2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

OSM vs HD Maps: Map Representations for Trajectory Prediction

Jing-Yan Liao[†], Parth Doshi[†], Zihan Zhang[†], David Paz[†], Henrik Christensen[†]

Abstract— **High Definition (HD) Maps have long been favored for their precise depictions of static road elements. However, their accessibility constraints and vulnerability to rapid environmental changes impede the widespread deployment of highly map-reliant autonomous driving tasks, such as motion forecasting. In this context, we propose to leverage OpenStreetMap (OSM) as a promising alternative to HD Maps for long-term motion forecasting. The contributions of this work are threefold: firstly, we extend the application of OSM to long-horizon forecasting, doubling the forecasting horizon compared to previous studies. Secondly, through an expanded observation landscape and the integration of intersection priors, our OSM-based approach exhibits competitive performance, narrowing the gap with HD-map-based models. Lastly, we conduct an exhaustive context-aware analysis, providing deeper insights in motion forecasting across diverse scenarios as well as conducting class-aware comparisons. This research not only advances long-term motion forecasting with coarse map representations but additionally offers a scalable solution within the domain of autonomous driving.**

Fig. 1. Spatial Evolution of a road network at the UC San Diego Campus in both satellite image [6] and vectorized centerline map [7]. The left side of the figure depicts the area prior to 2021, while the right side presents the current state of the road network. Below the satellite image, the corresponding vector map from the respective time period is displayed.

I. INTRODUCTION

Ensuring the safety of all road users is a fundamental mission of autonomous driving. To achieve this, self-driving systems must have a thorough understanding of their surroundings. In this context, motion forecasting emerges as a crucial component, serving the purpose of modeling agent behavior within a scene and predicting their future trajectories. Motion forecasting models [1]–[3] primarily rely on two key components: a map to depict the static elements in the environment, and past trajectories of agents within the scene, which account for the dynamic components. While past paths are typically well-defined as tracks, the pursuit of an optimal map representation [4], [5] remains an ongoing challenge.

HD Maps have emerged as the prevalent choice for map representation in motion forecasting applications due to their precision in conveying road markings, lane boundaries, and road geometry. This rich information contributes to the contextual and semantic understanding necessary for accurate predictions. However, the accessibility to HD Maps is notably constrained in practice, not to mention the considerable human and storage resources required for their creation and maintenance. This challenge hinders the large scale expansion of autonomous driving, especially in areas without up-to-date HD Map coverage. For instance, the UCSD on-campus autonomous driving practice consumes huge amount of human efforts on HD Map creation and maintenance. Unfortunately, this map quickly becomes outdated due to ongoing construction activities on campus, as depicted in Fig. 1. Notably, not only do the locations of road lane markings change, but parts of the road network itself undergo significant alterations within a short time frame, rendering the previously available HD Map obsolete and ineffective for autonomous driving applications. This example underscores the pressing need to explore alternative map representations that can accommodate dynamic real-world changes with increased efficiency and reduced manual labeling labor. Interestingly, coarse map, which is widely used in localization [8] [9] and navigation [10] [11]for autonomous driving, has been somewhat overlooked as a potential map representation for motion forecasting due to reliance on HD Map until recently [12]. Open-source maps like OSM [13] and proprietary maps such as Google maps and Apple maps offer scalability advantages over HD Maps. Although they lack precise lane-level information, coarse maps include information about road connectivity and intersections as shown in Fig. 2, which hold substantial potential for motion forecasting. Although proprietary maps might contains more detailed as well as standardized representation of the roads, we choose to conduct experiments with OSM due to its open-source nature.

Therefore in this work, we explore long horizon motion forecasting with OSM as a new map representation by utilizing the HiVT [1] architecture as a basis. We specifically

[†]Contextual Robotics Institute, University of California San Diego, La Jolla, CA 92093, USA

chose HiVT, HiVT-128 to be exact, for its good performance, efficient in memory usage, and its real-time capability that scales well as the number of agents in the scene grows. Our key contributions are summarized as follows:

- **OSM for long tail forecasting:** We extend the application of OSM for the more challenging long-term motion forecasting. In contrast to previous research [14], which focused on short-term forecasting within the Argoverse1 Motion Forecasting Dataset [15], where the models were tasked with projecting 3 seconds into the future, our study evaluates the performance of our model using the Argoverse2 Motion Forecasting Dataset [16]. This dataset doubles the forecasting horizon, providing a more complex and realistic testing ground to assess the effectiveness of our OSM-based approach.
- **Competitive performance of the OSM-based method:** Through the expansion of the observation landscape, our OSM-based method achieves substantial improvements. Notably, we narrow the performance gap between HD-map-based models and our OSM-based model, demonstrating the competitive potential of our approach.
- **Comprehensive Context-Aware Analysis:** We conduct an exhaustive context-aware comparison, both quantitative and qualitative, considering both map representations and agent types. Our evaluation encompasses the visualization of motion forecasting outcomes across various scenarios, including straight lines, intersections, and curved roads. Furthermore, we delve into class-aware analysis, exploring how different map representations impact agent behavior under various map contexts.

By introducing OSM as a map representation and rigorously assessing HiVT's performance in diverse scenarios, this research provides a fresh perspective on long-term motion forecasting. It offers valuable insights into scalability in the realm of autonomous driving and beyond.

II. RELATED WORK

A. Physics-based Motion Forecasting

Physics-based methods in motion forecasting adhere to the fundamental laws of physics. An example is the constant velocity (CV) model, assuming an object's recent motion determines its future trajectory. Another common kinematic model is the constant acceleration (CA) model, linking recent acceleration changes to the object's future path. Various studies [17] explore pedestrian and vehicle tracking using physics-based models, employing tools such as Kalman filters [18] and Particle Filters [19] to handle uncertainties and predict trajectories more accurately. However, these models exhibit limitations, notably the Markov assumption, which makes long-term motion forecasting challenging, and the lack of consideration for the surrounding environment, which leads to potential inaccuracies. Consequently, a paradigm shift towards map-based methods has emerged to incorporate contextual information, thereby improving the precision of motion forecasting.

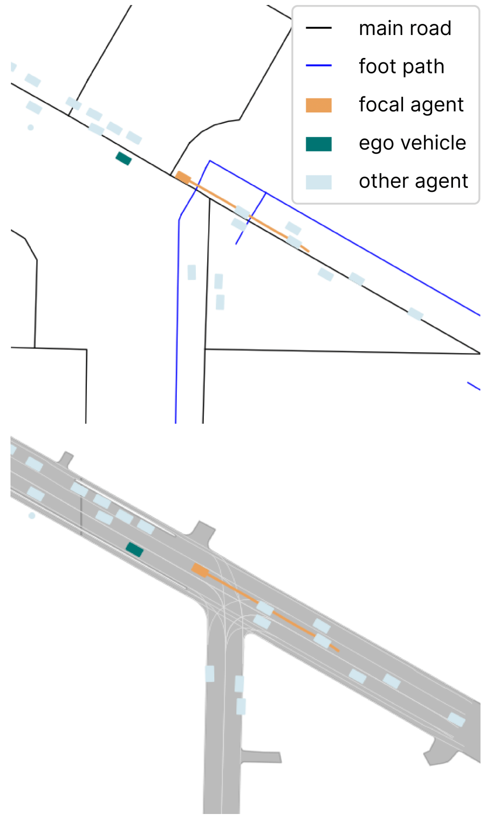

Fig. 2. This figure illustrates the distinctions between a OSM (top) and a HD map (down) in the same region. HD Map offers precise lane boundary delineations, lane connectivity (white lines), and driveable area (gray background), while OSM only provides road connectivity information.

B. Map-based Motion Forecasting

Initially, rasterized HD Maps [5], [20]–[23] gained popularity because of the impressive results achieved by Convolutional Neural Networks (CNNs) in computer vision. Previous works incorporate different features in the surrounding to provide rich and context-aware information for motion forecasting task. Some approaches entail the conversion of map elements (such as roads and crosswalks) into distinct layers with color-coded lane direction. Later, some studies have gone beyond simple rasterization to render more complex map features, including roadmaps, traffic lights, and speed limits, into bird's-eye view images. Afterward, rasterization techniques have been extended to encode semantic map details into a top-down spatial grid. This holistic approach aims to provide richer and more context-aware map representations, which are essential for enhancing the capabilities of motion forecasting models in making informed predictions. However, processing these rasterized representations with CNNs is computationally demanding and has limited observation landscape, which could lead to higher error in longer term motion forecasting. Recently, more research has shifted towards using vectorized HD Maps [24]–[29] due to their more efficient representation. For instance, VectorNet [4] started this trend by sampling key points from lane splines to simplify the map and encode it with graph neural networks. LaneGCN [2], on the other hand, builds lane graphs with centerline segments and uses graph convolutional networks to capture information. HiVT transforms map elements into relative positions to guarantee translational invariance. TPCN [30] describes maps as ordered point sets and leverages a pointcloud-based learning method to learn from the surrounding.

Although these methods introduce innovative ideas on how to better utilize HD maps, there is limited research emphasizing how motion forecasting leverages these map priors for various types of road users. It also remains unclear how coarse maps would impact motion forecasting capabilities, especially for newer, longer-term motion forecasting. Therefore in this work, we scrutinized over 20,000 scenarios in the Argoverse2 validation set, selectively extracting compelling examples and presenting our findings from visualizing the predictions.

III. APPROACH

This section provides a detailed account of the approach undertaken to integrate OSM into the HiVT model. The integration process involves two main steps: OSM data incorporation into the HiVT model and a comprehensive description of the OSM data format. Additionally, we present implementation details, including adjustments made to enhance the compatibility of the HiVT model with the Argoverse2 dataset and the preprocessing steps applied to OSM data.

A. Incorporating OSM data into HiVT

HiVT utilize vectorized HD map as a detailed scene representation. On the other hand, OSM, though with a sparser node distribution, shares a graph-based structure with the Argoverse2 HD Map. This inherently graph-based nature of OSM data makes it well-suited for integration with the HiVT model. The integration process can be summarized as follows:

1) **Boundary Extraction:** Our process initiates with the acquisition of the boundary formed by all agent tracks within each Argoverse2 scenario. The original agent coordinates, denoted as $a_i = (x_i, y_i)$, are initially provided in the city frame. We employ the Argoverse2 API to transform these coordinates into the WGS84 coordinate system. This transformed boundary then serves as the basis for downloading relevant OSM map data.
2) **Preprocessing for HiVT:** Following boundary extraction, we undertake a sequence of preprocessing steps to prepare the downloaded OSM data for seamless integration into the HiVT model. Firstly, we transform all OSM nodes from the WGS84 coordinate system into the city frame by the Argoverse2 API. Secondly, we perform interpolation on OSM ways to ensure uniformity in node-to-node distances. Finally, we transform the entire map into relative positions, similar to the HD Map preprocessing from HiVT. These steps ensure compatibility and coherence between the OSM-based data and the HiVT model.

B. OSM Data format

OpenStreetMap (OSM) serves as an extensive repository of diverse geographical attributes, encompassing various features such as road networks, structures, and facilities distributed across the Earth's surface. These geographical features are encapsulated within three fundamental data structures: nodes, ways, and relations.

Nodes, denoted as $n_i \in N$, encapsulate essential geographical information, including latitude, longitude, and a unique node identifier. Ways, represented as $w_i = \{n_j\} \in W$, constitute aggregations of nodes that collectively define contiguous segments of roadways. Relations, articulated as $r_i = \{n_j, w_k, r_l\} \in R$, serve to establish logical or geographic associations between disparate map objects. Each of these fundamental data structures accommodates supplementary metadata, such as road names or lane counts, stored as tags. This metadata enriches the core geographical information within the dataset.

For our study, which focuses on motion forecasting, we selectively extract nodes and ways from the OSM data, omitting metadata. This decision is motivated by the inconsistency in label attributes across different geographic locations, an inherent characteristic of crowd-sourced data. However, it is worth noting that future research endeavors may explore methods to incorporate this metadata into our representation to fully harness the potential of OSM data.

C. Implementation Details

The implementation phase involved a series of strategic adjustments aimed at enhancing the compatibility of the HiVT model with the Argoverse2 dataset and maximizing the effective utilization of OSM data.

1) HiVT Model Enhancement: To align the publicly available HiVT model with the unique characteristics of the Argoverse2 dataset, we introduced specific modifications. Notably, we restructured the dataloader to cater to the nuances of Argoverse2 data. Additionally, we adapted the architecture to incorporate agent class information as an additional input - a feature absent in the Argoverse 1 dataset. This architectural refinement empowers the model to incorporate agent class information, enriching its contextual understanding and facilitating an in-depth analysis of its performance across various agent types.

2) OSM Data Preprocessing: On the OSM front, we conducted essential preprocessing steps on the OSM data to ensure a fair and meaningful comparison with the HD Map provided by Argoverse2. This involved the interpolation of OSM nodes, the inclusion of intersection information, and transform to relative positions. The interpolation process traversed the entire OSM graph to maintain a uniform average distance of 1.5 meters between nodes, aligning closely with the average centerline segment length observed in Argoverse2. However, unlike HD maps equipped with ample accurate intersection labels, identifying intersections in OSM presented challenges due to inconsistencies in labeling. To address this, we employed a proximity-based approach, flagging all nodes within a 10-meter radius of specific markers

TABLE I
HIVT PERFORMANCE ON ARGOVERSE2 VALIDATION SET.

	observation landscape	minADE ↓	minFDE ↓	MR ↓	Inference Speed (Hz)	VRam usage (MB)
HD Map	100 m	0.943	1.934	0.287	6.94	4404
	125 m	0.928 (-1.5%)	1.869 (-3.3%)	0.278 (-3.1%)	7.05	6482
	150 m	**0.913** (-3.2%)	**1.804** (-6.7%)	**0.260** (-9.4%)	6.83	8934
OSM	100 m	1.366	3.215	0.430	8.16	3944
	125 m	1.060 (-22.4%)	2.301 (-22.4%)	0.320 (-25.6%)	8.03	5298
	150 m	**1.002** (-26.6%)	**2.134** (-33.6%)	**0.313** (-27.2%)	7.92	6726
No Map	N/A	1.663	4.119	0.471	13.19	3400

such as stop signs and traffic lights. This approach ensures a fair comparison with the HD map-based method, providing contextual information to HiVT that improves intersection-related results.

This integrated approach facilitates the seamless incorporation of OSM data into the HiVT, enabling further analysis for long-tail motion forecasting within urban landscapes.

IV. EXPERIMENTS

In this section, we present our comprehensive evaluation of the HiVT model on the publicly available Argoverse2 Motion Forecasting Dataset.

A. Dataset

Originally, the HiVT model underwent evaluation using the Argoverse 1 dataset. However, for reasons outlined below, we opted to develop a tailored implementation for the Argoverse2 dataset conversion.

- **Forecasting Horizon Expansion:** The Argoverse2 dataset presents a significant departure from its predecessor, notably in terms of forecasting horizons. With each scenario doubling to a duration of 11 seconds, the track history lengthens from 2 seconds to 5 seconds, while the forecasting horizon doubles to 6 seconds. This substantial increase in forecasting duration poses a more intricate challenge for motion forecasting, rendering it a valuable dataset for in-depth investigation.
- **Enhanced Class Information:** In contrast to Argoverse 1, which lacks class information, Argoverse2 introduces a richer classification scheme comprising 10 non-overlapping classes, encompassing both static and dynamic agents. The inclusion of detailed class information pertaining to agents permits a more nuanced analysis of forecasting behavior, thereby enriching our understanding of the motion forecasting task.
- **Diverse Scenarios:** The Argoverse2 dataset is collected from six distinct cities, providing a diverse range of scenarios and environments. This geographical diversity allows us to assess the scalability and adaptability of the OSM-based method across varied urban landscapes.

By undertaking experiments on the Argoverse2 Motion Forecasting Dataset, we aim to thoroughly examine the capabilities and performance of the HiVT model in addressing the challenges posed by extended forecasting horizons, enriched class information, and diverse real-world scenarios.

B. Metrics

The evaluation of our model's performance employs standard metrics for multi-modal motion forecasting, namely minimum Average Displacement Error (minADE), minimum Final Displacement Error (minFDE), and Miss Rate (MR). These metrics allow models to forecast up to 6 trajectories for each agent, aligning with the Argoverse2 motion forecasting evaluation protocol. MinADE quantifies the average $\mathcal{L}$-2 distance in meters between the best-predicted trajectory and the ground-truth trajectory across all future time steps, while minFDE measures the error specifically at the final future time step. MR represents the proportion of scenarios where the distance between the ground-truth endpoint and the best-predicted endpoint exceeds 2.0 meters. These metrics collectively provide a comprehensive assessment of our model's performance in the context of motion forecasting.

C. Model Training

For both HD Map and OSM data, our model underwent training for 100 epochs on two RTX 3090 GPUs with a batch size of 32. We adhered to the original training recipe outlined in the HiVT paper, maintaining consistency in hyperparameters.

These implementation details are pivotal in achieving the desired synergy between the HiVT model and OSM data. They serve as a foundation for a meaningful comparison between HD-map-based and OSM-based HiVT models in the context of long-term motion forecasting on the Argoverse2 dataset.

D. Quantitative Analysis

The validation results of our approach on the Argoverse2 dataset are presented in Table I. Initially, we examine the performance with the original observation landscape, which encompasses map information within a 100-meter radius around the agent. An evident performance gap becomes apparent between the HD-map-based method and our OSM-based approach in this configuration. In fact, the OSM-based method demonstrates only marginal superiority over methods devoid of map information. It is worth noting that all the HiVT models used in this analysis are HiVT-128.

However, a critical transformation occurs when we expand the observation landscape radius from 100 meters to 125 and ultimately to 150 meters. Here, significant changes in performance dynamics emerge. This expansion triggers significant shifts in performance dynamics. While the HD-map-based results exhibit a tendency to plateau, the OSM-based approach displays notable improvement as the observation landscape expands. Additionally, OSM-based method with 150 meters observation landscape exhibits remarkably comparable performance to the HD-map-based method within the first 40 frames, which means 4 seconds into the future as The Argoverse2 dataset provides data at a sampling rate of 10 Hz. We attribute this divergence in performance to the limited learning power of our light-weight model HiVT, which may reach a performance bottleneck as additional map information provides diminishing returns for learning. In contrast, the OSM-based approach, benefiting from less detailed initial information, capitalizes on an expanded observation landscape, enabling HiVT to better comprehend the overall road network in the surrounding environment.

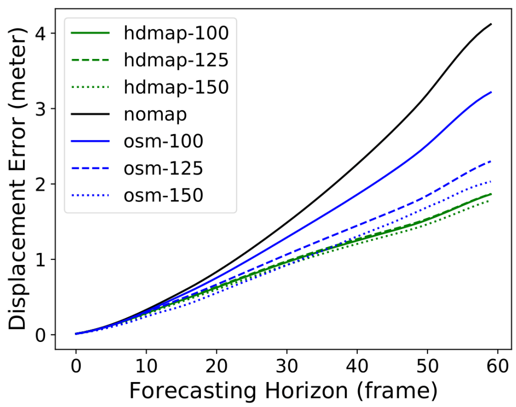

Fig. 3. By comparing the displacement error between various observation landscapes from 100 meters (indicated by the blue dashed line) to 150 meters (indicated by the blue solid line) reveals significant improvements in OSM-based method across the entire forecasting horizon.

Even with such a breakthrough in performance, we can still observe that there remains a slight gap in minFDE while the OSM-based method comes very close to matching the HD-map-based method in minADE. This highlights the need for more in-depth investigation. To achieve a more comprehensive understanding, we conducted a frame-by-frame analysis of displacement errors, considering that tracked agent information is provided at a rate of 10 frames per second, as depicted in Figure 3. It is evident that with additional OSM information, the error remains constrained to an almost linear increase over time, even performing really close to HD-map-based methods, emphasizing the benefits of incorporating more map data for motion forecasting. Furthermore, increasing observation landscape is more feasible for the OSM-based method, as it is essentially impossible to exceed the range of available OSM data.

Importantly, expanding the observation landscape in HiVT incurs minimal computational overhead during inference. Our evaluation, conducted on a Titan Xp GPU with a batch size of 32, reveals significantly less computational burden than HiVT after the observation landscape increase, both

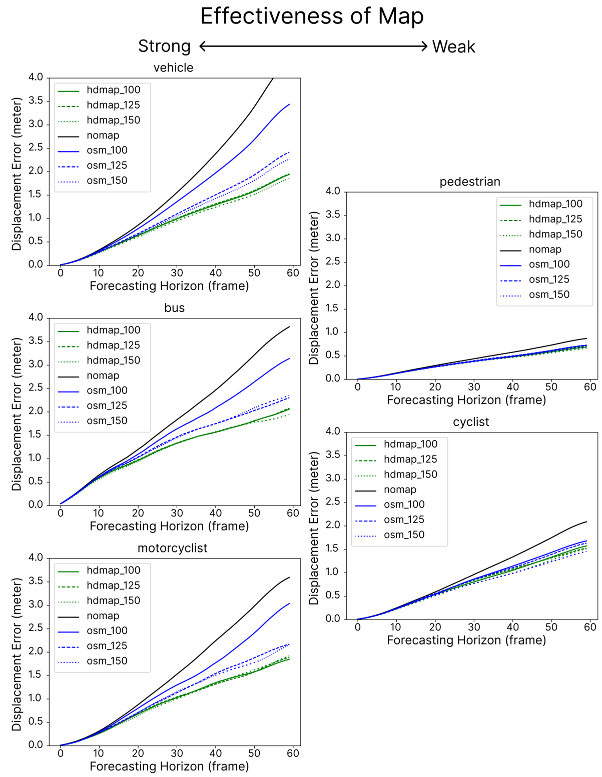

Fig. 4. In this figure, we show class-wise displacement error across forecasting horizon. We can tell that agents on the road (shown on the left), benefit from detailed lane-level information, while pedestrians and cyclists (shown on the right), demonstrate limited advancement in performance.

in terms of inference speed and GPU VRAM usage. This makes it an ideal choice for enhancing motion forecasting capabilities. It is essential to underscore that enlarging the observation landscape involves a trade-off between performance and memory consumption. Additionally, we find that expanding beyond 150 meters offers minimal benefits for two primary reasons: performance improvements reach a bottleneck, and memory consumption increases. Moreover, considering our focus on low-speed urban scenarios rather than highways, a 150-meter observation landscape provides more than sufficient foresight, translating to over 10 seconds into the future while traveling at 30 miles per hour.

E. Qualitative Analysis

In addition to our quantitative findings, we conducted a qualitative analysis with a focus on class-specific performance and visualization, shown in Fig. 4, to gain deeper insights into the influence of map data on motion forecasting for different agent types.

Within the context of our class-specific analysis, we observed that the forecasting performance for pedestrians and cyclists showed limited improvement when utilizing HD Map as oppose to OSM. Two primary factors contribute to this outcome. Firstly, the inherent low speed of both pedestrian and cyclist movements naturally constrains displacement errors to low levels. Secondly, the complex movement patterns exhibited by pedestrians and cyclists, which do not strictly

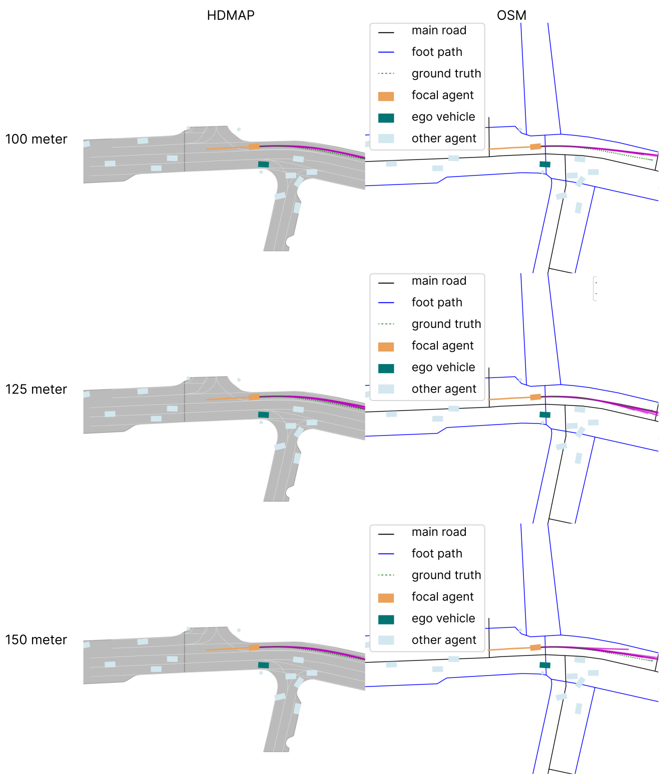

Fig. 5. Illustration of a section with slightly curved road.

adhere to lanes or road connectivity, result in both OSM and HD Map providing comparable information in such instances. Notably, the OSM-based method, incorporating a 150-meter observation landscape, outperformed all methods in predicting cyclist and pedestrian behavior. We attribute this success to the inclusion of bike lanes information in OSM, which may not be as comprehensively covered in Argoverse2 HD maps. Conversely, improvements in performance were evident for vehicle, motorcyclist, and bus when leveraging HD Map, underscoring the potential significance of lane-level information in certain scenarios.

However, our objective extends beyond the mere observation of performance disparities between OSM-based and HD-map-based methods. To uncover the underlying reasons for these disparities, scenarios were systematically categorized into three cases: straight and curved roads, and intersections. Two representative scenarios from the Argoverse2 validation set were meticulously selected for each case, totaling over 20,000 scenarios.

Drawing upon general knowledge, it was evident that lane-level information was primarily crucial at intersections for contextual understanding. To validate this hypothesis, scenarios representing straight and curved roads were visually examined, as shown in Fig. 5. The forecasting results, represented by purple lines with lower transparency indicating higher confidence, closely aligned with the ground truth trajectories depicted by green dotted lines. Additionally, minimal differences were observed between OSM-based and HD-map-based methods in these scenarios, indicating that road connectivity alone provides sufficient context for forecasting in straight and curved road scenarios. In contrast, in intersection scenarios where accurate identification of intersection locations is vital, the advantage of lane-level knowledge in HD Map became evident.

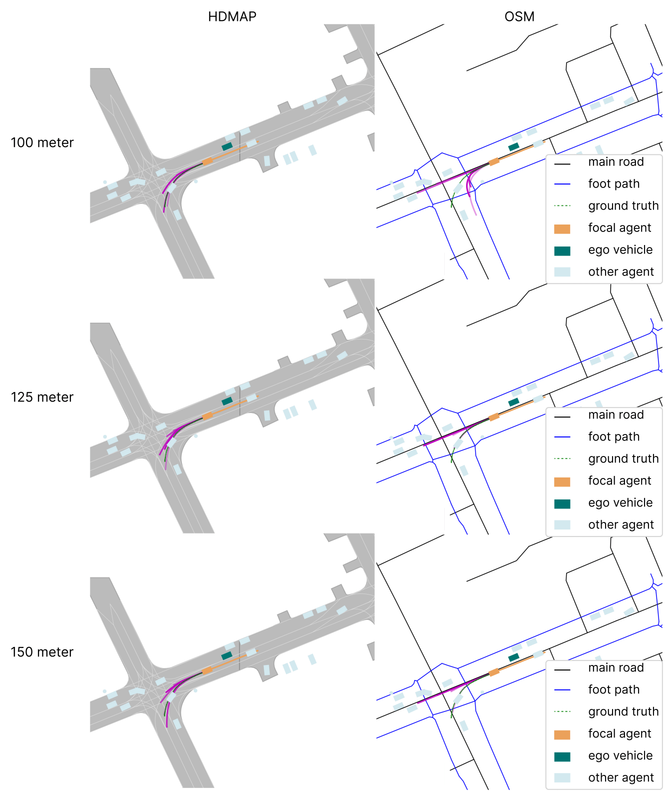

Fig. 6. Illustration of an intersection scenario.

Fig. 6 further illustrates this point, showing that as an agent approaches an intersection, HD-map-based methods exhibit a strong belief in turning behavior. The prediction is aligned with the agent's position to the left of the road, attributed to the availability of lane-level information. Conversely, the OSM-based method with a smaller observation landscape produces inaccurate predictions of turning behavior. However, when the observation landscape is increased to 150 meters, the OSM-based method shows comparable predictions at the intersection. While expanding the observation landscape in OSM-based methods aids in predicting turning behavior, the forecasts are less confident, highlighting the limitation of OSM-based methods due to the absence of lane-level information. This observation underscores the constraint of OSM-based methods in motion forecasting based on information extracted from OSM. While OSM-based methods may not provide highly confident predictions in turning behavior, the outcomes represent possible actions the agent might take given road connectivity. These results, although less confident, may be valuable for downstream tasks due to the prevailing multi-modality in the output, a direction we intend to explore in future work.

V. CONCLUSION

In summary, this research presents an approach using OpenStreetMap (OSM) for long-term prediction when coupled with HiVT. Through the expansion of the observation landscape and training our HiVT model with OSM data, we have achieved comparable results to HiVT trained with HD Map on the Argoverse2 dataset. Our exploration of map representations' impact on forecasting performance, with a focus on different agent types through our by-class analysis, has yielded valuable insights. Overall, our proposed methodology holds significant promise in advancing the scalability of autonomous driving systems by leveraging publicly accessible coarse map data. However, it is worth noting that the significance of lane-level information in motion forecasting remains crucial, especially in complex scenarios like intersections, where lane-level data continues to serve as a robust prior for predicting future trajectories.

REFERENCES

[1] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[2] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *ECCV*, 2020.

[3] Balakrishnan Varadarajan, Ahmed Hefny, Avikalp Srivastava, Khaled S. Refaat, Nigamaa Nayakanti, Andre Cornman, Kan Chen, Bertrand Douillard, Chi Pang Lam, Dragomir Anguelov, and Benjamin Sapp. Multipath++: Efficient information fusion and trajectory aggregation for behavior prediction, 2021.

[4] Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding HD maps and agent dynamics from vectorized representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11525–11533, 2020.

[5] Mayank Bansal, Alex Krizhevsky, and Abhijit Ogale. Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. 06 2019.

[6] Ersi. World imagery wayback. `https://livingatlas.arcgis.com/wayback/#active=4905&ext=-117.23339,32.87881,-117.23091,32.88101`.

[7] Henrik Christensen, David Paz, Hengyuan Zhang, Dominique Meyer, Hao Xiang, Yunhai Han, Yuhan Liu, Andrew Liang, Zheng Zhong, and Shiqi Tang. Autonomous vehicles for micro-mobility. *Springer - Autonomous and Intelligent Systems*, Nov 2021.

[8] Younghun Cho, Giseop Kim, Sangmin Lee, and Jee-Hwan Ryu. Openstreetmap-based lidar global localization in urban environment without a prior lidar map. *IEEE Robotics and Automation Letters*, 7(2):4999–5006, 2022.

[9] Philipp Ruchti, Bastian Steder, Michael Ruhnke, and Wolfram Burgard. Localization on openstreetmap data using a 3d laser scanner. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 5260–5265. IEEE, 2015.

[10] Matthias Hentschel and Bernardo Wagner. Autonomous robot navigation based on openstreetmap geodata. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 1645–1650. IEEE, 2010.

[11] KAFA Samah, S Ibrahim, N Ghazali, M Suffian, M Mansor, and WA Latif. Mapping a hospital using openstreetmap and graphhopper: A navigation system. *Bulletin of Electrical Engineering and Informatics*, 9(2):661–668, 2020.

[12] Jiaolong Xu, Liang Xiao, Dawei Zhao, Yiming Nie, and Bin Dai. Trajectory prediction for autonomous driving with topometric map. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8403–8408. IEEE, 2022.

[13] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . `https://www.openstreetmap.org`, 2017.

[14] Julian Schmidt, Julian Jordan, Franz Gritschneder, Thomas Monninger, and Klaus Dietmayer. Exploring navigation maps for learning-based motion prediction. *arXiv preprint arXiv:2302.06195*, 2023.

[15] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps, 2019.

[16] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James 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.

[17] Phillip Karle, Maximilian Geisslinger, Johannes Betz, and Markus Lienkamp. Scenario understanding and motion prediction for autonomous vehicles—review and comparison. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):16962–16982, 2022.

[18] A. Elnagar. Prediction of moving objects in dynamic environments using kalman filters. In *Proceedings 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation (Cat. No.01EX515)*, pages 414–419, 2001.

[19] Andreas Møgelmose, Mohan M. Trivedi, and Thomas B. Moeslund. Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations. In *2015 IEEE Intelligent Vehicles Symposium (IV)*, pages 330–335, 2015.

[20] Nemanja Djuric, Vladan Radosavljevic, Henggang Cui, Thi Nguyen, Fang-Chieh Chou, Tsung-Han Lin, Nitin Singh, and Jeff Schneider. Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2095–2104, 2020.

[21] Joey Hong, Benjamin Sapp, and James Philbin. Rules of the road: Predicting driving behavior with a convolutional model of semantic interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8454–8462, 2019.

[22] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, pages 683–700. Springer, 2020.

[23] Alexey Kamenev, Lirui Wang, Ollin Boer Bohan, Ishwar Kulkarni, Bilal Kartal, Artem Molchanov, Stan Birchfield, David Nistér, and Nikolai Smolyanskiy. Predictionnet: Real-time joint probabilistic traffic prediction for planning, control, and simulation. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8936–8942. IEEE, 2022.

[24] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele. Mtr++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying. *arXiv e-prints*, pages arXiv–2306, 2023.

[25] Junru Gu, Chen Sun, and Hang Zhao. Densetnt: End-to-end trajectory prediction from dense goal sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15303–15312, 2021.

[26] Nigamaa Nayakanti, Rami Al-Rfou, Aurick Zhou, Kratarth Goel, Khaled S. Refaat, and Benjamin Sapp. Wayformer: Motion forecasting via simple & efficient attention networks. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2980–2987, 2023.

[27] Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. Gohome: Graph-oriented heatmap output for future motion estimation. In *2022 international conference on robotics and automation (ICRA)*, pages 9107–9114. IEEE, 2022.

[28] Cong Zhao, Andi Song, Yuchuan Du, and Biao Yang. Trajgat: A map-embedded graph attention network for real-time vehicle trajectory imputation of roadside perception. *Transportation research part C: emerging technologies*, 142:103787, 2022.

[29] Zikang Zhou, Jianping Wang, Yung-Hui Li, and Yu-Kai Huang. Query-centric trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[30] Maosheng Ye, Tongyi Cao, and Qifeng Chen. Tpcn: Temporal point cloud networks for motion forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11318–11327, 2021.