Technical Report for the HetroD Challenge

Zhenyu Zhao, Zihan Zhai, Kexin Zhang, Jing He, Licheng Jiao, Xu Liu, Lingling Li
Key Laboratory of Intelligent Perception and Image Understanding, Xidian University
25171213940@stu.xidian.edu.cn

Abstract

This report describes the Kinematic method used by team winzzz for the HetroD Challenge v1.0. The method is a lightweight kinematic baseline that does not use neural networks or external pretrained models. It extracts motion priors for vehicles, two-wheelers, and pedestrians from the official training set and retrieves motion patterns similar to the current state of each target agent. For every agent, the method constructs a candidate bank containing a constant-velocity anchor trajectory and trajectories augmented with train-set motion residuals. It then selects 32 joint rollouts using oriented-box collision checks, type-specific valid-region constraints, and type-aware kinematic protection. On all 955 validation scenarios, our implementation achieved an overall HetroD score of 0.723261, with scores of 0.708954 for kinematic realism, 0.862053 for safety, and 0.795763 for cross-type interaction. Predictions for all 955 hidden test scenarios passed the official preflight validation with zero errors.

1. Introduction

The HetroD Challenge evaluates the generation of future joint trajectories from historical agent states and scene information. Unlike settings that focus only on motor vehicles, HetroD covers vehicles, two-wheelers, and pedestrians. These agent types differ substantially in speed, turning ability, valid motion regions, and interaction behavior, making heterogeneous traffic modeling an important component of the task.

The evaluation metrics include kinematic realism, collision safety, valid-region compliance, cross-type interaction, and multimodal coverage. Although constant-velocity extrapolation is smooth and stable, it cannot fully represent nonlinear behaviors such as turning, starting, stopping, and deceleration. Independent long-horizon extrapolation may also cause agents to collide or leave plausible motion regions.

In this submission, we use and extend a kinematic baseline based on train-set motion priors and explicit safety selection. The method extracts real motion residuals from the training data, builds a diverse candidate bank, and selects appropriate joint rollout combinations according to collision, valid-region, and type-aware kinematic criteria. The implementation is lightweight, deterministic under a fixed configuration, and does not require neural-network training.

Table 1. Dataset splits used in this submission.

Split	Scenarios	Future GT	Usage
Train	5,087	Yes	Motion priors
Validation	955	Yes	Evaluation
Test	955	No	Submission
Total	6,997	–	–

2. Dataset and Task

2.1. Dataset Splits

We use the HetroD Challenge v1.0 public data with schema version 1.3 [1]. The data contain 6,997 traffic scenarios divided as shown in Table 1. Future ground truth from the 5,087 training scenarios is used only to construct the motion-prior library. The 955 validation scenarios are used for model development and official metric evaluation. Future ground truth is hidden for the 955 test scenarios.

The scenarios include historical positions, velocities, headings, physical dimensions, object types, timestamps, map information, and the IDs of agents to be simulated. The three principal evaluation types are vehicles, two-wheelers, and pedestrians.

2.2. Prediction Target

The current time index is 10. The method reads observations from frames 0–10 and predicts frames 11–90, corresponding to 80 future steps at 0.1-second intervals, or an eight-second prediction horizon. For a scene containing N required agents, the submission tensor has shape

$$\mathbf{S} \in \mathbb{R}^{32\times N\times 80\times 4}, \tag{1}$$

where the last dimension represents global (x, y, z, yaw). The agent set must exactly match `metadata.required_agent_ids` for each test scenario.

2.3. Evaluation Metrics

The overall score combines kinematic realism, safety, cross-type interaction, and coverage. Kinematic realism measures the distributions of linear speed, linear acceleration, angular speed, and angular acceleration. Safety includes oriented-box collision safety and compliance with type-specific valid regions. Cross-type interaction evaluates the distance and timing of interactions between different agent types, while coverage measures whether the 32 rollouts represent multiple plausible futures. The main quality weights for kinematic realism, safety, and cross-type interaction are 0.30, 0.35, and 0.25, respectively. Coverage contributes a gated bonus with a maximum weight of 0.10 and is coupled with both kinematic realism and safety.

3. Method

3.1. Overview and State Estimation

The Kinematic pipeline consists of current-state estimation, motion-prior retrieval, per-agent candidate generation, and joint safety selection. For each required agent, we estimate position, velocity, acceleration, heading, and yaw rate using up to the five most recent valid observations. The final velocity combines a slope fitted to historical positions with the median recorded velocity:

$$\mathbf{v} = 0.7\mathbf{v}_{\mathrm{pos}} + 0.3\mathbf{v}_{\mathrm{record}}. \tag{2}$$

Type-specific static-speed, maximum-speed, and maximum-yaw-rate limits prevent noisy observations from producing extreme extrapolations. Every candidate bank retains a deterministic constant-velocity anchor,

$$\mathbf{p}_{\mathrm{CV}}(t) = \mathbf{p}_0 + t\mathbf{v}_0. \tag{3}$$

3.2. Motion Priors and Candidate Generation

To represent nonlinear motion, we extract local-frame future residuals from the training set. For a training trajectory, the positional residual is

$$\Delta\mathbf{p}_{\mathrm{prior}}(t) = \mathbf{p}_{\mathrm{GT}}^{\mathrm{local}}(t) - t\mathbf{v}_0^{\mathrm{local}}. \tag{4}$$

A corresponding heading residual is calculated relative to a damped yaw-rate reference. Each motion prior is indexed by agent type and four current-state features: speed, acceleration, yaw rate, and historical path extent. At inference time, the method first searches for a matching type and motion bucket. If necessary, it falls back to coarser same-type buckets and samples from the 64 nearest examples according to normalized feature distance. The retrieved positional and heading residuals are scaled by 0.8 and 0.7, respectively, and then transformed back to the global coordinate system. Together with the constant-velocity anchor trajectory, this process generates 64 internal candidate trajectories for each agent.

3.3. Joint Safety and Type-Aware Selection

Independently generated agent trajectories may collide when placed in the same joint rollout. We therefore represent each agent as an oriented rectangle using its observed length, width, and heading, and evaluate strict rectangle overlap between pairs of agents. A rectangular Hungarian assignment selects 32 trajectories from each 64-candidate bank while reducing the type-balanced joint collision objective. For efficiency, assignment costs are estimated every fourth future frame, but each proposed update is rechecked over all 80 frames before it can be accepted.

The selector also reads the type-specific valid-region polygons provided in the scene metadata. A sampled state is considered valid only if all four corners of the agent footprint lie within the corresponding polygon union. When collision safety remains unchanged, a low-weight valid-region cost acts as a secondary selection criterion.

Selection based only on valid-region compliance may favor unrealistically slow or straight candidate trajectories. To reduce this effect, we compare candidates using average speed, acceleration, yaw rate, and yaw acceleration. Full kinematic protection is applied to vehicles and two-wheelers, whereas pedestrians are given greater flexibility and are selected primarily according to collision and valid-region criteria. All agent types remain in the same joint assignment loop.

4. Training Procedure

The method does not train a neural network. Instead, training consists of a one-time offline construction of the motion-prior library. We processed all 5,087 official training scenarios and retained agents that were valid at the current time, had at least five valid history frames, and had all 80 future frames available. Their future trajectories were transformed into an agent-centered coordinate system, and positional and heading residuals were calculated relative to the kinematic baselines. This process produced 109,131 prior samples: 50,079 vehicle examples, 47,474 two-wheeler examples, and 11,578 pedestrian examples. The motion-prior library was constructed entirely from training-set ground truth. Validation future states were used only by the official evaluator, test future states were unavailable, and no external data or pretrained models were used.

5. Inference and Rollout Generation

At inference time, only observations at or before the current timestamp are read. Required test agents are taken directly from `metadata.required_agent_ids`. The method estimates each agent's current state, retrieves same-type priors, generates a 64-trajectory internal candidate bank, and jointly selects the final 32 rollouts. The selector uses two assignment passes, processes at most 16 highly collision-involved agents per pass, and verifies accepted updates on the complete future horizon.

The scenario ID and agent ID determine fixed random seeds, making inference reproducible under the same configuration. Predictions are stored as 32-bit floating-point values in one pickle file per scenario. Each file contains the ordered IDs of all required agents and their corresponding simulated states. The state array has shape $[32, N, 80, 4]$ and contains global (x, y, z, yaw) trajectories for future frames 11–90. With four parallel workers, generation for all 955 hidden test scenarios took approximately 36 minutes and 55 seconds in our environment. The final archive contains all 955 scenario files and 40,739 required agent trajectories. The official preflight validator reported 955 valid scenarios and zero errors.

6. Validation Results

6.1. Experimental Setup

We evaluate on all 955 validation scenarios using the official HetroD Evaluator version 0.8.0. Metrics are macro-averaged across locations, with agent-type balancing within each location. Future validation states are not read by the generator and are used only by the official evaluator.

6.2. Ablation Summary

The initial constant-velocity baseline achieves an overall score of 0.534894. Train-set residual priors provide realistic starting, stopping, turning, and nonuniform-speed behavior. Joint collision selection reduces unsafe combinations, while the expanded candidate bank provides alternatives when simple re-pairing cannot resolve a conflict. Valid-region selection reduces motion outside plausible type-specific areas, and the final type-aware kinematic protection preserves stable vehicle and two-wheeler motion while allowing pedestrians more flexible region-aware choices. With all components enabled, the final Kinematic method reaches 0.723261, an absolute improvement of 0.188367 over the constant-velocity baseline.

6.3. Final Validation Metrics

Table 2 reports the final results. The strong safety and valid-region scores indicate that joint oriented-box selection and map-region constraints effectively reduce implausible combinations. Kinematic realism and cross-type interaction also remain strong, showing that train-set motion residuals improve long-horizon behavior without neural-network training.

Table 2. Results of the final Kinematic method on all 955 validation scenarios.

Metric	Score
Overall HetroD score	**0.723261**
Kinematic realism	0.708954
Linear speed	0.683260
Linear acceleration	0.725609
Angular speed	0.574778
Angular acceleration	0.820991
Safety	0.862053
Collision safety	0.807825
Valid region	0.916281
Cross-type interaction	0.795763
Coverage	0.163481

7. Conclusion

This report describes the lightweight kinematic baseline used by team winzzz for the HetroD Challenge. The method combines train-only motion priors, oriented-box collision checking, type-specific valid-region constraints, and type-aware kinematic protection to generate 32 heterogeneous joint rollouts. It uses no external data, neural networks, or pretrained weights, and its deterministic generation process is straightforward to reproduce. The final method achieved an overall score of 0.723261 on all 955 validation scenarios, and predictions for all 955 test scenarios passed the official submission-format validation.

References

[1] Yu-Hsiang Chen, Wei-Jer Chang, Christian Kotulla, Thomas Keutgens, Steffen Runde, Tobias Moers, Christoph Klas, Wei Zhan, Masayoshi Tomizuka, and Yi-Ting Chen. HetroD: A high-fidelity drone dataset and benchmark for autonomous driving in heterogeneous traffic. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2026.