TrajTokFusion: Multi-Pool Candidate Selection for HetroD Simulation

Ying Tang* Bolian Peng Jinming Chai Xu Liu
Xidian University
y1nnngt@163.com
*Team leader

Abstract

This report describes the Y1nnng submission to the HetroD Challenge v1.0. The task requires realistic 8-second closed-loop rollouts for heterogeneous urban traffic scenes with vehicles, two-wheelers, and pedestrians. We propose TrajTokFusion, a multi-pool trajectory generation and selection pipeline. The method combines a TrajTok e32 branch with a large N128 stochastic candidate pool and a train-plus-validation fine-tuned E40 branch that provides complementary selected rollouts. A final re-selection stage chooses 32 rollouts per scenario using collision, valid-region, kinematic, and progress proxies without using hidden test labels. The method uses the official HetroD public data only. On public validation, this fusion configuration obtains an overall score of 0.733727431962.

1. Task Setting

HetroD evaluates closed-loop simulation for dense heterogeneous urban scenes. For each test scenario, a method receives ScenarioNet-compatible inputs and the official required-agent list. The required output is a set of 32 rollouts, each containing 80 future states for all required agents. The output horizon corresponds to timesteps 11–90, and each state is represented in global coordinates as (x, y, z, ψ).

The evaluation combines kinematic realism, safety, cross-type interaction, and coverage. These objectives are not always aligned: a very conservative rollout can reduce collisions but may also reduce motion realism, while a diverse stochastic sample can improve coverage but may include unsafe outliers. TrajTokFusion addresses this tension by separating stochastic generation from final candidate selection.

2. Method

2.1. Overview

TrajTokFusion is a multi-pool inference pipeline with three main components:

1. heterogeneous trajectory generation with TrajTok-style tokenized models;
2. candidate-pool fusion from complementary generator settings;
3. multi-objective re-selection to obtain the final 32 rollouts.

The goal is to exploit complementary candidate distributions. The N128 branch increases the chance of observing safe and realistic modes, while the train-plus-validation fine-tuned branch provides candidates adapted to the public validation distribution. The final selector then chooses rollouts that balance safety and motion quality.

2.2. Heterogeneous trajectory generation

The generator follows a TrajTok-style tokenized trajectory modeling design. Agent histories, agent-type information, and scene context are encoded jointly. The decoder samples future motion tokens and reconstructs continuous trajectories for all required agents. Vehicles, two-wheelers, and pedestrians share a unified prediction format, while their type information is preserved in the conditioning features.

The HetroD adapter preserves the official required-agent order and converts generated trajectories into the global-coordinate state format required by the challenge:

```
{
  "agent_id": agent_ids,
  "simulated_states": states
}
```

where states has shape $32 \times N \times 80 \times 4$ and N is the number of required agents.

Component	Setting
Main generator	TrajTok-style tokenized model
Branch A	e32, 128 sampled rollouts
Branch B	E40 fine-tuned, selected N64
Sampling temperature	0.9
Sampling parameter	$k = 16$
Final rollouts	32 per scenario
Selection objective	safety + kinematic + progress
External data	none

Table 1. Main settings of the TrajTokFusion multi-pool candidate.

2.3. Candidate-pool fusion

The final candidate pool combines two sources:

- an e32 TrajTok branch that samples 128 rollouts per scenario with temperature 0.9 and sampling parameter $k = 16$;
- a train-plus-validation fine-tuned E40 branch that supplies selected candidates from a 64-rollout inference pool.

The two branches are complementary. The e32 N128 pool improves inference-time search, while the fine-tuned branch contributes trajectories from a model adapted with all public labels. The fused pool is then re-ranked at the scenario level before submission.

2.4. Multi-objective re-selection

The re-selection stage does not use hidden test labels. It scores candidates with lightweight proxies:

- collision risk, penalizing close or overlapping agent interactions;
- valid-region risk, penalizing states likely to leave type-aware valid regions;
- kinematic plausibility, preferring smoother and more physically plausible motion;
- progress consistency, discouraging overly static or degenerate futures.

The selected 32 rollouts are the candidates with the best combined proxy scores. Unlike trajectory repair methods, this stage does not directly edit the trajectories. It only selects among generated candidates, which helps avoid artificial discontinuities in velocity, acceleration, or yaw.

3. Training

We use the official HetroD Challenge v1.0 public data only. The generator branches are trained or fine-tuned with the released public labels. Hidden test labels are not used, and we do not use external datasets, manual test annotations, or external pretrained trajectory models.

4. Validation

On public full-validation evaluation, the multi-pool TrajTokFusion configuration achieves:

- overall score: 0.733727431962;
- kinematic realism: 0.678138260114;
- safety: 0.906420581978;
- collision score: 0.834555084166;
- valid-region score: 0.978286079791;
- cross-type interaction: 0.822136434990;
- coverage: 0.122150756435.

The improvement over the N64 single-branch submitted configuration is mainly due to the enlarged N128 candidate pool and the complementary train-plus-validation branch. We also observed that purely safety-oriented selection can hurt kinematic realism, so the final re-selection includes kinematic and progress proxies in addition to collision and valid-region terms.

Before final submission, the test archive is checked with the official public preflight script to verify scenario coverage, file naming, required-agent IDs, output shape, finite states, and format validity. The preflight result is successful, with 955 required scenarios, 955 submitted files, 955 valid scenarios, 40,739 required-agent trajectories, and zero format errors.

5. Discussion

TrajTokFusion treats closed-loop traffic simulation as generation followed by structured candidate selection. This design is practical for HetroD because it keeps multimodality while reducing unsafe or invalid futures. The fusion stage is especially useful when different model settings have complementary strengths across scene density, agent type, and motion regime.

The main limitation is that the selection proxies are approximate. They improve the reliability of the candidate pool, but cannot fully replace a stronger interaction model trained with explicit closed-loop safety objectives. Future work should integrate differentiable interaction and kinematic regularization directly into training while keeping the multi-pool inference strategy.