Delving into Mapping Uncertainty for Mapless Trajectory Prediction

Zongzheng Zhang[1,2*], Xuchong Qiu[2*], Boran Zhang[1], Guantian Zheng[1], Xunjiang Gu[4],
Guoxuan Chi[1], Huan-ang Gao[1], Leichen Wang[2†], Ziming Liu[2], Xinrun Li[2],
Igor Gilitschenski[4], Hongyang Li[5], Hang Zhao[3], Hao Zhao[1✉]

Abstract— Recent advances in autonomous driving are moving towards mapless approaches, where High-Definition (HD) maps are generated online directly from sensor data, reducing the need for expensive labeling and maintenance. However, the reliability of these online-generated maps remains uncertain. While incorporating map uncertainty into downstream trajectory prediction tasks has shown potential for performance improvements, current strategies provide limited insights into the specific scenarios where this uncertainty is beneficial. In this work, we first analyze the driving scenarios in which mapping uncertainty has the greatest positive impact on trajectory prediction and identify a critical, previously overlooked factor: the agent's kinematic state. Building on these insights, we propose a novel Proprioceptive Scenario Gating that adaptively integrates map uncertainty into trajectory prediction based on forecasts of the ego vehicle's future kinematics. This lightweight, self-supervised approach enhances the synergy between online mapping and trajectory prediction, providing interpretability around where uncertainty is advantageous and outperforming previous integration methods. Additionally, we introduce a Covariance-based Map Uncertainty approach that better aligns with map geometry, further improving trajectory prediction. Extensive ablation studies confirm the effectiveness of our approach, achieving up to 23.6% improvement in mapless trajectory prediction performance over the state-of-the-art method using the real-world nuScenes driving dataset. Our code, data, and models are publicly available at https://github.com/Ethan-Zheng136/Map-Uncertainty-for-Trajectory-Prediction.

I. INTRODUCTION

Traditionally, High-Definition (HD) maps have served as a valuable resource for autonomous driving, supporting tasks in perception [1], [4]–[10], prediction [11], [12], and planning [2], [13]–[15]. These maps provide detailed, multi-layered representations of the environment, capturing road geometry (such as polylines and polygons), semantic attributes (e.g., road types), traffic signals, speed limits, and other critical information for static scene understanding. However, creating and maintaining HD maps is both time-consuming and costly, requiring extensive labeling and frequent updates to stay fresh. This dependency on pre-labeled data limits scalability, as HD maps are typically constrained to specific, pre-defined areas, restricting the broader deployment.

[1]Institute for AI Industry Research (AIR), Tsinghua University, China.
[2]Bosch Corporate Research, China.
[3]Institute for Interdisciplinary Information Sciences(IIIS), Tsinghua University, China.
[4]University of Toronto, Canada.
[5]University of HongKong, China.
∗ Equal contribution.
† Project lead.
✉ Corresponding to zhaohao@air.tsinghua.edu.cn

(a) **Motivation of the Proprioceptive Scenario Gating.**

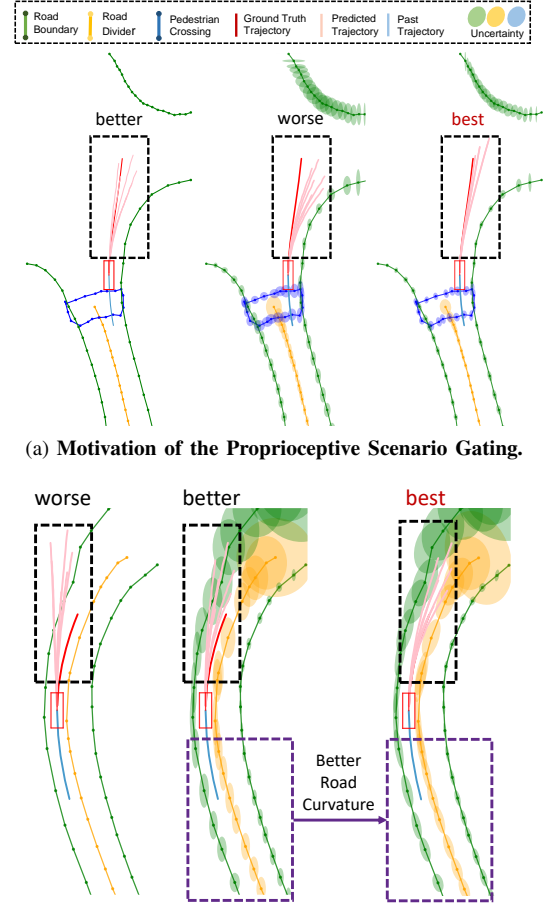(b) **Better road curvature with the Covariance-Based Uncertainty.**

Fig. 1: **Left:** Mapless trajectory prediction baseline (MapTRv2 [1] + HiVT [2]), **Middle:** Previous uncertainty integration method [3], **Right:** Ours. In (a), the comparison highlights that predictions enhanced with uncertainty occasionally underperform the baseline, underscoring the necessity of a gating mechanism to selectively incorporate uncertainty. In (b), our Covariance-based Map Uncertainty captures road curvature with good precision that the predicted trajectories align well with the ground truth, rendering the true trajectory invisible.

To address these challenges, recent research has shifted towards generating HD maps online directly from sensor data, creating either rasterized bird's-eye-view (BEV) maps [16]–[18] or vectorized representations with polylines [1], [7]–[9]. This "mapless driving" approach reduces reliance on pre-built, offline maps and minimizes the need for extensive labeling and maintenance. As the field progresses, the focus in mapless driving is evolving from perception (improving online map generation) to motion prediction, with an emphasis on advancing mapless trajectory prediction [3], [19].

In this context, recent studies have highlighted the potential of incorporating uncertainty [20]–[22] in online map generation, arising from factors such as occlusions by other vehicles and obstacles, as well as the varying distance to these elements, to enhance mapless trajectory prediction performance [3]. However, current methods offer limited insight into the specific scenarios where this uncertainty is most beneficial or where it performs poorly (Fig. 1(a)). In this work, we systematically examine the scenarios in which uncertainty modeling and integration yield the greatest gains in prediction accuracy. Our analysis reveals a key, previously overlooked factor: the agent's kinematic state. Specifically, we find that introducing map uncertainty improves trajectory prediction accuracy primarily in scenarios requiring abrupt changes in vehicle kinematics (e.g., transitioning from a straight to a curved path), while it may degrade performance when historical kinematics are the dominant predictor (e.g., during steady cruising on a straight path). These findings suggest that the effectiveness of online map uncertainty is highly scenario-dependent, prompting us to seek an adaptive approach to selectively apply map uncertainty in trajectory prediction based on scenario-specific conditions.

Specifically, we propose Proprioceptive Scenario Gating, a plug-and-play, lightweight, self-supervised module that adaptively integrates map uncertainty based on the kinematic state of the ego vehicle. By incorporating initial trajectory predictions from a dual parallel stream, this mechanism selectively combines uncertainty-enhanced and non-enhanced results through weighted integration, optimizing final trajectory predictions. Furthermore, aiming to help trajectory prediction methods make better use of road geometry for more accurate predictions, we introduce a Covariance-based Uncertainty Modeling approach to better align uncertainty with map geometry (Fig. 1(b)). Extensive ablation studies confirm the independent effectiveness of these two techniques, showing consistent performance improvements across various state-of-the-art mapper-predictor configurations.

In summary, our contributions are threefold: **First**, we systematically demonstrate that the effect of online map uncertainty on mapless trajectory prediction accuracy is scenario-dependent, with its impact varying based on the agent's driving kinematics. **Second**, we propose a lightweight, self-supervised Proprioceptive Scenario Gating mechanism that adaptively integrates map uncertainty, along with a Covariance-based Map Uncertainty approach that better aligns with map geometry. **Third**, our approach achieves up to a 23.6% improvement in trajectory prediction accuracy on the real-world nuScenes benchmark, outperforming existing methods for mapless trajectory prediction.

II. RELATED WORKS

A. Online Map Estimation

Online map estimation aims to directly predict the representations of static world elements around an autonomous vehicle from sensor data [23]. Early approaches primarily focused on producing 2D BEV rasterized semantic segmentations [16]–[18], [24].

Recent advances in vectorized map estimation have focused on improving BEV rasterization by decoding maps into polylines and polygons. Early methods relied on post-processing for map generation [6], [7], while later works introduced end-to-end learning to streamline the process and reduce information loss [8], [9]. Studies have also integrated lane perception, inner-instance information, and Euclidean geometry to enhance map representation [25], [26].

Further advancements have focused on enhancing localization accuracy and temporal consistency in map representations [5]. Several works have introduced techniques to improve localization and mitigate mapping errors, such as mask-guided instance decoding and anti-disturbance methods for jitter reduction [27], as well as hybrid representation learning within end-to-end frameworks [28]. Other studies have explored robustness against sensor corruptions and enhanced point set queries for map construction [29].

Despite advancements, the reliability of online-generated maps for downstream tasks remains uncertain. While Gu et al. [3] modeled uncertainty using a simple 2 single-variable Laplacian distribution, it limits flexibility for various road geometries. Building on this, we introduce a Covariance-based Map Uncertainty approach that captures more comprehensive uncertainty patterns, enhancing mapping reliability in downstream applications.
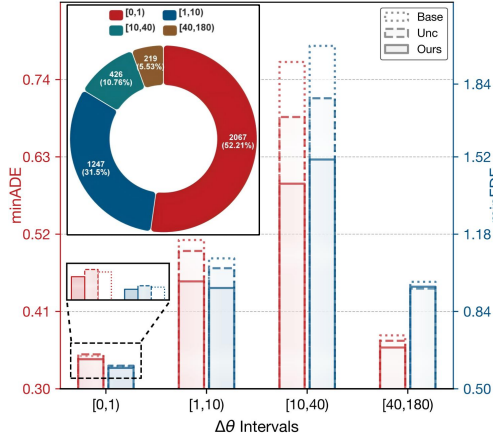
B. Map-Informed Trajectory Prediction

Trajectory prediction aims to forecast the future paths of agents, based on historical trajectories and scene context. Traditional models often rely on BEV maps encoded with Convolutional Neural Network (CNN) to represent the scene, combining with agent history to inform predictions [31]–[34]. While effective, these approaches suffer from data inefficiencies and spatial discretization, potentially limiting precision in complex scenes. To address these limitations, recent work has turned to vectorized map representations, which more efficiently capture detailed map elements relevant to agent motion [35]–[37]. Initial methods [13], [35], [37] introduced Graph Neural Networks (GNNs) to encode lane polylines, capturing map influence on agent behavior. Meanwhile, Transformer-based architectures capture spatial and temporal dependencies, with HiVT [2] and QCNet [14] further enhancing efficiency and robustness through query-centric designs and invariant scene representations.

While vectorized approaches have significantly advanced trajectory prediction, challenges remain in effectively integrating them with upstream modules like online mapping. Previous methods encode mapping uncertainty alongside vectorized point sets but lack systematic evaluation of when and where uncertainty should influence prediction [3]. Our work addresses these limitations by analyzing the specific conditions under which uncertainty integration is most beneficial and introducing a Proprioceptive Scenario Gating that adapts to varying contexts in real time. This adaptive framework enhances prediction accuracy, robustness, and interpretability in complex driving scenarios.
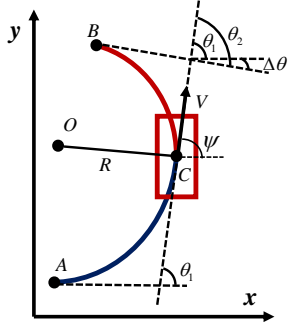
(a) **Vehicle's Particle Kinematic Model.** The red and blue lines indicate the future and the past 2s trajectory respectively. A and B denote the first and the last point of the trajectory. C is the center of the vehicle with a heading angle Ψ. O and R denote the instantaneous center and the radius 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 [30] 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 minADE and minFDE values indicate better trajectory prediction accuracy.

Fig. 2: **Preliminary Analysis of Map Uncertainty** for trajectory prediction based ego agent driving kinematic context.

III. METHOD

This section begins by analyzing where map uncertainty provides benefits for trajectory prediction and where it performs poorly, using the vehicle's particle kinematic model (Sec. III-A). Based on these insights, we provide an overview of our proposed framework (Sec. III-B). Then we introduce our Covariance-Based Map Uncertainty (Sec. III-C) and Proprioceptive Scenario Gating (Sec. III-D) to improve the current mapless driving paradigm. Finally, we present the overall loss function used to train our model (Sec. III-E).

A. Analysis of Map Uncertainty for Trajectory Prediction

We compare the baseline (MapTRv2 [1] + HiVT [2]) and the previous uncertainty method [3] for mapless driving to identify scenarios where uncertainty has the most significant impact on trajectory prediction performance.

In Fig. 2(a), according to the vehicle's particle kinematic model, we have:

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

where $\dot{\Psi}_1$ and $\dot{\Psi}_2$ denote the average angular velocities of the ego vehicle in the past and future intervals over the same time interval $\Delta t = 2$ s. Thus, θ_1 and θ_2 represent the total rotation experienced by the ego vehicle during the past and future 2 seconds. Then, we have

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

which quantifies the extent of change $\left|\dot{\Psi}_1 - \dot{\Psi}_2\right|$ in the ego vehicle's driving state between these two time periods. When $\Delta\theta$ tends to zero, it implies that the vehicle is either traveling in a straight line or making a steady turn.

As shown in Fig. 2(b), we divide the indicator $\Delta\theta$ values, which reflects curvature changes and vehicle state variations, into four intervals and analyze the number of scenarios and performance within each. Notably, 52.21% of the scenarios fall within the range of $[0, 1)$, where the baseline (minADE = 0.3466, minFDE = 0.5980) slightly outperforms the uncertainty method (minADE = 0.3490, minFDE = 0.6016), indicating that under consistent driving behavior, excluding uncertainty yields better results. In contrast, with larger future changes in vehicle kinematics, the uncertainty method results in lower prediction errors. As described in the following sections, our proposed method addresses these issues, demonstrating improved performance and achieving better results through experiments.

B. Architecture Overview

As shown in Fig. 3, our framework consists of a mapping module based on Covariance Map Uncertainty estimation and a trajectory prediction module with a Proprioceptive Scenario Gating to selectively leverage upstream uncertainty with vehicle kinematics.

First, we use online mapping methods including MapTR [4], MapTRv2 [1], MapTRv2-Centerline [1] and StreamMapNet [5] to regress map element vertices. For each vertex, uncertainty is modeled with a 2D Gaussian distribution. This uncertainty information is then encoded to refine downstream trajectory predictions. Specifically, we employ two representative models: HiVT [2], which is Transformer-based, and DenseTNT [13], which is GNN-based. During their context encoding stage, the map vertex coordinates (mean $\mu^{(i)}$ in uncertainty modeling) are first encoded via an MLP. For uncertainty modeling, we concatenate the mean $\mu^{(i)}$, covariance matrix $\Sigma^{(i)}$ into a single representation, which is then passed through the MLP. This process operates in two parallel streams—one that incorporates uncertainty and one that does not. The Proprioceptive Scenario Gating then dynamically combines the predicted trajectories output from these two streams by passing them through another MLP. This MLP generates weights for each stream, allowing the model to effectively merge the trajectories and produce a more optimal one.

C. Covariance-Based Uncertainty Modeling

Gu et al. [3] models uncertainty in each map element's vertex in the x and y directions using 2 single-variable Laplace distribution, assuming independence between x and
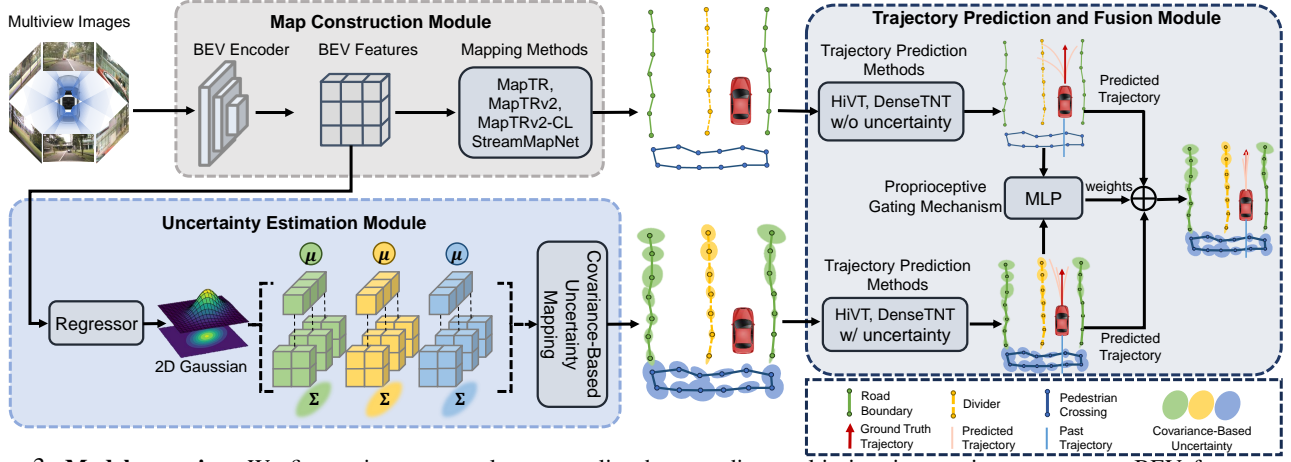
Fig. 3: **Model overview.** We first estimate map elements online by encoding multi-view images into a common BEV feature space to regress map element vertices. Each vertex's uncertainty is modeled using our proposed Covariance-based Uncertainty method, which leverages 2D Gaussian distribution templates. This uncertainty information, along with the original map vertices, is then passed to the downstream trajectory prediction module, which operates in two parallel streams: one that incorporates uncertainty and one that does not. Finally, the proposed Proprioceptive Scenario Gating (MLP network) dynamically adapts the optimal trajectory prediction based on the initial future trajectories prediction from these two streams.

y coordinates. However, in real-world scenarios, these two positional random variables are often correlated due to the directional alignment of the map element. Neglecting these correlations can lead to potentially incorrect modeling of unreliable map regions, which in turn may result in errant trajectory predictions.

To address this, we propose a 2D Gaussian distribution with a correlation coefficient to better represent the spread and orientation of joint uncertainty. Specifically, we use a probabilistic module that jointly predicts both uncertainty parameters (i.e., variance) and the x-y correlation coefficient. This extension enables the model to better capture the spatial relationship of uncertainties, offering a more comprehensive representation of map uncertainty that enhances both prediction accuracy and robustness.

Assume a map M with V vertices, where each vertex $v^{(i)}$ in 2D space follows a 2D Gaussian distribution with mean $\mu^{(i)}$ and covariance $\Sigma^{(i)}$. The joint probability density function is:

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

where $v^{(i)} = \begin{bmatrix} v_1^{(i)} & v_2^{(i)} \end{bmatrix}^T$ is the coordinate vector of the i-th vertex, and $\mu^{(i)} = \begin{bmatrix} \mu_1^{(i)} & \mu_2^{(i)} \end{bmatrix}^T$ is its mean vector. The covariance matrix for the i-th vertex is $\Sigma^{(i)} = \begin{bmatrix} \sigma_1^2, \rho\sigma_1\sigma_2; \rho\sigma_1\sigma_2, \sigma_2^2 \end{bmatrix}$, where σ_1 and σ_2 are the standard deviations in the x and y dimensions, and ρ denotes the correlation coefficient, capturing the linear dependency between dimensions.

Compared to modeling uncertainty with two single-variable Laplace distributions in the x and y directions (Fig. 1(b)), our covariance-based method better aligns with the geometry of map elements, such as road edge curvature. We choose a Gaussian distribution with covariance for uncertainty modeling choice as it theoretically outperforms

Laplace distribution due to its lighter tails, well-established mathematical, robust properties and reduced sensitivity to outliers.

D. Proprioceptive Scenario Gating

To address these context-dependent performance discrepancies, we introduce the Proprioceptive Scenario Gating mechanism to enhance trajectory prediction by selectively combining the strengths of both methods. This module enables context-sensitive scenario gating method for more accurate predictions, rather than indiscriminately incorporating uncertainty information.

Traditional exteroceptive gating approaches based on multi-view images or bird's-eye-view (BEV) data, rely on external sensors (e.g., cameras or LiDAR) to analyze surroundings from multiple perspectives. However, these methods do not consider the car agent inherent properties such as its kinematics. We therefore propose a simple proprioceptive gating method based on the vehicle's internal kinematic data, without relying on any additional environment information, effectively distinguishing scenario types based on trajectory characteristics. As explained in Sec III-B, the vehicle's trajectories inherently includes its kinematic data, which directly captures position and indirectly incorporates information such as translational velocity, angular velocity, and acceleration, playing a crucial role in distinguishing different scenario types.

Our approach follows a self-supervised paradigm. We employ a dual-encoder structure, where map data with and without uncertainty is processed by a trajectory prediction model to generate predictions $\mathbf{T}_{\text{base}} \in \mathbb{R}^{1\times512}$ and $\mathbf{T}_{\text{unc}} \in \mathbb{R}^{1\times512}$. These trajectories are flattened, padded and concatenated into an input tensor $\mathbf{T} \in \mathbb{R}^{1\times2}$, which is processed through the six-layer MLP network with sizes of 512, 256, 128, 64, 32, 2, reducing the tensor shape from $[1,512]$ to $[1,2]$:

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

The output $[v_{\text{base}}, v_{\text{unc}}]$ is transformed via temperature-scaled softmax to obtain weights $[w_{\text{base}}, w_{\text{unc}}]$ summing to 1, which serve as labels for training this module.

The gating mechanism processes predicted trajectories to extract and interpret kinematic features relevant to trajectory fusion. This training process enables the model to learn the relationship between kinematic features and optimal trajectory, allowing it to dynamically weight predicted trajectories.

E. Loss Functions

To train the covariance-based uncertainty model for online mapping, we modify the regression loss to a Negative Log-Likelihood (NLL) loss for a 2D Gaussian distribution:

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

where $\log((2\pi)^2 |\Sigma^{(i)}|)$ is the normalization term related to the determinant of the covariance matrix $\Sigma^{(i)}$, and $(v^{(i)} - \mu^{(i)})^T (\Sigma^{(i)})^{-1} (v^{(i)} - \mu^{(i)})$ represents the error term, measuring the Mahalanobis distance between predicted and actual positions while considering correlations between dimensions. To address instability issues reported in prior work [3] with Gaussian distributions during training, we also add a regularization term to prevent gradient explosion, stabilizing the training process and promoting smooth convergence.

For the Proprioceptive Scenario Gating, it is trained using a Mean Squared Error (MSE) loss function. This loss is computed based on the difference between the module's predicted scenario weights and the target weights, which reflect the optimal trajectory selection for each scenario.

IV. EXPERIENMENTS

We start with a description of the experimental setup (Sec.IV-A), then compare our method with state-of-the-art map-uncertainty integrated trajectory prediction methods (Sec.IV-B). Finally, we provide visualizations (Sec.IV-C) and an ablation study (Sec.IV-D).

A. Experiment Setup

Dataset. We evaluate our method on the nuScenes dataset [30] which is divided into 500 for training, 200 for validation, and 150 for testing. Each 20-second scene features keyframe annotations at 2 Hz, with six multi-view RGB cameras providing a 360° view around the ego-vehicle. We use the same trajectory processing method as in [3]. Each model then forecasts motion 3 seconds into the future based on 2 seconds of historical data.

Metrics. For trajectory prediction evaluation, we adopt standard metrics commonly used in recent prediction challenges [38]–[40]: minimum Average Displacement Error (minADE), minimum Final Displacement Error (minFDE), and Miss Rate (MR). Our model outputs six candidate trajectories per agent. The minADE metric computes the average Euclidean (ℓ_2) distance between the most accurate predicted trajectory and the ground truth (GT) trajectory

across all time steps, while minFDE measures distance at the final time step. MR quantifies the percentage of predictions where the endpoint of the best trajectory deviates from the GT endpoint by more than 2 meters.

Implementation Details. All models are trained on a single NVIDIA GeForce RTX 3090 GPU. For each mapping model, the learning rate is set to 1.5e-4, the regression loss weight is 0.03, and gradient norms are clipped at 3.

For the trajectory prediction task, HiVT is trained with a learning rate of 5.0e-4, a batch size of 32 and a temprature of 0.6. For DenseTNT, the learning rate remains at 5.0e-4, while the batch size and temperature are adjusted to 16 and 0.5, respectively. A dropout rate of 0.1 is applied consistently across all trajectory prediction models to prevent overfitting.

B. Main Results

To evaluate the impact of our uncertainty modeling strategy on downstream trajectory prediction, we conduct experiments across multiple combinations of online map estimation and prediction methods. Aiming to facilitate a more effective comparison with previous work, we select the same four mapping methods (MapTR [4], MapTRv2 [1], MapTRv2-Centerline [1] and StreamMapNet [5]) and two trajectory prediction models (HiVT [2] and DenseTNT [13]), resulting in eight combinations. Table. I shows consistent performance improvements across all combinations when using our formulation. With HiVT, our method improves minADE and minFDE by over 6% on MapTR and MapTRv2, and achieves a notable 13.6% improvement in MR on MapTRv2. For DenseTNT, the largest gains are observed on MapTRv2-Centerline, where minADE, minFDE, and MR improve by 19.4%, 10.8%, and 23.6%, respectively. Fig. 2(b) further illustrates that across different intervals of $\Delta\theta$, our method consistently improves trajectory prediction performance.

C. Visualizations

Qualitative results in Fig. 4 demonstrate that our method consistently improves trajectory prediction across various driving scenarios. In the **top left** of Fig. 4, the ego vehicle executes a steady circular turn through a complex scenario, where $\Delta\theta \approx 0$ indicates a stable driving state (Sec. III-B). The baseline (a) outperforms the previous uncertainty-based method (b), yet part of the predicted trajectory extends beyond the road boundary. In contrast, our method (c) aligns closely with the ground truth trajectory, remaining within the drivable area. In the **top right**, as the ego vehicle approaches a gentle curve, the previous method (b) marginally outperforms the baseline (a), but both predictions exceed the road boundary. Our method (c) follows the road curvature and avoids the vehicle ahead. The **Bottom left** visualizes the ego vehicle avoiding a truck while staying in its lane. Both the baseline and previous uncertainty-based method drift into the opposite lane, while our method successfully avoids the truck and maintains on its current lane. In the **bottom right**, the ego vehicle transitions from a turn to a straight path. The previous method underperforms compared to the

Prediction Method	HiVT [2]				DenseTNT [13]			
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓	FPS	minADE ↓	minFDE ↓	MR ↓	FPS
MapTR [4]	0.4015	0.8404	0.0960	17.91	1.1228	2.2151	0.3726	7.36
MapTR [4] + Unc [3]	0.3910	0.8049	0.0818	16.65	1.1946	2.2666	0.3848	7.75
MapTR [4] + 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 [1]	0.4017	0.8406	0.0959	15.66	1.3262	2.5687	0.4301	7.57
MapTRv2 [1] + Unc [3]	0.3913	0.8054	0.0819	14.98	1.3256	2.6390	0.4435	7.38
MapTRv2 [1] + 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 [1]	0.3789	0.7859	0.0865	14.02	0.8333	1.4752	0.1719	7.62
MapTRv2-CL [1] + Unc [3]	0.3674	0.7418	0.0739	14.03	0.9666	1.6439	0.2082	7.29
MapTRv2-CL [1] + 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 [5]	0.3963	0.8223	0.0923	14.17	1.0639	2.1430	0.3412	7.42
StreamMapNet [5] + Unc [3]	0.3899	0.8101	0.0861	13.69	1.0902	2.1412	0.3261	6.91
StreamMapNet [5] + 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: Quantitative results of trajectory prediction for representative mapping/prediction model combinations on the nuScenes [30] dataset under different settings. Our method incorporating the proposed Proprioceptive Scenario Gating and Covariance-based Map Uncertainty consistently improves prediction model performance, surpassing previous uncertainty integration methods across nearly all combinations. Additionally, we benchmark FPS on an NVIDIA RTX 3090, demonstrating that our method achieves comparable computational efficiency to the baseline while delivering superior predictive performance.

(a) MapTRv2+HiVT (b) Previous Uncertainty (c) Ours (d) GT

(a) MapTR+HiVT (b) Previous Uncertainty (c) Ours (d) GT

(a) MapTRv2-CL+DenseTNT (b) Previous Uncertainty (c) Ours (d) GT

(a) StreamMapNet+DenseTNT (b) Previous Uncertainty (c) Ours (d) GT
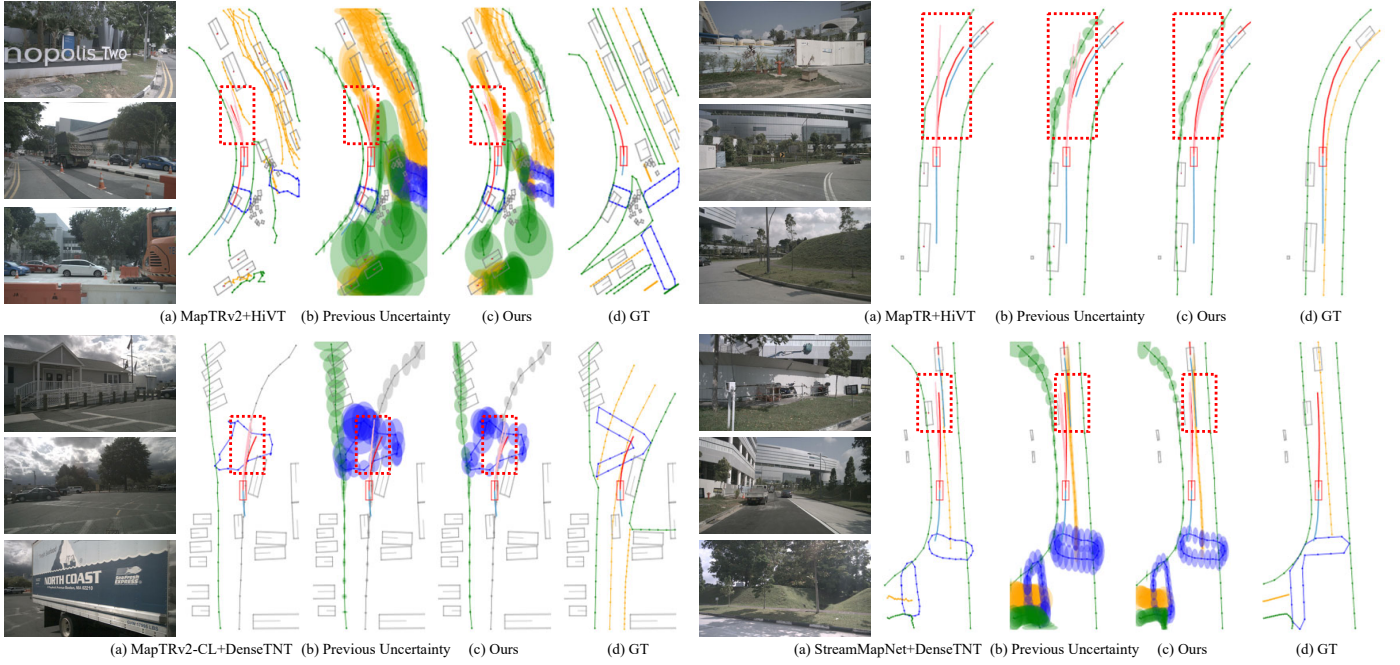
Fig. 4: Qualitative visualization showing various combinations of online mapping and trajectory prediction models. Our approach outperforms both the baseline and the state-of-the-art uncertainty modeling method [3] in all these driving scenarios.

baseline, while our method outperforms both, demonstrating the advantages of our gating mechanism in leveraging kinematic information. As shown in Table I, for DenseTNT, previous uncertainty-based methods generally underperform compared to the baseline. The **bottom-left** and **bottom-right** scenarios in Fig. 4 highlight that our method produces more precise and tightly clustered trajectories.

D. Ablations

1) Covariance-based Map Uncertainty: First, we evaluate the effectiveness of the proposed covariance-based uncertainty without using the Proprioceptive Scenario Gating, as shown in Table II. This formulation consistently enhances trajectory prediction performance across all mapping methods, offering richer, geometry-aligned uncertainty information to downstream tasks. The inclusion of covariance-based uncertainty significantly reduces minFDE and MR, with the most notable improvements observed in MapTRv2-Centerline (4.3% reduction in minFDE) and StreamMapNet (6.5% reduction in MR). We also explore the impact of modeling uncertainty with four different distributions, as shown in Table III. Our results indicate that whether using Laplace [3] or Gaussian distributions, incorporating covariance (correlation) into uncertainty modeling consistently outperforms the independent distribution approach, as it better aligns with road curvature. Moreover, covariance-based Gaussian distribution generally outperforms the covariance-based Laplace distribution. This further validates our choice of using joint Gaussian distribution for modeling uncertainty.

2) Proprioceptive Scenario Gating: We ablate the effectiveness of the proposed Proprioceptive Scenario Gating. Table IV shows that adaptively selecting the optimal

Prediction Method	HiVT [2]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [4]	0.4015	0.8404	0.0960
MapTR [4] + Unc [3]	0.3910	0.8049	0.0818
MapTR [4] + CovMat	**0.3851** (−1.5%)	**0.7812** (−2.9%)	**0.0815** (−0.4%)
MapTRv2 [1]	0.4017	0.8406	0.0959
MapTRv2 [1] + Unc [3]	0.3913	0.8054	0.0819
MapTRv2 [1] + CovMat	**0.3776** (−3.5%)	**0.7707** (−4.3%)	**0.0789** (−1.5%)
MapTRv2-CL [1]	0.3789	0.7859	0.0865
MapTRv2-CL [1] + Unc [3]	0.3674	0.7418	0.0739
MapTRv2-CL [1] + CovMat	**0.3662** (−0.3%)	**0.7411** (−0.1%)	**0.0728** (−1.5%)
StreamMapNet [5]	0.3963	0.8223	0.0923
StreamMapNet [5] + Unc [3]	0.3899	0.8101	0.0861
StreamMapNet [5] + CovMat	**0.3843** (−1.4%)	**0.8035** (−0.8%)	**0.0805** (−6.5%)

TABLE II: In general, leveraging our covariance-based uncertainty formulation (CovMat) for online mapping consistently improves prediction model performance across all combinations compared to the previous two single-variable Laplacian approach [3].

Prediction Method	HiVT [2]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [4]	0.4015	0.8404	0.0960
MapTR [4] + Bivariate Laplace w/o Covariance [3]	0.3910	0.8049	0.0802
MapTR [4] + Bivariate Gaussian w/o Covariance	0.3931	0.8079	0.0839
MapTR [4] + Bivariate Laplace w/ Covariance	0.3853	0.7862	**0.0803**
MapTR [4] + Bivariate Gaussian w/ Covariance (Ours)	**0.3851**	**0.7812**	0.0815

TABLE III: Investigating the impact of modeling uncertainty using four different probability distributions on trajectory prediction. The results show that the covariance-based joint Gaussian distribution provides the best modeling approach.

trajectory based on kinematic features improves prediction accuracy across all mapping-prediction combinations. By analyzing the values of w_{base} and w_{base}, we find that in stable vehicle states ($\Delta\theta \in [0,1)$), the average of w_{base} is 0.82, while in other cases, w_{base} has an average value of 0.37. This further demonstrates that our gating mechanism learns vehicle kinematic information from the trajectory and makes a more optimal trajectory.

3) Dynamic Proprioception vs. Exteroceptive Sensing: To validate the effectiveness of our dynamic proprioception-based approach, we compare it against exteroceptive sensing methods using the CLIP model [41] and ResNet50 [42]. Specifically, front camera images are inputs of the CLIP image encoder or ResNet50 to extract semantic features, which are then processed by a multi-layer perceptron to generate a weighted trajectory selection based on classification results. The road semantics are divided into four categories as ground truth: labels — *low-curvature turn*, *high-curvature turn*, *presence of other agents*, and *straight path* — each represented by one-hot encodings.

As shown in Table V, the CLIP-based and ResNet-

Prediction Method	HiVT [2]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [4]	0.4015	0.8404	0.0960
MapTR [4] + Unc [3]	0.3910	0.8049	0.0818
MapTR [4] + Gating	**0.3829** (−2.1%)	**0.7855** (−2.4%)	**0.0805** (−1.6%)
MapTRv2 [1]	0.4017	0.8406	0.0959
MapTRv2 [1] + Unc [3]	0.3913	0.8054	0.0819
MapTRv2 [1] + Gating	**0.3815** (−2.5%)	**0.7914** (−1.7%)	**0.0807** (−1.5%)
MapTRv2-CL [1]	0.3789	0.7859	0.0865
MapTRv2-CL [1] + Unc [3]	0.3674	0.7418	0.0739
MapTRv2-CL [1] + Gating	**0.3650** (−0.7%)	**0.7362** (−0.8%)	**0.0685** (−7.3%)
StreamMapNet [5]	0.3963	0.8223	0.0923
StreamMapNet [5] + Unc [3]	0.3899	0.8101	0.0861
StreamMapNet [5] + Gating	**0.3858** (−1.0%)	**0.7981** (−1.5%)	**0.0847** (−1.6%)

TABLE IV: Incorporating Gating method to dynamically leverage map uncertainty improves trajectory prediction consistently.

Prediction Method	HiVT [2]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [4]	0.4015	0.8404	0.0960
MapTR [4] + CLIP [41]	0.4200	0.8821	0.1040
MapTR [4] + ResNet50 [42]	0.4248	0.9008	0.1060
MapTR [4] + Ours	**0.3672**	**0.7395**	**0.0756**
MapTRv2 [1]	0.4017	0.8406	0.0959
MapTRv2 [1] + CLIP [41]	0.3988	0.8370	0.0969
MapTRv2 [1] + ResNet50 [42]	0.4191	0.8751	0.1041
MapTRv2 [1] + Ours	**0.3670**	**0.7538**	**0.0708**
MapTRv2-CL [1]	0.3789	0.7859	0.0865
MapTRv2-CL [1] + CLIP [41]	0.4433	0.9711	0.1124
MapTRv2-CL [1] + ResNet50 [42]	0.4211	0.9058	0.1052
MapTRv2-CL [1] + Ours	**0.3659**	**0.7404**	**0.0721**
StreamMapNet [5]	0.3963	0.8223	0.0923
StreamMapNet [5] + CLIP [41]	0.4191	0.9010	0.0926
StreamMapNet [5] + ResNet50 [42]	0.4199	0.9024	0.0988
StreamMapNet [5] + Ours	**0.3828**	**0.7981**	**0.0834**

TABLE V: Our Proprioceptive Scenario Gating outperforms the CLIP-based and ResNet-based exteroceptive alternatives in prediction accuracy across all mapping-prediction combinations.

Gating Method	CLIP [41]	ResNet50 [42]	Ours
Online HD Map Method	FPS	FPS	FPS
MapTR [4]	2.21	6.59	**52.26**
MapTRv2 [1]	2.15	6.51	**92.09**
MapTRv2-CL [1]	2.33	6.70	**87.35**
StreamMapNet [5]	2.42	6.91	**84.57**

TABLE VI: Our Proprioceptive Scenario Gating achieves FPS up to 30 times higher than the CLIP-based exteroceptive alternative approach and 10 times higher than the ResNet50-based exteroceptive method.

based exteroceptive approaches underperform our dynamic proprioception-based method across all models and metrics. In some cases, it even falls below baseline performance, suggesting that trajectory selection optimization based solely on semantic scene understanding may be insufficient. Additionally, Table VI highlights a substantial difference in FPS, with our method achieving speeds up to 30 times higher than CLIP and 10 times higher than ResNet50, making it more suitable for real-time deployment in vehicles.

E. Scalability

To validate the scalability of our proposed architecture, we replace the baseline (MapTR [4]) in the dual-stream structure with the BEV feature attention method [19], keeping the rest of the architecture unchanged. The results, as shown in Table VII, demonstrate that our approach achieves SOTA performance in trajectory prediction tasks, highlighting the strong generalization capability of our architecture.

Prediction Method	HiVT [2]		
Online HD Map Method	minADE ↓	minFDE ↓	MR ↓
MapTR [4] + Unc [3]	0.3910	0.8049	0.0818
MapTR [4] + BEV Feature Attention [19]	0.3617	0.7401	0.0720
MapTR [4] + CovMat	0.3851	0.7812	0.0815
MapTR [4] + Ours	**0.3494** (−3.5%)	**0.7028** (−5.6%)	**0.0653** (−9.3%)

TABLE VII: Repalcing the baseline (MapTR [4]) in our proposed dual-stream structure with a stronger alternative [19], showing significant improvements in prediction accuracy.

V. Conclusion

In this work, we explore the impact of map uncertainty for mapless trajectory prediction and identify the critical role of the vehicle's kinematic state in determining when such uncertainty is beneficial. Building on these insights, we introduce Proprioceptive Scenario Gating, a novel mechanism adaptively integrating map uncertainty into trajectory prediction based on forecasts of the vehicle's future kinematics. This self-supervised approach not only improves performance but also enhances interpretability by identifying scenarios where uncertainty contributes positively. Furthermore, we propose a Covariance-based Map Uncertainty modeling that better aligns with map geometry, leading to further performance improvements. Extensive experiments on the real-world data demonstrate the significant performance gains of our method.

References

[1] B. Liao, S. Chen, Y. Zhang, B. Jiang, Q. Zhang, W. Liu, C. Huang, and X. Wang, "MapTRv2: An end-to-end framework for online vectorized HD map construction," *IJCV*, pp. 1–23, 2023.

[2] Z. Zhou, L. Ye, J. Wang, K. Wu, and K. Lu, "HiVT: Hierarchical vector transformer for multi-agent motion prediction," in *CVPR*, pp. 8823–8833, 2022.

[3] X. Gu, G. Song, I. Gilitschenski, M. Pavone, and B. Ivanovic, "Producing and leveraging online map uncertainty in trajectory prediction," in *CVPR*, pp. 14521–14530, June 2024.

[4] B. Liao, S. Chen, X. Wang, T. Cheng, Q. Zhang, W. Liu, and C. Huang, "MapTR: Structured modeling and learning for online vectorized HD map construction," in *ICLR*, 2023.

[5] T. Yuan, Y. Liu, Y. Wang, Y. Wang, and H. Zhao, "StreamMapNet: Streaming mapping network for vectorized online HD map construction," in *WACV*, 2024.

[6] Q. Li, Y. Wang, Y. Wang, and H. Zhao, "HDMapNet: An online HD map construction and evaluation framework," in *ICRA*, pp. 4628–4634, IEEE, 2022.

[7] H. Dong, X. Zhang, J. Xu, R. Ai, W. Gu, H. Lu, J. Kannala, and X. Chen, "SuperFusion: Multilevel LiDAR-camera fusion for long-range HD map generation," *IEEE/CAA Journal of Automatica Sinica*, 2022.

[8] Y. Liu, Y. Yuantian, Y. Wang, Y. Wang, and H. Zhao, "VectorMapNet: End-to-end vectorized HD map learning," in *ICML*, PMLR, 2023.

[9] W. Ding, L. Qiao, X. Qiu, and C. Zhang, "PivotNet: Vectorized pivot learning for end-to-end HD map construction," in *ICCV*, 2023.

[10] Z. Zhang, X. Li, S. Zou, G. Chi, S. Li, X. Qiu, G. Wang, G. Zheng, L. Wang, H. Zhao, *et al.*, "Chameleon: Fast-slow neuro-symbolic lane topology extraction," *arXiv preprint arXiv:2503.07485*, 2025.

[11] Z. Yan, P. Li, Z. Fu, S. Xu, Y. Shi, X. Chen, Y. Zheng, Y. Li, T. Liu, C. Li, *et al.*, "Int2: Interactive trajectory prediction at intersections," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8536–8547, 2023.

[12] X. Zheng, L. Wu, Z. Yan, Y. Tang, H. Zhao, C. Zhong, B. Chen, and J. Gong, "Large language models powered context-aware motion prediction in autonomous driving," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 980–985, IEEE, 2024.

[13] J. Gu, C. Sun, and H. Zhao, "DenseTNT: End-to-end trajectory prediction from dense goal sets," in *ICCV*, 2021.

[14] Z. Zhou, J. Wang, Y.-H. Li, and Y.-K. Huang, "Query-centric trajectory prediction," in *CVPR*, pp. 17863–17873, 2023.

[15] K. Ding, B. Chen, Y. Su, H.-a. Gao, B. Jin, C. Sima, W. Zhang, X. Li, P. Barsch, H. Li, *et al.*, "Hint-ad: Holistically aligned interpretability in end-to-end autonomous driving," *arXiv preprint arXiv:2409.06702*, 2024.

[16] J. Philion and S. Fidler, "Lift, Splat, Shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D," in *ECCV*, 2020.

[17] Z. Liu, H. Tang, A. Amini, X. Yang, H. Mao, D. Rus, and S. Han, "BEVFusion: Multi-task multi-sensor fusion with unified bird's-eye view representation," in *ICRA*, 2023.

[18] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Y. Qiao, and J. Dai, "BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," in *ECCV*, 2022.

[19] X. Gu, G. Song, I. Gilitschenski, M. Pavone, and B. Ivanovic, "Accelerating online mapping and behavior prediction via direct bev feature attention," in *ECCV*, 2024.

[20] B. Tian, L. Luo, H. Zhao, and G. Zhou, "Vibus: Data-efficient 3d scene parsing with viewpoint bottleneck and uncertainty-spectrum modeling," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 194, pp. 302–318, 2022.

[21] G. Zhang, H.-a. Gao, Z. Jiang, H. Zhao, and Z. Zheng, "Ctrl-u: Robust conditional image generation via uncertainty-aware reward modeling," *arXiv preprint arXiv:2410.11236*, 2024.

[22] H. Zhao, M. Lu, A. Yao, Y. Guo, Y. Chen, and L. Zhang, "Pointly-supervised scene parsing with uncertainty mixture," *Computer Vision and Image Understanding*, vol. 200, p. 103040, 2020.

[23] X. Tang, K. Jiang, M. Yang, Z. Liu, P. Jia, B. Wijaya, T. Wen, L. Cui, and D. Yang, "High-definition maps construction based on visual sensor: A comprehensive survey," *IEEE Transactions on Intelligent Vehicles*, 2023.

[24] Y. B. Can, A. Liniger, D. P. Paudel, and L. V. Gool, "Structured bird's-eye-view traffic scene understanding from onboard images," in *ICCV*, 2021.

[25] T. Li, P. Jia, B. Wang, L. Chen, K. Jiang, J. Yan, and H. Li, "Lane-segnet: Map learning with lane segment perception for autonomous driving," *arXiv preprint arXiv:2312.16108*, 2023.

[26] Z. Zhang, Y. Zhang, X. Ding, F. Jin, and X. Yue, "Online vectorized hd map construction using geometry," *ECCV*, 2024.

[27] X. Liu, S. Wang, W. Li, R. Yang, J. Chen, and J. Zhu, "Mgmap: Mask-guided learning for online vectorized hd map construction," in *CVPR*, pp. 14812–14821, 2024.

[28] Y. Zhou, H. Zhang, J. Yu, Y. Yang, S. Jung, S.-I. Park, and B. Yoo, "Himap: Hybrid representation learning for end-to-end vectorized hd map construction," in *CVPR*, pp. 15396–15406, 2024.

[29] Z. Liu, X. Zhang, G. Liu, J. Zhao, and N. Xu, "Leveraging enhanced queries of point sets for vectorized map construction," in *ECCV*, pp. 461–477, Springer, 2025.

[30] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in *CVPR*, 2020.

[31] T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom, and E. M. Wolff, "CoverNet: Multimodal behavior prediction using trajectory sets," in *CVPR*, 2020.

[32] Y. Yuan, X. Weng, Y. Ou, and K. M. Kitani, "AgentFormer: Agent-aware transformers for socio-temporal multi-agent forecasting," in *ICCV*, pp. 9813–9823, 2021.

[33] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "HOME: Heatmap output for future motion estimation," in *ITSC*, 2021.

[34] B. Ivanovic, J. Harrison, and M. Pavone, "Expanding the deployment envelope of behavior prediction via adaptive meta-learning," in *ICRA*, 2023.

[35] J. Gao, C. Sun, H. Zhao, Y. Shen, D. Anguelov, C. Li, and C. Schmid, "VectorNet: Encoding HD maps and agent dynamics from vectorized representation," in *CVPR*, 2020.

[36] 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.

[37] H. Zhao, J. Gao, T. Lan, C. Sun, B. Sapp, B. Varadarajan, Y. Shen, Y. Shen, Y. Chai, C. Schmid, C. Li, and D. Anguelov, "TNT: Target-driveN Trajectory Prediction," in *CoRL*, 2020.

[38] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, "Argoverse: 3d tracking and forecasting with rich maps," in *CVPR*, 2019.

[39] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes, D. Ramanan, P. Carr, and J. Hays, "Argoverse 2: Next generation datasets for self-driving perception and forecasting," in *NIPSDB*, 2021.

[40] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *CVPR*, pp. 2446–2454, 2020.

[41] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *ICML*, pp. 8748–8763, PMLR, 2021.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, pp. 770–778, 2016.