

探索无地图轨迹预测中的不确定性映射

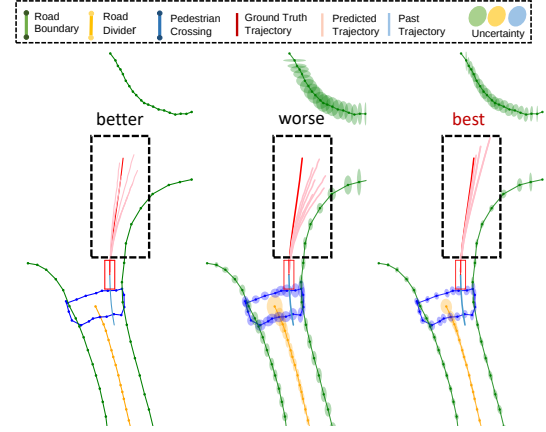
Zongzheng Zhang^{1,2*}, Xuchong Qiu^{2*}, Boran Zhang¹, Guantian Zheng¹, Xunjiang Gu⁴,
Guoxuan Chi¹, Huan-ang Gao¹, Leichen Wang^{2†}, Ziming Liu², Xinrun Li²,
Igor Gilitschenski⁴, Hongyang Li⁵, Hang Zhao³, Hao Zhao^{1✉}

Abstract—最近在自动驾驶方面的进展正朝着无地图的方法推进,其中高清(HD)地图直接从传感器数据在线生成,从而减少了对昂贵的标注和维护的需求。然而,这些在线生成地图的可靠性仍然不确定。尽管将地图不确定性融入下游轨迹预测任务显示出性能改进的潜力,但当前的策略对不确定性有利的特定场景提供的洞察力有限。在这项工作中,我们首先分析了映射不确定性对轨迹预测有最大积极影响的驾驶场景,并确定了一个关键的、先前被忽视的因素:代理的运动状态。基于这些见解,我们提出了一种新颖的自知情景门控算法,该算法根据自车未来运动学的预测,自适应地将地图不确定性整合到轨迹预测中。这种轻量级的自监督方法增强了在线映射和轨迹预测之间的协同作用,提供了不确定性有利的解释性,并且优于之前的集成方法。此外,我们介绍了一种基于协方差的地图不确定性方法,更好地与地图几何对齐,进一步改善了轨迹预测。广泛的消融研究证实了我们方法的有效性,在使用真实世界的 nuScenes 驾驶数据集的无地图轨迹预测性能上取得了高达 23.6 % 的改进。我们的代码、数据和模型在 <https://github.com/Ethan-Zheng136/Map-Uncertainty-for-Trajectory-Prediction> 上公开可用。

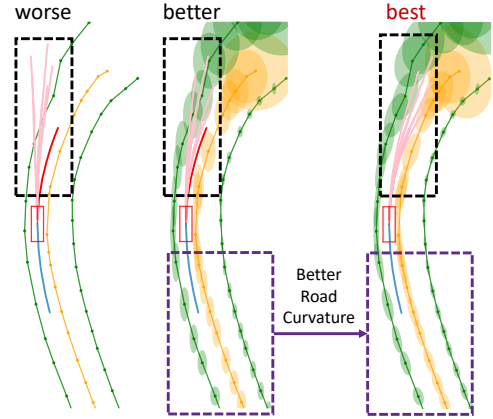
I. 介绍

传统上,高精度(HD)地图在自动驾驶中一直作为宝贵的资源,用于支持感知[?], [?], [?], [?], [?], [?], [?], [?], [?], [?]和规划[?], [?], [?], [?]等任务。这些地图提供了环境的详细、多层次的表示,包含道路几何(如折线和多边形)、语义属性(例如,路类型)、交通信号、速度限制以及其他对于静态场景理解至关重要的信息。然而,制作和维护高精度地图既耗时又昂贵,需要大量的标注并频繁进行更新以保持新鲜。这种对预标注数据的依赖限制了可扩展性,因为高精度地图通常限于特定的预定义区域,从而限制了更广泛的部署。

为了应对这些挑战,近期的研究逐渐转向直接从传感器数据在线生成高清地图,创建栅格化的鸟瞰视图(BEV)地图[?], [?], [?]或含折线的矢量化表示[?], [?], [?], [?]。这种“无地图驾驶”方法减少了对预先建立的离线地图的依赖,减少了广泛标注和维护的需要。随着该领域的发展,无地图驾驶的关注点从感知(改进在线地图生成)演变为运动预测,强调推进无地图轨迹预测[?], [?]。在此背景下,最近的研究强调了在在线地图生成时纳入不确定性[?], [?], [?]的潜力,这种不确定性来源于例如其他车辆和障碍物的遮挡,以及这些元素的不同距



(a) Motivation of the Proprioceptive Scenario Gating.



(b) Better road curvature with the Covariance-Based Uncertainty.

Fig. 1: 左: 无地图轨迹预测基线 (MapTRv2 [?] + HiVT [?]), 中: 之前的不确定性集成方法 [?], 右: 我们的。在 (a) 中, 对比显示增强了不确定性的预测偶尔比基线表现更差, 这强调了需要一个门控机制来有选择性地整合不确定性。在 (b) 中, 我们的基于协方差的地图不确定性能够很好地捕捉道路曲率, 使得预测轨迹能够与真实轨迹很好地对齐, 从而使真实轨迹变得不可见。

离, 以提高无地图轨迹预测的性能 [?]。然而, 当前的方法对在何种具体场景下这种不确定性最具效益或表现不佳的情况提供有限的洞见 (图 1 (a))。在这项工作中, 我们系统地研究了不确定性建模和整合在何种情境下能够最大程度提高预测准确性的场景。我们的分析揭示了一个关键的、此前被忽视的因素: 代理体的运动状态。具体而言, 我们发现, 在需要车辆运动状态剧烈变化的场景中 (例如, 从直线路径过渡到曲线路径), 引入地图不确定性主要能提高轨迹预测的准确性, 而在历史运动学是主要预测因素的场景中 (例如, 在直线路径上的稳定巡航), 可能会降低性能。这些发现表明, 在线地图不确

¹Institute for AI Industry Research (AIR), Tsinghua University, China.

²Bosch Corporate Research, China.

³Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, China.

⁴University of Toronto, Canada.

⁵University of HongKong, China.

* Equal contribution.

† Project lead.

✉ Corresponding to zhaohao@air.tsinghua.edu.cn

定性的有效性高度依赖于场景，促使我们寻求一种自适应的方法，根据特定情境条件选择性地在轨迹预测中应用地图不确定性。具体来说，我们提出了自体感知场景门控，这是一种即插即用的轻量级自监督模块，可以基于自车辆的运动状态自适应地整合地图不确定性。通过结合来自双并行流的初始轨迹预测，该机制通过加权整合选择性地结合增强和非增强的不确定性结果，优化最终的轨迹预测。此外，为了帮助轨迹预测方法更好地利用道路几何以获取更准确的预测，我们引入了一种基于协方差的不确定性建模方法，以更好地使不确定性与地图几何对齐（图 1 (b)）。大量消融实验证实了这两种技术的独立有效性，显示在各种最先进的映射器-预测器配置中表现出持续的性能提升。总之，我们的贡献有三点：首先，我们系统地证明了在线地图不确定性对无地图轨迹预测准确性的影响是情景依赖的，其影响根据代理的驾驶运动学而变化。其次，我们提出了一种轻量级自监督自体感知场景门控机制，自适应地整合地图不确定性，以及一种更好地与地图几何对齐的基于协方差的地图不确定性方法。第三，我们的方法在真实世界的 nuScenes 基准上实现了高达 23.6 % 的轨迹预测准确性提升，优于现有的无地图轨迹预测方法。

II. 相关工作

A. 在线地图估计

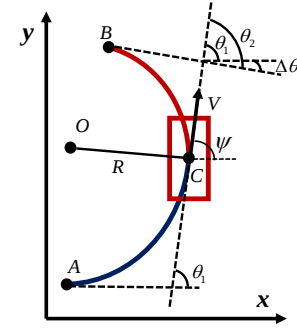
在线地图估计旨在直接从传感器数据中预测自动驾驶车辆周围静态世界元素的表示 [?]。早期的方法主要集中于生成 2D BEV 光栅化语义分割 [?], [?], [?], [?]。

矢量化地图估计的最新进展集中在通过将地图解码为折线和多边形来改进 BEV 栅格化。早期的方法依赖于后处理来生成地图 [?], [?]，而后来工作则引入了端到端学习，以简化流程并减少信息损失 [?], [?]。研究还融合了车道感知、实例内部信息和欧氏几何来增强地图表示 [?], [?]。

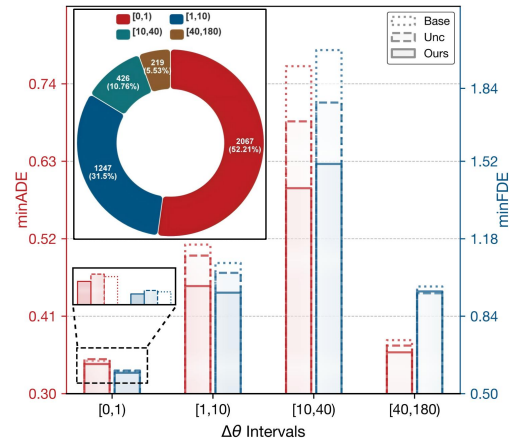
进一步的进展集中在提高地图表示中的定位精度和时间一致性 [?]。几项工作引入了技术来改善定位和减少映射错误，例如使用掩码引导的实例解码和减少抖动的抗干扰方法 [?]，以及端到端框架中的混合表示学习 [?]。其他研究则探讨了对传感器损坏的鲁棒性以及用于地图构建的增强点集查询 [?]。尽管取得了进展，在线生成地图在下游任务中的可靠性仍不确定。Gu 等人 [?] 使用一个简单的 2 单变量拉普拉斯分布来建模不确定性，但这限制了对各种道路几何形状的灵活性。在此基础上，我们引入了一种基于协方差的地图不确定性方法，该方法能够捕捉到更全面的 uncertainty 模式，从而增强下游应用中的映射可靠性。

B. 地图引导的轨迹预测

轨迹预测旨在根据历史轨迹和场景上下文预测代理的未来路径。传统模型通常依赖于使用卷积神经网络 (CNN) 编码的 BEV (鸟瞰图) 地图来表示场景，并结合代理历史信息来进行预测 [?], [?], [?], [?]。尽管有效，这些方法在数据效率和空间离散化方面存在问题，可能限制在复杂场景中的精度。为了应对这些限制，最近的研究转向了矢量化地图表示法，这种方法更加高效地捕获与代理运动相关的详细地图元素 [?], [?], [?]。最初的方法 [?], [?], [?] 引入了图神经网络 (GNN) 来编码车道折线，捕获地图对代理行为的影响。同时，基于 Transformer



(a) Vehicle's Particle Kinematic Model. The 红色 and 蓝色 lines indicate the future and the past 2s trajectory respectively. A and B denote the first and the last point of the trajectory. C is the center of the vehicle with a heading angle Ψ . O and R denote the instantaneous center and the radius of rotation. θ_1 and θ_2 represent the angle between the tangent and the velocity direction at A and B respectively.



(b) Statistics of the indicator $\Delta\theta$ on nuScenes dataset [?] and trajectory prediction performance. The upper left corner shows the distribution of the dataset across different $\Delta\theta$ intervals. The bar chart shows a comparison of the metrics among the base, previous uncertainty and our methods for each interval. Lower 最小平均位移误差 and 最小最终偏差误差 values indicate better trajectory prediction accuracy.

Fig. 2: 对轨迹预测中基于地图的不确定性进行的初步分析，以自我代理驾驶运动学背景为背景。

的架构捕捉空间和时间的依赖性，HiVT [?] 和 QCNet [?] 通过以查询为中心的设计和不变场景表示进一步增强了效率和鲁棒性。尽管矢量化方法显著推进了轨迹预测，挑战依然存在于如何有效地将它们与在线映射等上游模块集成。先前的方法在矢量化点集的同时编码了映射的不确定性，但缺乏对不确定性应该影响预测的时间和地点进行系统评估 [?]。我们的工作通过分析不确定性集成最有利的具体条件来解决这些限制，并引入了一种本体感知情景门控，以实时适应不同的上下文。这种自适应框架在复杂的驾驶场景中提高了预测的准确性、鲁棒性和可解释性。

III. 方法

本节首先分析地图不确定性在轨迹预测中提供优势的地方以及表现不佳的地方，使用车辆的粒子运动学模型 (第 ?? 节)。基于这些见解，我们概述了我们提出的框架 (第 ?? 节)。然后我们介绍我们的基于协方差的地图不确定性 (第 III-B 节) 和自身感知情景门控 (第 ?? 节)，以改进当前的无地图驾驶范式。最后，我们介绍用于训练

我们模型的整体损失函数（第 III-C 节）。

我们比较了基线（MapTRv2 [?] + HiVT [?]）和之前用于无地图驾驶的不确定性方法 [?]，以识别出不确定性对轨迹预测性能影响最显著的场景。

在图 2 (a) 中，根据车辆的粒子运动模型，我们有：

$$\theta_i = \dot{\Psi}_i \cdot \Delta t, i = 1, 2, \quad (1)$$

其中 $\dot{\Psi}_1$ 和 $\dot{\Psi}_2$ 分别表示自车过去和未来时间段内的平均角速度，这些时间段的长度为相同的时间间隔 $\Delta t = 2$ 秒。因此， θ_1 和 θ_2 表示自车在过去和未来 2 秒内经历的总旋转。然后，我们有

$$\Delta \theta = |\theta_1 - \theta_2| = |\dot{\Psi}_1 - \dot{\Psi}_2| \cdot \Delta t, \quad (2)$$

，它量化了这两个时间段内自车驾驶状态变化的程度 $|\dot{\Psi}_1 - \dot{\Psi}_2|$ 。当 $\Delta \theta$ 趋于零时，这表明车辆要么在直线行驶，要么在稳定转弯。

如图 2 (b) 所示，我们将表示曲率变化和车辆状态变化的指标 $\Delta \theta$ 的值分为四个区间，并分析每个区间内的场景数量和性能。值得注意的是，52.21 % 的场景位于 [0,1) 的范围内，在此范围内，基线（minADE = 0.3466，minFDE = 0.5980）略优于不确定性方法（minADE = 0.3490，minFDE = 0.6016），这表明在一致的驾驶行为下，排除不确定性可以获得更好的结果。相比之下，当车辆运动学的未来变化较大时，不确定性方法导致预测误差较低。如以下各节所述，我们提出的方法解决了这些问题，通过实验证明了更好的性能并取得了更佳的结果。

A. 架构概述

如图 3 所示，我们的框架由基于协方差地图不确定性估计的映射模块和具有本体感觉场景门控的轨迹预测模块组成，用于选择性地利用上游不确定性与车辆运动学。

首先，我们使用包括 MapTR [?]、MapTRv2 [?]、MapTRv2-Centerline [?] 和 StreamMapNet [?] 在内的在线映射方法来回归地图元素顶点。对于每个顶点，不确定性用二维高斯分布建模。然后将这种不确定性信息编码到下游轨迹预测的优化中。具体来说，我们采用了两个具有代表性的模型：HiVT [?]（基于 Transformer）和 DenseTNT [?]（基于 GNN）。在它们的上下文编码阶段，地图顶点坐标（不确定性建模中的均值 $\mu^{(i)}$ ）首先通过 MLP 编码。对于不确定性建模，我们将均值 $\mu^{(i)}$ 和协方差矩阵 $\Sigma^{(i)}$ 连接成一个单一表示，然后通过 MLP 处理。该过程在两个并行流中运行——一个包含不确定性，另一个不包含不确定性。然后，感知场景门控通过将这些流通过另一个 MLP 动态结合预测的轨迹。这个 MLP 为每个流生成权重，使得模型能够有效地合并轨迹并产生更优的结果。

B. 基于协方差的不确定性建模

Gu 等人在 [?] 中对每个地图元素的顶点在 x 和 y 方向上的不确定性建模，使用两个单变量拉普拉斯分布，并假设 x 和 y 坐标之间相互独立。然而，在真实场景中，由于地图元素的方向对齐，这两个位置随机变量通常是相关的。忽略这些相关性可能导致对不可靠地图区域的不正确建模，进而可能导致错误的轨迹预测。

为了解决这个问题，我们提出了使用具有相关系数的二维高斯分布来更好地描述关节不确定性的扩散和方向

性。具体来说，我们使用一个概率模块，该模块联合预测不确定性参数（即方差）和 $x - y$ 相关系数。此扩展使得模型能够更好地捕捉不确定性的空间关系，提供更全面的地图不确定性表示，从而提高预测的准确性和鲁棒性。

假设一个有 V 个顶点的地图 M ，在二维空间中，每个顶点 $v^{(i)}$ 服从一个均值为 $\mu^{(i)}$ 和协方差为 $\Sigma^{(i)}$ 的二维高斯分布。联合概率密度函数是：

$$f(M|\mu, \Sigma) = \prod_{i=1}^V \frac{1}{2\pi|\Sigma^{(i)}|^{1/2}} \exp \left(-\frac{1}{2} (v^{(i)} - \mu^{(i)})^T (\Sigma^{(i)})^{-1} (v^{(i)} - \mu^{(i)}) \right), \quad (3)$$

其中 $v^{(i)} = \begin{bmatrix} v_1^{(i)} & v_2^{(i)} \end{bmatrix}^T$ 是第 i 个顶点的坐标向量， $\mu^{(i)} = \begin{bmatrix} \mu_1^{(i)} & \mu_2^{(i)} \end{bmatrix}^T$ 是它的均值向量。第 i 个顶点的协方差矩阵是 $\Sigma^{(i)} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ ，其中 σ_1 和 σ_2 分别是在 x 和 y 维度上的标准差， ρ 表示相关系数，捕捉维度之间的线性依赖性。

与使用两个单变量拉普拉斯分布对 x 和 y 方向的不确定性进行建模相比（图 1 (b)），我们的基于协方差的方法更好地与地图元素的几何形状对齐，例如道路边缘曲率。我们选择具有协方差的高斯分布作为不确定性建模的选择，因为它理论上优于拉普拉斯分布，其原因是高斯分布具有较轻的尾部、成熟的数学性质、稳健的特性以及对异常值的敏感性较低。

为了解决这些上下文依赖的性能差异，我们引入了本体感知场景门控机制，通过选择性地结合两种方法的优点来增强轨迹预测。该模块能够进行上下文敏感的情景门控方法，从而实现更准确的预测，而不是不加区分地合并不确定性信息。

传统的外感门控方法基于多视图图像或鸟瞰图数据，依赖外部传感器（例如相机或 LiDAR）从多个角度分析周围环境。然而，这些方法没有考虑汽车代理的固有属性，例如其运动学。因此，我们提出了一种简单的自感门控方法，该方法基于车辆的内部运动学数据，不依赖任何额外的环境信息，有效地根据轨迹特征区分场景类型。如 Sec ?? 中所述，车辆的轨迹本质上包含其运动学数据，直接捕捉位置，间接包含诸如平移速度、角速度和加速度等信息，在区分不同场景类型中起着关键作用。

我们的方法遵循自监督范式。我们采用双编码器结构，使用和不使用不确定性的地图数据通过轨迹预测模型生成预测 $T_{base} \in \mathbb{R}^{1 \times 512}$ 和 $T_{unc} \in \mathbb{R}^{1 \times 512}$ 。这些轨迹被展平、填充并连接成输入张量 $T \in \mathbb{R}^{1 \times 2}$ ，然后通过六层 MLP 网络进行处理，其大小分别为 512、256、128、64、32、2，将张量形状从 [1,512] 缩小到 [1,2]：

$$T = \text{MLP}(\text{Flatten}(\text{Concat}(T_{base}, T_{unc}))) = [v_{base}, v_{unc}]. \quad (4)$$

输出 $[v_{base}, v_{unc}]$ 通过温度缩放的 softmax 变换以获得权重 $[w_{base}, w_{unc}]$ ，总和为 1，这些权重用作训练该模块的标签。

门控机制处理预测轨迹，以提取和解读与轨迹融合相关的运动学特征。这个训练过程使模型能够学习运动学特征与最优轨迹之间的关系，从而可以动态地对预测轨迹进行加权。

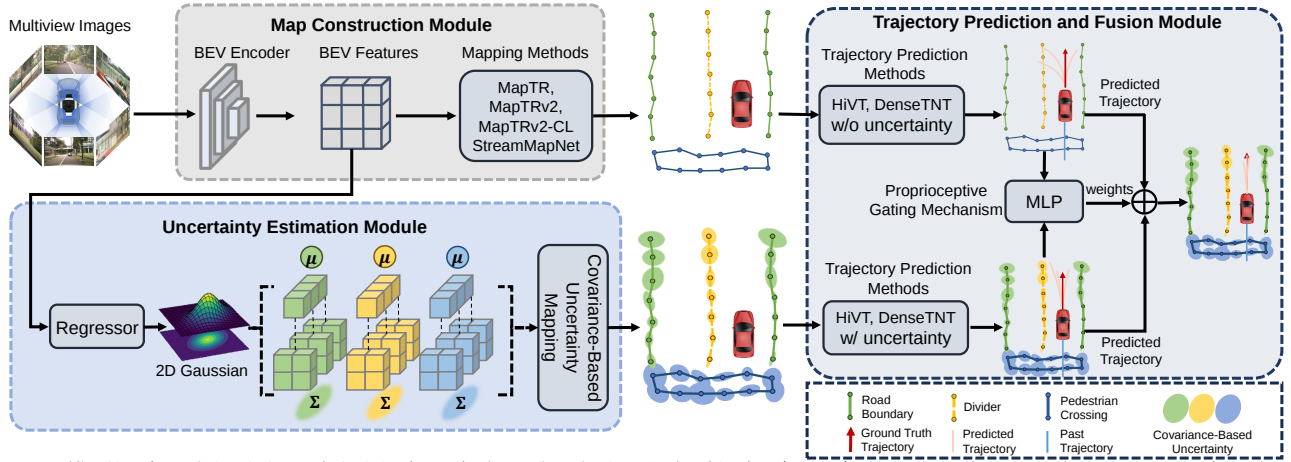


Fig. 3: 模型概述。我们首先通过将多视角图像编码到一个共同的鸟瞰视角（BEV）特征空间中，以回归地图元素顶点，在线估计地图元素。每个顶点的不确定性使用我们提出的基于协方差的不确定性方法建模，该方法利用二维高斯分布模板。这些不确定性信息以及原始地图顶点，然后被传递到下游的轨迹预测模块，该模块在两个并行流中运行：一个包含不确定性，一个则不包含。最终，所提出的本体感受场景门控（MLP 网络）根据这两个流产生的初始未来轨迹预测动态调整最佳轨迹预测。

C. 损失函数

为了训练用于在线映射的基于协方差的不确定性模型，我们将回归损失修改为二维高斯分布的负对数似然（NLL）损失：

$$\mathcal{L}_R(M|\mu, \Sigma) = \sum_{i=1}^V \left(\frac{1}{2} \log((2\pi)^2 |\Sigma^{(i)}|) + \frac{1}{2} (v^{(i)} - \mu^{(i)})^T (\Sigma^{(i)})^{-1} (v^{(i)} - \mu^{(i)}) \right), \quad (5)$$

其中 $\log((2\pi)^2 |\Sigma^{(i)}|)$ 是归一化项，与协方差矩阵 $\Sigma^{(i)}$ 的行列式相关， $(v^{(i)} - \mu^{(i)})^T (\Sigma^{(i)})^{-1} (v^{(i)} - \mu^{(i)})$ 表示误差项，测量预测位置与真实位置之间的马氏距离，同时考虑维度之间的相关性。针对以往工作中报道的高斯分布在训练过程中的不稳定问题 [?]，我们也加入了正则化项，以防止梯度爆炸，稳定训练过程，促进平滑收敛。

对于本体感觉情景门控，它是使用均方误差（MSE）损失函数进行训练的。该损失是基于模块预测的情景权重与目标权重之间的差异计算的，这些目标权重反映了每种情景的最佳轨迹选择。

IV. 实验

我们首先描述实验设置（第 IV-A 节），然后将我们的方法与最先进的地图不确定性集成轨迹预测方法进行比较（第 IV-B 节）。最后，我们提供可视化（第 IV-C 节）和一个消融研究（第 IV-D 节）。

A. 实验设置

数据集。我们在 nuScenes 数据集 [?] 上评估我们的方法，该数据集被划分为 500 个用于训练，200 个用于验证，150 个用于测试。每个 20 秒的场景以 2 Hz 的频率提供关键帧注释，六个多视角 RGB 摄像头提供了环绕自主车的 360° 视角。我们使用与 [?] 相同的轨迹处理方法。然后，每个模型基于 2 秒的历史数据预测未来 3 秒的运动。

指标。对于轨迹预测评估，我们采用最近预测挑战中常用的标准指标 [?, [?, [?]: 最小平均位移误差 (minADE)、最小最终位移误差 (minFDE) 和漏掉率 (MR)。我们的模型为每个智能体输出六条候选轨迹。minADE 指标计算在所有时间步长中最准确的预测轨迹和真实轨迹 (GT)

之间的平均欧几里得 (ℓ_2) 距离，而 minFDE 测量的是在最终时间步的距离。MR 量化的是最佳轨迹的终点与 GT 终点偏离超过 2 米的预测百分比。

执行细节。所有模型均在一台 NVIDIA GeForce RTX 3090 GPU 上训练。对于每个映射模型，学习率设定为 $1.5e-4$ ，回归损失权重为 0.03，梯度范数被截断为 3。

对于轨迹预测任务，HiVT 使用学习率 $5.0e-4$ 、批量大小 32 和温度 0.6 进行训练。对于 DenseTNT，学习率保持在 $5.0e-4$ ，而批量大小和温度分别调整为 16 和 0.5。在所有轨迹预测模型中一致应用 0.1 的丢弃率以防止过拟合。

B. 主要结果

为了评估我们的不确定性建模策略对下游轨迹预测的影响，我们在多种在线地图估计和预测方法的组合下进行了实验。为了更有效地与先前工作进行比较，我们选择了相同的四种映射方法 (MapTR [?], MapTRv2 [?], MapTRv2-Centerline [?] 和 StreamMapNet [?]) 以及两种轨迹预测模型 (HiVT [?] 和 DenseTNT [?])，从而得到八种组合。表格 I 显示，使用我们的公式后，所有组合的性能都有持续改善。使用 HiVT 时，我们的方法在 MapTR 和 MapTRv2 上将 minADE 和 minFDE 提高了超过 6%，并在 MapTRv2 上在 MR 上取得了显著的 13.6% 的改进。对于 DenseTNT，最大增益出现在 MapTRv2-Centerline，其中 minADE、minFDE 和 MR 分别提高了 19.4%、10.8% 和 23.6%。图 2 (b) 进一步说明，在不同的 $\Delta\theta$ 区间内，我们的方法始终如一地提高了轨迹预测性能。

C. 可视化

图 4 中的定性结果表明，我们的方法在各种驾驶场景中始终改进了轨迹预测。在图 4 的左上方，自车在一个复杂场景中执行稳定的圆形转弯，其中 $\Delta\theta \approx 0$ 表示稳定的驾驶状态（参见第 ?? 节）。基准方法 (a) 优于之前基于不确定性的方法 (b)，但部分预测轨迹超出了道路边界。相比之下，我们的方法 (c) 与真实轨迹高度一致，并保持在可驾驶区域内。在右上方，当自车接近一个缓慢的曲线时，之前的方法 (b) 略微优于基准方法 (a)，但两者的预测都超出了道路边界。我们的方法 (c) 则跟随道

Prediction Method	HiVT [?]				DenseTNT [?]			
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓	FPS	minADE ↓	minFDE ↓	MR ↓	FPS
MapTR [?]	0.4015	0.8404	0.0960	17.91	1.1228	2.2151	0.3726	7.36
MapTR [?] + Unc [?]	0.3910	0.8049	0.0818	16.65	1.1946	2.2666	0.3848	7.75
MapTR [?] + Ours	0.3672 (−6.1%)	0.7395 (−8.1%)	0.0756 (−7.6%)	15.08	1.0865 (−9.0%)	2.0969 (−7.5%)	0.3728 (−3.2%)	6.96
MapTRv2 [?]	0.4017	0.8406	0.0959	15.66	1.3262	2.5687	0.4301	7.57
MapTRv2 [?] + Unc [?]	0.3913	0.8054	0.0819	14.98	1.3256	2.6390	0.4435	7.38
MapTRv2 [?] + Ours	0.3670 (−6.2%)	0.7538 (−6.4%)	0.0708 (−13.6%)	13.65	1.1585 (−12.6%)	2.4566 (−7.4%)	0.3891 (−12.3%)	6.89
MapTRv2-CL [?]	0.3789	0.7859	0.0865	14.02	0.8333	1.4752	0.1719	7.62
MapTRv2-CL [?] + Unc [?]	0.3674	0.7418	0.0739	14.03	0.9666	1.6439	0.2082	7.29
MapTRv2-CL [?] + Ours	0.3659 (−0.4%)	0.7404 (−0.2%)	0.0721 (−2.4%)	13.47	0.7787 (−19.4%)	1.4662 (−10.8%)	0.1590 (−23.6%)	6.73
StreamMapNet [?]	0.3963	0.8223	0.0923	14.17	1.0639	2.1430	0.3412	7.42
StreamMapNet [?] + Unc [?]	0.3899	0.8101	0.0861	13.69	1.0902	2.1412	0.3261	6.91
StreamMapNet [?] + Ours	0.3828 (−1.8%)	0.7981 (−1.5%)	0.0834 (−0.3%)	13.26	0.9675 (−11.3%)	1.6883 (−21.2%)	0.2628 (−19.4%)	6.32

TABLE I: 轨迹预测的定量结果适用于在 nuScenes [?] 数据集的不同设置下的代表性映射/预测模型组合。我们的方法结合了提出的自由感觉场景门控和基于协方差的地图不确定性，始终提升预测模型性能，几乎在所有组合上超过了以前的不确定性集成方法。此外，我们在 NVIDIA RTX 3090 上对 FPS 进行了基准测试，结果表明我们的方法在实现与基线相当的计算效率的同时，提供了更优异的预测性能。

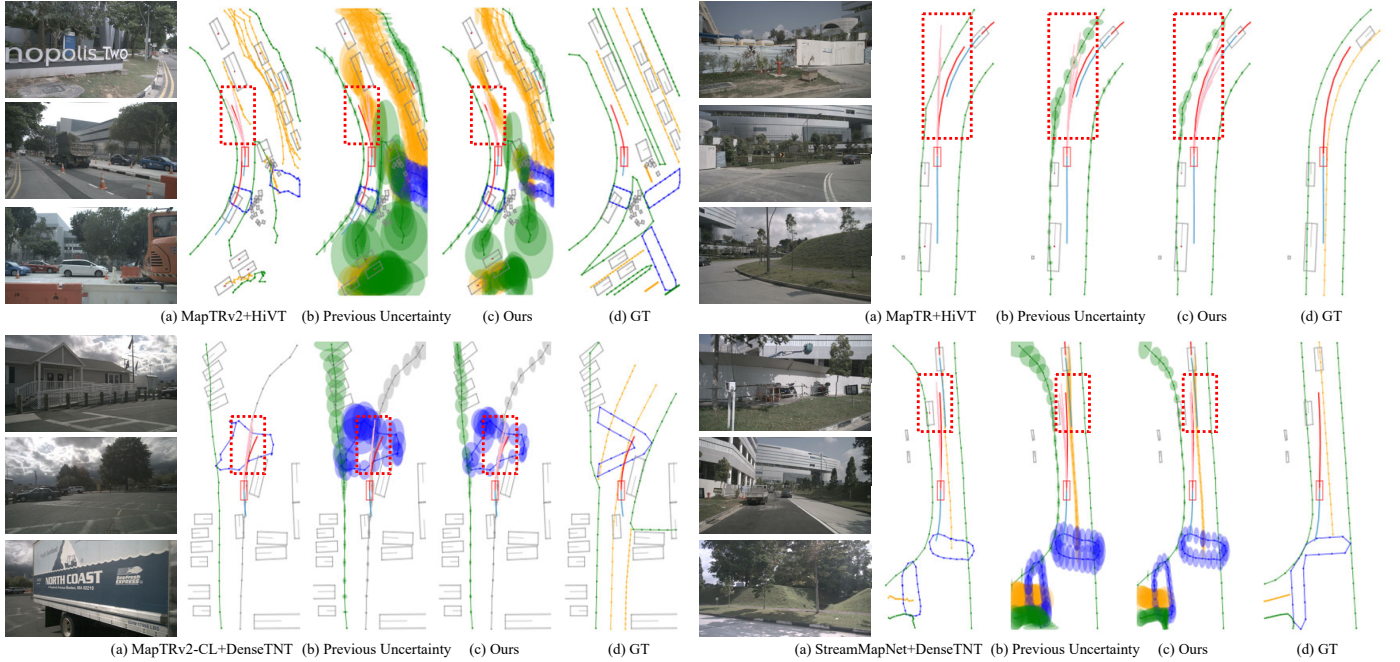


Fig. 4: 定性可视化展示了在线地图和轨迹预测模型的各种组合。在所有这些驾驶场景中，我们的方法都优于基线和最新的不确定性建模方法 [?]

路曲率并避开前方车辆。左下方的图像显示自行车在保持车道的同时避开了一辆卡车。基准方法和之前基于不确定性的方法都漂移到了对向车道，而我们的方法成功避开了卡车并保持在当前车道。右下方，自行车从一个转弯过渡到笔直的路径。之前的方法相较基准方法表现较差，而我们的方法优于两者，展示了我们门控机制在利用运动学信息方面的优势。如表 I 所示，对于 DenseTNT，之前的基于不确定性的方法相较基准方法通常表现较差。图 4 的左下方和右下方场景突显了我们的方法能够产生更精确且紧密集中的轨迹。

D. 消融实验

首先，我们评估了在不使用固有感知场景门控的情况下，所提出的基于协方差的不确定性效果，如表格 II 所示。该公式在所有映射方法中一致地提升了轨迹预测性能，为下游任务提供了更丰富、与几何对齐的不确定性信息。引入基于协方差的不确定性显著减少了 minFDE 和 MR，其中最显著的改进出现在 MapTRv2-Centerline

(minFDE 减少 4.3 %) 和 StreamMapNet (MR 减少 6.5 %)。我们还探讨了使用四种不同分布模拟不确定性的影响，如表格 III 所示。我们的结果表明，无论是使用 Laplace [?] 还是高斯分布，将协方差（相关性）纳入不确定性建模中始终优于独立分布方法，因为它更好地与道路曲率对齐。此外，基于协方差的高斯分布通常优于基于协方差的拉普拉斯分布。这进一步验证了我们选择使用联合高斯分布来建模不确定性的理由。

1) 本体感觉场景门控：我们研究了所提出的本体感受场景门控的有效性。表 IV 显示，基于运动学特征自适应选择最优轨迹提高了所有映射-预测组合的预测准确性。通过分析 w_{base} 和 w_{base} 的值，我们发现，在稳定车辆状态 ($\Delta\theta \in [0, 1)$) 中， w_{base} 的平均值为 0.82，而在其他情况下， w_{base} 的平均值为 0.37。这进一步证明，我们的门控机制从轨迹中学习了车辆的运动学信息，并做出了更优的轨迹。

为了验证我们基于动态自身感觉的方法的有效性，我们使用 CLIP 模型 [?] 和 ResNet50 [?] 将其与外感测

Prediction Method	HiVT [?]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [?]	0.4015	0.8404	0.0960
MapTR [?] + Unc [?]	0.3910	0.8049	0.0818
MapTR [?] + CovMat	0.3851 (-1.5%)	0.7812 (-2.9%)	0.0815 (-0.4%)
MapTRv2 [?]	0.4017	0.8406	0.0959
MapTRv2 [?] + Unc [?]	0.3913	0.8054	0.0819
MapTRv2 [?] + CovMat	0.3776 (-3.5%)	0.7707 (-4.3%)	0.0789 (-1.5%)
MapTRv2-CL [?]	0.3789	0.7859	0.0865
MapTRv2-CL [?] + Unc [?]	0.3674	0.7418	0.0739
MapTRv2-CL [?] + CovMat	0.3662 (-0.3%)	0.7411 (-0.1%)	0.0728 (-1.5%)
StreamMapNet [?]	0.3963	0.8223	0.0923
StreamMapNet [?] + Unc [?]	0.3899	0.8101	0.0861
StreamMapNet [?] + CovMat	0.3843 (-1.4%)	0.8035 (-0.8%)	0.0805 (-6.5%)

TABLE II: 一般来说，利用我们的基于协方差的不确定性公式（CovMat）进行在线映射，相较于之前的两种单变量拉普拉斯方法 [?]，在所有组合中都能持续提高预测模型的性能。

Prediction Method	HiVT [?]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [?]	0.4015	0.8404	0.0960
MapTR [?] + Bivariate Laplace w/o Covariance [?]	0.3910	0.8049	0.0802
MapTR [?] + Bivariate Gaussian w/o Covariance	0.3931	0.8079	0.0839
MapTR [?] + Bivariate Laplace w/ Covariance	0.3853	0.7862	0.0803
MapTR [?] + Bivariate Gaussian w/ Covariance (Ours)	0.3851	0.7812	0.0815

TABLE III: 研究采用四种不同概率分布对轨迹预测中的模型不确定性进行影响分析。结果表明，基于协方差的联合高斯分布提供了最佳的建模方法。

方法进行比较。具体来说，前摄像头图像是 CLIP 图像编码器或 ResNet50 的输入，用于提取语义特征，然后通过多层感知器处理，以根据分类结果生成加权轨迹选择。道路语义被划分为四个类别作为真实标记：低曲率转弯、高曲率转弯、存在其他代理和直线路径，每个类别都由单热编码表示。

如表 ?? 所示，基于 CLIP 和 ResNet 的外感知方法在所有模型和指标上均不如我们基于动态本体感觉的方法。在某些情况下，它甚至低于基线性能，这表明仅基于语义场景理解进行轨迹选择优化可能是不够的。此外，表 V 强调了 FPS 的显著差异，我们的方法实现的速度最高可达 CLIP 的 30 倍和 ResNet50 的 10 倍，使其更适合于车辆的实时部署。

Prediction Method	HiVT [?]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [?]	0.4015	0.8404	0.0960
MapTR [?] + Unc [?]	0.3910	0.8049	0.0818
MapTR [?] + Gating	0.3829 (-2.1%)	0.7855 (-2.4%)	0.0805 (-1.6%)
MapTRv2 [?]	0.4017	0.8406	0.0959
MapTRv2 [?] + Unc [?]	0.3913	0.8054	0.0819
MapTRv2 [?] + Gating	0.3815 (-2.5%)	0.7914 (-1.7%)	0.0807 (-1.5%)
MapTRv2-CL [?]	0.3789	0.7859	0.0865
MapTRv2-CL [?] + Unc [?]	0.3674	0.7418	0.0739
MapTRv2-CL [?] + Gating	0.3650 (-0.7%)	0.7362 (-0.8%)	0.0685 (-7.3%)
StreamMapNet [?]	0.3963	0.8223	0.0923
StreamMapNet [?] + Unc [?]	0.3899	0.8101	0.0861
StreamMapNet [?] + Gating	0.3858 (-1.0%)	0.7981 (-1.5%)	0.0847 (-1.6%)

TABLE IV: 结合门控方法以动态利用地图的不确定性能够持续改进轨迹预测。

Gating Method	CLIP [?]	ResNet50 [?]	Ours
Online HD Map Method	FPS	FPS	FPS
MapTR [?]	2.21	6.59	52.26
MapTRv2 [?]	2.15	6.51	92.09
MapTRv2-CL [?]	2.33	6.70	87.35
StreamMapNet [?]	2.42	6.91	84.57

TABLE V: 我们的本体感受场景门控的 FPS 比基于 CLIP 的外部感知替代方法高达 30 倍，比基于 ResNet50 的外部感知方法高 10 倍。

Prediction Method	HiVT [?]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [?] + Unc [?]	0.3910	0.8049	0.0818
MapTR [?] + BEV Feature Attention [?]	0.3617	0.7401	0.0730
MapTR [?] + CovMat	0.3851	0.7812	0.0815
MapTR [?] + Ours	0.3494 (-3.5%)	0.7028 (-5.6%)	0.0653 (-9.3%)

TABLE VI: 用一个更强的替代方案 [?] 替换我们提出的双流结构中的基线 (MapTR [?])，显示出预测准确性的显著提高。

E. 可扩展性

为了验证我们所提架构的可扩展性，我们在双流结构中用 BEV 特征注意方法 [?] 替换了基线方法 (MapTR [?])，保留架构的其余部分不变。如表 VI 所示，结果表明我们的方法在轨迹预测任务中达到了 SOTA 性能，突显了我们架构强大的泛化能力。

V. 结论

在这项工作中，我们探讨了地图不确定性对无地图轨迹预测的影响，并确定了车辆运动状态在决定这种不确定性何时有利中的关键作用。基于这些见解，我们引入了一种新的机制——自感知场景门控，它根据对车辆未来运动学的预测，自适应地将地图不确定性整合到轨迹预测中。该自监督方法不仅提高了性能，还通过识别不确定性有积极贡献的场景来增强可解释性。此外，我们提出了一种基于协方差的地图不确定性建模，它更好地与地图几何对齐，从而进一步提高了性能。在真实世界数据上的大量实验证明了我们方法的显著性能提升。