IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

M3TR: A Generalist Model for Real-World HD Map Completion

Fabian Immel[1], Richard Fehler[1], Frank Bieder[1], Jan-Hendrik Pauls[2] and Christoph Stiller[2]

***Abstract*—Autonomous vehicles rely on HD maps for their operation, but offline HD maps eventually become outdated. For this reason, online HD map construction methods use live sensor data to infer map information instead. Research on real map changes shows that oftentimes entire parts of an HD map remain unchanged and can be used as a prior. We therefore introduce M3TR (Multi-Masking Map Transformer), a generalist approach for HD map completion both with and without offline HD map priors. As a necessary foundation, we address shortcomings in ground truth labels for Argoverse 2 and nuScenes and propose the first comprehensive benchmark for HD map completion. Unlike existing models that specialize in a single kind of map change, which is unrealistic for deployment, our Generalist model handles all kinds of changes, matching the effectiveness of Expert models. With our map masking as augmentation regime, we can even achieve a $+1.4$ mAP improvement without a prior. Finally, by fully utilizing prior HD map elements and optimizing query designs, M3TR outperforms existing methods by $+4.3$ mAP while being the first real-world deployable model for offline HD map priors.** https://github.com/immel-f/m3tr

***Index Terms*—Intelligent Transportation Systems; Deep Learning for Visual Perception; Computer Vision for Transportation**

I. Introduction

In order to drive safely, autonomous vehicles need to understand the geometry and topology of the roads as well as the traffic rules that apply to them. Current systems employ detailed semantic high-definition (HD) maps that provide this rich knowledge, but are primarily created using offline SLAM approaches. However, maintaining such offline HD maps to account for changes in no time is infeasible. Therefore, recent advances in computer vision aim to perceive HD map information with onboard sensors [1]–[6].

This task of online vectorized *HD map construction* uses sensor data, *e.g.* from cameras or LiDAR sensors, to detect vectorized map elements (lane markings, road borders, *etc.*) with their semantic meaning. When compared to offline HD planning maps however, online HD map construction models still miss important information. Examples for this are lane marking types (solid vs. dashed) [1]–[6], lane centerlines [1]–[3], [5], [6] and 3D instances [1]–[3], [5], [6].

Recent work [7] demonstrated that maps become outdated only gradually, leaving parts of offline map information still valid for use as a prior. Map changes, like construction sites, usually invalidate semantically coherent elements like specific lanes, while leaving the rest unchanged (see also Fig. 2). Combined with methods for detecting valid map elements [8], [9], which are outside of the scope of this work, this leads to the situation where map perception models must fill in invalid areas using the remaining HD map and sensor data, a task which we refer to as *HD map completion*.

Existing work incorporating prior information falls short for three main reasons: Detection transformer queries are used to provide vectorized priors to the model [10], but fail to fully utilize all map information. Furthermore, current approaches lack a clear task definition and evaluation metric that differentiates prior map elements and those needing to be perceived online. Finally, previous models specialize on a single map prior type that is assumed to be known in advance. Since any part of an offline HD map could change, this is unrealistic for real-world deployment.

Contributions

To address these points, we present M3TR (Multi-Masking Map Transformer), a generalist HD map completion model with the following contributions:

- A new HD map completion benchmark for models with prior offline HD map information. This includes semantically richer labels and the first metric that explicitly focuses on the performance for elements *without* a prior.
- We propose a novel query design to incorporate map priors on a point query and query set level that considerably improves detection performance on the Argoverse 2 and nuScenes datasets by up to +4.3 mAP.
- We introduce a novel training regime which yields a single model that can make use of any HD map prior. This *Generalist* model achieves performance on par with specialized models without needing to know which kind of map information is available, even improving performance without a prior by up to 1.4 mAP.

II. Related Work

Related work can be grouped into two main categories: Common online HD map construction methods without priors and methods that use prior vectorized map information.

Manuscript received: May, 27, 2025; Revised August, 22, 2025; Accepted September, 25, 2025.

This paper was recommended for publication by Editor Abhinav Valada upon evaluation of the Associate Editor and Reviewers' comments.

This work was supported by the German Federal Ministry of Education and Research (BMBF) within the project HAIBrid (FKZ 01IS21096D) and by the just better DATA (jbDATA) project supported by the German Federal Ministry for Economic Affairs and Climate Action of Germany (BMWK) and the European Union, grant number 19A23003H. We thank our research partner Mercedes-Benz AG for the fruitful collaboration. We also gratefully acknowledge financial support and computing resources provided by the Helmholtz Association's Initiative and Networking Fund on HAICORE@FZJ.

[1]Fabian Immel, Richard Fehler and Frank Bieder are with the FZI Research Center for Information Technology, Germany {immel, fehler, bieder}@fzi.de

[2]Jan-Hendrik Pauls and Christoph Stiller are with the Institute of Measurement and Control Systems, Karlsruhe Institute of Technology, Germany {jan-hendrik.pauls, stiller}@kit.edu

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

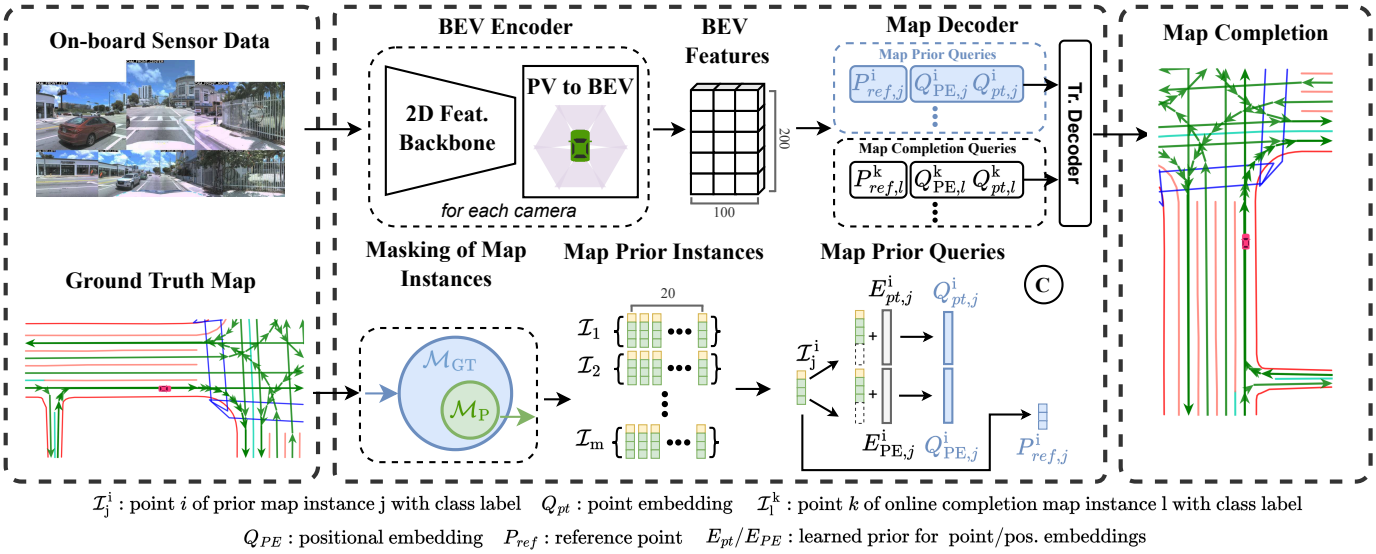


Fig. 1: Overview of the model architecture of M3TR and the investigated point query encoder designs. For our evaluated task of HD map completion, we mask out instances from the ground truth map $\mathcal{M}_{\mathrm{GT}}$ to create a map prior $\mathcal{M}_{\mathrm{P}}$. Using $\mathcal{M}_{\mathrm{P}}$, we try to reconstruct $\mathcal{M}_{\mathrm{GT}}$. The map prior instances are supplied to the model as queries, using the shown point query design and the detection query set design, latter of which is further illustrated in Fig. 4.

A. Online HD Map Construction without Priors

Detection transformer (DETR) [11] based architectures can be used to provide vectorized map element detections, handling HD map polyline and polygon elements in their original sparse representation. To detect map elements in the surrounding scene, a bird's eye view (BEV) feature grid, representing a fixed environment area, is generated by transforming 2D image features with methods proposed in general 3D object detection and BEV segmentation [12]–[14]. MapTR [3], [4] implements fast detection for complete map elements by modifying the original object queries of the transformer decoder to represent polylines and polygons with a fixed number of points. This enables fast parallel transformer decoding in contrast to early autoregressive approaches like VectorMapNet [2].

Recent contributions in the online HD map construction task show two significant improvements to the MapTR and MapTRv2 baselines, concentrating on query design and formulation. The first set [6], [15] improves single-shot detection performance by utilizing complete map element shapes and masks in the detection query representation.

The second set [5], [16] extends the single-shot detection task to the temporal and spatial context of past time steps.

B. Online HD Map Construction with Priors

Methods discussed in the previous section take only sensor data into account. In real-world autonomous systems, maps ranging from navigation maps to HD maps are used for at least routing, extending to motion prediction, path planning, and other driving tasks. Since this map becomes outdated piece by piece, online HD map construction that utilizes still up-to-date parts of maps as an optional prior is an attractive solution from an application perspective.

MapEX [10] was among the first to propose a detection query design allowing for both existing map element transformer queries as prior and regular learned transformer queries used for detecting unknown map elements. We use it as a baseline, but improve not only its evaluation scheme, but also the query design and model capabilities.

PriorDrive [17] proposes a HD map construction framework which integrates either SD navigation maps, incomplete HD maps or online constructed HD maps from previous drives at the same location. SMERF [18] incorporates a SD map prior by first encoding SD map elements with a transformer encoder and fusing them with the BEV feature grid via cross attention, showing improvements on the OpenLaneV2 [19] dataset detection and topology metrics. The approach of [10] to incorporate existing map prior with varying degradation levels was extended by [20] and consecutively [8] to use heavily modified map prior inputs to simulate outdated and incorrect map priors. This expands the training task to map verification, change detection, and map update, showing a significant sim-to-real gap on real public [7] or proprietary [20] data.

The limitations of prior work, mentioned in Sec. I, are addressed in the following section along with our improvements.

III. The HD Map Completion Benchmark

In this section we describe the novel HD map completion benchmark. Sec. III-A discusses our improved ground truth while Sec. III-B presents the HD map completion task.

A. Improved Ground Truth Maps

Most recent HD map construction models are trained using labels that have largely been unchanged since VectorMapNet [2] despite having major shortcomings.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

The labels do not include information from the Argoverse 2 HD maps that is necessary for autonomous driving. Additionally, issues in the label generation algorithms introduce errors into the ground truth instances and a geographic overlap leads to leakage between training and evaluation data. Fixes to these issues have been proposed in different works [4], [5], [21], [22], however they are scattered and not united in one single ground truth set.

As a foundation for the proposed HD map completion benchmark, we combine these improvements into one label set, together with a novel separation of dashed and solid dividers, leveraging [23] for the label generation in Argoverse 2. A comparison with previously used labels is listed in Tab. I.

B. The HD Map Completion Task

Our focus in this work is on using offline HD maps as priors that became outdated and thus partially invalid. This idea is based on the largest public dataset of real map changes, Trust but Verify [7]. When applying existing work to outdated offline HD maps, three open issues arise: Map changes in public datasets are not labeled on a point level and rare, requiring map priors to be derived synthetically [7]. Unfortunately, the map change generation schemes of previous approaches [8], [10], [17] follow assumptions that are not applicable for outdated offline HD maps. Real changes do not occur at random, but rather follow a local pattern with semantic correlation [20], that only affects specific elements and leaves most elements unchanged. This was also observed in [7], where changes remove, modify or add semantically coherent elements rather than randomly drop/add elements or apply noise.

To better align our synthetic priors with real priors, we define adapted map prior scenarios $\mathcal{S}_p = (\mathcal{M}_p, \mathcal{D})$ consisting of a map prior $\mathcal{M}_p$ and sensor data $\mathcal{D}$. Map priors $\mathcal{M}_p$ are derived from the complete ground truth map $\mathcal{M}_{\mathrm{GT}}$ using the scenario specific prior generator P_p which masks out or selects only specific map elements:

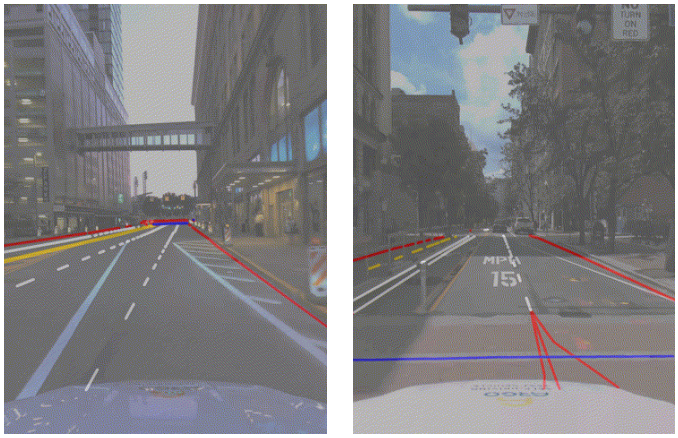
$$\mathcal{M}_p = P_p(\mathcal{M}_{\mathrm{GT}}). \tag{1}$$

TABLE I: Features of labels on Argoverse 2 used in various state of the art approaches and in our proposed ground truth.

Method	Divider Types	Lane Centerl.	3D Instances	Fixed GT Artifacts	Geo. Split
VectorMapNet [2]	-	-	-	-	-
MapTRv2 [4], [21]	-	✓	✓	-	-
StreamMapNet [16]	-	-	-	-	✓
MapTracker [5]	-	-	-	✓	✓
MapEX [10]	-	-	-	-	-
PriorDrive [17]	-	-	-	-	-
M3TR (Ours)	✓	✓	✓	✓	✓

TABLE II: Systematic map prior scenarios $\boldsymbol{\mathcal{S}}$ defined in this work.

Name	**Description**
$\mathcal{S}_{\overline{\mathrm{EL}}}$	Ego lane is masked out.
$\mathcal{S}_{\overline{\mathrm{ER}}}$	Ego road is masked out.
$\mathcal{S}_{\mathrm{BD}}$	Only road boundaries are provided as prior.
$\mathcal{S}_{\mathrm{CL}}$	Only lane center lines are provided as prior.
$\mathcal{S}_{\varnothing}$	No map prior.



(a) Ex. for $\mathcal{S}_{\overline{\mathrm{EL}}}$: Own lane blocked. (b) Ex. for $\mathcal{S}_{\overline{\mathrm{ER}}}$: New bike lane.

Fig. 2: Visualization of map changes from [7], with the outdated map reprojected into the image. Real map changes translate easily into the proposed map prior scenarios.

The task of the model is to reconstruct the complete map from the given partial prior and the sensor information.

MapEX [10] has already begun moving in this direction by including $\mathcal{S}_{\mathrm{BD}}$ as a scenario, however the other scenarios include modifications like point-level noise that are not applicable for outdated offline HD maps, but only for maps perceived in previous time steps.

We show in Sec. V that the semantic class of map prior has a strong influence on the model performance and thus propose to separate map prior scenarios semantically. This enables a systematic investigation to guide efforts in data collection or map maintenance. The prior scenarios are listed in Tab. II and visualized in Fig. 6. Fig. 2 shows that real map changes [7] can easily be categorized into the proposed scenarios. Fig. 2a has the own lane become blocked, resulting in invalidated elements akin to $\mathcal{S}_{\overline{\mathrm{EL}}}$. In Fig. 2b, a bike lane is added that causes the ego road to become invalid, similar to $\mathcal{S}_{\overline{\mathrm{ER}}}$.

The scenarios assume that it is known beforehand which elements are no longer valid, a task for which separate proposed solutions exist [8], [9]. In turn however, this also brings a large benefit: These map prior scenarios help reduce the sim-to-real gap that occurs with artificial map changes [20]. Synthetically generated map changes are often not realistic in conjunction with what the sensors actually observe in the real world, creating inconsistencies between the simulated map modifications and the sensor data. We avoid these inconsistencies, as the reconstruction task is indifferent to whether semantically coherent elements are masked synthetically or if elements become masked due to real map changes.

C. A Prior-Aware HD Map Completion Metric

To measure map completion performance, we need to solve an issue already pointed out by [8]: current evaluation metrics do not differentiate between map elements that are available as prior and those that need to be perceived online [10], [17], [20]. However, transformer models quickly learn to pass

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

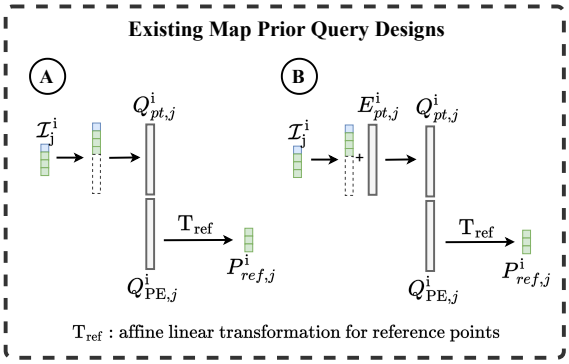


Fig. 3: Visualization of existing baseline map prior point query designs [10]. We improve upon existing queries by integrating the prior information in all parts of the query.

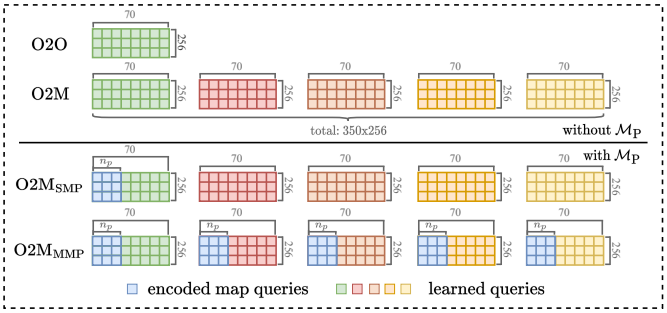


Fig. 4: Visualization of different detection query set designs with and without map prior $\mathcal{M}_p$. The set of queries are matched to ground truth map elements in either a one-to-one (O2O) or one-to-many (O2M) fashion. Compared to the baseline $\mathrm{O2M_{SMP}}$ query set design for map priors, we propose a tiling $\mathrm{O2M_{MMP}}$ design.

through prior elements almost identically and, if known as prior, any downstream application would prefer the map prior over the corresponding, possibly noisy prediction. Hence, we propose to focus on exactly those map elements which are unknown to the model at inference time.

The standard evaluation metric for methods with vectorized output [2], [4], [5] is the mean average precision (mAP), using the Chamfer distance with thresholds of $\tau \in \{0.5\mathrm{m}, 1.0\mathrm{m}, 1.5\mathrm{m}\}$. The mAP is averaged across the average precision (AP) of the individual label classes: dashed dividers, solid dividers, road boundaries, lane centerline paths and pedestrian crossings, with the class specific AP averaged across the Chamfer distance thresholds τ. Analogously, to evaluate completion performance, we define the mean average *completion* precision, $\mathrm{mAP}^{\mathcal{C}}$, which uses not the entire map $\mathcal{M}_{\mathrm{GT}}$, but only the map elements $\mathcal{M}_{\overline{p}} = \mathcal{M}_{\mathrm{GT}} \setminus \mathcal{M}_p$ which are missing in the specific scenario.

IV. A Deployable HD Map Completion Model

This section describes the M3TR (Multi-Masking Map Transformer) model itself. Sec. IV-B presents our *Generalist* training regime, Sec. IV-C how to use map masking as augmentation, and Sec. IV-A the novel map prior query design.

A. Query Design

In recent work, queries of the detection transformer have emerged as the main way to supply the model with prior information [5], [8], [10], [17], [20], [24]. How exactly these queries are composed and where they are inserted is often neglected and not described in detail. BEV detection transformer queries consist of different sub-elements and map prior queries can be composed in many different ways and inserted at multiple points, making the available option space quite large. We explore that option space to incorporate prior map knowledge on two architectural levels, the *point query design* and the *query set design*, and use MapEX [10] as our baseline.

Point Query Design: A fixed set of points is used as map decoder queries per map element to be predicted. Each point query consists of two vectors which are concatenated: the point embedding Q_{pt} and the positional embedding Q_{PE}.

Most HD map construction transformers use learned point embeddings since they assume no prior knowledge about map elements. To improve upon this, we compare the two existing approaches on how to encode map prior information into the point queries Ⓐ and Ⓑ, depicted in Fig. 3, with our novel approach Ⓒ, shown in Fig. 1.

While in Ⓐ, the baseline proposed in MapEX [10], the zero-padded point information is directly used as point embedding Q_{pt}, we propose to combine it with a two-part learned prior embedding E_{pt} in Ⓒ. This makes use of the prior information, but provides a learnable degree of freedom for the model.

Ⓑ, a learned embedding design also explored in [10], differs from Ⓒ in the positional embedding Q_{PE}. It is formed by either a sum of zero-padded point information for Ⓐ and Ⓑ or a learned prior embedding E_{PE} for Ⓒ.

To each query also belongs a reference point on the BEV grid, P_{ref}, which guides the deformable cross-attention in the decoder. In the previous designs Ⓐ and Ⓑ, it is generated from the positional embeddings with a linear projection. To improve upon this, in Ⓒ we propose to directly define it based on the map prior point information.

Query Set Design: MapTRv2 [4] proposed one-to-many (O2M) matching, a source of significant performance gains compared to the original MapTR one-to-one (O2O) matching. We explore two possible ways to adapt it to map prior information which are depicted in Fig. 4.

In the $\mathrm{O2M_{SMP}}$ (*Single Map Prior*) query set design only a single repetition of queries makes use of map prior queries while auxiliary queries are purely learned, like in the original MapTRv2. In contrast, the $\mathrm{O2M_{MMP}}$ (*Multiple Map Prior*) design includes map prior information in a tiling fashion, once for every ground truth repetition. This allows the incorporation of map prior knowledge in the auxiliary queries as well.

We follow MapEX [10] for the loss, including the pre-attribution of map prior instances during the Hungarian assignment, which we extend to the tiled $\mathrm{O2M_{MMP}}$ map prior queries. Outside of instances related to map priors, the MapEX loss is equivalent to the loss of the MapTRv2 base architecture.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

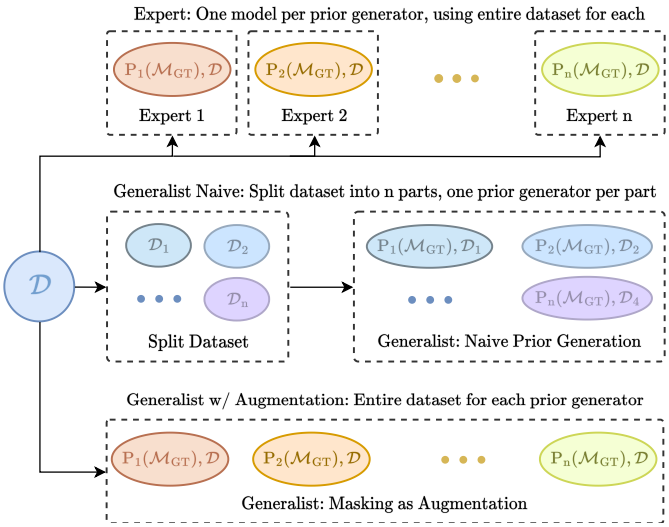
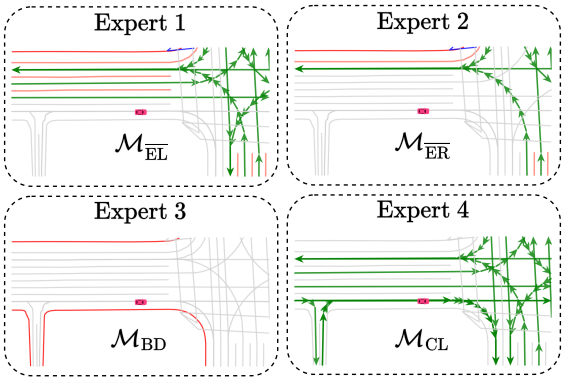
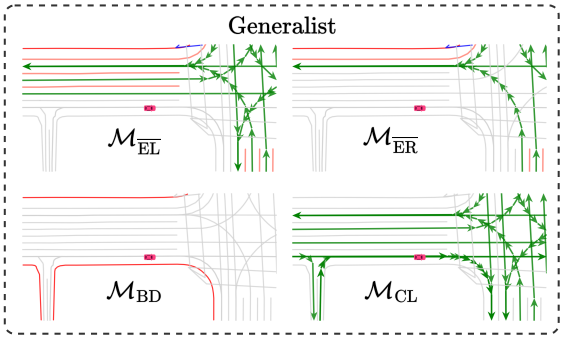


Fig. 5: Visualization of the different training regimes for variable map priors investigated in this work. Compared to previous expert training regimes and a naive *Generalist* prior generation, our masking as augmentation leverages all available data for a *Generalist* model with improved performance.



(a) One expert model per map prior scenario.



(b) One *Generalist* model for all map prior scenarios.

Fig. 6: Visualization of previous expert models vs. the *Generalist* model proposed in this work. The map prior scenarios $\mathcal{S}_p$ are listed in Tab. II.

B. Generalist and Expert Models

Previous works train one model for each map prior scenario $\mathcal{S}_p$ which is unrealistic for deployment in real autonomous systems. All models would need to be readily available in GPU memory and the suitable model would need to be correctly selected by a not yet existing oracle that identifies the available prior category. We refer to these models as *Experts* and instead propose a *Generalist* model that can exploit arbitrary parts of HD maps as a prior. Instead of only one, the *Generalist* is trained on *all* scenarios $\boldsymbol{\mathcal{S}} = \cup \mathcal{S}_p$. As we show below, while needing no extra memory or compute, it is on par with specialized *Experts*. We visualize the distinction in Fig. 6.

C. Map Masking as Augmentation

To train a *Generalist* with synthetically derived map priors, various training regimes are conceivable. As depicted in Fig. 5, to derive n map prior scenarios, one could naively split the dataset $\mathcal{D}$ into n equal disjoint smaller datasets $\mathcal{D}_i$ and use each part to derive one kind of prior scenario $\mathcal{S}_i$:

$$\mathcal{S}_i = (\mathcal{M}_{p_i}, \mathcal{D}_i) = (P_{p_i}(\mathcal{M}_{\text{GT}}), \mathcal{D}_i). \tag{2}$$

Instead, we propose to use the synthetic prior scenarios as augmentation for generic HD map construction. This means that the entire dataset $\mathcal{D}$ is used to derive each map prior scenario $\mathcal{S}_p^{\#}$ and, hence, the augmented scenario set $\boldsymbol{\mathcal{S}}^{\#}$:

$$\mathcal{S}_i^{\#} = (\mathcal{M}_p, \mathcal{D}) = (P_p(\mathcal{M}_{\text{GT}}), \mathcal{D}) \tag{3}$$

$$\boldsymbol{\mathcal{S}}^{\#} = \cup_p \mathcal{S}_p^{\#}. \tag{4}$$

This exploits the entire combinatorial variety of map prior categories and leads to an n-fold increase in training data, promising greater generalization performance.

V. Experiments

We conduct experiments on the Argoverse 2 and nuScenes datasets to validate our method, with Argoverse 2 as the main dataset. Sec. V-A elaborates on the choices of dataset and metric, Sec. V-B on the implementation and Sec. V-C discusses the performance of M3TR in comparison with existing baselines.

A. Dataset and Metric

As mentioned above, we validate our method on the Argoverse 2 [25] and nuScenes datasets [26], the standard public datasets for HD map construction. Both Argoverse 2 and nuScenes contain 1000 driving sequences, covering 17 km² and 5 km², respectively [22]. Since nuScenes has only 40,000 samples compared to 158,000 for Argoverse 2, contrary to most existing work, we regard Argoverse 2 as our primary dataset for evaluation. As discussed in Sec. III-A, we use a novel kind of ground truth that resolves a number of problems compared to the labels used in [2]–[4], [16].

As our metric, we use the mean average *completion* precision mAP^{C} defined in Sec. III-B to compare the methods.

To simulate real use cases with various priors, we compare expert model groups against a single generalist model using mean performance per class across scenarios. This assumes, for the benefit of the experts, a perfect oracle for prior-to-model assignment and no mixing of prior categories.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE III: Comparison of methods over map prior scenarios on the Argoverse 2 data set, with the geographical split from [22]. Only elements not in the map prior are evaluated. $\mathcal{M}_{\overline{\text{EL}}}$: Ego lane is masked. $\mathcal{M}_{\overline{\text{ER}}}$: Ego road is masked. $\mathcal{M}_{\text{BD}}$: Only road boundaries as prior. $\mathcal{M}_{\text{CL}}$: Only centerlines as prior. $\mathcal{M}_{\varnothing}$: No map prior. $\mathcal{O}(|\mathcal{M}_{\text{P}}|)$ indicates how the deployment effort scales with the number of map prior types. *: Re-implemented by the authors, as code was not publicly available at the time of publication. †: For no map prior both methods are equivalent to MapTRv2, as changes are only made regarding map priors. ‡: For real world inference, the method would additionally need a not yet existing scenario to expert assignment oracle, as indicated by the last column. FPS calculated on a NVIDIA RTX 6000 Ada GPU.

Dataset: Argoverse 2		*$AP^{\mathcal{C}}$ = AP for Masked Elements only*							$\mathcal{O}(\lvert\mathcal{M}_{\text{P}}\rvert)$	**Var. Prior**
Method	**Map Prior**	$\text{AP}^{\mathcal{C}}_{\text{dsh}}$	$\text{AP}^{\mathcal{C}}_{\text{sol}}$	$\text{AP}^{\mathcal{C}}_{\text{bou}}$	$\text{AP}^{\mathcal{C}}_{\text{cen}}$	$\text{AP}^{\mathcal{C}}_{\text{ped}}$	**mAP$^{\mathcal{C}}$**	**vs. [10]**	**FPS / VRAM**	**w/o Oracle**
MapTRv2 [4]	$\mathcal{M}_{\varnothing}$	37.9	55.0	49.7	48.2	41.7	**46.5**	†+0.0	-	-
MapEX* [10] Models	$\mathcal{M}_{\overline{\text{EL}}}$	45.3	64.5	53.4	52.8	44.9	**52.2**	-	$\mathcal{O}(\text{n})$ < 14.3 FPS ‡ > 16.3 GB ‡	✗
	$\mathcal{M}_{\overline{\text{ER}}}$	41.5	62.4	54.9	55.3	45.5	**51.9**	-		
	$\mathcal{M}_{\text{BD}}$	37.7	56.0	-	50.6	44.5	**47.2**	-		
	$\mathcal{M}_{\text{CL}}$	43.2	61.8	58.1	-	42.8	**51.5**	-		
	$\mathcal{M}_{\varnothing}$	†37.9	†55.0	†49.7	†48.2	†41.7	†**46.5**	-		
	Mean	**41.1**	**59.9**	**54.0**	**51.7**	**43.9**	**49.9**	-		
M3TR Expert Models	$\mathcal{M}_{\overline{\text{EL}}}$	51.7	69.4	56.3	55.4	49.7	**56.5**	**+4.3**	$\mathcal{O}(\text{n})$ < 14.3 FPS ‡ > 16.3 GB ‡	✗
	$\mathcal{M}_{\overline{\text{ER}}}$	44.8	66.5	57.0	57.8	48.7	**55.0**	**+3.1**		
	$\mathcal{M}_{\text{BD}}$	40.2	57.3	-	54.7	49.2	**50.2**	**+3.0**		
	$\mathcal{M}_{\text{CL}}$	45.1	63.2	61.1	-	48.6	**55.0**	**+3.5**		
	$\mathcal{M}_{\varnothing}$	†37.9	†55.0	†49.7	†48.2	†41.7	†**46.5**	†+0.0		
	Mean	**43.9**	**62.3**	**56.0**	**54.0**	**47.5**	**52.6**	**+2.7**		
M3TR Generalist	$\mathcal{M}_{\overline{\text{EL}}}$	48.8	67.8	59.5	54.8	51.8	**56.5**	**+4.3**	$\mathcal{O}(1)$ 14.3 FPS 3.3 GB	✓
	$\mathcal{M}_{\overline{\text{ER}}}$	45.7	64.4	57.0	56.9	51.1	**55.0**	**+3.1**		
	$\mathcal{M}_{\text{BD}}$	41.2	57.3	-	53.0	48.0	**49.9**	**+2.7**		
	$\mathcal{M}_{\text{CL}}$	42.5	59.3	57.4	-	45.6	**51.2**	-0.3		
	$\mathcal{M}_{\varnothing}$	40.4	55.4	50.3	49.4	43.9	**47.9**	**+1.4**		
	Mean	**43.7**	**60.8**	**56.0**	**53.5**	**48.1**	**52.1**	**+2.2**		

TABLE IV: Results *without* map masking as augmentation as ablation on the Argoverse 2 data set.

$\mathcal{M}_p$	$\text{AP}^{\mathcal{C}}_{\text{dsh}}$	$\text{AP}^{\mathcal{C}}_{\text{sol}}$	$\text{AP}^{\mathcal{C}}_{\text{bou}}$	$\text{AP}^{\mathcal{C}}_{\text{cen}}$	$\text{AP}^{\mathcal{C}}_{\text{ped}}$	**mAP$^{\mathcal{C}}$**	**vs. [10]**
$\mathcal{M}_{\overline{\text{EL}}}$	49.4	69.3	56.9	54.8	49.7	**56.0**	**+3.8**
$\mathcal{M}_{\overline{\text{ER}}}$	41.4	65.6	55.8	56.2	48.6	**54.4**	**+2.5**
$\mathcal{M}_{\text{BD}}$	42.7	58.4	-	52.7	46.4	**49.7**	**+2.7**
$\mathcal{M}_{\text{CL}}$	42.7	59.9	55.4	-	43.6	**50.4**	-1.1
$\mathcal{M}_{\varnothing}$	40.5	56.0	49.0	48.4	41.8	**47.2**	**+0.7**
Mean	**43.3**	**61.8**	**54.3**	**53.0**	**46.0**	**51.5**	**+1.6**

TABLE V: Comparison of map query encoders for the map prior scenario $\mathcal{M}_{\overline{\text{EL}}}$ on the Argoverse 2 dataset.

Map Query Enc.			**AP$^{\mathcal{C}}$**				
Point Enc.	O2M_{MMP}	**mAP$^{\mathcal{C}}$**	dsh.	sol.	bou.	cen.	ped.
A [10]	—	52.2	45.3	64.5	53.4	52.8	44.9
B	—	52.4 **(+0.2)**	46.8	65.3	53.0	52.5	44.5
C	—	53.5 **(+1.3)**	48.4	66.7	55.4	50.6	46.5
C	✓	56.5 **(+4.3)**	51.7	69.4	56.3	55.4	49.7

B. Implementation Details and Baseline

We base our code and the model architecture on the MapTRv2 [4] framework and re-implement MapEX [10] as a baseline. Public code for [10] was not available at the time of writing and information about some of the query design particulars discussed in Sec. IV-A is not present in the paper. We therefore selected the variants Ⓐ and O2M_{SMP} for the point query and query set design in our re-implementation. All models use ResNet50 [27] as the image backbone and parameters unrelated to map priors are left unchanged from the MapTRv2 base for fair comparison. We also follow one of the label modalities of MapTRv2 and use 3D map instances for Argoverse 2 as mentioned in Tab. I.

All models are trained until convergence, *i.e.* for 24 / 110 epochs for experts and 54 / 224 epochs for the generalist on Argoverse 2 / nuScenes respectively, with the best checkpoint shown. The generalist was trained on nine map prior scenarios for Argoverse 2 and seven for nuScenes. The four scenarios not explicitly shown are missing only centerlines / pedestrian crossings / road borders / dividers.

C. Map Completion Performance

As we view Argoverse 2 as our main dataset for evaluation, we investigate more map prior scenarios on it than on nuScenes. We first discuss the results on Argoverse 2 along with ablations on the map query encoder and the map masking as augmentation. Then we present the slightly reduced set of experiments on the nuScenes dataset.

Results on Argoverse 2: Tab. III shows the performance of the M3TR expert and generalist variants as well as a MapEX expert baseline for five selected map prior scenarios on Argoverse 2. All methods using map priors show enhanced average precision compared to the prior-less scenario, with varying benefit depending on the supplied map prior. A qualitative example of this can be seen in Fig. 7.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE VI: Comparison of methods and masking scenarios on the nuScenes data set, with the geographical split from [22]. Only elements not in the map prior are evaluated. $\mathcal{M}_{\text{BD}}$: Only road boundaries as prior. $\mathcal{M}_{\text{CL}}$: Only centerlines as prior. $\mathcal{M}_{\varnothing}$: No map prior. $\mathcal{O}(|\mathcal{M}_{\text{P}}|)$ indicates how the deployment effort scales with the number of map prior types. The last column indicates whether the method could handle variable priors without a not yet existing scenario to expert assignment oracle. *: Re-implemented by the authors, as code was not publicly available at the time of publication. †: For no map prior both expert methods are equivalent, as changes in the base architecture are only made regarding map priors.

Dataset: nuScenes		*$AP^{\mathcal{C}}$ = AP for Masked Elements only*								
Method	**Map Prior**	$\mathbf{AP}^{\mathcal{C}}_{\text{dsh}}$	$\mathbf{AP}^{\mathcal{C}}_{\text{sol}}$	$\mathbf{AP}^{\mathcal{C}}_{\text{bou}}$	$\mathbf{AP}^{\mathcal{C}}_{\text{cen}}$	$\mathbf{AP}^{\mathcal{C}}_{\text{ped}}$	$\mathbf{mAP}^{\mathcal{C}}$	**vs. [10]**	$\mathcal{O}(\lvert\mathcal{M}_{\text{P}}\rvert)$	**Var. Prior w/o Oracle**
MapTRv2 [4]	$\mathcal{M}_{\varnothing}$	12.5	19.1	32.4	29.1	21.6	**22.9**	†+0.0	-	-
MapEX* [10] Models	$\mathcal{M}_{\text{BD}}$	13.2	21.1	-	31.0	22.0	**21.9**	-	$\mathcal{O}(\text{n})$	✗
	$\mathcal{M}_{\text{CL}}$	16.6	26.0	39.8	-	23.4	**26.4**	-		
	$\mathcal{M}_{\varnothing}$	†12.5	†19.1	†32.4	†29.1	†21.6	†**22.9**	-		
	Mean	**14.1**	**22.1**	**36.1**	**30.1**	**22.3**	**23.7**	-		
M3TR Expert Models	$\mathcal{M}_{\text{BD}}$	15.3	26.7	-	34.9	28.3	**26.3**	**+4.4**	$\mathcal{O}(\text{n})$	✗
	$\mathcal{M}_{\text{CL}}$	23.1	33.2	46.6	-	27.8	**32.5**	**+6.1**		
	$\mathcal{M}_{\varnothing}$	†12.5	†19.1	†32.4	†29.1	†21.6	†**22.9**	†+0.0		
	Mean	**17.0**	**26.3**	**39.5**	**32.0**	**25.9**	**27.2**	**+3.5**		
M3TR Generalist	$\mathcal{M}_{\text{BD}}$	14.5	23.2	-	32.7	24.7	**23.8**	**+1.9**	$\mathcal{O}(1)$	✓
	$\mathcal{M}_{\text{CL}}$	15.3	24.4	38.6	-	24.4	**25.7**	-0.7		
	$\mathcal{M}_{\varnothing}$	12.4	20.0	31.8	29.8	23.4	**23.5**	**+0.6**		
	Mean	**14.1**	**22.5**	**35.2**	**31.3**	**24.2**	**24.3**	**+0.5**		

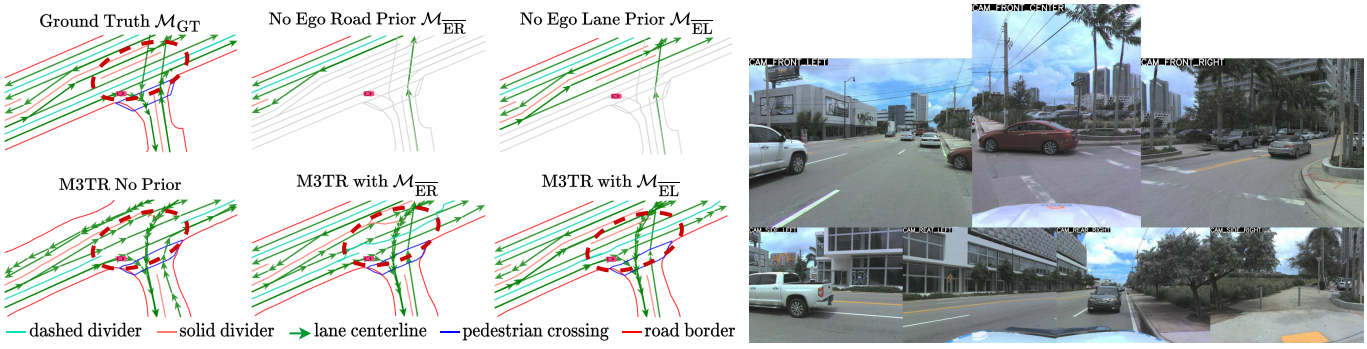


Fig. 7: Example of the M3TR generalist model on the same sample from Argoverse 2 with different map priors. The more information available, the better the model can reconstruct elements not contained in the prior set.

For almost all scenarios, the M3TR experts variants substantially improve the prediction performance compared to the MapEX baseline. Except for the $\mathcal{M}_{\text{CL}}$ scenario, the generalist model matches the performance of the *Expert* models in their expert scenarios as well.

In the more realistic use case with varying map priors, the generalist likewise outperforms the baseline in the average of all scenarios, while using a fifth of the VRAM and without an oracle for perfect prior scenario to expert assignment. Such an assignment system does not exist yet and poses a major obstacle for real-world use of expert model ensembles.

The generalist also improves performance in the no prior scenario ($\mathcal{M}_{\varnothing}$), without architectural changes relevant for this scenario compared to the MapTRv2 base. This suggests that the various prior scenarios function as augmentation that aids learning even without any prior.

Ablations on Argoverse 2: The ablation in Tab. IV highlights that using map masking as augmentation, *i.e.* deriving training data for each map prior scenario from the entire dataset, is effective. Compared to the naive prior generation (Tab. IV), the generalist model with augmentation in Tab. III performs +0.6 $\text{mAP}^{\mathcal{C}}$ better.

Tab. V compares various map prior encoding modalities, using point encoder names from Fig. 1. Compared to the baseline encoder from MapEX [10], Ⓐ, our proposed query design Ⓒ shows significantly improved performance. Encoder Ⓑ, which skips modifying positional queries and reference points, has only a partial performance increase as a result. Including map priors in the one-to-many queries (O2M_{MMP}) further boosts performance.

Results on nuScenes: Table VI shows the results on nuScenes with a reduced set of map prior scenarios. The expert performance gains over the baseline exceed those on Argoverse 2, though the *Generalist* shows reduced performance compared to M3TR *Experts*. Notably, the *Generalist* still outperforms MapTRv2 without a prior, confirming the effectiveness of map masking as augmentation.

With the general decrease in $\text{mAP}^{\mathcal{C}}$ compared to Argoverse 2, we hypothesize that nuScenes' smaller sample count hampers generalization, consistent with observations in [22].

VI. Conclusion

This work proposes M3TR, a generalist approach for HD map construction with variable map priors.

We introduce improved ground truth and define a new HD map completion benchmark, including a systematic set of prior scenarios for outdated HD maps and a metric that focuses on the elements not given as a map prior. Our systematic examination of query design fully incorporates prior map information, yielding up to +4.3 $\text{mAP}^{\boldsymbol{C}}$ compared to the MapEX [10] baseline on Argoverse 2. Training with partially masked maps also serves as effective augmentation, improving performance even without priors. Finally, our Generalist model handles all map prior scenarios while matching the performance of specialized *Experts*, requiring only constant memory and no knowledge of which map information is available. This makes M3TR the first real-world deployable model for HD map construction with offline HD map priors.

References

[1] 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 (ICRA)*, 2022, pp. 4628–4634. 1

[2] Y. Liu, T. Yuan, Y. Wang, Y. Wang, and H. Zhao, "VectorMapNet: End-to-end vectorized HD map learning," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 22 352–22 369. 1, 2, 3, 4, 5

[3] 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 *The Eleventh International Conference on Learning Representations*, 2022. 1, 2, 5

[4] 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*, Oct 2024. [Online]. Available: https://doi.org/10.1007/s11263-024-02235-z 1, 2, 3, 4, 5, 6, 7

[5] J. Chen, Y. Wu, J. Tan, H. Ma, and Y. Furukawa, "Maptracker: Tracking with strided memory fusion for consistent vector hd mapping," in *Computer Vision – ECCV 2024*, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds. Cham: Springer Nature Switzerland, 2025, pp. 90–107. 1, 2, 3, 4

[6] 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 (CVPR)*, June 2024, pp. 15 396–15 406. 1, 2

[7] J. Lambert and J. Hays, "Trust, but verify: Cross-modality fusion for hd map change detection," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, J. Vanschoren and S. Yeung, Eds., vol. 1. Curran, 2021. 1, 2, 3

[8] L. Wild, L. Ericson, R. Valencia, and P. Jensfelt, "Exelmap: Explainable element-based hd-map change detection and update," in *Proceedings of the ECCV 2024 2nd Workshop on Vision-Centric Autonomous Driving (VCAD)*, 2024. [Online]. Available: https://arxiv.org/abs/2409.10178 1, 2, 3, 4

[9] J.-H. Pauls, T. Strauss, C. Hasberg, M. Lauer, and C. Stiller, "Hd map verification without accurate localization prior using spatio-semantic 1d signals," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 680–686. 1, 3

[10] R. Sun, L. Yang, D. Lingrand, and F. Precioso, "Mind the map! accounting for existing map information when estimating online hdmaps from sensor data," *arXiv preprint arXiv:2311.10517*, 2024. [Online]. Available: https://arxiv.org/abs/2311.10517 1, 2, 3, 4, 6, 7, 8

[11] 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

[12] J. Philion and S. Fidler, "Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*. Springer, 2020, pp. 194–210. 2

[13] 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 *European conference on computer vision*. Springer, 2022, pp. 1–18. 2

[14] S. Chen, T. Cheng, X. Wang, W. Meng, Q. Zhang, and W. Liu, "Efficient and robust 2d-to-bev representation learning via geometry-guided kernel transformer," *arXiv preprint arXiv:2206.04584*, 2022. 2

[15] S. Choi, J. Kim, H. Shin, and J. W. Choi, "Mask2map: Vectorized hd map construction using bird's eye view segmentation masks," *arXiv preprint arXiv:2407.13517*, 2024. 2

[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 (WACV)*, January 2024, pp. 7356–7365. 2, 3, 5

[17] S. Zeng, X. Chang, X. Liu, Z. Pan, and X. Wei, "Driving with prior maps: Unified vector prior encoding for autonomous vehicle mapping," *arXiv preprint arXiv:2409.05352*, 2024. [Online]. Available: https://arxiv.org/abs/2409.05352 2, 3, 4

[18] 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 *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 4029–4035. 2

[19] 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," in *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. 2

[20] S. M. Bateman, N. Xu, H. C. Zhao, Y. Ben Shalom, V. Gong, G. Long, and W. Maddern, "Exploring real world map change generalization of prior-informed hd map prediction models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2024, pp. 4568–4578. 2, 3, 4

[21] B. Liao, S. Chen, B. Jiang, T. Cheng, Q. Zhang, W. Liu, C. Huang, and X. Wang, "Lane graph as path: Continuity-preserving path-wise modeling for online lane graph construction," in *Computer Vision – ECCV 2024*, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds. Cham: Springer Nature Switzerland, 2025, pp. 334–351. 3

[22] A. Lilja, J. Fu, E. Stenborg, and L. Hammarstrand, "Localization is all you evaluate: Data leakage in online mapping datasets and how to fix it," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 22 150–22 159. 3, 5, 6, 7

[23] F. Immel, R. Fehler, F. Bieder, and C. Stiller, "Generation of training data from hd maps in the lanelet2 framework," *arXiv preprint arXiv:2407.17409*, 2024. [Online]. Available: https://arxiv.org/abs/2407.17409 3

[24] 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," *arXiv preprint arXiv:2311.04079*, 2023. 4

[25] 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 *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021)*, 2021. 5

[26] 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 (CVPR)*, June 2020. 5

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 6