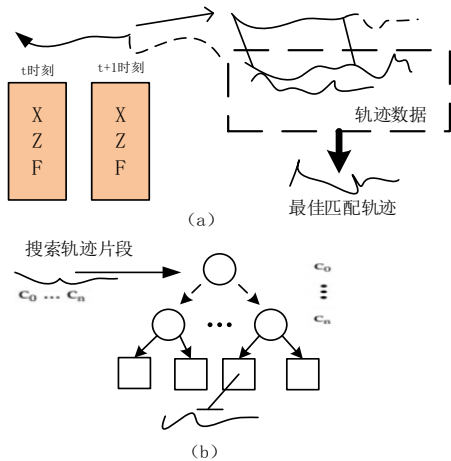


图 7. 物理遍历搜索示意图 (此图来源于文献【1】)

### 4.3 基于光流信息表征的穿越预测

光流法预测行人轨迹使用的是神经网络生成未来的光流，利用生成的光流在当前帧或者之前的帧中的图像上进行采集特征，得到未来预测图像。Patra 的工作<sup>[48]</sup>利用编码器和 LSTM 网络模型结构实现光流预测，同时在历史帧图像中进行采样从而生成未来预测帧图像。比如，光流信息表征的行人穿越检测的方法<sup>[49]</sup>利用深度体素流 (Deep Voxel Flow, 建成 DVF) 进行三维光流信息的提取，模型进行光流预测，然后在连续的两帧图像上进行同步采样，生成未来预测的帧图像。这种方法利用光流进行采样的方式生成未来图像，保持了十分良好的预测效果。现阶段的光流预测法一般来说首先估计光流，第二步根据估计的光流来计算之前帧图像采样坐标，最后通过迁移像素值的方式生成未来的预测图像帧。但是光流法也有弊端，当存在遮挡、模糊等情况下

时，光流预测法效果欠佳。因为光流法的算法仅是进行像素的移植，无法生成像素，且光流信息表征的行人穿越检测只能预测未来短时的视频帧信号。无法做到长时的预测。

### 4.4 基于深度学习的穿越预测

在行人轨迹预测任务中，当前的主流方法为利用社会力模型 (social force model) 去表征目标群之间或者目标与场景之间的上下文关系。比如著名的 social-LSTM 模型<sup>[52]</sup>将行人周围一定范围的目标进行位置关系编码，实现了准确的轨迹预测。

预测行人穿越行为的研究可以分为学习预测人与空间的相互作用或人与人的相互作用。后者模拟场景的动态内容，即行人如何相互作用。GAN 的重点是后者：学习预测人与人之间的交互。这个主题的现有工作以及 RNN 在层序预测和生成模型方面的进展良好。

轨迹预测是一个多模态的问题，模型的生成可以与时间序列数据一并使用来模拟未来的轨迹，在一种叫做 Social-GAN<sup>[50]</sup>中有三个关键组件：生成器、汇集模块和鉴别器。

对抗网络 GAN 所建立的一个学习框架，实际上就是生成模型和判别模型之间的一个模仿游戏。生成器的目的，就是要尽量去模仿、建模和学习真实数据的分布规律；而判别模型则是要判别自己所得到的一个输入数据，究竟是来自于真实的数据分布还是来自于一个生成模型。判别器通过这两个内部模型之间不断的竞争，从而提高两个模型的生成能力和判别能力。对抗网络框架示意如图 8 所示。

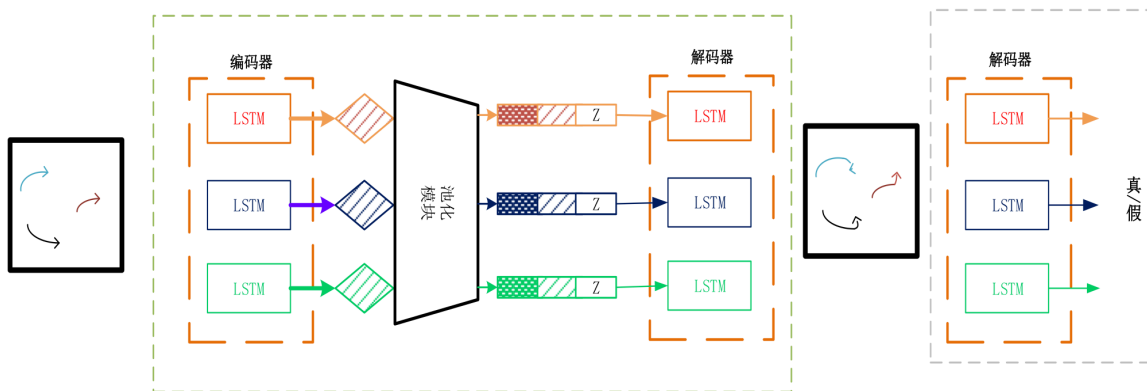


图 8. 基于 GAN 的行人轨迹预测方法

### 4.5 小结

行人轨迹预测的几种方法具体如表 3 所示。行人轨迹的预测着力于穿越行人的初始阶段，结合多

尺度行人检测等技术在行人发生穿越行为的开始进行预测，从而得到行人在未来时刻的位置信息等。这类方法不仅仅用于穿越行人检测上还用在自动驾驶领域内各类问题中。

表3 预测方法详情

方法	时间	模型	数据集	数据类型
DeepVoxel Flow	2017	CNN	UCD-101 <sup>[1]</sup>	RGB
Social-LSTM	2016	LSTM	ETH UNIV ZARA1 <sup>[50]</sup>	RGB
Social GAN	2018	GAN	ETH UNIV ZARA1 <sup>[50]</sup>	RGB
Trajectory Forecasting	2020	Seq2Seq	ETH UCV <sup>[50]</sup>	RGB

## 5 现阶段研究新思路

未来的行人检测技术的研究需要仍需要在以下几个方面时效性、多目标、多尺度，复杂环境和遮挡问题。当前的穿越行人检测技术大多数面向车载端，这是自动驾驶智能车发展的必然方向。自动驾驶当前的目标是建立高效准确的穿越行人检测方法，而在一些复杂场景天气下，尤其是利用视觉信号进行行人穿越检测时，存在很大的挑战。行人穿越检测很重要，为了更好的检测行人穿越，现阶段的主流新思路有如下几个：

(1) 驾驶员注意预测：驾驶员会注意到可能发生穿越的行人，结合自己的驾驶经验，会进行关键目标选择<sup>[51]</sup>，将这种机制融合到算法中可能会促进行人穿越检测性能提升。

(2) 联合注意角度：驾驶员视觉注意和行人视觉注意角度在发生穿越时，会产生交互，这种交互信息的利用可以有望提高行人穿越判别准确度。

(3) 场景几何结构使用：行人意图结合道路结构几何信息，是研究行人穿越的关键信息，比如站

在路边的行人与道路的角度越靠近 90° 穿越的概率越大。

(3) 多传感融合：从车载传感角度出发在视频输入设备的基础上融合多种传感器技术，进行检测，比如融合红外设备、3D 激光雷达等。

(4) 车路协同技术：另外随着国家布局车联网技术，车路协同技术得以大力发展，一些基于车路协同的穿越行人检测方法得到重视。其中，可以利用路侧及路上布设的新式传感器，进行行人穿越信号的识别，并及时发送给自主车辆。由路上的设备来进行感知（或者做一些边缘计算工作），这样，昂贵的传感器成本转移到道路基础设施上，给个人用户带来便利。

(5) 车联网技术：伴随着国家布局的车辆网技术的发展，在单车智能自动驾驶中，在车上装摄像头、激光雷达、毫米波雷达等传感器，让车能感知到周围的情况，自动做出反应。但是，单车智能存在不可避免的视觉盲区。假如在此区域内出现突然穿出的行人，则对于车辆来说，行人出现时减速已

## 参考文献

- [1] 公安部交通管理局，中华人民共和国道路交通事故统计年报 2019 版，2020.
- [2] Kumar P, Perrollaz M, Lefevre S, et al. Learning-based approach for online lane change intention prediction[C]// International Conference on Intelligent Vehicles Symposium (IV). 2013: 797-802.
- [3] Wang Z, Mülling K, Deisenroth M P, et al. Probabilistic movement modeling for intention inference in human-robot interaction[J]. The International Journal of Robotics Research, 2013, 32(7): 841-858.
- [4] Hu Y, Zhan W, Tomizuka M. Probabilistic prediction of vehicle semantic intention and motion[C]// International Conference on Intelligent Vehicles Symposium (IV).2018: 307-313.
- [5] Xue J R, Fang J W, Zhang P. A survey of scene understanding by event reasoning in autonomous driving[J]. International Journal of Automation and Computing, 2018, 15(3): 249-266.
- [6] A. T. Schulz and R. Stiefelhagen, A controlled interactive multiple model filter for combined pedestrian intention

- recognition and path prediction[J].in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp.173–178.
- [7] S. Köhler, M. Goldhammer, K. Zindler, K. Doll, and K. Dietmeyer, Stereo-vision-based pedestrian’s intention detection in a moving vehicle[J].in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 2317–2322.
- [8] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmuëdderich, Pedestrian crossing prediction using multiple context-based models[C] in 17th International IEEE Conference on Intelligent Transportation Systems, Oct 2014, pp. 378–385.
- [9] J.F.P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, Context-based pedestrian path prediction[C] in European Conference on Computer Vision, 2014, pp. 618–633.
- [10] X. Li, L. Li, F. Flohr, J. Wang, H. Xiong, M. Bernhard, S. Pan, D. M. Gavrila, and K. Li, A unified framework for concurrent pedestrian and cyclist detection[J] IEEE Transactions on Intelligent Transportation Systems, 18(2): 269–281, 2017.
- [11] S. Ferguson, B. Luders, R. C. Grande, and J. P. How, Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions[J] in Algorithmic Foundations of Robotics XI: Selected Contributions of the Eleventh International Workshop on the Algorithmic Foundations of Robotics, H. L. Akin, N.M. Amato, V. Isler, and A.F. van der Stappen, Eds. Cham: Springer International Publishing, 2015, pp. 161–177.
- [12] R. Furuhashi and K. Yamada, Estimation of street crossing intention from a pedestrian’s posture on a sidewalk using multiple image frames[C] in The First Asian Conference on Pattern Recognition, Nov 2011, pp. 17–21.
- [13] J. Hariyono and K. Jo, Detection of pedestrian crossing road: A study on pedestrian pose recognition[J] Neurocomputing, 234: 144 – 153, 2017.
- [14] F. Schneemann and P. Heinemann, Context-based detection of pedestrian crossing intention for autonomous driving in urban environments[C] in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, pp. 2243–2248.
- [15] C. Englund, J. Kovaceva, M. Lindman, and J.-F. Grönvall, Using random forests for data mining data and drowsy driver classification using FOT data[M] in Proceedings of On the Move to Meaningful Internet Systems: OTM 2012, ser. Part II, R. Meersman, H. Panetto, T. Dillon, S. Rinderle-Ma, P. Dadam, X. Zhou, S. Pearson, A. Ferscha, S. Bergamaschi, and I.F. Cruz, Eds., vol. 5766. Rom, Italy: Springer Verlag, 9 2012, pp. 748–758
- [16] C. Englund, M. Nilsson, and A. V. oronov, The application of data mining techniques to model visual distraction of bicyclists[M] Expert Systems with Applications, vol. 52, pp. 99–107, 2016.
- [17] L. Devroye, L. Györfi, A. Krzyżak, and G. Lugosi, On the Strong Universal Consistency of Nearest Neighbor Regression Function Estimates[M] The Annals of Statistics, 1994.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, [J] in Advances in neural information processing systems, 2012, pp. 1097–1105.
- [19] Varytimidis D, Alonso-Fernandez F, Duran B, et al. Action and intention recognition of pedestrians in urban traffic[J]. arXiv preprint arXiv:1810.09805, 2018: 676-682.
- [20] Raul Q M, Ignacio P A, David F L, et al. Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2018:1-12.
- [21] Wang J, Fleet D, Hertzmann A. Erratum: "Gaussian process dynamical models for human motion" (IEEE Transactions on Pattern Analysis and Machine Intelligence (292)) [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 30.
- [22] Volz B, Behrendt K, Mielenz H, et al. A data-driven approach for pedestrian intention estimation[C] // 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2016.
- [23] Jiang X, Li P, Li Y, et al. Graph Neural Based End-to-end Data Association Framework for Online Multiple-Object Tracking[J]. arXiv preprint arXiv:1907.05315, 2019.
- [24] Ma C, Li Y, Yang F, et al. Deep association: End-to-end graph-based learning for multiple object tracking with



- conv-graph neural network[C]Proceedings of the 2019 on International Conference on Multimedia Retrieval. 2019: 253-261.
- [25] Li J, Gao X, Jiang T. Graph Networks for Multiple Object Tracking[C].2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2020.
- [26] Brasó G, Leal-Taixé L. Learning a neural solver for multiple object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 6247-6257.
- [27] Weng X, Wang Y, Man Y, et al. GNN3DMOT: Graph Neural Network for 3D Multi-Object Tracking with Multi-Feature Learning[J]. arXiv preprint arXiv:2006.07327, 2020.
- [28] Weng X, Yuan Y, Kitani K. Joint 3d tracking and forecasting with graph neural network and diversity sampling[J]. arXiv preprint arXiv:2003.07847, 2020.
- [29] Wang Y, Weng X, Kitani K. Joint Detection and Multi-Object Tracking with Graph Neural Networks[J]. arXiv preprint arXiv:2006.13164, 2020.
- [30] Wang, X., Turetken, E., Fleuret, F., et al.: Tracking interacting objects using intertwined flows[J], IEEE Trans. Pattern Anal. Mach. Intell., 2016, 38, pp. 2312–2326
- [31] Butt, A.A., Collins, R.T.:Multi-target tracking by Lagrangian relaxation to min-cost network flow[J] Proc. IEEE Conf. Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 1846–1853
- [32] Berclaz, J., Fleuret, F., Turetken, E., et al.: ‘Multiple object tracking using k-shortest paths optimization[J], IEEE Trans. Pattern Anal. Mach. Intell., 2011,33, (9), pp. 1806–1819.
- [33] Tang, S., Andres, B., Andriluka, M., et al.: ‘Subgraph decomposition for multi-target tracking[J]. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 5033–5041
- [34] Dehghan, A., Modiri-Assari, S., Shah, M.: ‘GMMCP tracker: globally optimal generalized maximum multi clique problem for multiple object tracking[C]Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 4091–4099.
- [35] Milan, A., Schindler, K., Roth, S. Detection-and trajectory-level exclusion in multiple object tracking[J].Proc. IEEE Conf. Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 3682–3689.
- [36] Wang, B., Wang, L., Shuai, B., et al. Joint learning of convolutional neural networks and temporally constrained metrics for tracklet association[J] Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 2016, pp. 1–8.
- [37] Leal-Taixé, L., Canton-Ferrer, C., Schindler, K. Learning by tracking: siamese CNN for robust target association[J] Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 2016, pp. 33–40.
- [38] Son, J., Baek, M., Cho, M., et al.: Multi-object tracking with quadruplet convolutional neural networks[C]. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 5620–5629.
- [39] Ma, C., Yang, C., Yang, F., et al.: Trajectory factory: tracklet cleaving and reconnection by deep siamese BiGRU for multiple object tracking[R] arXiv preprint arXiv:1804.04555, 2018.
- [40] Tang, S., Andres, B., Andriluka, M., et al. Multi-person tracking by multicut and deep matching[J]. European Conf. on Computer Vision, Amsterdam, 2016, pp. 100–111.
- [41] Schuster, S., Verneza, P., Choi, W., et al. Deep network flow for multi-object tracking [C]. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017, pp. 2730–2739.
- [42] Wojke, N., Bewley, A., Paulus, D. Simple online and realtime tracking with a deep association metric[C]. Proc. Int. Conf. on Image Processing, Beijing, China, 2017, pp. 3645–3649.
- [43] Fang, K., Xiang, Y., Li, X., et al. Recurrent autoregressive networks for online multi-object tracking[C]. Proc. IEEE Winter Conf. on Applications of Computer Vision, Lake Tahoe, NV, USA, 2018, pp. 466–475.
- [44] Sadeghian, A., Alahi, A., Savarese, S. Tracking the untrackable: learning to track multiple cues with long-term dependencies[C]. Proc. IEEE Int. Conf. Computer Vision, Venice, Italy, 2017, pp. 300–311.
- [45] Chen, L., Ai, H., Shang, C., et al. Online multi-object tracking with convolutional neural networks[C]. Proc. Int.

- Conf. on Image Processing, Beijing, China, 2017, pp. 645–649.
- [46] Chu, Q., Ouyang, W., Li, H., et al. Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism[C]. Proc. IEEE Int. Conf. Computer Vision, Venice, Italy, 2017, pp. 4846–4855.
- [47] Kim, C., Li, F., Rehg, J.M. Multi-object tracking with neural gating using bilinear LSTM[J]. European Conf. on Computer Vision, Munich, Germany, 2018, pp. 200–215.
- [48] Patraucean V, Handa A, Cipolla R. Spatio-temporal video autoencoder with differentiable memory[J]. CoRR, vol. abs/1511.06309v5. Sep, 2016.
- [49] Liu Z, Yeh R A, Tang X, et al. Video frame synthesis using deep voxel flow[C] International Conference on Computer Vision (ICCV). 2017: 4463-4471.
- [50] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, Socialgan: Socially acceptable trajectories with generative adversarial networks[C]. CVPR, 2018. 4023-4362.
- [51] Fang J, Yan D, Qiao J, Xue J, and Yu H. DADA: Driver Attention Prediction in Driving Accident Scenarios [J], IEEE Transactions on Intelligent Transportation Systems (IEEE-TITS), DOI:10.1109/TITS.2020.3044678, 2020
- [52] Alahi A, Goel K, Ramanathan V, et al. Social LSTM: Human trajectory prediction in crowded spaces[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2016: 961–971.
- [53] Ji Y, Yang Y, Shen H T, et al. View-invariant Action Recognition via Unsupervised Attention Transfer (ANT)[J]. Pattern Recognition, 2021:107807.
- [54] Zhang Z, Peng H. Deeper and Wider Siamese Networks for Real-Time Visual Tracking[J]. 2019.
- [55] Liu Z, Yeh R A, Tang X, et al. Video Frame Synthesis using Deep Voxel Flow[J]. 2017.
- [56] Zhu Y, Ren D, Fan M, et al. Robust Trajectory Forecasting for Multiple Intelligent Agents in Dynamic Scene[J], 2020
- [57] 张丹, 单海军, 王哲, 吴陈炜, 无人系统之“眼”——计算机视觉技术与应用浅析[J]. 无人系统技术, 2019, 02:1-11.
- [58] 张朝阳, 李星宏, 5G 在无人驾驶汽车中的应用前景分析[J]. 无人系统技术, 2019, 02:65-68.
- [59] V. Karasev and S. Soatto, “Intent-Aware Long-Term Prediction of Pedestrian Motion,” in IEEE Int. Conf. Robot. Autom., 2016.