FSGMap: Frequency-Space Priors Gated Mixture-of-Experts for HD Map Construction

1st Gaosheng Sun
School of Artificial Intelligence
Anhui University
Hefei, China
wa24301156@stu.ahu.edu.cn

2nd Luyang Ye
School of Artificial Intelligence
Anhui University
Hefei, China
w124301206@stu.ahu.edu.cn

3rd Yanxu Su*
School of Artificial Intelligence
Anhui University
Hefei, China
yanxu.su@ahu.edu.cn

***Abstract*—High-definition (HD) maps are crucial for autonomous driving. While recent online map construction methods have gained popularity, they often suffer from critical failures in complex scenarios, including directional distortion, missing map elements, and inaccurate keypoint localization. These issues stem from a uniform processing approach that fails to distinguish between heterogeneous map characteristics, such as sharp geometric boundaries and smooth semantic regions. In this paper, we propose FSGMap, a novel framework based on Frequency-Space Priors Gated Mixture-of-Experts. Our approach introduces Frequency-Space Dual Prior Modulation (FSDM) to generate orientation-aware priors and semantic-aware spectral composition priors. To account for prediction reliability under ambiguous frequency characteristics, an Uncertainty-Aware Guidance (UAG) mechanism is integrated to adaptively guide map queries based on entropy-based reliability hints. Furthermore, we design a Prior Refinement Mixture-of-Experts (PR-MoE) that employs hierarchical routing to enable coarse-to-fine expert collaboration. These structured priors and adaptive refinement significantly enhance the model's ability to capture precise topology and reliable geometry. FSGMap achieves exceptional results on the challenging nuScenes dataset, outperforming several widely recognized approaches and improving the baseline model MapTRv2 by 4.5 mAP.**

***Index Terms*—Autonomous driving, Computer vision, HD map construction**

I. Introduction

High-definition (HD) maps provide critical geometric and semantic priors for autonomous driving, supporting localization, motion prediction, and planning. While traditional HD maps rely on expensive offline surveys and manual annotation, recent research has shifted towards *online HD map construction*. This paradigm leverages onboard multi-view sensors to generate vectorized map elements in real-time, significantly enhancing the scalability and adaptability of autonomous vehicles in dynamic environments.

Despite the progress made by DETR-like architectures [1] in lifting image features to the bird's-eye-view (BEV) space [2]–[4], existing methods, e.g., [5]–[8], still struggle in complex driving scenes. As illustrated in Fig. 1, current frameworks often suffer from three critical failures: (*i*) **Directional Distortion**, where predicted polylines exhibit inconsistent orientations or deviate from the actual road curvature; (*ii*) **Missing Map Elements**, where certain instances fail to be detected due to feature interference or environmental clutter; and (*iii*) **Inaccurate Keypoint Localization**, resulting in jittery boundaries that lack the precision required for safe navigation. We observe that these issues fundamentally stem from a "one-size-fits-all" processing approach that ignores the heterogeneous characteristics of map elements. Specifically, geometric boundaries require sharp, orientation-aware high-frequency cues, while semantic regions depend on stable, low-frequency context. Treating them uniformly leads to feature entanglement and degraded performance.

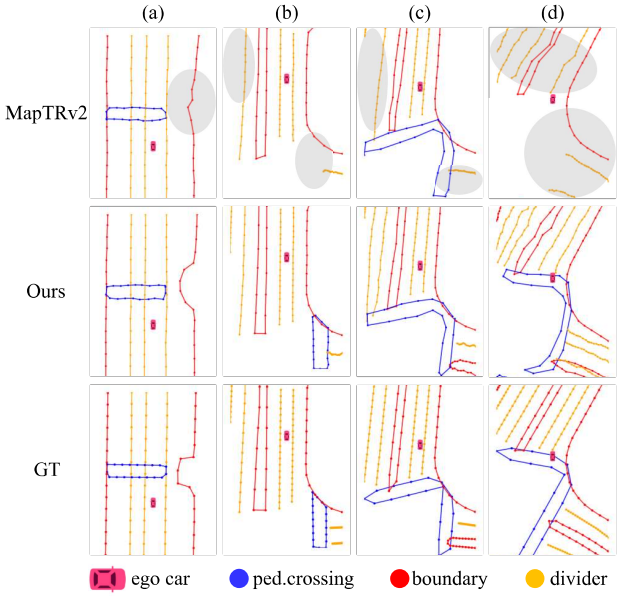

Fig. 1. Examples of previous failures (MapTRv2) and our improved results. Compared to the baseline, which suffers from directional distortion, missing map elements, and inaccurate keypoint localization, our proposed FSGMap performs better in disentangling geometric and semantic cues, producing consistent orientations and precise boundaries (highlighted in gray areas).

To address the above challenges, we propose **FSGMap**, a **Map** network built upon the **F**requency-**S**pace **G**ated Mixture-of-Experts. FSGMap is motivated by the insight that explicitly disentangling frequency-domain priors can provide stronger inductive biases for geometry and semantics. The framework

*Corresponding author

introduces three collaborative components: *Frequency-Space Dual Prior Modulation (FSDM)*, *Uncertainty-Aware Guidance (UAG)*, and *Prior Refinement Mixture-of-Experts (PR-MoE)*.

Specifically, FSDM leverages frequency spatial information to generate orientation-aware geometric priors and spectral composition priors. This allows the model to process sharp directional cues and smooth textures separately, directly addressing directional distortion. Regions with ambiguous frequency characteristics often occur in complex intersections and under sensor noise or occlusions. To account for prediction reliability in such regions, UAG computes entropy-based uncertainty signals to condition expert routing. Finally, PR-MoE implements a hierarchical expert collaboration strategy. By applying coarse-to-fine prior refinement at both inter-group and intra-group levels, the decoder can adaptively allocate specialized experts for different map elements, ensuring precise keypoint localization and reducing element omission.

The main contributions are summarized as follows:

- We propose **FSGMap**, a novel framework that leverages frequency-space priors to address common failures such as directional distortion, missing elements, and inaccurate localization in online HD map construction.
- To explicitly disentangle geometric and semantic representations, we introduce **Frequency-Space Dual Prior Modulation (FSDM)** to generate orientation-aware priors and semantic-aware spectral composition priors, and **Uncertainty-Aware Guidance (UAG)** to provide structured inductive biases for adaptive expert routing.
- A **Prior Refinement Mixture-of-Experts** decoder is further designed, where hierarchical routing enables fine-grained expert collaboration and significantly improves map accuracy and reliability.

II. Related Work

A. Vectorized Map Construction

HD map construction has evolved from raster-based post-processing methods like HDMapNet [9] to efficient end-to-end vectorized paradigms. VectorMapNet [10] pioneered this shift using a DETR-based decoder and auto-regressive optimization. Subsequently, MapTR [5] and MapTRv2 [6] established a unified one-stage framework with hierarchical query embeddings, serving as the foundation for modern approaches. Follow-up works, including HIMap [7], MapQR [8], InstaGraM [11] and SafeMap [12], further refine internal query interactions and instance-point correlations.

To enhance stability and reliability, other streams of research incorporate external information or temporal cues. Methods like P-MapNet [13], SMERF [14], and SDTagNet [15] exploit SD maps, with SDTagNet specifically leveraging text-annotated semantic features for far-range perception. Others, such as StreamMapNet [16], TICMapNet [17], MapTracker [18] and InteractionMap [19], utilize temporal fusion to accumulate geometric details. However, existing works overlook the potential of frequency-domain information for directional priors and boundary localization. We address these gaps by introducing structured frequency-component priors to promote context-aware predictions.

B. Priors for HD Map Construction

Priors provide crucial initialization and regularization for map learning. Existing methods generally utilize semantic or structural priors. For semantic guidance, MGMap [20], BeMapNet [21], and Bi-Mapper [22] employ masks or dual-stream architectures to refine feature localization and boundary smoothness. Structurally, HeightMapNet [23] integrates height priors in BEV space to improve spatial comprehension. DAMap [24] introduces the distance-aware focal loss and hybrid loss scheme to assign appropriate labels for matching samples. HoMap [25] enriches BEV features and models relationships between instances using high-order statistics.

In contrast to prior works that expand capacity via auxiliary branches or heavy additional modules, our method focuses on compact and effective prior encoding, which will be instantiated through a gated expert framework described next.

C. Mixture-of-Experts

Mixture of Experts (MoE) scales model capacity by dynamically routing inputs to specialized sub-networks, referred to as experts, via a sparse gating mechanism [26]–[28]. In HD map construction, MapExpert [29] applies a sparse MoE with a top-k gating strategy. Like traditional MoE approaches, MapExpert relies on a learned router network and an auxiliary load-balancing loss to prevent expert collapse.

In contrast, FSGMap advances the MoE paradigm by explicitly conditioning the routing process on structured priors. Instead of relying solely on a learned router and auxiliary load-balancing objectives, we inject uncertainty estimates and frequency-domain cues into the gating function, enabling experts to specialize in distinct and explainable sub-domains, such as high-uncertainty regions or high-frequency structural boundaries. This prior-conditioned routing mechanism leads to a more stable and interpretable mapping framework.

III. Method

A. Overview

Fig. 2 presents the overall framework of **FSGMap**. Given N_c synchronized multi-view images $\{I_i\}_{i=1}^{N_c}$, our framework outputs a set of N_m vectorized map elements $\{M_j\}_{j=1}^{N_m}$. Each map element is defined by a semantic class label c and an ordered polyline consisting of N_p control points $\{(x_k, y_k)\}_{k=1}^{N_p}$.

As illustrated in Fig. 2, the pipeline consists of sequential processing stages starting from the map encoder to the transformer decoder. First, the map encoder, utilizing a ResNet50 backbone, extracts multi-scale image features from the input images. These features are then projected into a unified BEV representation $F_{\text{BEV}} \in \mathbb{R}^{H \times W \times C}$ via an LSS-based PV-to-BEV view transformer.

Subsequently, the BEV features are processed by the Transformer Decoder. Following the attention layers of MapTRv2, the decoder integrates a Frequency-Space Encoder equipped with the FSDM and UAG modules, followed by a PR-MoE

module which replaces the standard Feed-Forward Network (FFN). Specifically, the FSDM module leverages frequency sub-bands and spatial reference points to generate orientation-aware priors and semantic-aware spectral composition priors. The UAG mechanism quantifies the uncertainty of these priors to guide the PR-MoE, enabling uncertainty-aware expert routing for final map element prediction. We detail the FSDM, UAG, and PR-MoE in Sec. III-B, Sec. III-C, and Sec. III-D, respectively.

B. Frequency-Space Dual Prior Modulation

Map elements in HD map construction inherently exhibit heterogeneous characteristics across frequency scales. Sharp geometric structures, such as lane boundaries, are dominated by high-frequency edge information, while semantic regions, like pedestrian crossing, are characterized by low-frequency continuity. Processing these distinct hints within a unified feature space often leads to feature entanglement, compromising both geometric precision and semantic consistency. To address this, we introduce the FSDM module. FSDM leverages Discrete Wavelet Transforms (DWT) to construct interpretable, frequency-aware priors that disentangle these characteristics before expert routing.

Given the backbone feature map, we first apply a DWT to decompose the representation into four sub-bands: the low-frequency approximation $\mathbf{F}_{LL}$, and high-frequency details capturing horizontal ($\mathbf{F}_{LH}$), vertical ($\mathbf{F}_{HL}$), and diagonal ($\mathbf{F}_{HH}$) variations. We align these spectral features with the object queries $\mathbf{q} \in \mathbb{R}^{B\times N\times D}$ by performing bilinear sampling $\mathcal{S}(\cdot)$ at the normalized reference points $\mathbf{p}$. The reference points are initialized following the standard DETR-style query formulation. This yields a set of location-specific spectral features $\mathcal{F} = \{\mathbf{f}_k \mid k \in \{LL, LH, HL, HH\}\}$, where $\mathbf{f}_k = \mathcal{S}(\mathbf{F}_k, \mathbf{p})$ represents the feature vector sampled from the corresponding sub-band $\mathbf{F}_k$.

The FSDM then generates priors via an attention mechanism, as illustrated in Fig. 3. Notably, the reference points $\mathbf{p}$ themselves are treated as a **spatial prior**, providing explicit spatial weight that complement the frequency-domain priors.

Orientation-Aware Prior. Lane boundaries exhibit strong directionality. To capture this, we first compute an orientation weight $\boldsymbol{\alpha} \in \mathbb{R}^{B\times N\times 3}$ to resolve the directional ambiguity among the high-frequency components. $\mathbf{q}_s$ denotes a small lightweight projected query embedding, obtained by applying a linear projection to the decoder query $\mathbf{q}$ to reduce computational overhead during prior estimation. By concatenating the projected query $\mathbf{q}_s$ with the directional sub-bands, we obtain:

$$\boldsymbol{\alpha} = \text{Softmax}\left(\text{MLP}_{\text{orient}}\left([\mathbf{q}_s, \mathbf{f}_{LH}, \mathbf{f}_{HL}, \mathbf{f}_{HH}]\right)\right). \tag{1}$$

These weights are used to synthesize a direction-aware high-frequency descriptor, $\mathbf{f}_{HF}^{agg}$, via weighted aggregation:

$$\mathbf{f}_{HF}^{agg} = \sum_{k\in\{LH,HL,HH\}} \boldsymbol{\alpha}_k \cdot \mathbf{f}_k. \tag{2}$$

Spectral Composition Prior. To determine the balance between semantic context and geometric detail, we compute a composition weight $\boldsymbol{\omega} \in \mathbb{R}^{B\times N\times 2}$. This weight arbitrates between the low-frequency context $\mathbf{f}_{LL}$ and the aggregated high-frequency details $\mathbf{f}_{HF}^{agg}$:

$$\boldsymbol{\omega} = \text{Softmax}\left(\text{MLP}_{\text{comp}}\left([\mathbf{q}_s, \mathbf{f}_{LL}, \mathbf{f}_{HF}^{agg}]\right)\right). \tag{3}$$

The resulting priors $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$ explicitly encode the directional and spectral preferences of each map element query, providing a structured inductive bias for the subsequent gating mechanism.

C. Uncertainty-Aware Guidance

While FSDM provides strong priors, static routing based solely on these priors may fail when the frequency characteristics are ambiguous. To account for prediction reliability during expert routing, we introduce the UAG mechanism.

UAG computes entropy-based uncertainty hints to characterize the uncertainty in the frequency-space prior distributions. We calculate the Shannon entropy for both the orientation weights $\boldsymbol{\alpha}$ and the composition weights $\boldsymbol{\omega}$ to measure the ambiguity in direction and frequency scale, respectively:

$$\mathcal{H}_{\text{orient}} = -\sum_{i=1}^{3} \boldsymbol{\alpha}_i \log(\boldsymbol{\alpha}_i+\epsilon), \quad \mathcal{H}_{\text{comp}} = -\sum_{j=1}^{2} \boldsymbol{\omega}_j \log(\boldsymbol{\omega}_j+\epsilon), \tag{4}$$

where ϵ is a small constant for numerical stability. The final uncertainty hint $\mathcal{H}_{\text{hint}}$ is the average of these two entropy terms:

$$\mathcal{H}_{\text{hint}} = \frac{1}{2}\left(\mathcal{H}_{\text{orient}} + \mathcal{H}_{\text{comp}}\right). \tag{5}$$

A higher $\mathcal{H}_{\text{hint}}$ indicates that the model is uncertain about the element's dominant direction or spectral composition. This scalar hint is injected into the MoE gating network as a continuous conditioning signal. Instead of imposing hard rules, this uncertainty-aware conditioning allows the gating network to learn an adaptive routing strategy: When uncertainty is low, the gate follows the FSDM priors; when uncertainty is high, the gate can distribute computation across diverse experts to improve representation stability.

D. Prior Refinement Mixture-of-Experts

In traditional mixture-of-experts (MoE) models, experts are often routed based on a fixed query-based routing mechanism, which can lead to inefficiencies when dealing with complex map elements. To address this, we propose the PR-MoE mechanism. This hierarchical routing design follows a coarse-to-fine prior refinement paradigm, where global structural and uncertainty-aware hints guide expert group selection, followed by fine-grained adaptation within each group.

Prior Hint Vector Encoder. The Prior Hint Vector Encoder effectively fuses heterogeneous prior hints into a compact latent representation to guide the routing process. Let $\mathcal{P}_{freq} = \{\mathbf{w}_{ori}, \mathbf{w}_{comp}\}$ denote the frequency-domain weights that capture orientation and composition information from the FSDM module, $\mathbf{p}_{sp}$ denote the spatial position hints, and u_{ent} represent the prediction uncertainty measured by entropy

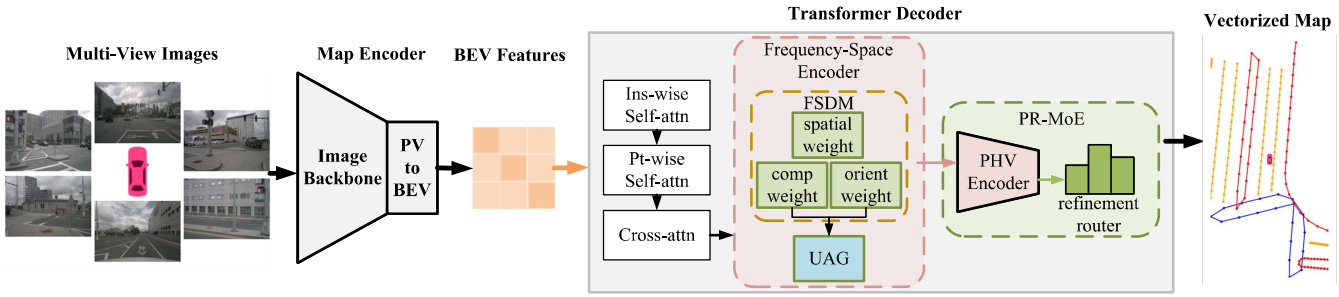

Fig. 2. Overview of FSGMap framework. Our approach follows an online vectorized HD map construction paradigm. Distinct from the baseline MapTRv2, we propose three key components: (1) Frequency-Space Dual Modulation (FSDM) module in a Frequency-Space Encoder; (2) Uncertainty-Aware Guidance (UAG) mechanism in the Frequency-Space Encoder; (3) Prior Refinement Mixture-of-Experts (PR-MoE), which replaces the standard feed-forward network (FFN) to enable uncertainty-aware expert routing.

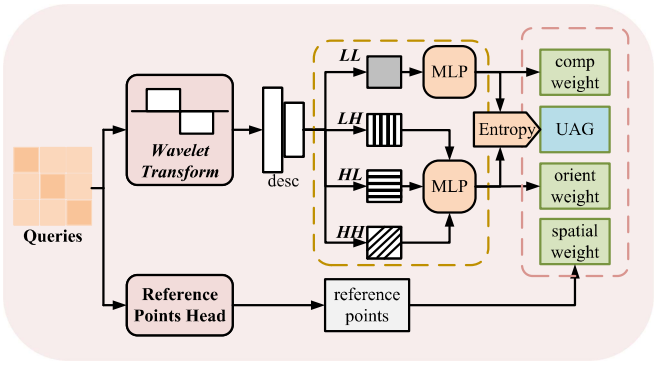

Fig. 3. Frequency-space encoder with Frequency-Space Dual Prior Modulation (FSDM) and Uncertainty-Aware Guidance (UAG). The *decs* denotes dimensionality decrease, *comp weight* indicates the spectral composition weight between high- and low-frequency components, and *orient weight* indicates orientation-aware weight.

from the UAG module. These inputs are processed through independent projection heads and then fused:

$$\begin{aligned} \mathbf{z}_{freq} &= \phi_{ori}(\mathbf{w}_{ori}) \oplus \phi_{comp}(\mathbf{w}_{comp}), \\ \mathbf{z}_{sp} &= \phi_{sp}(\mathbf{p}_{sp}), \\ \mathbf{z}_{ent} &= \phi_{ent}(u_{ent}), \\ \mathbf{h}_{prior} &= \mathbf{W}_p[\mathbf{z}_{freq}; \mathbf{z}_{sp}; \mathbf{z}_{ent}] + \mathbf{b}_p, \end{aligned} \tag{6}$$

where $\phi_{(\cdot)}$ denotes lightweight MLP encoders, $\oplus$ represents feature concatenation, and $[\cdot;\cdot]$ denotes concatenation along the channel dimension. The resulting $\mathbf{h}_{prior} \in \mathbb{R}^{d_h}$ serves as a comprehensive descriptor of map element characteristics.

Group Mixture-of-Experts with Refinement Router. As illustrated in Fig. 4, the core of PR-MoE is the hierarchical injection of the hint vector into the routing process. We introduce a *Prior Refinement Bias* that modulates the gating logits in a coarse-to-fine manner. First, we derive the **inter-group bias** $\mathbf{b}_{inter} \in \mathbb{R}^G$ and an **intra-group refinement bias** $\mathbf{b}_{intra} \in \mathbb{R}^E$:

$$\mathbf{b}_{inter} = \sigma(\text{MLP}_g(\mathbf{h}_{prior})), \quad \mathbf{b}_{intra} = \sigma(\text{MLP}_f(\mathbf{h}_{prior})), \tag{7}$$

where σ is the $\tanh$ activation function. To align the group-level bias with individual experts, we define a mapping $\mathcal{M}(\cdot)$ that expands $\mathbf{b}_{inter}$ to the expert dimension. The final expert bias $\boldsymbol{\beta} \in \mathbb{R}^E$ is:

$$\boldsymbol{\beta} = \mathcal{M}(\mathbf{b}_{inter}) + \lambda \cdot \mathbf{b}_{intra}, \tag{8}$$

where λ is the refinement scale. The gating network then computes the routing weights $g(\mathbf{x})$ by conditioning on both the input query $\mathbf{x}$ and $\mathbf{h}_{prior}$:

$$g(\mathbf{x}) = \text{Softmax}\left(\frac{\mathbf{W}_g[\mathbf{x}; \mathbf{h}_{prior}] + \boldsymbol{\beta} + \xi}{\tau}\right), \tag{9}$$

where τ is the temperature and ξ is the exploration noise. $g(\mathbf{x})$ denotes the pre-selection routing probabilities over all experts produced by the softmax gating network. Each expert $E_i(\cdot)$ utilizes a gated feed-forward network with SiLU activation. The final output $\mathbf{y}$ is the weighted sum of the top-k selected experts:

$$\begin{aligned} E_i(\mathbf{x}) &= (\text{SiLU}(\mathbf{x}\mathbf{W}_1^{(i)}) \odot (\mathbf{x}\mathbf{W}_2^{(i)}))\mathbf{W}_3^{(i)}, \\ \mathbf{y} &= \sum_{i \in \text{Top-}k} g(\mathbf{x})_i \cdot E_i(\mathbf{x}) + \mathbf{x}, \end{aligned} \tag{10}$$

where $\odot$ denotes element-wise multiplication. The top-k experts are selected according to the highest values in $g(\mathbf{x})$, and the corresponding weights $g(\mathbf{x})_i$ are renormalized over the selected experts before aggregation. This hierarchical routing ensures that expert selection is both structurally guided and contextually refined.

E. Training Loss

Inspired by MapTRv2 [6], we supervise training with three loss terms: a one-to-one instance prediction loss $\mathcal{L}_{\text{one2one}}$, an auxiliary one-to-many loss $\mathcal{L}_{\text{one2many}}$, and a dense foreground segmentation loss $\mathcal{L}_{\text{dense}}$.

Specifically, $\mathcal{L}_{\text{one2one}}$ employs the Hungarian algorithm with Manhattan distance for bipartite matching, incorporating classification, point-wise regression, and edge direction losses. $\mathcal{L}_{\text{one2many}}$ introduces an additional prediction branch to generate multiple hypotheses per instance, improving convergence

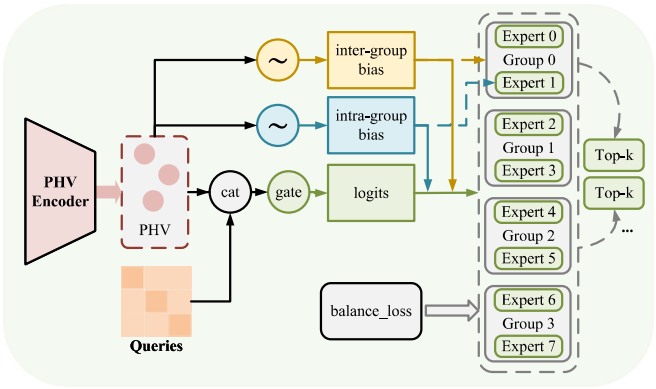

Fig. 4. Prior Refinement Mixture-of-Experts(PR-MoE). $\sim$ denotes an MLP with an activation function, and *PHV* stands for the Prior Hint Vector.

and training stability. $\mathcal{L}_{\text{dense}}$ consists of BEV and perspective-view segmentation losses to provide dense foreground supervision.

To encourage balanced expert utilization in the Mixture-of-Experts architecture, we further introduce an auxiliary expert balance loss $\mathcal{L}_{\text{aux}}$ following MapExpert [29]. The overall training objective is defined as:

$$\mathcal{L} = \lambda_o \mathcal{L}_{\text{one2one}} + \lambda_m \mathcal{L}_{\text{one2many}} + \lambda_d \mathcal{L}_{\text{dense}} + \lambda_a \mathcal{L}_{\text{aux}}, \quad (11)$$

where λ_o, λ_m, λ_d and λ_a are loss weights.

IV. Experiments

A. Dataset and Evaluation Metric

We evaluate our method on two widely adopted benchmarks for HD vectorized map construction, nuScenes [30] and Argoverse 2 [31]. The nuScenes dataset contains 1,000 urban driving scenes captured with a six-camera surround-view system and provides 2D vectorized map elements as ground truth. Argoverse 2 consists of 1,000 driving sequences of 15 seconds each, collected using seven multi-view cameras, and offers high-quality 3D vectorized map annotations for detailed road geometry representation.

For a fair comparison, we follow the standard metric used in previous works [6], [8] and evaluate three kinds of map elements—pedestrian crossing (*ped*), lane divider (*div*), and road boundary (*bou*). "EB0" and "R50" correspond to the backbones Efficient-B0 and ResNet50. With the ego-coordination as the center, the perception ranges are $[-15\,\text{m}, 15\,\text{m}]$ for the X-axis and $[-30\,\text{m}, 30\,\text{m}]$ for the Y-axis. We adopt the mean Average Precision (mAP) to evaluate the map vectorization quality based on the chamfer distance. A candidate is considered as True-Positive (TP) only if the chamfer distance to ground truth is less than certain thresholds ($t \in T, T = \{0.5, 1.0, 1.5\}$). The final AP metric is calculated by averaging across all thresholds and all classes.

B. Implementation Detail

We follow the implementation of MapTRv2, using $N_{\text{ins}} = 100$ instance queries, each predicting $N_p = 20$ points. The Transformer decoder is configured with $N_{\text{head}} = 8$ attention heads and an embedding dimension of $D = 256$. For training, we adopt the AdamW optimizer with a cosine annealing learning rate schedule, starting from 3.5×10^{-4}. The model is trained for 24 epochs on nuScenes and 6 epochs on Argoverse 2. The per-GPU batch size is set to 4, and all experiments are conducted on 2 NVIDIA RTX 4090 GPUs. We fix $\tau = 1.0$ in all experiments and observe stable performance. For loss weights, $\lambda_a = 0.001$, and the others are same as MapTRv2.

C. Main Results

We compare **FSGMap** with several recent camera-based online HD map construction methods on the nuScenes dataset, as summarized in Tab. I. FSGMap is designed as a single-frame online mapping framework. Temporal fusion methods, e.g., StreamMapNet, MapTracker, are not included in the comparison due to their reliance on multi-frame memory and additional latency. Under the same experimental protocol, FSGMap consistently outperforms existing approaches in terms of detection accuracy. In particular, our method achieves a **4.5 mAP** improvement over the strong baseline MapTRv2, demonstrating the effectiveness of the proposed frequency–space gated mixture-of-experts design.

In addition, when evaluated on a single RTX 4090 GPU, our model runs at **15.6 FPS**, meeting the real-time requirements of online mapping systems. In comparison with several widely recognized methods, FSGMap exhibits superior performance under a similar parameter budget. Moreover, FSGMap achieves performance comparable to MapQR with a significantly reduced number of parameters, only **55.0M**.

Meanwhile, we also evaluate our framework on the Argoverse 2 dataset. Argoverse 2 provides a 3D vectorized map, which has extra height information compared to the nuScenes dataset. As shown in Tab. II, in terms of 2D vectorized map construction, FSGMap achieves 69.0 mAP, which is 1.6 mAP higher than MapTRv2. In terms of 3D vectorized map construction, FSGMap achieves 66.2 mAP, which is 1.5 mAP higher than MapTRv2. The experiments on Argoverse 2 dataset demonstrates the generalization ability of FSGMap in terms of vectorized map construction.

These results indicate that our method achieves a favorable trade-off between accuracy, efficiency, and model compactness, making it well suited for practical deployment in autonomous driving systems.

D. Ablation Study

Effectiveness of Key Components. We first evaluate the contribution of each key component in FSGMap by progressively introducing them into the baseline. As shown in Tab. III, incorporating PR-MoE improves 3.3 mAP, validating its effectiveness in collaborative refinement of map instances through specialized experts. The inclusion of spatial, orientation, and composition weights in FSDM demonstrates the benefit of jointly modeling spatial and frequency-domain representations for capturing geometric structures. Adding UAG further boosts performance, suggesting that uncertainty-guided expert routing

TABLE I
QUANTITATIVE RESULTS ON NUSCENES DATASET

Method	Reference	Backbone	Epoch	AP_{ped}	AP_{div}	AP_{bou}	mAP	FPS	Params(M)
HDMapNet [9]	ICRA2022	EB0	30	14.4	21.7	33.0	23.0	–	–
MapTRv2 [6]	IJCV2024	R50	24	59.8	62.4	62.4	61.5	23.6	40.3
MGMap [20]	CVPR2024	R50	24	61.8	65.0	67.5	64.8	20.6	55.9
MapQR [8]	ECCV2024	R50	24	63.4	68.0	67.7	66.4	20.2	125.3
HeightMapNet [23]	WACV2025	R50	24	60.4	62.6	62.2	61.7	–	–
DAMap+MapTRv2 [24]	ICCV2025	R50	24	58.5	64.7	65.1	62.8	21.1	52.0
HoMap [25]	TIV2025	R50	24	65.5	62.5	66.6	64.8	–	–
MapTRv2+ours	–	R50	24	64.5	67.8	65.6	66.0(**+4.5**)	15.6	55.0

TABLE II
QUANTITATIVE RESULTS ON ARGOVERSE 2 DATASET

Method	dim	AP_{ped}	AP_{div}	AP_{bou}	mAP
MapTRv2 [6]	2d	62.9	72.1	67.1	67.4
MapTRv2+Ours	2d	65.5	73.2	68.2	69.0
MapTRv2 [6]	3d	60.7	68.9	64.5	64.7
MapTRv2+Ours	3d	62.1	71.0	65.5	66.2

TABLE III
ABLATION STUDY ON KEY COMPONENTS OF FSGMAP

FSDM			UAG	PR-MoE	mAP
spatial	*orientation*	*composition*			
					61.5
				✓	64.8
✓				✓	65.0
✓	✓			✓	65.4
✓	✓	✓		✓	65.8
✓	✓	✓	✓	✓	66.0

TABLE IV
ABLATION STUDY ON PR-MOE INTERNAL DESIGN

Refinement Bias	Grouping Strategy	Top-k / Experts	mAP
	No Grouping	2 / 8	65.4
✓	Unbalanced Grouping (1–2–2–3)	2 / 8	65.6
✓	Unbalanced Grouping (1–1–2–4)	2 / 8	65.1
✓	Average Grouping (2–2–2–2)	2 / 8	66.0

TABLE V
ABLATION STUDY ON BALANCE LOSS WEIGHT AND EXPLORATION NOISE

$\xi = 0$		$\lambda_a = 0.001$	
λ_a	mAP	ξ	mAP
0	64.3	0	65.5
0.001	65.5	0.001	65.7
0.005	64.9	0.01	**66.0**
0.01	63.5	0.1	64.4

provides more stable expert selection when frequency priors become ambiguous.

PR-MoE Design Variants. We further analyze different design choices of PR-MoE, including the refinement-bias and expert grouping strategies. As reported in Tab. IV, removing the refinement-bias leads to a clear performance drop, highlighting its role as a structured prior for guiding expert refinement. We also compare different expert grouping strategies. For unbalanced grouping, experts are deterministically partitioned into groups of unequal sizes, e.g., 1–2–2–3. The average grouping strategy consistently outperforms unbalanced grouping, indicating that balanced expert collaboration is critical for stable and effective refinement. Unless otherwise stated, the number of experts is fixed to 8 and top-k routing is set to 2, following MapExpert [29].

Effect of balance loss weight and exploration noise. We analyze the impact of the balance loss weight λ_a and the exploration noise ξ used in expert routing. As shown in Tab. V, introducing a small balance loss significantly improves expert utilization and overall performance, while overly large λ_a leads to performance degradation due to excessive regularization. Similarly, a moderate amount of exploration noise facilitates diversified expert activation and improves robustness, whereas larger noise values destabilize routing decisions. Based on these observations, we adopt $\lambda_a = 0.001$ and $\xi = 0.01$ in all experiments.

V. CONCLUSION AND FUTURE WORK

In this paper, we present **FSGMap**, which effectively addresses directional distortion, missing map elements, and localization inaccuracies in online HD mapping by leveraging frequency-space priors through Frequency-Space Priors Gated Mixture-of-Experts. By achieving a 4.5 mAP improvement over the MapTRv2 baseline on the nuScenes dataset, our approach demonstrates superior precision and more stable structural representations. Future research will extend this framework to temporal and multi-modal fusion to further enhance topological consistency and reliability in complex driving environments.

REFERENCES

[1] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*. Springer, 2020, pp. 213–229.

[2] Y. Li, Z. Ge, G. Yu, J. Yang, Z. Wang, Y. Shi, J. Sun, and Z. Li, "BEVDepth: Acquisition of reliable depth for multi-view 3D object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 1477–1485.

[3] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Q. Yu, and J. Dai, "BEVFormer: Learning bird's-eye-view representation from lidar-camera via spatiotemporal transformers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 3, pp. 2020–2036, 2025.

[4] T. Liang, H. Xie, K. Yu, Z. Xia, Z. Lin, Y. Wang, T. Tang, B. Wang, and Z. Tang, "BEVFusion: A simple and robust lidar-camera fusion framework," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10 421–10 434, 2022.

[5] 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 *International Conference on Learning Representations*, 2023.

[6] 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," *International Journal of Computer Vision*, vol. 133, no. 3, pp. 1352–1374, 2025.

[7] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15 396–15 406.

[8] Z. Liu, X. Zhang, G. Liu, J. Zhao, and N. Xu, "Leveraging enhanced queries of point sets for vectorized map construction," in *European Conference on Computer Vision*. Springer, 2024, pp. 461–477.

[9] Q. Li, Y. Wang, Y. Wang, and H. Zhao, "HDMapNet: An online HD map construction and evaluation framework," in *2022 International Conference on Robotics and Automation*. IEEE, 2022, pp. 4628–4634.

[10] Y. Liu, T. Yuan, Y. Wang, Y. Wang, and H. Zhao, "VectorMapNet: End-to-end vectorized HD map learning," in *International Conference on Machine Learning*. PMLR, 2023, pp. 22 352–22 369.

[11] J. Shin, H. Jeong, F. Rameau, and D. Kum, "InstaGraM: Instance-level graph modeling for vectorized HD map learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 2, pp. 1889–1899, 2025.

[12] X. Hao, L. Kong, R. Yin, P. Wang, J. Zhang, Y. Diao, and S. Zhao, "SafeMap: Robust HD map construction from incomplete observations," in *Forty-second International Conference on Machine Learning*, 2025.

[13] 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, no. 10, pp. 8539–8546, 2024.

[14] 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 *IEEE International Conference on Robotics and Automation*, 2024, pp. 4029–4035.

[15] F. Immel, J.-H. Pauls, R. Fehler, F. Bieder, J. Merkert, and C. Stiller, "SDTagNet: Leveraging text-annotated navigation maps for online HD map construction," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[16] T. Yuan, Y. Liu, Y. Wang, Y. Wang, and H. Zhao, "StreamMapNet: Streaming mapping network for vectorized online HD map construction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 7356–7365.

[17] W. Qiu, S. Pang, H. Zhang, J. Fang, and J. Xue, "TICMapNet: A tightly coupled temporal fusion pipeline for vectorized HD map learning," *IEEE Robotics and Automation Letters*, vol. 9, no. 12, pp. 11 289–11 296, 2024.

[18] J. Chen, Y. Wu, J. Tan, H. Ma, and Y. Furukawa, "MapTracker: Tracking with strided memory fusion for consistent vector HD mapping," in *European Conference on Computer Vision*. Springer, 2024, pp. 90–107.

[19] K. Wu, C. Yang, and Z. Li, "InteractionMap: Improving online vectorized HDMap construction with interaction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025, pp. 17 176–17 186.

[20] X. Liu, S. Wang, W. Li, R. Yang, J. Chen, and J. Zhu, "MGMap: Mask-guided learning for online vectorized HD map construction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 14 812–14 821.

[21] L. Qiao, W. Ding, X. Qiu, and C. Zhang, "End-to-end vectorized HD-map construction with piecewise Bezier curve," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 218–13 228.

[22] S. Li, K. Yang, H. Shi, J. Zhang, J. Lin, Z. Teng, and Z. Li, "BiMapper: Holistic BEV semantic mapping for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7034–7041, 2023.

[23] W. Qiu, S. Pang, H. Zhang, J. Fang, and J. Xue, "HeightMapNet: Explicit height modeling for end-to-end HD map learning," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2025, pp. 6022–6031.

[24] J. Dong, C. Li, Y. Lin, J. Fu, S. Zhou, and N. Zheng, "DAMap: Distance-aware mapnet for high quality HD map construction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 5285–5294.

[25] Y. Cai, W. Dong, Z. Liu, H. Wang, and L. Chen, "HoMap: End-to-end vectorized HD map construction with high-order modeling," *IEEE Transactions on Intelligent Vehicles*, vol. 10, no. 5, pp. 2987–2997, 2025.

[26] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," *arXiv preprint arXiv:1701.06538*, 2017.

[27] C. Riquelme, J. Puigcerver, B. Mustafa, M. Neumann, R. Jenatton, A. Susano Pinto, D. Keysers, and N. Houlsby, "Scaling vision with sparse mixture of experts," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8583–8595, 2021.

[28] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *Journal of Machine Learning Research*, vol. 23, no. 120, pp. 1–39, 2022.

[29] D. Zhang, D. Chen, P. Zhi, Y. Chen, Z. Yuan, C. Li, R. Zhou, Q. Zhou *et al.*, "MapExpert: Online HD map construction with simple and efficient sparse map element expert," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 14 745–14 753.

[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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 621–11 631.

[31] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. Kaesemodel Pontes, D. Ramanan, P. Carr, and J. Hays, "Argoverse 2: Next generation datasets for self-driving perception and forecasting," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, J. Vanschoren and S. Yeung, Eds., vol. 1, 2021.