SEPT: Standard-Definition Map Enhanced Scene Perception and Topology Reasoning for Autonomous Driving

SEPT: 用于自动驾驶的标准定义地图增强场景感知与拓扑推理

Muleilan Pei, Jiayao Shan, Peiliang Li, Jieqi Shi, Jing Huo, Yang Gao, and Shaojie Shen

裴穆磊岚，单佳尧，李沛良，石杰琦，霍晶，高扬，沈少杰

Abstract-Online scene perception and topology reasoning are critical for autonomous vehicles to understand their driving environments, particularly for mapless driving systems that endeavor to reduce reliance on costly High-Definition (HD) maps. However, recent advances in online scene understanding still face limitations, especially in long-range or occluded scenarios, due to the inherent constraints of onboard sensors. To address this challenge, we propose a Standard-Definition (SD) Map Enhanced scene Perception and Topology reasoning (SEPT) framework, which explores how to effectively incorporate the SD map as prior knowledge into existing perception and reasoning pipelines. Specifically, we introduce a novel hybrid feature fusion strategy that combines SD maps with Bird's-Eye-View (BEV) features, considering both rasterized and vectorized representations, while mitigating potential misalignment between SD maps and BEV feature spaces. Additionally, we leverage the SD map characteristics to design an auxiliary intersection-aware keypoint detection task, which further enhances the overall scene understanding performance. Experimental results on the large-scale OpenLane-V2 dataset demonstrate that by effectively integrating SD map priors, our framework significantly improves both scene perception and topology reasoning, outperforming existing methods by a substantial margin.

摘要-在线场景感知与拓扑推理对于自动驾驶车辆理解其驾驶环境至关重要，尤其对于旨在减少对昂贵高精度 (HD) 地图依赖的无地图驾驶系统。然而，近期在线场景理解的进展仍面临限制，特别是在远距离或遮挡场景中，因车载传感器的固有限制。为解决此挑战，我们提出了标准定义 (SD) 地图增强的场景感知与拓扑推理 (SEPT) 框架，探讨如何将 SD 地图作为先验知识有效融入现有感知与推理流程。具体而言，我们引入了一种新颖的混合特征融合策略，结合 SD 地图与鸟瞰视图 (BEV) 特征，兼顾栅格化与矢量化表示，同时缓解 SD 地图与 BEV 特征空间之间的潜在错位。此外，我们利用 SD 地图特性设计了辅助的路口感知关键点检测任务，进一步提升整体场景理解性能。在大规模 OpenLane-V2 数据集上的实验结果表明，通过有效整合 SD 地图先验，我们的框架显著提升了场景感知与拓扑推理能力，较现有方法有大幅度优势。

Index Terms-Computer vision for transportation, deep learning for visual perception, intelligent transportation systems.

关键词-交通计算机视觉，视觉感知深度学习，智能交通系统。

I. INTRODUCTION

一、引言

SCENE understanding is essential for autonomous vehicles, facilitating critical downstream tasks such as accurate motion prediction and decision-making. High-Definition (HD) maps play a pivotal role in this process, providing rich geometric and semantic information, as well as topology relationships. However, HD maps present significant challenges, including high annotation costs, scalability limitations, and ongoing maintenance demands [1], which underscore the increasing need for online scene perception and topology reasoning [2].

> 场景理解对于自动驾驶车辆至关重要，支持准确的运动预测与决策等关键下游任务。高精度 (HD) 地图在此过程中发挥核心作用，提供丰富的几何与语义信息及拓扑关系。然而，HD 地图存在高昂标注成本、扩展性受限及持续维护需求等重大挑战 [1]，凸显了在线场景感知与拓扑推理日益增长的需求 [2]。

In recent years, vision-centric mapless driving approaches (i.e., driving without HD maps) have made significant strides [3], [4], especially within advanced driver assistance systems. These methods aim to reduce the heavy reliance on HD maps by leveraging onboard sensors to perceive the complex scene structure of driving environments in real time. Specifically, with multi-view images as input, a variety of tasks need to be addressed, including lane segment detection, traffic element recognition, and scene topology reasoning [5], [6].

> 近年来，以视觉为核心的无地图驾驶方法 (即不依赖 HD 地图驾驶) 取得显著进展 [3]，[4]，尤其在高级驾驶辅助系统中。这些方法旨在通过车载传感器实时感知复杂驾驶环境的场景结构，减少对 HD 地图的高度依赖。具体而言，以多视角图像为输入，需要解决多项任务，包括车道段检测、交通元素识别及场景拓扑推理 [5]，[6]。

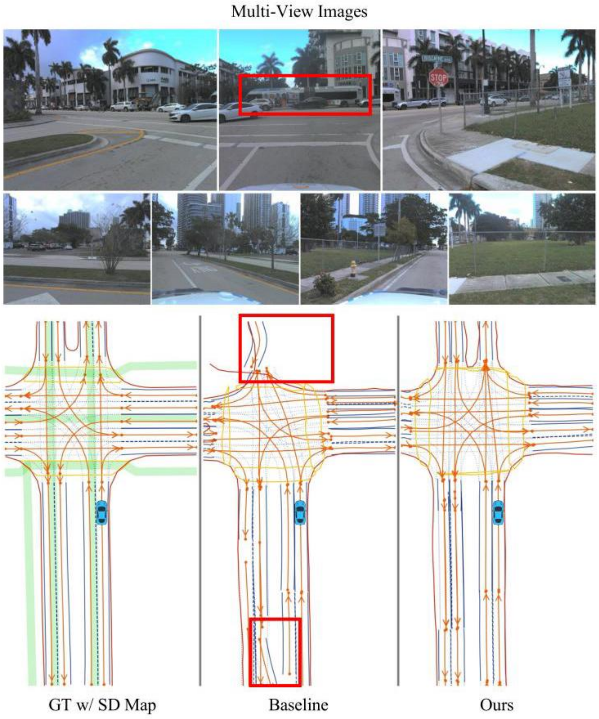

Fig. 1. Illustration of the SD map prior for enhancing online scene understanding in long-distance and occlusion scenarios. In this example, the front view is severely obstructed by a bus at the intersection (highlighted in the red box in the front-view image), and the left-back zone is distant. The baseline (LaneSegNet [6]) fails to correctly perceive the road structure (indicated by the red boxes in the top and bottom of the middle visualization), whereas our SEPT framework accurately predicts the road layout with the augmentation of the SD map prior. The Ground Truth (GT) of lane segments is shown in the left figure, with the green line representing the SD map.

图 1. 标准定义地图先验用于增强远距离及遮挡场景在线理解的示意图。示例中，路口前视被一辆公交车严重遮挡 (前视图红框标注)，左后方区域较远。基线方法 (LaneSegNet [6]) 未能正确感知道路结构 (中间可视化上下红框所示)，而我们的 SEPT 框架在 SD 地图先验增强下准确预测了道路布局。车道段真实标注 (GT) 见左图，绿色线条表示 SD 地图。

Nevertheless, due to the inherent limitations of onboard sensors, such as constrained perception range and restricted field of view, fully mapless driving systems often struggle to accurately reconstruct far-seeing or occluded road conditions. Given that human drivers typically perceive the surrounding scenarios by combining observations with navigation maps, also known as Standard-Definition (SD) maps [7], integrating SD maps as additional prior knowledge of road structures offers a promising solution to complement onboard sensory inputs. In general, the SD map provides a centerline skeleton of road networks without detailed and high-precision annotations [8], making it lightweight, scalable, easily accessible, and low-cost [9]. This basic geographic and road-level topological information can effectively augment online sensing capabilities, thereby enhancing scene perception and topology reasoning, particularly in long-distance or occlusion scenarios, as demonstrated in Fig. 1.

然而，由于车载传感器固有限制，如感知范围受限和视野狭窄，纯无地图驾驶系统常难以准确重建远距离或遮挡的道路状况。鉴于人类驾驶员通常结合导航地图 (即标准定义 (SD) 地图 [7]) 与观察感知周围场景，将 SD 地图作为道路结构的额外先验知识整合，成为补充车载传感输入的有力方案。一般而言，SD 地图提供道路网络的中心线骨架，缺乏详细高精度标注 [8]，因而轻量、可扩展、易获取且成本低廉 [9]。这种基础的地理及道路级拓扑信息能有效增强在线感知能力，提升场景感知与拓扑推理，尤其在远距离或遮挡场景中，如图 1 所示。

Despite the substantial potential benefits of SD maps, the effective integration of such map priors into current online perception and reasoning paradigms remains an ongoing challenge. Existing approaches typically rely on relatively simple encoding strategies to represent SD maps, either in a rasterized [8] or vectorized [10] format. Each representation has distinct advantages: dense rasterization preserves spatial positional information and fine-grained local details, while sparse vectorization captures complex geometry and topology more efficiently. However, most

This work was supported in part by the Hong Kong PhD Fellowship Scheme, and in part by the HKUST-DJI Joint Innovation Laboratory. (Corresponding author: Jieqi Shi.)

本工作部分由香港博士研究生奖学金计划资助，部分由香港科技大学-大疆联合创新实验室支持。(通讯作者: 石杰琦)

Muleilan Pei and Shaojie Shen are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China (email: mpei@ust.hk; eeshaojie@ust.hk).

裴穆磊岚与沈少杰隶属于中国香港科技大学电子与计算机工程系 (邮箱:mpei@ust.hk；eeshaojie@ust.hk)。

Jiayao Shan and Peiliang Li are with Zhuoyu Technology, Shenzhen, China (email: jiayao.shan@zyt.com; peiliang.li@zyt.com)

单佳尧与李沛良隶属于中国深圳卓宇科技 (邮箱:jiayao.shan@zyt.com；peiliang.li@zyt.com)。

Jieqi Shi, Jing Huo, and Yang Gao are with State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China (email: isjieqi@nju.edu.cn; huojing@nju.edu.cn; gaoy@nju.edu.cn).

石杰琦、霍晶与高扬隶属于中国南京大学新型软件技术国家重点实验室 (邮箱:isjieqi@nju.edu.cn；huojing@nju.edu.cn；gaoy@nju.edu.cn)。

methods either focus on one representation or combine the two in a simplistic manner [11], which limits effective feature extraction and results in suboptimal utilization or information loss from the SD map. To address this gap, we encode the SD map using a hybrid representation and propose a lightweight yet effective fusion module to augment the Bird's-Eye-View (BEV) features with SD map priors. Additionally, inherent inaccuracies in GPS signals often cause weak spatial misalignment between the SD map and BEV space [12]. While previous works tend to neglect this artifact or dismiss it as noise, we introduce a feature alignment mechanism to resolve this issue. Specifically, for rasterization, we design a feature transformation network that dynamically modulates the features through predicting a scaling factor and bias term for each feature channel; for vectorization, we adopt a cross-attention mechanism [13] that adaptively attends to corresponding features, ensuring better alignment with the BEV feature space.

尽管 SD 地图具有显著的潜在优势，但将此类地图先验有效整合到当前的在线感知与推理范式中仍然是一个持续的挑战。现有方法通常依赖相对简单的编码策略来表示 SD 地图，采用栅格化 [8] 或矢量化 [10] 格式。每种表示各有优势: 密集栅格化保留空间位置信息和细粒度局部细节，而稀疏矢量化则更高效地捕捉复杂的几何形状和拓扑结构。然而，大多数方法要么专注于单一表示，要么以简单方式结合两者 [11]，这限制了有效特征提取，导致 SD 地图信息利用不足或信息丢失。为弥补这一不足，我们采用混合表示对 SD 地图进行编码，并提出一个轻量且高效的融合模块，以 SD 地图先验增强鸟瞰图 (BEV) 特征。此外，GPS 信号固有的不准确性常导致 SD 地图与 BEV 空间之间存在轻微的空间错位 [12]。以往工作往往忽视这一现象或将其视为噪声，而我们引入了特征对齐机制来解决该问题。具体而言，对于栅格化，我们设计了一个特征变换网络，通过预测每个特征通道的缩放因子和偏置项动态调节特征；对于矢量化，我们采用了交叉注意力机制 [13]，自适应地关注对应特征，确保与 BEV 特征空间更好地对齐。

Moreover, existing approaches overlook the importance of topological road structures in driving scenes. For example, intersections, including cross, merge, or diverge nodes, serve as critical topological attributes that signify changes in road networks. Such keypoints can be effectively identified from SD map priors, which provide valuable characteristics about road structures. To leverage this information, we introduce an auxiliary task focused on recognizing the distribution of road intersections derived from SD maps. This task enables BEV features to capture crucial road topology, thereby enhancing overall driving scene understanding.

此外，现有方法忽视了驾驶场景中道路拓扑结构的重要性。例如，交叉口，包括交叉、汇合或分岔节点，是标志道路网络变化的关键拓扑属性。这些关键点可通过 SD 地图先验有效识别，提供有关道路结构的宝贵特征。为利用这些信息，我们引入了一个辅助任务，专注于识别源自 SD 地图的道路交叉口分布。该任务使 BEV 特征能够捕捉关键的道路拓扑，从而提升整体驾驶场景理解能力。

In summary, the primary contributions of this letter are as follows: (1) We propose a novel hybrid fusion strategy for SD maps that combines both rasterized and vectorized representations, ensuring effective alignment with BEV features for improved synergy. (2) We introduce an auxiliary Intersection-aware KeyPoint Detection (IKPD) task conditioned on the SD map prior, further enhancing scene understanding capabilities. (3) Extensive experiments on the large-scale OpenLane-V2 dataset demonstrate that our SD map-enhanced framework, termed SEPT, significantly improves both scene perception and topology reasoning performance.

综上所述，本信的主要贡献包括:(1) 提出了一种结合栅格化与矢量化表示的 SD 地图新型混合融合策略，确保与 BEV 特征的有效对齐以提升协同效果；(2) 引入了基于 SD 地图先验的辅助交叉口感知关键点检测 (IKPD) 任务，进一步增强场景理解能力；(3) 在大规模 OpenLane-V2 数据集上的大量实验表明，我们的 SD 地图增强框架 SEPT 显著提升了场景感知和拓扑推理性能。

II. RELATED WORK

II. 相关工作

A. Online Scene Perception

A. 在线场景感知

Online HD map construction relies on the accurate perception of scene elements. Pioneering efforts have focused on laneline detection [14], [15] to capture road geometry, or centerline perception [2], [16] to recognize lane connectivity. Given the intertwined nature of these two representations, a comprehensive mapping format, lane segment [6], has been proposed to seamlessly integrate both geometric 3D lanelines and topological 3D lane centerlines, along with areas defined by road boundaries and pedestrian crossings. Additionally, traffic element recognition has also been extensively explored in the literature [17], [18] for driving scene understanding, including the detection of traffic lights, road signs, and their associated semantic attributes. Despite advances in detecting these map elements, current online scene perception systems still struggle with occlusions and long-range scenarios. To address these limitations, our work leverages SD map priors, serving as essential complementary prompts with the potential to improve performance in these challenging conditions.

在线高清地图构建依赖于对场景元素的准确感知。早期工作主要聚焦于车道线检测 [14]，[15] 以捕捉道路几何，或中心线感知 [2]，[16] 以识别车道连通性。鉴于这两种表示的紧密关联，提出了综合映射格式——车道段 [6]，无缝整合几何三维车道线和拓扑三维车道中心线，以及由道路边界和人行横道定义的区域。此外，文献中也广泛探讨了交通元素识别 [17]，[18]，包括交通信号灯、道路标志及其相关语义属性的检测，以促进驾驶场景理解。尽管在检测这些地图元素方面取得进展，当前在线场景感知系统仍面临遮挡和远距离场景的挑战。为克服这些限制，我们的工作利用 SD 地图先验，作为重要的补充提示，有望提升在这些复杂条件下的性能。

B. Scene Topology Reasoning

B. 场景拓扑推理

Scene topology information is significant for downstream trajectory prediction [19] and behavior planning [20] tasks, as it provides the topological relationships among lanes and between lanes and traffic elements. Nevertheless, research on topology reasoning has been limited until the emergence of the OpenLane-V2 benchmark [5], which utilizes adjacency matrices to characterize topological connectivity. Most existing methods rely on Multi-Layer Perceptrons (MLPs) [21] or Graph Neural Networks (GNNs) [2] to learn these connection relationships, or incorporate spatial position encoding [22] to enhance reasoning capabilities. These methods, however, are prone to disruption by endpoint shift issues. To address this, the calculation of geometric distance and semantic similarity [23] has been proposed to mitigate such effects. Moreover, since SD maps inherently contain the topological structure of driving scenes, recent works [24] have explored leveraging this prior knowledge to further improve topology reasoning.

场景拓扑信息对于下游轨迹预测 [19] 和行为规划 [20] 任务至关重要，因为它提供了车道之间及车道与交通元素之间的拓扑关系。然而，拓扑推理的研究较为有限，直到 OpenLane-V2 基准 [5] 的出现，该基准利用邻接矩阵表征拓扑连通性。大多数现有方法依赖多层感知机 (MLP)[21] 或图神经网络 (GNN)[2] 学习这些连接关系，或结合空间位置编码 [22] 以增强推理能力。但这些方法易受端点偏移问题干扰。为此，提出了几何距离和语义相似度计算 [23] 以缓解该影响。此外，由于 SD 地图本质上包含驾驶场景的拓扑结构，近期工作 [24] 探索利用该先验知识进一步提升拓扑推理。

C.SD Map Prior for Autonomous Driving

C. 自动驾驶的 SD 地图先验

SD maps, such as Google Maps, are widely used for urban navigation and have recently garnered increasing attention in autonomous driving tasks. Previous studies have primarily concentrated on leveraging SD map priors to enhance online map construction, particularly in long-range scenarios [25]. These methods typically involve rasterizing SD maps [7] and employing Convolutional Neural Networks (CNNs) to extract features. However, the intrinsic weak alignment between SD maps and BEV features remains a challenge [1], leading to the adoption of attention mechanisms [8]. Recent advances in topology reasoning have also incorporated SD maps by vectorizing them into polylines and using Transformer [10] or GNN [9] architectures to improve online lane topology understanding. To fully exploit both representations, a concurrent approach [11] combines these two distinct streams; however, its fusion strategy remains overly simplistic, limiting effectiveness. Considering the existing constraints in SD map utilization, our work further explores their potential by developing a powerful hybrid fusion module and introducing an auxiliary intersection-aware keypoint forecasting task.

SD 地图，如谷歌地图，广泛应用于城市导航，近年来在自动驾驶任务中受到越来越多的关注。以往研究主要集中于利用 SD 地图先验来提升在线地图构建，尤其是在远距离场景中 [25]。这些方法通常涉及将 SD 地图栅格化 [7]，并采用卷积神经网络 (CNN) 提取特征。然而，SD 地图与鸟瞰视图 (BEV) 特征之间的内在弱对齐仍然是一个挑战 [1]，因此引入了注意力机制 [8]。最近在拓扑推理方面的进展也通过将 SD 地图矢量化为折线，并使用 Transformer[10] 或图神经网络 (GNN)[9] 架构来提升在线车道拓扑理解。为了充分利用这两种表示形式，一种并行方法 [11] 结合了这两条不同的路径；但其融合策略过于简单，限制了效果。鉴于现有 SD 地图利用的局限性，我们的工作通过开发强大的混合融合模块并引入辅助的路口感知关键点预测任务，进一步挖掘其潜力。

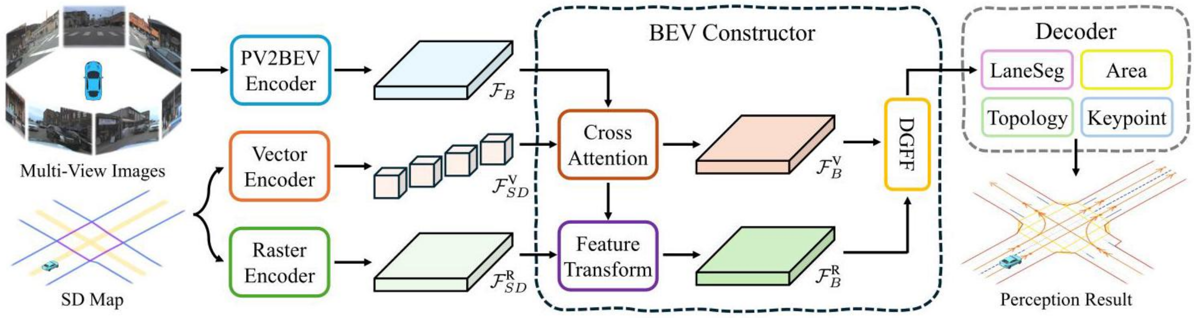

Fig. 2. Overview of the SEPT architecture, demonstrating how it enhances the existing perception and reasoning model for online scene understanding through the integration of the SD map prior.

图 2. SEPT 架构概览，展示了如何通过整合 SD 地图先验，增强现有的感知与推理模型，实现在线场景理解。

III. METHODOLOGY

三、方法论

A. Task Statement

A. 任务描述

The online driving scene understanding task involves both scene perception and topology reasoning, using multi-view images and the corresponding SD map priors as inputs. Scene perception includes detecting lane segments, drivable areas, and traffic elements. To be specific, lane segments comprise directed lane centerlines, left and right lane boundaries, and their associated line types (e.g., non-visible, solid, dashed). Drivable areas are represented by undirected curves or closed polygons corresponding to road boundaries and pedestrian crossings. Traffic elements encompass traffic lights and road signs visible in the front view, together with their relevant attributes. For topology reasoning, the goal is to infer the topological relationships among lane segments and between lane segments and traffic elements. This topological information is typically modeled as a lane graph, where nodes represent lane segments or traffic elements, and edges signify connectivity relationships. An adjacency matrix is employed to characterize the lane graph.

在线驾驶场景理解任务包括场景感知和拓扑推理，输入为多视角图像及对应的 SD 地图先验。场景感知涵盖车道段、可行驶区域及交通元素的检测。具体而言，车道段包括有向的车道中心线、左右车道边界及其对应的线型 (如不可见、实线、虚线)。可行驶区域由无向曲线或闭合多边形表示，对应道路边界和人行横道。交通元素包括前视图中可见的交通信号灯和道路标志及其相关属性。拓扑推理的目标是推断车道段之间以及车道段与交通元素之间的拓扑关系。该拓扑信息通常建模为车道图，节点代表车道段或交通元素，边表示连接关系。邻接矩阵用于描述车道图。

B. Framework Overview

B. 框架概述

The overall pipeline of our SEPT framework is illustrated in Fig. 2, which improves the baseline model by incorporating SD map priors. Specifically, given multi-view images, the PV2BEV encoder first extracts visual information via the image backbone and then transforms the Perspective-View (PV) features into the BEV feature, denoted as $\mathcal{F}_B$, by view transformation. Additionally, the SD map prior is encoded in two distinct formats: rasterized features $\mathcal{F}_{SD}^{\mathrm{R}}$ and vectorized features $\mathcal{F}_{SD}^{\mathrm{V}}$, through a hybrid SD map encoding approach. These two representations are then leveraged to augment the BEV feature through a Feature Transformation (FT) module and a cross-attention network, respectively, producing the enhanced BEV features $\mathcal{F}_B^{\mathrm{R}}$ and $\mathcal{F}_B^{\mathrm{V}}$. A lightweight yet effective Dual Gated Feature Fusion (DGFF) module is employed to fuse these two augmented features, generating the final enhanced BEV feature $\mathcal{F}_B^{\mathrm{SD}}$. This feature is consequently decoded to address various subtasks by different heads, such as the lane segment head, area head, topology head, etc. Notably, we also introduce an additional keypoint head for an auxiliary task, which detects road intersections from SD maps, further enhancing scene understanding capabilities.

我们 SEPT 框架的整体流程如图 2 所示，通过引入 SD 地图先验提升基线模型。具体地，给定多视角图像，PV2BEV 编码器首先通过图像主干网络提取视觉信息，然后通过视角变换将透视视图 (PV) 特征转换为鸟瞰视图 (BEV) 特征，记为 $\mathcal{F}_B$ 。此外，SD 地图先验通过混合 SD 地图编码方法被编码为两种不同格式: 栅格化特征 $\mathcal{F}_{SD}^{\mathrm{R}}$ 和矢量化特征 $\mathcal{F}_{SD}^{\mathrm{V}}$ 。这两种表示随后分别通过特征变换 (FT) 模块和交叉注意力网络增强 BEV 特征，生成增强的 BEV 特征 $\mathcal{F}_B^{\mathrm{R}}$ 和 $\mathcal{F}_B^{\mathrm{V}}$ 。采用轻量且高效的双门控特征融合 (DGFF) 模块融合这两种增强特征，生成最终的增强 BEV 特征 $\mathcal{F}_B^{\mathrm{SD}}$ 。该特征随后由不同的头部解码以完成各子任务，如车道段头、区域头、拓扑头等。值得注意的是，我们还引入了一个额外的关键点头作为辅助任务，用于从 SD 地图检测路口，进一步提升场景理解能力。

C. Hybrid SD Map Encoding and Fusion

C. 混合 SD 地图编码与融合

To fully leverage SD map priors, we introduce a hybrid encoding approach, utilizing both rasterized and vectorized formats. These two representations are incorporated to enhance the BEV feature while ensuring implicit alignment between them. In addition, we design an efficient and effective fusion strategy to seamlessly integrate these features, thereby improving overall performance. Herein, let the BEV feature be represented as $\mathcal{F}_B \in \mathbb{R}^{H\times W\times C}$, where H and W correspond to the spatial dimensions of the BEV perception range, and C denotes the feature dimension.

为充分利用 SD 地图先验，我们引入了一种混合编码方法，结合栅格化和矢量化两种格式。这两种表示被用来增强 BEV 特征，同时确保它们之间的隐式对齐。此外，我们设计了一种高效且有效的融合策略，实现这些特征的无缝整合，从而提升整体性能。这里，设 BEV 特征表示为 $\mathcal{F}_B \in \mathbb{R}^{H\times W\times C}$ ，其中 H 和 W 对应 BEV 感知范围的空间维度，C 表示特征维度。

1) Vectorized SD Map Encoding: Given raw polylines of SD maps, we begin by uniformly resampling these sequences to obtain M segments. For each segment, we further evenly sample a fixed number of points. Fol-

lowing the structure of the classical vectorized method, SMERF [10], we then vectorize the SD map and extract the initial vectorized feature $\mathcal{F}_{SD}^{\mathrm{V}} \in \mathbb{R}^{M\times C}$ using a Transformer-based encoder model. In this paradigm, spatial misalignment between the SD map tokens and the BEV space can be mitigated through a multihead cross-attention mechanism. Here, the BEV feature acts as query tokens, while the SD map tokens serve as keys and values. This enables the BEV queries to adaptively aggregate relevant SD map tokens conditioned on a learnable attention distribution. As a result, we obtain implicitly aligned BEV features $\mathcal{F}_{B}^{\mathrm{V}} \in \mathbb{R}^{H\times W\times C}$, complemented by the vectorized SD map priors.

1) 向量化 SD 地图编码: 给定原始的 SD 地图折线，我们首先对这些序列进行均匀重采样以获得 M 段。对于每个段，我们进一步均匀采样固定数量的点。遵循经典向量化方法 SMERF [10] 的结构，随后我们对 SD 地图进行向量化，并使用基于 Transformer 的编码器模型提取初始向量化特征 $\mathcal{F}_{SD}^{\mathrm{V}} \in \mathbb{R}^{M\times C}$ 。在此范式中，SD 地图标记与 BEV 空间之间的空间错位可以通过多头交叉注意力机制得到缓解。这里，BEV 特征作为查询标记，而 SD 地图标记作为键和值。这使得 BEV 查询能够根据可学习的注意力分布自适应地聚合相关的 SD 地图标记。因此，我们获得了隐式对齐的 BEV 特征 $\mathcal{F}_{B}^{\mathrm{V}} \in \mathbb{R}^{H\times W\times C}$ ，并辅以向量化的 SD 地图先验。

2) Rasterized SD Map Encoding: We first rasterize the SD map into an $H \times W$ canvas with a binary representation, where each grid cell is assigned a value of 1 if occupied by a polyline, and 0 otherwise. Different road types, such as crosswalks and sidewalks, are encoded as separate channels. The original SD map features are then extracted using CNNs, yielding the rasterized feature $\mathcal{F}_{SD}^{\mathrm{R}} \in \mathbb{R}^{H\times W\times C}$. Note that this feature may be weakly misaligned with the BEV space. To address this, motivated by the T-Net in PointNet [26], we introduce a Feature Transformation (FT) module to align $\mathcal{F}_{SD}^{\mathrm{R}}$ with $\mathcal{F}_{B}^{\mathrm{V}}$ at the feature level. Specifically, we first project both features along the channel dimension and compute their feature difference $\mathcal{F}_{\triangle} \in \mathbb{R}^{H\times W\times C}$, which represents a form of calibration error. We then apply a max-pooling operation on $\mathcal{F}_{\triangle}$ to obtain the global context vector $\mathcal{F}_{\triangle}^{\text{Global}} \in \mathbb{R}^{C}$. Next, we leverage Feature-wise Linear Modulation (FiLM) [27] to predict the scaling factor $\gamma \in \mathbb{R}^{C}$ and the bias term $\beta \in \mathbb{R}^{C}$ for each feature channel. Finally, these transformation parameters are applied to the rasterized feature $\mathcal{F}_{SD}^{\mathrm{R}}$, resulting in the enhanced BEV feature $\mathcal{F}_{B}^{\mathrm{R}} \in \mathbb{R}^{H\times W\times C}$, with implicit spatial alignment, as follows:

2) 光栅化 SD 地图编码: 我们首先将 SD 地图光栅化为一个 $H \times W$ 画布，采用二值表示，其中每个网格单元若被折线占据则赋值为 1，否则为 0。不同的道路类型，如人行横道和人行道，被编码为独立的通道。随后使用卷积神经网络 (CNN) 提取原始 SD 地图特征，得到光栅化特征 $\mathcal{F}_{SD}^{\mathrm{R}} \in \mathbb{R}^{H\times W\times C}$ 。注意，该特征可能与 BEV 空间存在弱错位。为解决此问题，受 PointNet [26] 中 T-Net 的启发，我们引入特征变换 (FT) 模块，在特征层面对齐 $\mathcal{F}_{SD}^{\mathrm{R}}$ 与 $\mathcal{F}_{B}^{\mathrm{V}}$ 。具体地，我们首先沿通道维度投影两者并计算它们的特征差异 $\mathcal{F}_{\triangle} \in \mathbb{R}^{H\times W\times C}$ ，该差异代表一种校准误差。然后对 $\mathcal{F}_{\triangle}$ 进行最大池化操作以获得全局上下文向量 $\mathcal{F}_{\triangle}^{\text{Global}} \in \mathbb{R}^{C}$ 。接着，我们利用特征线性调制 (FiLM)[27] 预测每个特征通道的缩放因子 $\gamma \in \mathbb{R}^{C}$ 和偏置项 $\beta \in \mathbb{R}^{C}$ 。最后，将这些变换参数应用于光栅化特征 $\mathcal{F}_{SD}^{\mathrm{R}}$ ，得到隐式空间对齐的增强 BEV 特征 $\mathcal{F}_{B}^{\mathrm{R}} \in \mathbb{R}^{H\times W\times C}$ ，具体如下:

$$\mathcal{F}_{B}^{\mathrm{R}} = \gamma \odot \mathcal{F}_{SD}^{\mathrm{R}} + \beta, \tag{1}$$

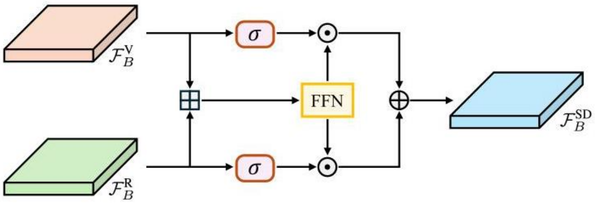
$\mathcal{F}_{B}^{\mathrm{V}}$
σ
FFN
σ
$\mathcal{F}_{B}^{\mathrm{R}}$
$\mathcal{F}_{B}^{\mathrm{SD}}$

Fig. 3. The hybrid feature fusion process of the DGFF module.

图 3. DGFF 模块的混合特征融合过程。

where $\odot$ denotes the Hadamard (element-wise) product, and all operations follow the broadcasting mechanism.

其中 $\odot$ 表示 Hadamard(元素级) 乘积，所有操作均遵循广播机制。

3) Dual Gated Feature Fusion: After obtaining the two BEV features augmented by rasterized and vectorized SD map features, it is essential to design an effective fusion strategy to combine these distinct features, as the characteristics of the two branches may differ significantly. To this end, we propose a lightweight yet powerful fusion network called the Dual Gated Feature Fusion (DGFF) module, which leverages the gated attention mechanism to aggregate the dual-branch features. As depicted in Fig. 3, the two features $\mathcal{F}_B^{\mathrm{R}}$ and $\mathcal{F}_B^{\mathrm{V}}$ are first concatenated along the feature dimension and passed through a feed-forward network to produce a fused feature $\mathcal{F}_B^{\mathrm{R}\boxplus\mathrm{V}} \in \mathbb{R}^{H\times W\times C}$, as follows:

3) 双门控特征融合: 在获得由光栅化和向量化 SD 地图特征增强的两种 BEV 特征后，设计有效的融合策略以结合这两种不同特征至关重要，因为这两条分支的特性可能存在显著差异。为此，我们提出了一种轻量且强大的融合网络——双门控特征融合 (DGFF) 模块，该模块利用门控注意力机制聚合双分支特征。如图 3 所示，两个特征 $\mathcal{F}_B^{\mathrm{R}}$ 和 $\mathcal{F}_B^{\mathrm{V}}$ 首先在特征维度上拼接，并通过前馈网络生成融合特征 $\mathcal{F}_B^{\mathrm{R}\boxplus\mathrm{V}} \in \mathbb{R}^{H\times W\times C}$ ，具体如下:

$$\mathcal{F}_B^{\mathrm{R}\boxplus\mathrm{V}} = \mathrm{FFN}\left(\mathcal{F}_B^{\mathrm{R}} \boxplus \mathcal{F}_B^{\mathrm{V}}\right), \tag{2}$$

where $\boxplus$ denotes concatenation along the feature dimension, and $\mathrm{FFN}(\cdot)$ represents the feed-forward network. Next, an element-wise gating mechanism is performed on each input stream using the sigmoid function, enabling the model to adaptively weight the contributions of rasterized and vectorized features. This is expressed as:

其中 $\boxplus$ 表示沿特征维度的拼接，$\mathrm{FFN}(\cdot)$ 代表前馈网络。接下来，使用 sigmoid 函数对每个输入流执行逐元素门控机制，使模型能够自适应地权衡栅格化和矢量化特征的贡献。表达式如下:

$$w_{\mathrm{R}} = \sigma\left(\mathcal{F}_B^{\mathrm{R}}\right),\ w_{\mathrm{V}} = \sigma\left(\mathcal{F}_B^{\mathrm{V}}\right), \tag{3}$$

where $\sigma(\cdot)$ is the element-wise sigmoid function, producing attention weights for each input stream. Although simpler, the gating mechanism introduces nonlinearity and adaptability, allowing the model to capture richer feature interactions without increasing the number of learnable parameters. This strikes a balance between representational capacity and efficiency. Finally, two parallel projection networks refine each gated feature before merging them via a weighted averaging operation, generating the final enhanced BEV feature $\mathcal{F}_B^{\mathrm{SD}} \in \mathbb{R}^{H\times W\times C}$. This process can be formulated by the following expression:

其中 $\sigma(\cdot)$ 是逐元素 sigmoid 函数，为每个输入流生成注意力权重。尽管结构更简单，门控机制引入了非线性和适应性，使模型能够捕捉更丰富的特征交互，而无需增加可学习参数数量。这在表示能力和效率之间取得了平衡。最后，两个并行投影网络对每个门控特征进行细化，然后通过加权平均操作合并，生成最终增强的 BEV 特征 $\mathcal{F}_B^{\mathrm{SD}} \in \mathbb{R}^{H\times W\times C}$ 。该过程可用以下表达式表示:

$$\mathcal{F}_B^{\mathrm{SD}} = \mu \cdot \mathrm{Proj}_{\mathrm{R}} \left(w_{\mathrm{R}} \odot \mathcal{F}_B^{\mathrm{R} \boxplus \mathrm{V}} \right) + \nu \cdot \mathrm{Proj}_{\mathrm{V}} \left(w_{\mathrm{V}} \odot \mathcal{F}_B^{\mathrm{R} \boxplus \mathrm{V}} \right), \tag{4}$$

where $\mathrm{Proj}_{\mathrm{R}}$ and $\mathrm{Proj}_{\mathrm{V}}$ are the parallel projection networks. μ and ν are hyperparameters for balancing each term.

其中 $\mathrm{Proj}_{\mathrm{R}}$ 和 $\mathrm{Proj}_{\mathrm{V}}$ 是并行投影网络。μ 和 ν 是用于平衡各项的超参数。

Overall, with the support of the DGFF module, our hybrid SD map encoding and fusion strategy can adaptively fuse heterogeneous feature representations, empowering the model to emphasize the most informative components from each branch. This substantially boosts the representation capabilities of both rasterized and vectorized SD map priors, while maintaining efficiency and expressiveness.

总体而言，在 DGFF 模块的支持下，我们的混合 SD 地图编码与融合策略能够自适应融合异构特征表示，使模型能够强调每个分支中最具信息量的部分。这显著提升了栅格化和矢量化 SD 地图先验的表示能力，同时保持了效率和表现力。

D. Intersection-Aware Keypoint Detection

D. 路口感知关键点检测

To further enhance the BEV feature representation and improve understanding of road topology and geometry, we introduce an Intersection-Aware Keypoint Detection (IKPD) task. This auxiliary task helps the model capture road interaction patterns by detecting the road interaction distribution.

为了进一步增强 BEV 特征表示并提升对道路拓扑和几何结构的理解，我们引入了路口感知关键点检测 (IKPD) 任务。该辅助任务通过检测道路交互分布，帮助模型捕捉道路交互模式。

1) Intersection Generation: The first step in implementing the IKPD task involves identifying the intersection points of roads from SD maps, which will serve as the ground truth for supervision. Since the SD map prior provides the essential skeleton of road networks, intersection locations (e.g., merging, diverging, and crossing points) can be easily extracted. However, intersection points are typically sparsely distributed across the scene, and directly using these points as supervision can lead to class imbalance during training. Additionally, due to intrinsic positional biases relative to the ground-truth HD maps in certain scenarios, the intersection points derived from the SD map may not perfectly align with the finer details of the road structure. To mitigate this issue, we represent the keypoints as Gaussian distributions, similar to the approach used in confidence-based keypoint detection [28], [29]. Specifically, for each scene, we construct a heatmap $\mathcal{H} \in \mathbb{R}^{H \times W \times 1}$ to model the ground-truth distribution of road intersections. Each intersection is represented as a Gaussian distribution centered at its location, with a certain radius reflecting the spatial uncertainty.

1) 路口生成: 实现 IKPD 任务的第一步是从 SD 地图中识别道路的路口点，这些点将作为监督的真实标签。由于 SD 地图先验提供了道路网络的基本骨架，路口位置 (如汇合、分叉和交叉点) 可以轻松提取。然而，路口点通常在场景中分布稀疏，直接使用这些点作为监督会导致训练中的类别不平衡。此外，由于某些场景中相对于真实 HD 地图存在固有的位置信偏差，SD 地图导出的路口点可能无法与道路结构的细节完全对齐。为缓解此问题，我们将关键点表示为高斯分布，类似于基于置信度的关键点检测方法 [28]，[29]。具体而言，对于每个场景，我们构建一个热图 $\mathcal{H} \in \mathbb{R}^{H\times W\times 1}$ 来模拟道路路口的真实分布。每个路口以其位置为中心，采用一定半径的高斯分布，反映空间不确定性。

2) IKPD Head: Given the augmented BEV feature $\mathcal{F}_B^{\text{SD}}$, we aim to design a lightweight network capable of effectively decoding the road intersection heatmap, thereby enriching the BEV feature with crucial geometric and topological information about the road structure. The IKPD head follows an efficient design paradigm that emphasizes both local feature extraction and global context reasoning. Specifically, the BEV feature is first passed through a series of CNN blocks with residual connections. Each residual block comprises depthwise separable convolutions (i.e., a depthwise convolution followed by a pointwise convolution) [30], which decouple spatial and channel-wise operations for computational efficiency. Dilated convolutions are also incorporated for capturing broader spatial context information. After each convolution, the output feature is refined with the Squeeze-and-Excitation (SE) [31] attention, which recalibrates the channel-wise features by computing global statistics and adaptively weighting the importance of each channel. This allows the IKPD head to prioritize the most relevant features for keypoint detection, improving its ability to focus on critical patterns. Consequently, the final output is produced through a 1×1 convolution followed by a sigmoid activation function, generating a heatmap that represents the distribution of road intersections.

2) IKPD 头: 给定增强的 BEV 特征 $\mathcal{F}_B^{\text{SD}}$ ，我们旨在设计一个轻量级网络，有效解码道路路口热图，从而丰富 BEV 特征中关于道路结构的关键几何和拓扑信息。IKPD 头遵循高效设计范式，强调局部特征提取与全局上下文推理。具体来说，BEV 特征首先通过一系列带残差连接的卷积神经网络 (CNN) 模块。每个残差模块包含深度可分离卷积 (即先深度卷积后逐点卷积)[30]，该结构将空间和通道操作解耦以提高计算效率。还引入了空洞卷积以捕获更广泛的空间上下文信息。每次卷积后，输出特征通过 Squeeze-and-Excitation(SE)[31] 注意力机制进行细化，该机制通过计算全局统计量并自适应加权各通道的重要性，重新校准通道特征。这使 IKPD 头能够优先关注关键特征，提升关键点检测的能力。最终输出通过 1×1 卷积和 sigmoid 激活函数生成，形成表示道路路口分布的热图。

E. Training Objectives

E. 训练目标

Following the baseline approaches [2], [6], the supervision is applied to each head to optimize distinct training objectives, including detection losses for lane segments, areas, and traffic elements, denoted as $\mathcal{L}_{\text{DET}}$, and topology reasoning losses, denoted as $\mathcal{L}_{\text{TOP}}$. Our proposed SEPT framework does not modify the baseline loss functions but introduces an additional loss term for the auxiliary IKPD head, denoted as $\mathcal{L}_{\text{IKPD}}$. Given the road intersection distribution is sparse and imbalanced, we employ focal loss [28] to supervise the keypoint heatmap training. The overall loss $\mathcal{L}$ for SEPT is formulated as:

继基线方法 [2]，[6] 之后，监督信号被应用于每个头部以优化不同的训练目标，包括车道段、区域和交通元素的检测损失，记为 $\mathcal{L}_{\text{DET}}$ ，以及拓扑推理损失，记为 $\mathcal{L}_{\text{TOP}}$ 。我们提出的 SEPT 框架并未修改基线损失函数，而是为辅助 IKPD 头引入了额外的损失项，记为 $\mathcal{L}_{\text{IKPD}}$ 。鉴于路口分布稀疏且不平衡，我们采用焦点损失 (focal loss)[28] 来监督关键点热图的训练。SEPT 的总体损失 $\mathcal{L}$ 定义如下:

$$\mathcal{L} = \mathcal{L}_{\text{DET}} + \mathcal{L}_{\text{TOP}} + \mathcal{L}_{\text{IKPD}}. \tag{5}$$

TABLE I

Quantitative results on the OLV2 validation split, benchmarked using `OLS` . All metrics follow the higher-the-better criterion. The official ranking metric is shaded in gray, and the best results are indicated in bold. A ”-” denotes the absence of relevant DATA.

在 OLV2 验证集上的定量结果，基于 `OLS` 进行基准测试。所有指标均遵循越高越好的原则。官方排名指标以灰色阴影标出，最佳结果以粗体显示。“-” 表示缺少相关数据。

Method	DET_l ↑	DET_t ↑	v1.0			v1.1			Params
			TOP_{ll} ↑	TOP_{lt} ↑	OLS↑	TOP_{ll} ↑	TOP_{lt} ↑	OLS↑	
TopoNet [2]	28.6	48.6	4.1	20.3	35.6	10.9	23.8	39.8	62.6M
w/ OLV2 [9]	27.9	48.1	5.1	20.9	36.1	-	-	-	75.9M
w/ OSMG [9]	30.0	47.6	5.4	21.3	36.7	-	-	-	64.6M
w/ OSMR [9]	30.6	44.6	7.7	22.9	37.7	-	-	-	75.9M
w/ SMERF [10]	33.4	48.6	7.5	23.4	39.4	15.4	25.4	42.9	65.8M
w/ SEPT (Ours)	34.2 (+5.6)	49.8 (+1.2)	8.3 (+4.2)	23.8 (+3.5)	40.4 (+4.8)	19.5 (+8.6)	27.5 (+3.7)	45.2 (+5.4)	70.4M
TopoLogic [23]	29.2	46.5	18.0	20.6	40.9	23.6	24.2	43.4	61.8M
w/ SMERF [10]	31.0	48.7	21.2	22.4	43.3	26.9	26.2	45.7	65.1M
w/ SEPT (Ours)	34.3 (+5.1)	48.9 (+2.4)	25.1 (+7.1)	25.1 (+4.5)	45.8 (+4.9)	31.2 (+7.6)	29.7 (+5.5)	48.4 (+5.0)	69.6M

方法	DET_l ↑	DET_t ↑	v1.0			v1.1			参数
			TOP_{ll} ↑	TOP_{lt} ↑	OLS↑	TOP_{ll} ↑	TOP_{lt} ↑	OLS↑	
TopoNet [2]	28.6	48.6	4.1	20.3	35.6	10.9	23.8	39.8	62.6M
使用 OLV2 [9]	27.9	48.1	5.1	20.9	36.1	-	-	-	75.9M
使用 OSMG [9]	30.0	47.6	5.4	21.3	36.7	-	-	-	64.6M
使用 OSMR [9]	30.6	44.6	7.7	22.9	37.7	-	-	-	75.9M
使用 SMERF [10]	33.4	48.6	7.5	23.4	39.4	15.4	25.4	42.9	65.8M
使用 SEPT(本方法)	34.2 (+5.6)	49.8 (+1.2)	8.3 (+4.2)	23.8 (+3.5)	40.4 (+4.8)	19.5 (+8.6)	27.5 (+3.7)	45.2 (+5.4)	70.4M
TopoLogic [23]	29.2	46.5	18.0	20.6	40.9	23.6	24.2	43.4	61.8M
使用 SMERF [10]	31.0	48.7	21.2	22.4	43.3	26.9	26.2	45.7	65.1M
使用 SEPT(本方法)	34.3 (+5.1)	48.9 (+2.4)	25.1 (+7.1)	25.1 (+4.5)	45.8 (+4.9)	31.2 (+7.6)	29.7 (+5.5)	48.4 (+5.0)	69.6M

IV. EXPERIMENTS AND RESULTS

四、实验与结果

A. Experiment Setups

A. 实验设置

1) Real-World Dataset: We train and evaluate our proposed approach on the large-scale OpenLane-V2 (OLV2) dataset [5], which, to the best of our knowledge, is currently the only benchmark for both scene perception and topology reasoning in autonomous driving. All experiments in our work are conducted on the primary subset of OLV2, subset_A, built upon the Argoverse 2 [32] dataset with additional annotations for lane segments, traffic elements, and lane topology, etc. The subset A comprises 1 k scenes collected from six cities, with 2Hz multi-view images and optional SD map information (including three categories: roads, crosswalks, and sidewalks) extracted from the OpenStreetMap (OSM) [33]. The training set consists of approximately 27 k frames, while the validation set contains around 4.8k frames.

1) 真实世界数据集: 我们在大规模 OpenLane-V2(OLV2) 数据集 [5] 上训练和评估所提出的方法。据我们所知, 该数据集目前是自动驾驶中场景感知和拓扑推理的唯一基准。我们所有的实验均在 OLV2 的主子集 subset_A 上进行，该子集基于 Argoverse 2 [32] 数据集构建，附加了车道段、交通元素和车道拓扑等标注。子集 A 包含从六个城市采集的 1 k 个场景，拥有 2Hz 多视角图像和可选的 SD 地图信息 (包括道路、人行横道和人行道三类)，这些信息来自 OpenStreetMap(OSM)[33]。训练集约有 27 k 帧，验证集约有 4.8k 帧。

2) Evaluation Metrics: We evaluate the performance of perception and reasoning using the official metrics provided by OLV2 [5]. There are two primary benchmark categories, each with distinct evaluation metrics: OLV2 Score (OLS) and OLV2 UniScore (OLUS). Both scores are averages derived from various metrics across different subtasks. The main distinction between them is that OLS focuses exclusively on lane centerline perception, while OLUS emphasizes lane segment perception. Specifically, OLS includes four sub-metrics: DET_l , $\mathrm{DET}_t, \mathrm{TOP}_{ll}$, and TOP_{lt} . DET_l measures the mean average precision (mAP) for lane centerline detection, based on the Fréchet distance with match thresholds of 1.0, 2.0, and 3.0. DET_t represents the mAP for traffic element recognition, conditioned on the average Intersection over Union (IoU) with a match threshold of 0.75 across various traffic attributes. TOP_{ll} and TOP_{lt} measure mAP for topology among lane centerlines and between lane centerlines and traffic elements, using the adjacency matrix. Notably, there are two versions for calculating TOP scores: V1.0 and V1.1, with the V1.0 calculation containing a potential loophole issue [22]. The OLS score is computed as the average of these four metrics, given by:

2) 评估指标: 我们使用 OLV2 官方提供的指标 [5] 评估感知和推理性能。主要有两类基准指标，分别是 OLV2 分数 (OLS) 和 OLV2 统一分数 (OLUS)，两者均为不同子任务多项指标的平均值。主要区别在于 OLS 专注于车道中心线感知，而 OLUS 强调车道段感知。具体而言，OLS 包含四个子指标: DET_l 、$\mathrm{DET}_t, \mathrm{TOP}_{ll}$ 和 TOP_{lt} 。DET_l 衡量基于 Fréchet 距离 (匹配阈值为 1.0、2.0 和 3.0) 的车道中心线检测的平均精度均值 (mAP)。DET_t 表示基于平均交并比 (IoU) 且匹配阈值为 0.75 的多种交通属性的交通元素识别 mAP。TOP_{ll} 和 TOP_{lt} 分别衡量车道中心线之间及车道中心线与交通元素之间的拓扑 mAP，采用邻接矩阵表示。值得注意的是，TOP 分数的计算有两个版本:V1.0 和 V1.1，其中 V1.0 版本存在潜在漏洞 [22]。OLS 分数为这四个指标的平均值，计算公式为:

$$\mathrm{OLS} = \frac{1}{4}\left[\mathrm{DET}_l + \mathrm{DET}_t + f\left(\mathrm{TOP}_{ll}\right) + f\left(\mathrm{TOP}_{lt}\right)\right], \tag{6}$$

where f represents the square root function.

其中 f 表示平方根函数。

In contrast, OLUS encompasses five sub-metrics: DET_{ls} , $\mathrm{DET}_a, \mathrm{DET}_{te}, \mathrm{TOP}_{lsls}$, and TOP_{lste} , covering detection for lane segments, areas, and traffic elements, as well as topology reasoning among lane segments and between lane segments and traffic elements. These metrics follow a similar calculation procedure to OLS, with the addition of DET$_a$, which is measured using Chamfer distance.

相比之下，OLUS 包含五个子指标: DET$_{ls}$ 、DET$_a$, DET$_{te}$, TOP$_{lsls}$ 和 TOP$_{lste}$ ，涵盖车道段、区域和交通元素的检测，以及车道段之间和车道段与交通元素之间的拓扑推理。这些指标的计算方法与 OLS 类似，额外包含使用 Chamfer 距离测量的 DET$_a$ 。

3) Implementation Details: We select two representative high-performance models as baselines: TopoNet [2] for OLS and LaneSegNet [6] for OLUS. To ensure a fair comparison, we retain the official implementations of both baseline models, incorporating only the modules designed specifically for the SD map prior into the codebase. All models employ the default ResNet-50 backbone, with BEV feature dimensions set to $H = 200$ and $W = 100$. We train our model on eight GPUs with a total batch size of 8 . The training configuration, including the learning rate and optimizer, remains consistent with baseline settings, and all models share the same hyperparameters.

3) 实现细节: 我们选择两个代表性的高性能模型作为基线:TopoNet [2] 用于 OLS，LaneSegNet [6] 用于 OLUS。为保证公平比较，我们保留两个基线模型的官方实现，仅将专门设计的 SD 地图先验模块整合进代码库。所有模型均采用默认的 ResNet-50 主干网络，BEV 特征维度设置为 $H = 200$ 和 $W = 100$ 。我们在八块 GPU 上训练模型，总批量大小为 8。训练配置 (包括学习率和优化器) 与基线保持一致，所有模型共享相同的超参数。

B. Comparison with State-of-the-Art

B. 与最先进方法的比较

We compare our SEPT framework with other state-of-the-art methods that incorporate SD maps as input on the OLV2 benchmark, using the OLS evaluation metric. TopoNet [2] serves as the baseline for centerline perception. TopoNet with OLV2, OSMG, and OSMR [9] represent approaches utilizing rasterized SD maps from OLV2, augmented with full OSM attributes (e.g., stop signs, speed limits), and graph-based SD maps with OSM augmentation, respectively. SMERF is a classical method that enhances lane-topology understanding with SD maps in a vectorized representation. We present results for both v1.0 and v1.1 metrics to provide a comprehensive comparison, as shown in the upper group of Tab. I. Since the v1.1 metric is a recent update, the results for TopoNet and SMERF on v1.1 have been reevaluated using their official checkpoints, while others are not available. Compared to the baseline, our method significantly improves performance across all subtasks by effectively integrating SD map priors without introducing excessive parameters. Specifically, we achieve a 4.8 OLS improvement for v1.0 and a 5.4 OLS improvement for v1.1 , with a notable 8.6TOP_{ll} increase in topology reasoning, outperforming other existing methods augmented with SD maps. Moreover, we also apply our SEPT framework to the recent model, TopoLogic [23], which is designed to enhance lane topology reasoning. As shown in the lower group of Tab. I, our SEPT consistently improves performance across all subtasks, achieving a 4.9 OLS gain for v1.0 and a 5.0 OLS gain for v1.1.

我们在 OLV2 基准上，使用 OLS 评估指标，将我们的 SEPT 框架与其他将 SD 地图作为输入的先进方法进行了比较。TopoNet [2] 作为中心线感知的基线方法。TopoNet 结合 OLV2、OSMG 和 OSMR [9] 分别代表利用 OLV2 栅格化 SD 地图、增强了完整 OSM 属性 (如停车标志、限速) 以及基于图的 SD 地图并进行 OSM 增强的方法。SMERF 是一种经典方法，利用矢量化表示的 SD 地图增强车道拓扑理解。我们展示了 v1.0 和 v1.1 两个版本的指标结果，以提供全面比较，如表 I 上半部分所示。由于 v1.1 指标是近期更新，TopoNet 和 SMERF 在 v1.1 上的结果已使用其官方检查点重新评估，其他方法则无此数据。与基线相比，我们的方法通过有效整合 SD 地图先验，在所有子任务上显著提升性能，且未引入过多参数。具体而言，我们在 v1.0 上实现了 4.8 的 OLS 提升，在 v1.1 上实现了 5.4 的 OLS 提升，拓扑推理方面有显著的 8.6TOP$_{ll}$ 提升，优于其他基于 SD 地图增强的现有方法。此外，我们还将 SEPT 框架应用于近期设计用于增强车道拓扑推理的模型 TopoLogic [23]。如表 I 下半部分所示，我们的 SEPT 在所有子任务上持续提升性能，v1.0 获得 4.9 的 OLS 增益，v1.1 获得 5.0 的 OLS 增益。

TABLE II

Quantitative results on the OLV2 validation split with map element buckets, benethmarked using OLUS.

在 OLV2 验证集上，基于地图元素分桶，使用 OLUS 进行基准测试的定量结果。

Method	Raster	Vector	IKPD	DET$_{ls}$ ↑	DET$_{a}$ ↑	DET $_{te}$ ↑	TOP$_{lsls}$ ↑	TOP $_{\text{lste}}$ ↑	OLUS ↑	Params
LaneSegNet [6]				30.9	20.0	36.7	25.6	20.8	36.7	61.8M
w/ Raster Only	✓			33.8 (+2.9)	28.1 (+8.1)	38.1 (+1.4)	27.5 (+1.9)	21.8 (+1.0)	39.9 (+3.2)	65.5M
w/ Vector Only		✓		35.3 (+4.4)	22.3 (+2.5)	39.2 (+2.3)	30.2 (+4.6)	22.6 (+1.8)	39.9 (+3.2)	65.8M
w/ Hybrid Fusion	✓	✓		35.8 (+4.9)	28.2 (+8.2)	39.6 (+2.9)	31.0 (+5.4)	22.7 (+1.9)	41.4 (+4.7)	69.8M
w/ SEPT	✓	✓	✓	38.4 (+7.5)	29.0 (+9.0)	40.0 (+3.3)	32.2 (+6.6)	23.8 (+3.0)	42.6 (+5.9)	70.4M

方法	栅格	矢量	IKPD	DET$_{ls}$ ↑	DET$_{a}$ ↑	DET $_{te}$ ↑	TOP$_{lsls}$ ↑	TOP $_{\text{lste}}$ ↑	OLUS ↑	参数
LaneSegNet [6]				30.9	20.0	36.7	25.6	20.8	36.7	61.8M
仅使用栅格	✓			33.8 (+2.9)	28.1 (+8.1)	38.1 (+1.4)	27.5 (+1.9)	21.8 (+1.0)	39.9 (+3.2)	65.5M
仅使用矢量		✓		35.3 (+4.4)	22.3 (+2.5)	39.2 (+2.3)	30.2 (+4.6)	22.6 (+1.8)	39.9 (+3.2)	65.8M
混合融合	✓	✓		35.8 (+4.9)	28.2 (+8.2)	39.6 (+2.9)	31.0 (+5.4)	22.7 (+1.9)	41.4 (+4.7)	69.8M
使用 SEPT	✓	✓	✓	38.4 (+7.5)	29.0 (+9.0)	40.0 (+3.3)	32.2 (+6.6)	23.8 (+3.0)	42.6 (+5.9)	70.4M

TABLE III

EFFECT OF FT MODULE ON RASTERIZED SD MAP ENCODING.

FT 模块对光栅化 SD 地图编码的影响。

FT	DET$_{ls}$	DET$_{a}$	DET$_{te}$	TOP$_{lsls}$	TOP $_{\text{lste}}$	OLUS
✘	32.2	24.5	36.7	26.7	21.2	38.2
✓	33.8	28.1	38.1	27.5	21.8	39.9

傅里叶变换	DET$_{ls}$	DET$_{a}$	DET$_{te}$	TOP$_{lsls}$	TOP $_{\text{lste}}$	奥卢斯
✘	32.2	24.5	36.7	26.7	21.2	38.2
✓	33.8	28.1	38.1	27.5	21.8	39.9

In addition, we further evaluate our approach using the latest and more challenging OLUS evaluation metric, which focuses on lane segment perception and provides a more comprehensive assessment of the map element bucket. We directly apply our SEPT framework to LaneSegNet [6], a leading method for driving scene topology. As shown in Tab. II, even without further adaptation, our framework significantly enhances the baseline performance across all five subtasks, resulting in an overall improvement of 5.9 OLUS. This underscores the effectiveness and generalizability of our proposed framework.

此外，我们进一步使用最新且更具挑战性的 OLUS 评估指标对我们的方法进行评估，该指标侧重于车道段感知，并提供对地图元素桶的更全面评估。我们直接将 SEPT 框架应用于 LaneSegNet [6]，这是一种领先的驾驶场景拓扑方法。如表 II 所示，即使没有进一步的适配，我们的框架也显著提升了基线模型在所有五个子任务上的表现，总体提升了 5.9 OLUS 分数。这凸显了我们所提框架的有效性和泛化能力。

C. Ablation Study

C. 消融研究

We conduct ablation studies to validate the effectiveness of each proposed component of SEPT as well as FT and DGFF modules using the OLV2 validation split, employing the OLUS evaluation metric for a comprehensive assessment. LaneSegNet [6] serves as the baseline model.

我们使用 OLV2 验证集划分，采用 OLUS 评估指标进行全面评估，开展消融研究以验证 SEPT 各个提出组件以及 FT 和 DGFF 模块的有效性。LaneSegNet [6] 作为基线模型。

1) Component Study of SEPT: We conduct an in-depth analysis of the contributions of each proposed component, as shown in Tab. II.

1) SEPT 组件研究: 我们对各个提出组件的贡献进行了深入分析，如表 II 所示。

Hybrid Fusion. Compared to the baseline, integrating the SD map prior, whether in rasterized or vectorized representation, improves performance across all metrics, with both formats yielding comparable results. Specifically, the rasterized SD map significantly enhances area detection by 8.1DET_a , as rasterization captures more spatial information. In contrast, the vectorized format improves the lane segment detection and topology reasoning between lane segments by 4.4DET_{ls} and $4.6\text{TOP}_{\text{lals}}$, respectively, as vectorization better preserves the geometry and topology of the road structure. This highlights that both representations have distinct advantages and should complement each other, underscoring the superiority of employing our hybrid fusion strategy.

混合融合。与基线相比，整合 SD 地图先验，无论是栅格化还是矢量化表示，都提升了所有指标的性能，且两种格式的结果相当。具体而言，栅格化 SD 地图显著提升了区域检测能力 8.1DET_a ，因为栅格化捕捉了更多的空间信息。相反，矢量化格式分别提升了车道段检测和车道段间拓扑推理能力 4.4DET_{ls} 和 $4.6\text{TOP}_{\text{lals}}$ ，因为矢量化更好地保留了道路结构的几何形状和拓扑关系。这表明两种表示各有优势，应相辅相成，凸显了采用我们混合融合策略的优越性。

TABLE IV
COMPARISON OF DIFFERENT FEATURE FUSION STRATEGIES.

不同特征融合策略的比较。

Fusion Strategy	DET$_{ls}$	DET$_{a}$	DET$_{te}$	TOP$_{lsls}$	TOP$_{\mathrm{lste}}$	OLUS
Addition	35.4	23.9	36.8	29.5	21.7	39.4
Concatenation	35.7	24.3	37.7	29.9	22.0	39.9
Cross-Attention	35.5	25.5	38.3	30.0	22.5	40.3
DGFF (Ours)	35.8	28.2	39.6	31.0	22.7	41.4

融合策略	DET$_{ls}$	DET$_{a}$	DET$_{te}$	TOP$_{lsls}$	TOP$_{\mathrm{lste}}$	OLUS
加法	35.4	23.9	36.8	29.5	21.7	39.4
拼接	35.7	24.3	37.7	29.9	22.0	39.9
交叉注意力	35.5	25.5	38.3	30.0	22.5	40.3
DGFF(我们的方法)	35.8	28.2	39.6	31.0	22.7	41.4

TABLE V

COMPARISON OF DIFFERENT FUSION WEIGHTS.

不同融合权重的比较。

Raster	Vector	DET$_{ls}$	DET$_{a}$	DET$_{te}$	TOP$_{lsls}$	TOP $_{lste}$	OLUS
0.2	0.8	35.5	27.1	39.2	30.1	22.4	40.8
0.5	0.5	35.8	28.2	39.6	31.0	22.7	41.4
0.8	0.2	35.2	27.5	38.9	30.1	22.3	40.7

光栅	矢量	DET$_{ls}$	DET$_{a}$	DET$_{te}$	TOP$_{lsls}$	TOP $_{lste}$	OLUS
0.2	0.8	35.5	27.1	39.2	30.1	22.4	40.8
0.5	0.5	35.8	28.2	39.6	31.0	22.7	41.4
0.8	0.2	35.2	27.5	38.9	30.1	22.3	40.7

IKPD. As shown in the last two rows of Tab. II, incorporating the auxiliary IKPD task enables our method to fully exploit the potential of the SD map prior, leading to further improvements across all subtasks in scene perception and topology reasoning.

IKPD。如表 II 最后两行所示，加入辅助 IKPD 任务使我们的方法能够充分利用 SD 地图先验的潜力，从而在场景感知和拓扑推理的所有子任务中均取得进一步提升。

2) Effect of the FT Module: We first investigate the impact of spatial alignment in the rasterized SD map encoding process. In the rasterized SD map-only configuration, we remove the proposed FT module from the pipeline. As shown in Tab. III, the model with the FT module consistently outperforms its counterpart across all metrics. This highlights the importance of aligning the SD map and BEV feature space, with feature-level alignment effectively mitigating associated challenges.

2) FT 模块的作用: 我们首先研究了栅格化 SD 地图编码过程中空间对齐的影响。在仅使用栅格化 SD 地图的配置中，我们从流程中移除了所提的 FT 模块。如表 III 所示，带有 FT 模块的模型在所有指标上均持续优于其对应版本。这凸显了对齐 SD 地图与 BEV 特征空间的重要性，特征级对齐有效缓解了相关挑战。

3) Effect of the DGFF Module: We further evaluate the fusion capability of the proposed DGFF module. Given the two BEV features $\mathcal{F}_B^{\mathrm{R}}$ and $\mathcal{F}_B^{\mathrm{V}}$, enhanced by rasterized and vectorized SD maps, respectively, we replace the DGFF with various fusion strategies, including element-wise addition, concatenation (followed by FFN), and cross-attention. As shown in Tab. IV, simple combinations of these features either hinder overall performance or fail to achieve a synergistic effect. While cross-attention improves performance, it introduces significant computational overhead. In contrast, our DGFF module is both more efficient and effective, utilizing a gated attention mechanism to integrate the two features and deliver superior performance. Additionally, we also explore the impact of different combination weights (i.e., μ and ν in Eq. 4) for the gated rasterized and vectorized features. From Tab. V, we observe that a balanced combination of the two features yields the best performance, with the rasterized term slightly outperforming in area detection and the vectorized term excelling in lane segment detection. These results are consistent with the findings discussed in Section IV-C1.

3) DGFF 模块的作用: 我们进一步评估了所提 DGFF 模块的融合能力。给定由栅格化和矢量化 SD 地图分别增强的两个 BEV 特征 $\mathcal{F}_B^{\mathrm{R}}$ 和 $\mathcal{F}_B^{\mathrm{V}}$，我们用多种融合策略替代 DGFF，包括逐元素相加、拼接 (后接前馈网络) 和交叉注意力。如表 IV 所示，这些特征的简单组合要么阻碍整体性能，要么未能实现协同效应。虽然交叉注意力提升了性能，但带来了显著的计算开销。相比之下，我们的 DGFF 模块更高效且更有效，利用门控注意力机制融合两种特征，表现更优。此外，我们还探讨了门控栅格化和矢量化特征的不同组合权重 (即公式 4 中的 μ 和 ν) 的影响。从表 V 观察到，两种特征的均衡组合效果最佳，其中栅格化项在区域检测中略优，矢量化项在车道段检测中表现更佳。这些结果与第四章 C1 节的发现一致。

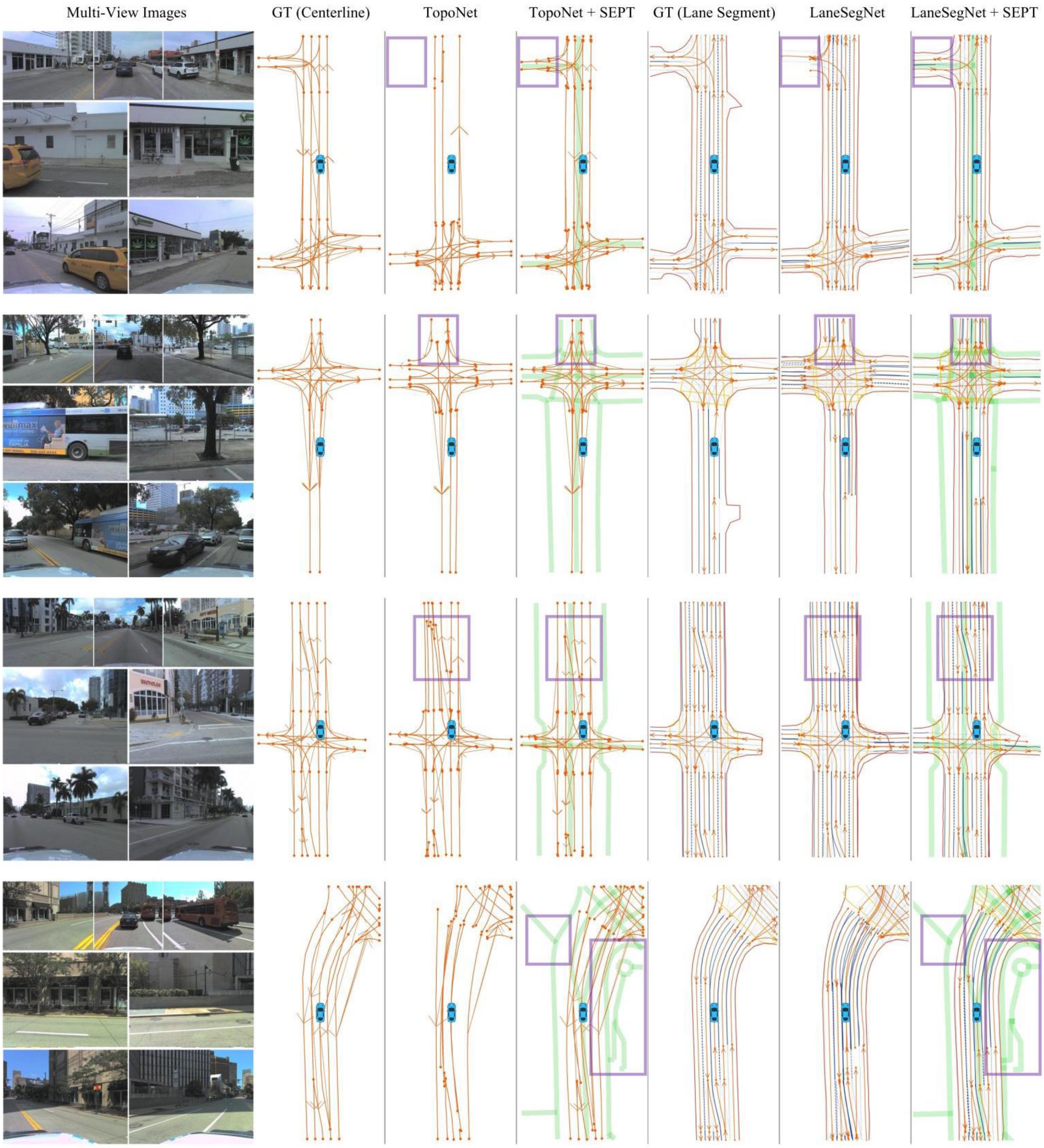

Fig. 4. Qualitative comparisons between the baselines with and without our SEPT module on the OLV2 validation split. From left to right, the figure presents the multi-view images, the ground truth (GT) for centerline perception and topology, the results of the baseline (TopoNet) with and without SEPT, the GT for lane segment perception and topology, and the results of the baseline (LaneSegNet) with and without SEPT. The green line indicates the corresponding SD map prior.

图 4. 在 OLV2 验证集上，带有和不带 SEPT 模块的基线模型的定性比较。图中从左至右依次展示多视角图像、车道中心线感知与拓扑的真实标注 (GT)、带有和不带 SEPT 的基线 (TopoNet) 结果、车道段感知与拓扑的真实标注 (GT) 以及带有和不带 SEPT 的基线 (LaneSegNet) 结果。绿色线条表示对应的 SD 地图先验。

D. Qualitative Results

D. 定性结果

We present visualizations from the OLV2 validation split to demonstrate the improvements brought by our proposed SEPT framework over two baseline models: TopoNet [2] for lane centerline perception and LaneSegNet [6] for lane segment perception. As illustrated in Fig. 4, we highlight several representative cases. In the first row, we demonstrate that in long-range scenarios, the baseline models either fail to detect or inaccurately predict the left turn (highlighted in purple boxes), whereas our model successfully identifies it. In the second row, due to occlusion from a front vehicle, our approach notably enhances the prediction of the occluded road structure at the intersection. The third row illustrates how effectively incorporating the SD map improves lane topology reasoning. The final row presents an interesting case where the SD map provides outdated information, resulting in incorrect prior knowledge. However, our SEPT framework demonstrates the ability to prioritize online perception, yielding a correct prediction that aligns with current observations. This shows that our model strikes an effective balance between onboard sensing and SD map priors. Overall, our SEPT framework significantly improves both scene perception and topology reasoning. More qualitative results can be found in our supplementary video.

我们展示了 OLV2 验证集上的可视化结果，以证明所提 SEPT 框架相较于两个基线模型的改进效果:TopoNet [2] 用于车道中心线感知，LaneSegNet [6] 用于车道段感知。如图 4 所示，我们突出展示了若干典型案例。第一行展示了在远距离场景中，基线模型要么未能检测到左转 (紫色框标注)，要么预测不准确，而我们的模型成功识别。第二行中，由于前车遮挡，我们的方法显著提升了交叉口处被遮挡路况的预测。第三行展示了有效融合 SD 地图如何改善车道拓扑推理。最后一行呈现了一个有趣案例，SD 地图提供了过时信息，导致先验知识错误，但我们的 SEPT 框架能够优先考虑在线感知，给出与当前观测一致的正确预测。这表明我们的模型在车载感知与 SD 地图先验之间实现了有效平衡。总体而言，SEPT 框架显著提升了场景感知和拓扑推理能力。更多定性结果可见补充视频。

V. CONCLUSION

V. 结论

In this letter, we present SEPT, a novel framework that integrates SD map priors into existing perception and reasoning models to enhance online scene understanding. SEPT effectively mitigates the misalignment issue in both rasterized and vectorized SD map representations and leverages the DGFF module to fuse these features for synergistic improvement. To further exploit the potential of SD maps, we introduce the auxiliary IKPD task, which enhances the model in capturing road interaction patterns. We apply our SEPT to two baseline methods and validate it on the large-scale OLV2 benchmark. The significant performance gains demonstrate the superiority of our framework. Future work will focus on incorporating additional SD map prior information, such as lane numbers and road directions, to further enhance scene perception and topology reasoning for mapless driving. REFERENCES

本文提出了 SEPT，一种将 SD 地图先验整合进现有感知与推理模型以增强在线场景理解的新框架。SEPT 有效缓解了栅格化和矢量化 SD 地图表示中的错位问题，并利用 DGFF 模块融合这些特征，实现协同提升。为进一步挖掘 SD 地图潜力，我们引入了辅助 IKPD 任务，增强模型捕捉道路交互模式的能力。我们将 SEPT 应用于两个基线方法，并在大规模 OLV2 基准上验证，显著的性能提升证明了框架的优越性。未来工作将聚焦于引入更多 SD 地图先验信息，如车道数量和道路方向，以进一步提升无地图驾驶的场景感知与拓扑推理能力。参考文献

[1] J. Li, P. Jia, J. Chen, J. Liu, and L. He, ”Local map construction methods with sd map: A novel survey,” arXiv preprint arXiv:2409.02415, 2024.

J. Li, P. Jia, J. Chen, J. Liu, 和 L. He, “基于 SD 地图的局部地图构建方法: 一项新颖综述,” arXiv 预印本 arXiv:2409.02415, 2024.

[2] T. Li, L. Chen, H. Wang, Y. Li, J. Yang, X. Geng, S. Jiang, Y. Wang, H. Xu, C. Xu et al., ”Graph-based topology reasoning for driving scenes,” arXiv preprint arXiv:2304.05277, 2023.

T. Li, L. Chen, H. Wang, Y. Li, J. Yang, X. Geng, S. Jiang, Y. Wang, H. Xu, C. Xu 等，“基于图的驾驶场景拓扑推理”，arXiv 预印本 arXiv:2304.05277，2023 年。

[3] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang et al., ”Planning-oriented autonomous driving,” in CVPR, 2023, pp. 17853-17862.

Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang 等，“面向规划的自动驾驶”，发表于 CVPR，2023 年，第 17853-17862 页。

[4] B. Jiang, S. Chen, Q. Xu, B. Liao, J. Chen, H. Zhou, Q. Zhang, W. Liu, C. Huang, and X. Wang, ”Vad: Vectorized scene representation for efficient autonomous driving,” in Proc. of the IEEE Intl. Conf. Comput. Vis. (ICCV), 2023, pp. 8340-8350.

B. Jiang, S. Chen, Q. Xu, B. Liao, J. Chen, H. Zhou, Q. Zhang, W. Liu, C. Huang, 和 X. Wang，“VAD: 用于高效自动驾驶的矢量化场景表示”，发表于 IEEE 国际计算机视觉会议 (ICCV) 论文集，2023 年，第 8340-8350 页。

[5] H. Wang, T. Li, Y. Li, L. Chen, C. Sima, Z. Liu, B. Wang, P. Jia, Y. Wang, S. Jiang et al., ”Openlane-v2: A topology reasoning benchmark for unified 3d hd mapping,” Advances in Neural Information Processing Systems, vol. 36, 2024.

H. Wang, T. Li, Y. Li, L. Chen, C. Sima, Z. Liu, B. Wang, P. Jia, Y. Wang, S. Jiang 等，“OpenLane-v2: 统一三维高清地图的拓扑推理基准”，《神经信息处理系统进展》(Advances in Neural Information Processing Systems)，第 36 卷，2024 年。

[6] T. Li, P. Jia, B. Wang, L. Chen, K. Jiang, J. Yan, and H. Li, ”Lanesegnet: Map learning with lane segment perception for autonomous driving,” arXiv preprint arXiv:2312.16108, 2023.

T. Li, P. Jia, B. Wang, L. Chen, K. Jiang, J. Yan, 和 H. Li，“LaneSegNet: 基于车道段感知的地图学习用于自动驾驶”，arXiv 预印本 arXiv:2312.16108，2023 年。

[7] T. Ort, J. M. Walls, S. A. Parkison, I. Gilitschenski, and D. Rus, ”Maplite 2.0: Online hd map inference using a prior sd map,” IEEE Robotics and Automation Letters, vol. 7, no. 3, pp. 8355-8362, 2022.

T. Ort, J. M. Walls, S. A. Parkison, I. Gilitschenski, 和 D. Rus，“MapLite 2.0: 利用先验低精度地图的在线高清地图推断”，《IEEE 机器人与自动化快报》，第 7 卷第 3 期，第 8355-8362 页，2022 年。

[8] Z. Jiang, Z. Zhu, P. Li, H.-a. Gao, T. Yuan, Y. Shi, H. Zhao, and H. Zhao, ”P-mapnet: Far-seeing map generator enhanced by both sdmap and hdmap priors,” IEEE Robotics and Automation Letters, vol. 9, pp. 8539-8546, 2024.

Z. Jiang, Z. Zhu, P. Li, H.-a. Gao, T. Yuan, Y. Shi, H. Zhao, 和 H. Zhao，“P-MapNet: 结合低精度地图和高清地图先验的远视地图生成器”，《IEEE 机器人与自动化快报》，第 9 卷，第 8539-8546 页，2024 年。

[9] H. Zhang, D. Paz, y. Guo, A. Das, X. Huang, K. Haug, H. Christensen, and L. Ren, ”Enhancing online road network perception and reasoning with standard definition maps,” in Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots Syst.(IROS), 2024.

H. Zhang, D. Paz, Y. Guo, A. Das, X. Huang, K. Haug, H. Christensen, 和 L. Ren，“利用低精度地图增强在线道路网络感知与推理”，发表于 IEEE/RSJ 国际智能机器人系统会议 (IROS)，2024 年。

[10] K. Z. Luo, X. Weng, Y. Wang, S. Wu, J. Li, K. Q. Weinberger, Y. Wang, and M. Pavone, ”Augmenting lane perception and topology understanding with standard definition navigation maps,” in Proc. of the IEEE Intl. Conf. on Robot. and Autom. (ICRA), 2024, pp. 4029-4035.

K. Z. Luo, X. Weng, Y. Wang, S. Wu, J. Li, K. Q. Weinberger, Y. Wang, 和 M. Pavone，“利用低精度导航地图增强车道感知与拓扑理解”，发表于 IEEE 国际机器人与自动化会议 (ICRA)，2024 年，第 4029-4035 页。

[11] S. Yang, M. Jiang, Z. Fan, X. Xie, X. Tan, Y. Li, E. Ding, L. Wang, and J. Wang, ”Toposd: Topology-enhanced lane segment perception with sdmap prior,” arXiv preprint arXiv:2411.14751, 2024.

S. Yang, M. Jiang, Z. Fan, X. Xie, X. Tan, Y. Li, E. Ding, L. Wang, 和 J. Wang，“TopoSD: 结合低精度地图先验的拓扑增强车道段感知”，arXiv 预印本 arXiv:2411.14751，2024 年。

[12] H. Wu, Z. Zhang, S. Lin, X. Mu, Q. Zhao, M. Yang, and T. Qin, ”Maplocnet: Coarse-to-fine feature registration for visual re-localization in navigation maps,” in Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots Syst.(IROS). IEEE, 2024, pp. 13 198-13 205.

H. Wu, Z. Zhang, S. Lin, X. Mu, Q. Zhao, M. Yang, 和 T. Qin，“MapLocNet: 用于导航地图视觉重定位的粗到细特征配准”，发表于 IEEE/RSJ 国际智能机器人系统会议 (IROS)，IEEE，2024 年，第 13198-13205 页。

[13] A. Vaswani, ”Attention is all you need,” Advances in Neural Information Processing Systems, 2017.

> A. Vaswani，“Attention is all you need(注意力机制即一切)”，《神经信息处理系统进展》，2017 年。

[14] B. Liao, S. Chen, X. Wang et al., ”Maptr: Structured modeling and learning for online vectorized hd map construction,” in International Conference on Learning Representations, 2023.

> B. Liao, S. Chen, X. Wang 等，“MapTR: 用于在线矢量化高清地图构建的结构化建模与学习”，发表于国际学习表征会议，2023 年。

[15] Q. Li, Y. Wang, Y. Wang, and H. Zhao, ”Hdmapnet: An online hd map construction and evaluation framework,” in Proc. of the IEEE Intl. Conf. on Robot. and Autom. (ICRA), 2022, pp. 4628-4634.

> Q. Li, Y. Wang, Y. Wang, 和 H. Zhao，“Hdmapnet: 一个在线高清地图构建与评估框架,”发表于 IEEE 国际机器人与自动化会议 (ICRA)，2022 年，第 4628-4634 页。

[16] Z. Xu, Y. Liu, Y. Sun, M. Liu, and L. Wang, ”Centerlinedet: Centerline graph detection for road lanes with vehicle-mounted sensors by transformer for hd map generation,” in Proc. of the IEEE Intl. Conf. on Robot. and Autom. (ICRA), 2023, pp. 3553-3559.

> Z. Xu, Y. Liu, Y. Sun, M. Liu, 和 L. Wang，“Centerlinedet: 基于变换器的车载传感器道路车道中心线图检测，用于高清地图生成,”发表于 IEEE 国际机器人与自动化会议 (ICRA)，2023 年，第 3553-3559 页。

[17] T. Langenberg, T. Lüddecke, and F. Wörgötter, ”Deep metadata fusion for traffic light to lane assignment,” IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 973-980, 2019.

> T. Langenberg, T. Lüddecke, 和 F. Wörgötter，“用于交通灯到车道分配的深度元数据融合,”IEEE 机器人与自动化快报，卷 4，第 2 期，2019 年，第 973-980 页。

[18] Y. Liu, T. Wang, X. Zhang, and J. Sun, ”Petr: Position embedding transformation for multi-view 3d object detection,” in European Conference on Computer Vision. Springer, 2022, pp. 531-548.

> Y. Liu, T. Wang, X. Zhang, 和 J. Sun，“Petr: 多视角三维目标检测的位置嵌入变换,”发表于欧洲计算机视觉会议，Springer 出版社，2022 年，第 531-548 页。

[19] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, ”Learning lane graph representations for motion forecasting,” in ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part II 16. Springer, 2020, pp. 541-556.

> M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, 和 R. Urtasun，“用于运动预测的车道图表示学习,”发表于 2020 年第 16 届欧洲计算机视觉会议 (ECCV)，英国格拉斯哥，2020 年 8 月 23-28 日，会议论文集第二部分，Springer 出版社，2020 年，第 541-556 页。

[20] H. Liu, L. Chen, Y. Qiao, C. Lv, and H. Li, ”Reasoning multi-agent behavioral topology for interactive autonomous driving,” in Annual Conference on Neural Information Processing Systems, 2024.

H. Liu, L. Chen, Y. Qiao, C. Lv, 和 H. Li, “多智能体行为拓扑推理用于交互式自动驾驶, ”发表于神经信息处理系统年会，2024 年。

[21] Y. B. Can, A. Liniger, D. P. Paudel, and L. Van Gool, ”Structured bird’s-eye-view traffic scene understanding from onboard images,” in Proc. of the IEEE Intl. Conf. Comput. Vis. (ICCV), 2021, pp. 15661-15670.

Y. B. Can, A. Liniger, D. P. Paudel, 和 L. Van Gool, “基于车载图像的结构化鸟瞰交通场景理解, ”发表于 IEEE 国际计算机视觉会议 (ICCV)，2021 年，第 15661-15670 页。

[22] D. Wu, J. Chang, F. Jia, Y. Liu, T. Wang, and J. Shen, ”Topomlp: A simple yet strong pipeline for driving topology reasoning,” in The Twelfth International Conference on Learning Representations, 2024.

D. Wu, J. Chang, F. Jia, Y. Liu, T. Wang, 和 J. Shen, “Topomlp: 一个简单而强大的驾驶拓扑推理流程, ”发表于第十二届国际表征学习会议，2024 年。

[23] Y. Fu, W. Liao, X. Liu, H. Xu, Y. Ma, Y. Zhang, and F. Dai, ”Topologic: An interpretable pipeline for lane topology reasoning on driving scenes,” Advances in Neural Information Processing Systems, vol. 37, pp. 61658-61676, 2024.

Y. Fu, W. Liao, X. Liu, H. Xu, Y. Ma, Y. Zhang, 和 F. Dai, “Topologic: 一个可解释的驾驶场景车道拓扑推理流程, ”神经信息处理系统进展，卷 37，2024 年，第 61658-61676 页。

[24] Z. Ma, S. Liang, Y. Wen, W. Lu, and G. Wan, ”Roadpainter: Points are ideal navigators for topology transformer,” in European Conference on Computer Vision. Springer, 2025, pp. 179-195.

Z. Ma, S. Liang, Y. Wen, W. Lu, 和 G. Wan, “Roadpainter: 点是拓扑变换器的理想导航者, ”发表于欧洲计算机视觉会议，Springer 出版社，2025 年，第 179-195 页。

[25] H. Wu, Z. Zhang, S. Lin, T. Qin, J. Pan, Q. Zhao, C. Xu, and M. Yang, ”Blos-bev: Navigation map enhanced lane segmentation network, beyond line of sight,” in 2024 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2024, pp. 3212-3219.

H. Wu, Z. Zhang, S. Lin, T. Qin, J. Pan, Q. Zhao, C. Xu, 和 M. Yang, “Blos-bev: 导航地图增强的车道分割网络，超越视线范围, ”发表于 2024 年 IEEE 智能车辆研讨会 (IV)，IEEE，2024 年，第 3212-3219 页。

[26] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, ”Pointnet: Deep learning on point sets for 3 d classification and segmentation,” in CVPR,2017, pp. 652-660.

C. R. Qi, H. Su, K. Mo, 和 L. J. Guibas, “Pointnet: 基于点集的深度学习用于分类和分割, ”发表于 CVPR，2017 年，第 652-660 页。

[27] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, ”Film: Visual reasoning with a general conditioning layer,” in Proceedings of the AAAI conference on artificial intelligence, vol. 32, no. 1, 2018.

E. Perez, F. Strub, H. De Vries, V. Dumoulin, 和 A. Courville,“Film: 具有通用条件层的视觉推理,”发表于 AAAI 人工智能会议论文集，卷 32，第 1 期，2018 年。

[28] H. Law and J. Deng, ”Cornernet: Detecting objects as paired keypoints,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 734-750.

H. Law 和 J. Deng，“Cornernet: 将目标检测为成对关键点，”发表于欧洲计算机视觉会议 (ECCV), 2018 年，第 734-750 页。

[29] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, ”Centernet: Keypoint triplets for object detection,” in CVPR, 2019, pp. 6569-6578.

K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, 和 Q. Tian，“Centernet: 用于目标检测的关键点三元组,”发表于 CVPR，2019 年，第 6569-6578 页。

[30] F. Chollet, ”Xception: Deep learning with depthwise separable convolutions,” in CVPR, 2017, pp. 1251-1258.

F. Chollet，“Xception: 基于深度可分离卷积的深度学习,”发表于 CVPR，2017 年，第 1251-1258 页。

[31] J. Hu, L. Shen, and G. Sun, ”Squeeze-and-excitation networks,” in CVPR, 2018, pp. 7132-7141.

J. Hu, L. Shen, 和 G. Sun，“挤压与激励网络 (Squeeze-and-excitation networks),”发表于 CVPR，2018 年，第 7132-7141 页。

[32] B. Wilson, W. Qi, T. Agarwal et al., ”Argoverse 2: Next generation datasets for self-driving perception and forecasting,” arXiv preprint arXiv:2301.00493, 2023.

B. Wilson, W. Qi, T. Agarwal 等，“Argoverse 2: 面向自动驾驶感知与预测的下一代数据集，”arXiv 预印本 arXiv:2301.00493，2023 年。

[33] M. Haklay and P. Weber, ”Openstreetmap: User-generated street maps,” IEEE Pervasive computing, vol. 7, no. 4, pp. 12-18, 2008.

M. Haklay 和 P. Weber，“OpenStreetMap: 用户生成的街道地图，”IEEE 普适计算，卷 7，第 4 期，第 12-18 页，2008 年。