Infrastructure-Assisted Defensive Forecasting of Latent Traffic Hazards from Uncalibrated CCTV

Je-Seok Ham[1,2][0000−0002−9438−5460], Hyunseo Lee[3], Yongseon Lee[4], Kwanyong Park[5], Jinyoung Moon[1], and Changick Kim[2]

[1] Electronics and Telecommunications Research Institute (ETRI)
jsham@etri.re.kr
[2] Korea Advanced Institute of Science and Technology (KAIST)
[3] Kyungpook National University (KNU)
[4] Seoul National University of Science and Technology
[5] University of Seoul

Abstract. Safe autonomous driving requires anticipating potentially hazardous vehicle motions beyond the limited field of view of ego-centric sensors. Existing urban CCTV cameras provide a complementary infrastructure-side perspective, but their use for motion forecasting is hindered by unknown camera parameters, perspective distortion, and noisy vehicle tracks. To address these challenges, we present NAD-Traj, an infrastructure-assisted trajectory forecasting framework designed for uncalibrated urban CCTV. We introduce a calibration-free vehicle trajectory extraction pipeline and a robust graph-based prediction model equipped with a GRU-based temporal refinement module and a noise-aware loss function. Using this pipeline, we construct NAD-Traj-DB, containing 7,364 traffic scenes, over 528K vehicle-track instances, and 21.4M trajectory points from urban intersections. Experimental results on NAD-Traj-DB and the public V2X-Seq benchmark demonstrate that our approach reduces the Miss Rate (MR) by 23.6% relative to the strongest evaluated baseline. These results highlight the potential of uncalibrated urban CCTV to serve as a promising infrastructure-side perception and forecasting framework for future defensive-driving systems.

1 Introduction

Developing autonomous systems that can anticipate potentially hazardous road-user motions remains a core challenge for intelligent transportation systems [6, 10]. Traditional autonomous vehicles rely primarily on ego-centric sensors, such as LiDAR and onboard cameras, whose fields of view can be severely restricted by occlusion at complex urban intersections [2, 11, 13]. As a result, surrounding vehicles performing nonlinear maneuvers, including abrupt lane changes [4] and sharp turns, may remain partially or fully hidden until they approach a potential conflict area. Anticipating such motions [5, 7, 12] before they become directly observable is therefore important for defensive driving.

Existing monocular urban closed-circuit television (CCTV) cameras [14, 15, 22] offer a scalable and cost-effective complementary perspective by providing broad coverage of urban intersections [8, 9]. However, extracting reliable motion information from these infrastructure streams remains challenging because of substantial observation uncertainty [18]. Unlike data collected using carefully calibrated autonomous-driving platforms, real-world CCTV cameras often lack explicit camera parameters [1], resulting in perspective distortion, variations in target scale, and temporal tracking jitter. Most existing trajectory forecasting approaches [19, 20] are developed using calibrated, map-aligned trajectories obtained from controlled platforms and are not explicitly designed for uncalibrated infrastructure video.

To address these challenges, we present NAD-Traj, an infrastructure-assisted trajectory forecasting framework designed for uncalibrated urban CCTV. Our framework first employs a trajectory-guided homography estimator that predicts local geometric constraints from tracked vehicles and aggregates them to recover a globally consistent bird's-eye-view (BEV) transformation without explicit camera parameters. We then introduce a graph-based multi-modal forecasting model that captures agent–agent, agent–lane, and scene-level interactions. A gated recurrent unit (GRU)-based residual refinement module and a noise-aware loss function further reduce sensitivity to temporal jitter and spatial outliers. Through these contributions, our framework transforms noisy infrastructure observations into stable and kinematically plausible trajectory forecasts that can complement ego-centric perception for defensive driving.

2 Proposed Framework: NAD-Traj

We propose the **N**oise-**A**ware **D**efensive **Traj**ectory Prediction model (NAD-Traj), an infrastructure-assisted forecasting framework designed to support defensive driving under observation uncertainty, as shown in Fig. 1. The framework addresses two challenges: recovering physically consistent trajectories from uncalibrated infrastructure and forecasting future motion from noisy observations.

2.1 Calibration-Free Infrastructure Perception

Given a sequence of uncalibrated 2D bounding boxes extracted using YOLOv8x-VisDrone [17] and ByteTrack [23], we estimate an image-to-ground homography through a two-stage trajectory-guided pipeline. First, due to the lack of ground-truth geometric annotations in real-world CCTV, a Transformer encoder is pretrained on a synthetic dataset. It processes 8-dimensional local features extracted from tracked bounding boxes to predict local 2×2 Jacobians. During deployment, the estimator is frozen, and the model robustly infers these local Jacobians directly from uncalibrated CCTV observations. To aggregate these noisy local predictions, we employ a RANSAC-based optimization that iteratively samples pairs of predicted Jacobians to compute candidate homographies. The final globally consistent transformation is selected by maximizing a consensus score across all local predictions, effectively handling outliers and measurement noise. The

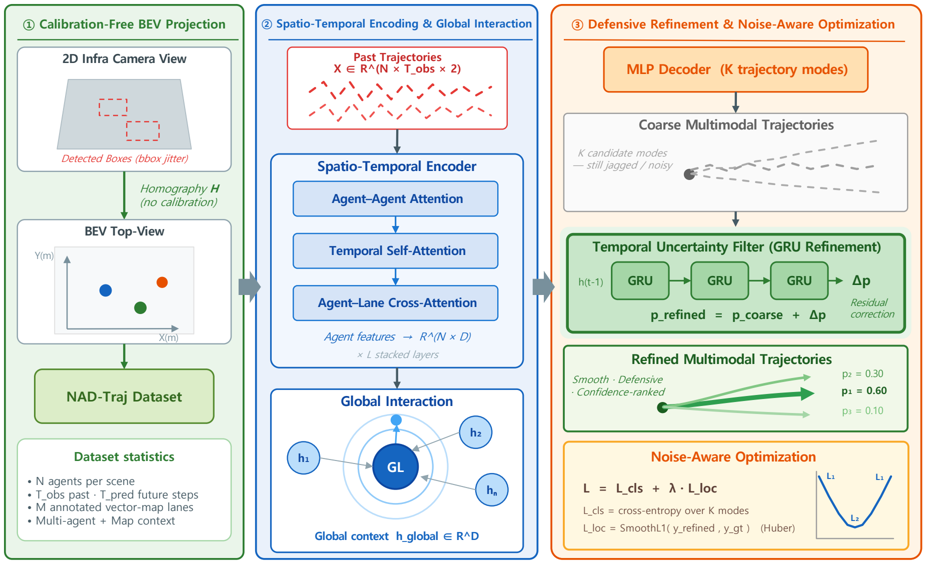

Fig. 1: Overview of the proposed Noise-Aware Defensive Trajectory Prediction model (NAD-Traj).

resulting global homography $\hat{\mathbf{H}}$ projects an image coordinate $\mathbf{p}_t$ into a bird's-eye-view (BEV) coordinate $\mathbf{P}_t$:

$$\mathbf{P}_t = \mathcal{H}(\mathbf{p}_t; \hat{\mathbf{H}}), \tag{1}$$

where $\mathcal{H}(\cdot)$ denotes the homography transformation.

2.2 Defensive Forecasting under Uncertainty

Although the homography provides a consistent BEV representation, the reconstructed trajectories may still contain bounding-box drift, tracking jitter, and minor geometric distortions. We therefore introduce a graph-based multi-modal trajectory forecasting model designed for noisy infrastructure observations.

Hierarchical Context Graph: We represent agents and lane elements as a spatio-temporal graph. Agent-agent, temporal, and agent-lane encoders extract spatial interactions, motion dynamics, and topological constraints from the vectorized lane context. Crucially, a Global Interaction Token aggregates information from all encoded agents to model intersection-level context and long-range dependencies. This process yields a contextual representation $\mathbf{c}_v$ for each target agent v.

Temporal Uncertainty Filter: A multi-modal MLP decoder generates K coarse future trajectories, denoted as $\check{\mathbf{p}}_{t,m}^{(v)}$, together with their confidence scores. Because observation noise may propagate through direct decoding and produce temporally inconsistent outputs, we apply a GRU-based residual refinement module to each predicted mode:

$$\mathbf{s}_{t,m}^{(v)} = \mathrm{GRU}\left([\check{\mathbf{p}}_{t,m}^{(v)}; \mathbf{c}_v], \mathbf{s}_{t-1,m}^{(v)}\right), \qquad \Delta\mathbf{p}_{t,m}^{(v)} = \mathrm{MLP}(\mathbf{s}_{t,m}^{(v)}), \tag{2}$$

where $\mathbf{s}_{t,m}^{(v)}$ represents the hidden state of the GRU at time step t, and $\mathbf{c}_v$ is the dense contextual feature for agent v. The final refined trajectory $\hat{\mathbf{p}}_{t,m}^{(v)}$ is obtained through residual correction:

$$\hat{\mathbf{p}}_{t,m}^{(v)} = \check{\mathbf{p}}_{t,m}^{(v)} + \Delta\mathbf{p}_{t,m}^{(v)}. \tag{3}$$

This formulation preserves the coarse motion hypothesis while mitigating temporal jitter and discontinuities.

Noise-Aware Optimization: Standard loss functions are notoriously vulnerable to CCTV observation noise. To correct minor temporal fluctuations while preventing massive spatial outliers from hijacking the gradient, we optimize the network using a noise-aware Smooth L1 regression loss:

$$SmoothL1(\beta) = \begin{cases} \frac{1}{2\tau}\beta^2, & \text{if } |\beta| < \tau \\ |\beta| - \frac{1}{2}\tau, & \text{otherwise} \end{cases}$$

Here, $\beta = \hat{\mathbf{p}}_{t,m}^{(v)} - \mathbf{p}_t^{*(v)}$ denotes a scalar coordinate-wise prediction error between $\hat{\mathbf{p}}_{t,m}^{(v)}$ and the ground-truth position $\mathbf{p}_t^{*(v)}$, and τ controls the transition between quadratic and linear penalization. This formulation applies an L_2 penalty for small deviations to allow fine-grained adjustments, then switches to an L_1 penalty for large errors so that catastrophic infrastructure tracking flaws do not dominate the gradient.

Because the network generates K diverse multi-modal predictions to account for future uncertainty, we adopt a Best-of-K assignment to compute the localization loss over the prediction horizon, T_{fwd}:

$$\mathcal{L}_{loc} = \min_{m\in\{1,\ldots,K\}} \sum_{t=1}^{T_{fwd}} \mathrm{SmoothL1}\left(\hat{\mathbf{p}}_{t,m}^{(v)} - \mathbf{p}_t^{*(v)}\right)$$

This Winner-Takes-All formulation ensures that only the most accurate predicted mode is penalized and refined. By doing so, the model preserves the multi-modal diversity of potential defensive maneuvers while applying the robust Huber penalty to the best candidate to effectively filter out infrastructure-induced noise.

Let m^* denote the mode selected by the localization objective and $\pi_{m^*}^{(v)}$ its predicted confidence. The classification loss is defined as:

$$\mathcal{L}_{\text{cls}} = -\log \pi_{m^*}^{(v)}. \tag{4}$$

The final training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda\mathcal{L}_{\text{loc}}, \tag{5}$$

Table 1: Validation of the geometric trajectory extraction on the public BrnoCompSpeed benchmark.

Method	Absolute Error (km/h) ↓		Relative Error (%) ↓	
	Average	Median	Average	Median
FullACC [3]	8.59	8.45	11.41	10.89
Learned + RANSAC [15]	2.15	1.60	2.07	2.65
NAD-Traj Extraction (Ours)	1.71	1.19	2.09	1.52

where λ balances mode classification and trajectory regression. This equilibrium allows the model to learn robust defensive behaviors even in highly noisy environments.

3 Experiments

3.1 Dataset and Implementation Details

We evaluate NAD-Traj on the proposed NAD-Traj-DB and the public V2X-Seq benchmark [21]. NAD-Traj-DB contains 528,501 vehicle-tracked instances and approximately 21.4 million trajectory points collected from three uncalibrated urban CCTV intersections. The forecasting subset contains 19,859 dynamically informative target agents. We use 6,080 scenes for training and 1,284 scenes for evaluation. To prevent data leakage from overlapping vehicle tracks, the training and evaluation sets are strictly separated by non-overlapping temporal windows for each intersection.

For all experiments, the models observe 4.0 s of trajectory history and predict 4.0 s into the future. The decoder generates $K = 6$ trajectory modes. NAD-Traj is trained with an initial learning rate of 1×10^{-3} and a cosine-annealing schedule. The loss weight and Smooth L1 transition threshold are set to $\lambda = 1.1$ and $\tau = 1.0$, respectively. All compared methods use the same observation and prediction horizons.

3.2 Evaluation Metrics

We evaluate the predictive performance using three standard metrics: minimum Average Displacement Error (minADE), minimum Final Displacement Error (minFDE), and Miss Rate (MR). Given K predicted trajectories, minADE and minFDE measure the minimum average and final Euclidean displacement errors over the predicted modes, respectively. MR is the fraction of samples whose minFDE exceeds $\epsilon = 2.0\,\mathrm{m}$:

$$\mathrm{MR} = \frac{1}{N} \sum_{i=1}^{N} \nVdash \left(\mathrm{minFDE}_i > \epsilon\right). \tag{6}$$

Table 2: Quantitative performance comparison across different observation uncertainty domains. To emphasize defensive forecasting reliability against latent hazards, the Miss Rate (MR), is highlighted.

Method	Uncalibrated CCTV (NAD-Traj-DB)			Cooperative V2X (V2X-Seq)		
	minADE ↓	minFDE ↓	**MR ↓**	minADE ↓	minFDE ↓	**MR ↓**
TNT [24]	2.182	3.698	0.484	7.38	15.27	0.72
HiVT [25]	1.033	1.543	<u>0.204</u>	1.28	2.11	0.31
V2X-Graph [16]	<u>0.996</u>	<u>1.517</u>	0.212	<u>1.17</u>	<u>2.03</u>	<u>0.29</u>
NAD-Traj (Ours)	**0.943**	**1.288**	**0.162**	**1.13**	**1.86**	**0.27**

Table 3: Ablation study on the components of NAD-Traj. The integration of the Temporal Uncertainty Filter (GRU) and Noise-Aware Optimization (Smooth L1 Loss) is critical for preventing failures (MR) under severe infrastructure noise.

Global Interaction Token	Temporal Uncertainty Filter	Noise-Aware Optimization	minADE ↓	minFDE ↓	MR ↓
-	-	-	1.112	1.771	0.242
✓	✓	-	1.098	1.689	0.235
✓	-	✓	1.033	1.430	<u>0.174</u>
-	✓	✓	<u>0.997</u>	<u>1.395</u>	<u>0.174</u>
✓	✓	✓	**0.943**	**1.288**	**0.162**

3.3 Quantitative and Qualitative Evaluation

Geometric Extraction Accuracy: Before evaluating trajectory forecasting, we validate the calibration-free extraction module on the public BrnoCompSpeed benchmark [18], which provides physical vehicle-speed annotations. As shown in Table 1, our extraction module achieves mean and median absolute errors of 1.71 and 1.19 km/h, respectively, outperforming the evaluated automatic calibration baselines. These results support the geometric consistency of the reconstructed BEV trajectories.

Robustness to Observation Uncertainty: Table 2 compares the predictive performance of NAD-Traj against representative baseline models, including TNT [24], HiVT [25], and V2X-Graph [16]. Our evaluations demonstrate that NAD-Traj achieves the lowest MR of 0.162, marking a substantial 23.6% reduction in prediction failures compared to the representative models V2X-Graph (MR: 0.212). Baseline architectures designed primarily for cleaner sensor observations show degraded forecasting accuracy in the uncalibrated infrastructure setting. The significant error reduction achieved by our model validates its capacity to translate noisy infrastructure observations into more accurate and robust trajectory forecasts.

To verify the framework's resilience beyond isolated single-camera noise, we extended our evaluation to the public V2X-Seq benchmark under the cooperative perception (PP-VIC) setting. NAD-Traj consistently outperformed all evaluated baselines on the V2X-Seq benchmark, pushing the minFDE down to 1.86 and the MR to 0.27. Ultimately, this confirms that our uncertainty-aware architecture does not merely overfit to local CCTV noise but serves as a promising

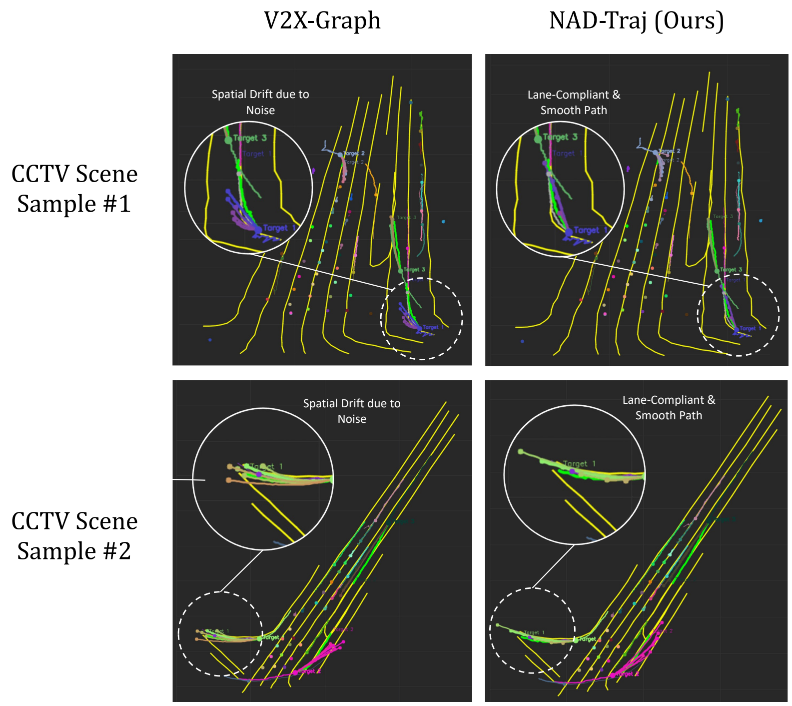

Fig. 2: Qualitative results comparison between V2X-Graph (left) and our proposed NAD-Traj (right).

infrastructure-side perception and forecasting framework for future defensive-driving systems across diverse, heterogeneous infrastructure deployments.

Component Ablation Study: To validate the individual contributions of our proposed uncertainty-handling mechanisms, we conduct a comprehensive ablation study as summarized in Table 3. Substituting the noise-aware formulation with a standard loss drastically degrades overall performance—most notably causing the Miss Rate (MR) to surge to 0.235. Furthermore, bypassing the GRU-based temporal uncertainty filter increases the minFDE to 1.430. Ultimately, this confirms that integrating all components proves essential for neutralizing infrastructure-induced noise and securing safe, defensive predictive behaviors.

Trajectory Visualization: To intuitively illustrate our framework's robustness against infrastructure uncertainty, we visualize the predicted multi-modal trajectories in Fig. 2. In scenarios featuring complex right-turn maneuvers, which frequently act as latent hazards for oncoming ego-vehicles, the baseline V2X-Graph exhibits severe structural vulnerabilities. Conversely, NAD-Traj intelligently filters out this observation noise, maintaining strictly lane-compliant and kinematically smooth future trajectories.

4 Conclusion

In this paper, we presented an infrastructure-assisted framework designed for the defensive forecasting of latent traffic hazards under severe observation uncertainty. By eliminating the reliance on explicit camera calibration and integrating noise-aware learning mechanisms (temporal uncertainty filter and Smooth L1 optimization), our model successfully translates highly noisy CCTV video streams into reliable, lane-compliant trajectory predictions. Experimental results demonstrate that this approach significantly reduces critical prediction failures (Miss Rate). Ultimately, this work illustrates that existing, uncalibrated urban camera networks can be effectively leveraged to provide a promising infrastructure-side perception and forecasting framework for future defensive-driving systems.

5 Acknowledgement

This work was supported by IITP grant funded by the Korea government (MSIT) (No. RS-2020-II200004, Development of Previsional Intelligence based on Long-term Visual Memory Network).

References

1. Bartl, V., Špaňhel, J., Dobeš, P., Juranek, R., Herout, A.: Automatic camera calibration by landmarks on rigid objects: V. bartl et al. Machine Vision and Applications **32**(1), 2 (2021)
2. Chen, Z., Chen, P., Yang, S., Li, L., Wu, J., Sun, R., Wang, Z., Yu, G.: Multimodal vehicle trajectory prediction via hierarchical attention and raster-vector maps encoding in unstructured road environments. Engineering Applications of Artificial Intelligence **167**, 113819 (2026)
3. Dubská, M., Herout, A., Juránek, R., Sochor, J.: Fully automatic roadside camera calibration for traffic surveillance. IEEE Transactions on Intelligent Transportation Systems **16**(3), 1162–1171 (2014)
4. Gao, K., Li, X., Chen, B., Hu, L., Liu, J., Du, R., Li, Y.: Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. IEEE Transactions on Intelligent Transportation Systems **24**(6), 6203–6216 (2023)
5. Ham, J.S., Bae, K., Moon, J.: Mcip: Multi-stream network for pedestrian crossing intention prediction. In: European Conference on Computer Vision. pp. 663–679. Springer (2022)
6. Ham, J.S., Huang, J., Jiang, P., Moon, J., Kwon, Y., Saripalli, S., Kim, C.: Multimodal understanding with gpt-4o to enhance generalizable pedestrian behavior prediction. Computers and Electrical Engineering **129**, 110741 (2026)
7. Ham, J.S., Kim, D.H., Jung, N., Moon, J.: Cipf: Crossing intention prediction network based on feature fusion modules for improving pedestrian safety. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3666–3675 (2023)
8. Ham, J.S., Kim, D.H., Moon, J.: Watchoutped: A dataset and model for vulnerable pedestrian anticipation in surveillance videos. Knowledge-Based Systems **316**, 113352 (2025)

9. Huang, J., Gautam, A., Choi, J., Saripalli, S.: Ultra-wideband technology for improved detection of vulnerable road users in urban settings: Dataset and evaluation. IEEE Transactions on Intelligent Vehicles (2024)
10. Huang, J., Liu, H., Wang, X., Lin, Q., Song, G., Fan, P., Yu, L.: Real-world autonomous driving: Exploring energy dynamics in gradually evolving traffic flows. IEEE Transactions on Intelligent Transportation Systems (2026)
11. Huang, Y., Du, J., Yang, Z., Zhou, Z., Zhang, L., Chen, H.: A survey on trajectory-prediction methods for autonomous driving. IEEE transactions on intelligent vehicles **7**(3), 652–674 (2022)
12. Kothari, P., Kreiss, S., Alahi, A.: Human trajectory forecasting in crowds: A deep learning perspective. IEEE Transactions on Intelligent Transportation Systems **23**(7), 7386–7400 (2021)
13. Lin, L., Li, W., Bi, H., Qin, L.: Vehicle trajectory prediction using lstms with spatial–temporal attention mechanisms. IEEE Intelligent Transportation Systems Magazine **14**(2), 197–208 (2021)
14. Na, K.I., Park, B.: Real-time 3d multi-pedestrian detection and tracking using 3d lidar point cloud for mobile robot. ETRI Journal **45**(5), 836–846 (2023)
15. Revaud, J., Humenberger, M.: Robust automatic monocular vehicle speed estimation for traffic surveillance. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4551–4561 (2021)
16. Ruan, H., Yu, H., Yang, W., Fan, S., Nie, Z.: Learning cooperative trajectory representations for motion forecasting. Advances in Neural Information Processing Systems **37**, 13430–13457 (2024)
17. Shamrai, M.: Yolov8x-visdrone. `https://huggingface.co/mshamrai/yolov8x-visdrone` (2023), hugging Face model repository, accessed May 20, 2024
18. Sochor, J., Juránek, R., Špaňhel, J., Maršík, L., Širokỳ, A., Herout, A., Zemčík, P.: Comprehensive data set for automatic single camera visual speed measurement. IEEE Transactions on Intelligent Transportation Systems **20**(5), 1633–1643 (2018)
19. Yang, J., Liu, J.w.: Vehicle trajectory prediction model based on graph attention kolmogorov-arnold networks and multiple attention. Engineering Applications of Artificial Intelligence **159**, 111804 (2025)
20. Yu, H., Luo, Y., Shu, M., Huo, Y., Yang, Z., Shi, Y., Guo, Z., Li, H., Hu, X., Yuan, J., et al.: Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21361–21370 (2022)
21. Yu, H., Yang, W., Ruan, H., Yang, Z., Tang, Y., Gao, X., Hao, X., Shi, Y., Pan, Y., Sun, N., et al.: V2x-seq: A large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5486–5495 (2023)
22. Zhang, X., Feng, Y., Angeloudis, P., Demiris, Y.: Monocular visual traffic surveillance: A review. ieee transactions on intelligent transportation systems **23**(9), 14148–14165 (2022)
23. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X.: Bytetrack: Multi-object tracking by associating every detection box. In: European conference on computer vision. pp. 1–21. Springer (2022)
24. Zhao, H., Gao, J., Lan, T., Sun, C., Sapp, B., Varadarajan, B., Shen, Y., Shen, Y., Chai, Y., Schmid, C., et al.: Tnt: Target-driven trajectory prediction. In: Conference on robot learning. pp. 895–904. PMLR (2021)
25. Zhou, Z., Ye, L., Wang, J., Wu, K., Lu, K.: Hivt: Hierarchical vector transformer for multi-agent motion prediction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8823–8833 (2022)