Article

Network Dreamer: Bridging Scenario-Level and Network-Level Traffic Simulation via Star Pattern Matching

Xingnan Zhou [1] **and Ciprian Alecsandru** [1,*]

[1] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC H3G 1M8, Canada
* Correspondence: ciprian.alecsandru@concordia.ca (C.A.)

Version February 12, 2026 submitted to Sustainability

Abstract: Autonomous driving datasets contain rich driving scenarios, yet GPS anonymization disconnects them from the road network context required for sustainable transportation planning. We propose *Network Dreamer*, a GPS-free localization framework that places Waymo Open Motion Dataset (WOMD) scenarios onto OpenStreetMap networks using a 48-dimensional "star pattern" descriptor encoding intersection geometry, arm angles, lengths, road types, and neighbor topology. Star pattern matching achieves 90.0% top-1 accuracy in synthetic evaluation but degrades to 33.5–55.3% under realistic noise—quantifying the WOMD–OSM domain gap. Route-level matching, chaining star patterns along SDC trajectories with inter-distance constraints across all 1,000 WOMD v1.3.1 shards (70,541 scenarios), achieves 51.9% match rate (17,645 routes), validated by 1,597 chain repeats with sub-2-meter consistency, 100% type agreement, and 100% spatial accuracy against satellite imagery (50 scenarios). Matched routes yield corridor-level traffic parameters across 990 road segments (187.4 km), including travel time reliability indices and speed transition zones directly usable for microsimulation calibration. A safety overlay reveals WOMD under-samples signalized intersections (4.3% vs. 17.8% city-wide), a bias discoverable only through GPS-free localization (SDG 3.6).

Keywords: autonomous driving; traffic simulation; road network; star pattern matching; topological fingerprint; OpenStreetMap; Waymo Open Motion Dataset; sustainable transportation; traffic safety; surrogate safety measures; urban planning; map conflation; SDG

---

1. Introduction

Sustainable urban transportation demands simulation tools that faithfully represent the interplay between individual driving behavior and network-level traffic dynamics [1,2]. Yet the dominant paradigm in AV simulation operates at the *scenario level*—generating or replaying short driving episodes within small spatial extents—leaving a gap between these local snapshots and the network-wide perspective that urban planners require [3,4].

The Waymo Open Motion Dataset (WOMD) [5] contains 103,354 real-world scenarios spanning 570+ hours across six U.S. cities, each capturing 9.1 seconds of multi-agent behavior with detailed road graph information. Generative approaches such as Scenario Dreamer [6], SLEDGE [7], and TrafficGen [8] can now synthesize driving scenarios with impressive fidelity. However, each scenario exists in isolation, disconnected from the broader road network—a 64 m snippet tells nothing about upstream origins or downstream dispersal [9,10]. Network-level simulators such as SUMO [11] can model city-scale dynamics but rely on manually calibrated demand models. The gap between rich local data with no global context and global simulation with impoverished local detail is a significant barrier to data-driven sustainable transportation research.

In this paper, we propose **Network Dreamer**, a framework that bridges this gap by (1) probabilistically localizing fragmented driving scenarios onto real-world road networks and (2) extracting empirical traffic flow parameters from the matched data. Our approach addresses several interrelated challenges:

Challenge 1: Localization without GPS. WOMD scenarios do not provide absolute geographic coordinates, only local road graph structures. How can we determine *where* on a city's road network each scenario was captured? We develop a systematic framework for GPS-free scenario localization that adapts intersection neighborhood descriptors from the GIS map conflation tradition [12,13] into a compact 48-dimensional "star pattern" representation encoding the center intersection's geometry together with the angles, lengths, road types, and neighbor topology of its approach arms. The key insight is that while individual intersections are often structurally ambiguous (thousands of 4-way intersections exist in a city), the *local neighborhood context*—the specific configuration of arms radiating outward plus the properties of adjacent intersections—creates a highly discriminative fingerprint, and critically, this descriptor alone must provide all discriminative power in the absence of GPS coordinates. Our star pattern achieves 90.0% top-1 matching accuracy on San Francisco and 79.6% on Phoenix, a 225× improvement over single-intersection matching baselines.

Challenge 2: Traffic Flow Extraction at Scale. Given probabilistic localization, how do we aggregate the behavioral data from thousands of short scenarios into meaningful network-level traffic flow parameters? We develop a systematic pipeline to extract turning ratios, speed distributions, headway patterns, and gap acceptance parameters from matched scenarios, with confidence intervals that account for both matching uncertainty and sampling variability.

The contributions of this paper are as follows:

1. **Problem-level:** The **first systematic study of GPS-free WOMD-to-OSM scenario localization**, formulating and addressing the problem of localizing driving scenarios from the Waymo Open Motion Dataset onto public road networks without any geographic coordinates.
2. **System-level:** A **complete pipeline from topology extraction to traffic flow analysis**, including systematic extraction of turning ratios, speed distributions, and behavioral parameters with uncertainty quantification, connecting AV-scale trajectory data to network-level traffic simulation. The pipeline is validated at scale on 56,797 scenarios from 236 TFRecord shards of WOMD v1.3.1, encompassing 3.35 million vehicles and 1.02 million approach observations.
3. **Algorithm-level:** An **adaptation of established GIS intersection features into a 48-dimensional star pattern descriptor** that achieves 90.0% top-1 accuracy on San Francisco and 79.6% on Phoenix—a 225× improvement over single-intersection baselines—by combining intersection degree, approach arm angles, edge lengths, road types, and neighbor topology [12,13] into a fixed-length representation enabling efficient brute-force matching at city scale.
4. A **comprehensive ablation and discriminability analysis** demonstrating that arm angles are the most discriminative single feature (0.4% to 40.8% top-1), and that the progressive addition of arm lengths, road types, and neighbor context achieves 90.0% accuracy with the full 48D star pattern. We also characterize noise robustness and per-intersection-type performance.
5. A **quantitative sustainability analysis** outlining the potential for improved emission estimation accuracy through data-calibrated simulation (to be validated in future work) and mapping contributions to specific UN Sustainable Development Goal targets (SDG 3.6, 9.1, 11.2, 11.6, 13.2).

The remainder of this paper is organized as follows. Section 2 reviews related work on traffic simulation, scenario generation, map matching, and GIS map conflation. Section 3 presents the materials and methods, including the topological fingerprint, the star pattern matching algorithm, discriminability analysis, traffic flow extraction, ground truth validation strategy, and evaluation metrics. Section 4 reports experimental results, with the star pattern matching results as the primary contribution. Section 5 discusses implications, sustainability analysis, limitations, ethical considerations, and future directions. Section 6 concludes.

2. Literature Review

2.1. Autonomous Driving Datasets and Simulators

WOMD [5] provides 103,354 scenarios with lane-level road graphs, 91 timesteps at 10 Hz, and up to 128 tracked agents per scenario. Replay-based simulators (Waymax [14], ScenarioNet [15]) faithfully reconstruct recorded scenarios but are limited to original spatial extents. Generative simulators—SLEDGE [7], CTG++ [16], SceneDiffuser++ [17], TrafficSim [18], TrafficBots [19], Trajeglish [20]—push beyond replay, though SceneDiffuser++ relies on Waymo's non-public geo-referenced data.

2.2. Scenario Generation via Generative Models

TrafficGen [8] pioneered autoregressive traffic generation. Scenario Dreamer [6] combines vectorized latent diffusion with autoregressive trajectory modeling within 64 m regions; its scene extrapolation extends to 500 m routes but without network-level consistency. DriveArena [21] uses OSM road networks with learned behavior models, LCSim [22] targets large-scale controllable simulation, and GraphWalker [23] generates road networks from trajectory data. These works point toward the convergence of local generation and network-level simulation—the gap Network Dreamer addresses.

2.3. Map Matching, Map Conflation, and Road Network Reconstruction

Map conflation. Node-based matching using intersection topology is well-established: Yang et al. [12] proposed pattern-based node matching, Li and Goodchild [13] developed structure-based junction matching, and Walter and Fritsch [24] and Song et al. [25] established foundational geometric/topological approaches. Recently, SP-GEM [26] and RNEM [27] applied learned graph embeddings. In autonomous driving, InterKey [28] uses cross-modal intersection keypoints and Badino et al. [29] pioneered topometric localization. **Probabilistic approaches:** Newson and Krumm [30] established HMM-based map matching; Fan et al. [31] and Zhang et al. [32] showed that learned embeddings improve network matching. Neural map construction (MapTRv2 [33], GNMap [34]) focuses on *building* maps, whereas our task is *matching* an existing local graph to a known network.

How our approach differs. Unlike traditional conflation, which assumes geo-referenced coordinates in both sources, we have *zero* geographic reference (no proximity-based filtering possible). We must bridge WOMD's lane-level graph against OSM's road-level representation (Section 3.2.2), exploit dynamic behavioral context (signals, trajectories), and output probability distributions for downstream uncertainty propagation (Section 3.5).

2.4. Network-Level Traffic Simulation

Network-level simulators (e.g., SUMO [11]) require realistic calibration data [35,36], yet AV datasets lack the geographic context to map observations to network elements. Geo-referenced datasets (highD [37], exiD [38], rounD [39]) could serve as cross-validation targets (Section 3.7).

2.5. Summary of Research Gap

Simulation is critical for sustainable transportation planning [40,41], yet evaluating AV impacts requires network-level capabilities grounded in real behavioral data [42,43]. Table 1 summarizes Network Dreamer's positioning.

Table 1. Comparison of related approaches along key dimensions.

Approach	Scope	Data-Driven	Network	Generative	GPS-Free	Probabilistic
Waymax [14]	Scenario	Yes (replay)	No	No	N/A	N/A
ScenarioNet [15]	Scenario	Yes (replay)	No	No	N/A	N/A
SLEDGE [7]	Scenario	Yes	No	Yes	N/A	N/A
Scenario Dreamer [6]	Scenario	Yes	No	Yes	N/A	N/A
SceneDiffuser++ [17]	City-scale	Yes	Partial	Yes	No	N/A
DriveArena [21]	Network	Partial	Yes	Partial	N/A	N/A
LCSim [22]	Network	Partial	Yes	Yes	N/A	N/A
SUMO [11]	Network	No	Yes	No	N/A	N/A
GraphWalker [23]	Network	Yes	No	Yes	N/A	N/A
Fan et al. [31]	Matching	N/A	N/A	N/A	No	No
Network Dreamer	**Network**	**Yes**	**Yes**	**No**	**Yes (90%)**	**Yes**

No existing approach combines data-driven behavioral modeling, network-level context, GPS-free localization, and probabilistic matching. Network Dreamer fills this gap.

3. Materials and Methods

3.1. Overview

Figure 1 illustrates the overall architecture of Network Dreamer, which consists of three interconnected phases:

- **Phase 1 (Sections 3.2–3.4):** Road network backbone construction via topological fingerprint and star pattern extraction from both WOMD scenarios and OSM networks. The star pattern (Section 3.4.3) encodes each intersection as a 48-dimensional feature vector capturing center properties, arm geometry, and neighbor context.
- **Phase 2 (Section 3.5):** Probabilistic topological matching with discriminability analysis, producing calibrated location probability distributions. Star pattern matching achieves 90.0% top-1 accuracy on San Francisco and 79.6% on Phoenix.
- **Phase 3 (Section 3.6):** Traffic flow pattern extraction from matched scenarios, with confidence intervals accounting for matching uncertainty.

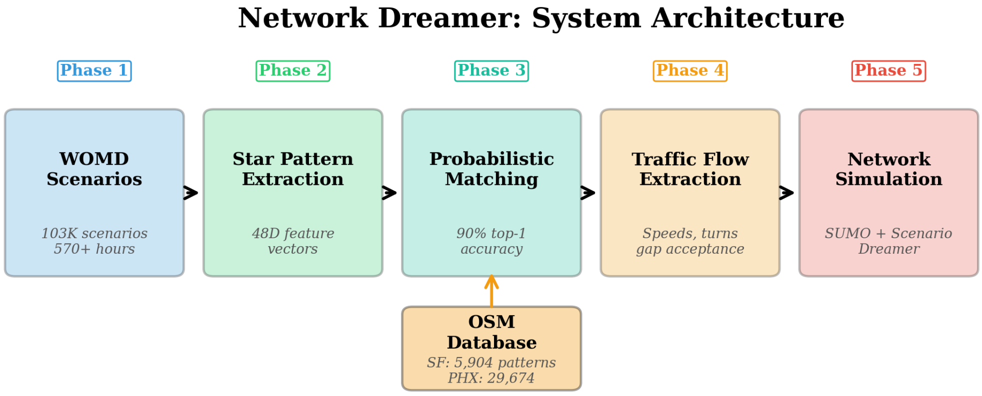

Figure 1. System overview of Network Dreamer. The three-stage pipeline takes fragmented WOMD scenarios and an OSM road network as input, performs star pattern extraction and probabilistic topological matching to localize scenarios, and extracts traffic flow parameters with uncertainty quantification.

3.2. Data Sources and Preprocessing

3.2.1. Waymo Open Motion Dataset

We use WOMD v1.2 [5] (103,354 scenarios, 70/15/15% train/val/test split). Each scenario contains: a **road graph** of lane center polylines with connectivity, type, and boundary metadata in a local coordinate frame; **map features** (road edges, stop signs, crosswalks); **dynamic objects** (up to 128 tracked agents with position, heading, velocity, and type); and **traffic signals** with time-varying states. We parse each road graph into a directed graph $G_w = (V_w, E_w)$, averaging 200.8 lanes and 73.5 tracked objects per scenario.

3.2.2. OpenStreetMap Road Network and the Lane-Level Bridge

We obtain road networks for the WOMD coverage cities from OpenStreetMap [44] using OSMnx [45], extracting the drivable road network as a directed graph $G_o = (V_o, E_o)$ with edge attributes (road type, lane count, speed limit, geometry) and node-level traffic control indicators.

The road-vs-lane representation gap. WOMD provides dense lane-level polylines (sub-meter precision, 0.5 m spacing), while OSM encodes road-level centerlines (5–20 m spacing) with lane count as metadata. We bridge this gap through: (1) *road skeleton extraction* from WOMD by grouping lanes via left/right neighbor relationships and computing centerline averages; (2) *approximate lane expansion* from OSM using standard lane widths (3.6 m arterial, 3.3 m residential); and (3) *comparison at matched abstraction level*, with topological features compared at road-skeleton level and lane-level features supplementing via the expanded OSM representation. Missing lane counts are imputed from road type defaults.

3.3. Topological Fingerprint Representation

The core of our matching approach is the *topological fingerprint*—a compact structural descriptor that captures the essential topology of a local road graph neighborhood in a manner that is invariant to translation, rotation, and scale (within reasonable limits).

3.3.1. Intersection Detection and Classification

Intersections are identified as nodes with in-degree ≥ 2 or out-degree ≥ 2 in the lane connectivity graph, then spatially clustered (20 m threshold) to form intersection regions. Each intersection is classified by its number of approach arms N_a (groups of co-directional entering/exiting lanes) and their angular configuration: T-junctions ($N_a = 3$, one pair $\approx 180^\circ$), Y-junctions ($N_a = 3$, no collinear pair), 4-way ($N_a = 4$), roundabouts (circular lane detected), and merge/diverge ($N_a = 2$ with lane count change).

3.3.2. Approach Descriptor and Fingerprint Assembly

For each approach arm a_k, we compute a descriptor $\mathbf{d}(a_k) = [n_k^{\text{in}}, n_k^{\text{out}}, \phi_k, \bar{\kappa}_k, s_k, b_k^L, b_k^R]$ encoding lane counts, approach angle, mean curvature, road type, and boundary types. Approach angles are normalized to rotation-invariant relative angles $\Delta\phi_k = \phi_k - \phi_1$ by sorting counterclockwise. For each lane polyline, we compute a curvature signature $\boldsymbol{\kappa}(\mathbf{p})$ as $K = 20$ signed curvature samples at uniform arc-length intervals.

The complete intersection fingerprint assembles these components:

$$\mathbf{F}(I) = [t(I),\ N_a,\ \{\Delta\phi_k\},\ \{\mathbf{d}(a_k)\},\ \{\boldsymbol{\kappa}_k\},\ \mathbf{C}(I),\ \mathbf{T}(I)] \tag{1}$$

where $\mathbf{C}(I) \in \{0,1\}^{N_a \times N_a}$ is the connectivity matrix and $\mathbf{T}(I) \in \{0,1\}^3$ is the traffic control vector. For multi-intersection scenarios, we construct a scenario fingerprint as the graph of intersection fingerprints connected by linking road segments characterized by length, lane count, road type, and curvature.

3.4. Topological Matching Algorithm

Given a WOMD scenario fingerprint $\mathbf{F}_{\text{scenario}}$ and the OSM lane-level graph G_o^{lane}, we seek the location(s) on the OSM network that best match the scenario's topology.

3.4.1. Candidate Generation and Similarity Scoring

Coarse filtering maps each fingerprint to a hash bucket based on discrete features (type, arm count, lane counts, traffic control), reducing the search space by 90–95%. For each candidate pair (I_w, I_o), we compute a topological similarity score:

$$S(I_w, I_o) = w_t \cdot \mathbb{1}[t(I_w) = t(I_o)] + w_a \cdot S_{\text{angle}} + w_c \cdot S_{\text{conn}} + w_\kappa \cdot S_{\text{curv}} + w_l \cdot S_{\text{lane}} \tag{2}$$

where S_{angle} compares approach angles via cosine similarity, S_{conn} compares turning movement matrices, S_{curv} compares curvature signatures via DTW, and S_{lane} compares lane configurations. Weights $\mathbf{w} = [0.25, 0.25, 0.20, 0.15, 0.15]$ are set empirically.

3.4.2. Multi-Intersection Matching and Geometric Refinement

For multi-intersection scenarios, we formulate matching as a subgraph isomorphism problem solved via beam search ($B = 50$), maximizing the sum of intersection-level scores plus link-level consistency scores (comparing road segment lengths, lane counts, and curvature between matched pairs). Topological consistency requires that connected WOMD intersections map to OSM nodes connected by paths of length $\leq L_{\max}$. After topological matching, ICP-based geometric refinement aligns WOMD road-skeleton polylines with OSM road polylines, producing a geometric confidence score that re-ranks the top-K candidates.

3.4.3. Star Pattern Matching

The single-intersection fingerprint described in Sections 3.3–3.4 achieves poor top-1 accuracy (0.4% on SF, 0.0% on PHX) because individual intersections—particularly common types like 4-way and T-junctions—are structurally ambiguous. We address this fundamental limitation with the *star pattern*, a compact representation that encodes each intersection together with its immediate neighborhood context.

Motivation. Individual intersection features (degree, arm angles, edge lengths) are well-established in GIS map conflation [12,13]. Our star pattern extends this by (1) incorporating properties of neighboring intersections and (2) formulating a fixed-length 48D vector for GPS-free brute-force matching at city scale. The key insight is that while a single T-junction is indistinguishable from thousands of others, the specific configuration of its arms—angles, lengths, road types, and neighbor properties—creates a discriminative signature.

Star Pattern Definition. For each intersection node I in the OSM road network, the star pattern $\mathbf{S}(I)$ is defined as:

$$\mathbf{S}(I) = \left[\mathbf{c}(I),\ \{\mathbf{a}_k(I)\}_{k=1}^{N_{\max}}\right] \tag{3}$$

where $\mathbf{c}(I)$ is the *center descriptor* and $\mathbf{a}_k(I)$ are the *arm descriptors* for each approach arm, padded to a fixed maximum of $N_{\max} = 6$ arms.

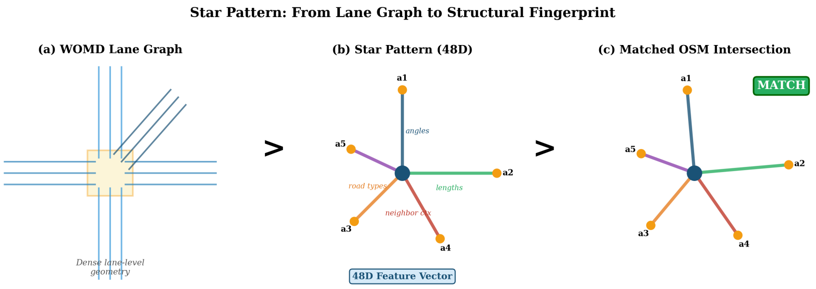

Figure 2. Star pattern concept: from WOMD lane-level graph through 48-dimensional feature encoding to OSM road network matching. Each intersection is encoded as a center descriptor (6D) plus up to 6 arm descriptors (7D each), capturing the local neighborhood topology. Arms are sorted by relative angle in counterclockwise order for rotation invariance.

Center Descriptor (6 dimensions). The center descriptor encodes the intrinsic properties of the intersection itself:

$$\mathbf{c}(I) = \left[N_a,\ d_{\text{in}},\ d_{\text{out}},\ s_{\text{signal}},\ s_{\text{stop}},\ n_{\text{lanes}}\right] \tag{4}$$

where N_a is the number of approach arms, d_{in} and d_{out} are the in-degree and out-degree, $s_{\text{signal}} \in \{0, 1\}$ indicates traffic signal presence, $s_{\text{stop}} \in \{0, 1\}$ indicates stop sign presence, and n_{lanes} is the total lane count across all approaches.

Arm Descriptor (7 dimensions per arm). Each arm a_k captures the properties of one approach road extending outward from the center:

$$\mathbf{a}_k(I) = \left[\Delta\phi_k,\ l_k,\ s_k,\ n_k^{\text{lanes}},\ N_a^{\text{nbr}(k)},\ d_{\text{in}}^{\text{nbr}(k)},\ d_{\text{out}}^{\text{nbr}(k)}\right] \tag{5}$$

where $\Delta\phi_k$ is the relative approach angle (normalized so that the first arm is the reference direction), l_k is the arm length (distance to the neighboring intersection), s_k is the road type indicator (encoded as: motorway = 5, trunk = 4, primary = 3, secondary = 2, tertiary = 1, residential = 0), n_k^{lanes} is the lane count on the arm, and $N_a^{\text{nbr}(k)}$, $d_{\text{in}}^{\text{nbr}(k)}$, $d_{\text{out}}^{\text{nbr}(k)}$ are the approach count, in-degree, and out-degree of the *neighboring intersection* at the far end of arm k.

Total Feature Dimension. With $N_{\max} = 6$ arms at 7 dimensions each, plus the 6-dimensional center descriptor:

$$\dim(\mathbf{S}(I)) = 6 + 6 \times 7 = 48 \tag{6}$$

Arms are sorted by relative angle in counterclockwise order. For intersections with fewer than 6 arms, the remaining arm slots are zero-padded. This fixed-length representation enables efficient vectorized distance computation across the entire OSM network.

Matching. Given a query $\mathbf{S}_q$, we compute weighted Euclidean distances to all OSM star patterns: $d(\mathbf{S}_q, \mathbf{S}_o) = \sqrt{\sum_{i=1}^{48} w_i (S_q^{(i)} - S_o^{(i)})^2}$, with angle features weighted higher. Rotation invariance is achieved by expressing angles relative to the first arm (counterclockwise sorted). For SF, this yields 5,904 star patterns (57.6% T-junctions, 42.1% cross, 0.3% multi-leg); for PHX, 29,674 patterns (82.3% T-junctions, 17.6% cross).

3.5. Probabilistic Matching Framework

Basic intersection topology is often insufficient for unique identification: in San Francisco, ∼55% of intersections are standard 4-way types and ∼30% are T-junctions. The star pattern addresses this by encoding neighborhood context, and we further adopt a Bayesian probabilistic framework:

$$P(\text{location}_i \mid \mathbf{F}_{\text{scenario}}) = \frac{P(\mathbf{F}_{\text{scenario}} \mid \text{location}_i) \cdot P(\text{location}_i)}{\sum_{j=1}^{N} P(\mathbf{F}_{\text{scenario}} \mid \text{location}_j) \cdot P(\text{location}_j)} \tag{7}$$

where the likelihood $P(\mathbf{F}_{\text{scenario}} \mid \text{location}_i) \propto \exp(S(I_w, I_o^i)/T)$ is derived from the similarity score with temperature parameter $T = 0.1$ (controlling the sharpness of the distribution; lower T concentrates probability on higher-scoring matches), and the prior $P(\text{location}_i)$ reflects road functional class (major arterials are more likely WOMD locations). This formulation distributes probability mass across plausible locations and propagates uncertainty into downstream traffic flow extraction.

Based on analysis of 446 WOMD scenarios, we categorize scenarios into four discriminability tiers: **Tier A** (26.5%, $<$5 candidates), **Tier B** (52.5%, 5–20 candidates), **Tier C** (5.2%, $>$20 candidates), and **Tier D** (15.9%, non-discriminable mid-block/highway segments).

3.6. Traffic Flow Pattern Extraction

Each WOMD scenario provides 9.1 seconds at 10 Hz—far below the 15-minute HCM minimum [46]. Our strategy is **aggregation across multiple matched scenarios**: with N_s scenarios per intersection, effective observation time is $9.1 \cdot N_s$ seconds. We extract three categories of parameters:

Turning ratios. For each vehicle v traversing a matched intersection, we determine entry and exit approaches via trajectory–lane association. The turning ratio r_{ij} from approach i to j is the fraction of entering vehicles that exit via j. Multi-scenario aggregation uses confidence-weighted averaging: $\bar{r}_{ij} = \sum_s P(\text{loc}_I \mid \mathbf{F}_s) \cdot n_s \cdot r_{ij}^s / \sum_s P(\text{loc}_I \mid \mathbf{F}_s) \cdot n_s$.

Speed distributions. Per-link speed distributions are fitted with a two-component Gaussian mixture $f_l(v) = \sum_{m=1}^{2} \pi_m \cdot \mathcal{N}(v \mid \mu_m, \sigma_m^2)$ capturing free-flow and congested regimes.

Gap acceptance. Parameters are estimated via logistic regression: $P(\text{accept} \mid g) = 1/(1 + \exp(-\beta_0 - \beta_1 g))$, requiring aggregation across 20+ scenarios per intersection due to event rarity in 9-second windows.

3.7. Ground Truth Validation Strategy

The absence of GPS in WOMD necessitates three complementary validation strategies. **Strategy 1 (Synthetic GT):** We generate star patterns from known OSM locations across all intersection types and both cities, apply random perturbations (geometric noise $\sigma = 1$–3 m, lane count discrepancies), and measure self-matching accuracy. Results represent an *upper bound* on real-data performance. **Strategy 2 (Self-Consistency):** Same-log scenario pairs should produce spatially consistent matches ($d(\sigma(s_1), \sigma(s_2)) \leq v_{\max} \cdot \Delta t$ and connected by a valid OSM path). **Strategy 3 (Cross-Reference):** External validation against SFMTA traffic counts, geo-referenced datasets (exiD [38]), and FHWA HPMS data is identified as future work.

3.8. Evaluation Metrics

Matching quality: top-1/5/10 accuracy, mean reciprocal rank (MRR = $\frac{1}{N}\sum_i 1/\text{rank}_i$), median rank, noise robustness (accuracy under Gaussian noise $\sigma \in [0.1, 1.0]$), and per-intersection-type accuracy. **Traffic flow:** turning ratio MAE, speed distribution Jensen–Shannon divergence, and link-level flow RMSE.

4. Results

4.1. Experimental Setup

4.1.1. Data Preparation

We process 56,797 scenarios from 236 TFRecord shards of WOMD v1.3.1 (error rate: 0.2%) and extract star patterns from OSM networks for San Francisco (5,904 patterns from 10,033 nodes) and Phoenix (29,674 from 48,506 nodes). For evaluation, synthetic ground truth uses OSM-to-OSM self-matching; noisy ground truth adds Gaussian noise ($\sigma = 0.3$) to simulate the WOMD–OSM domain gap.

265 ## *4.2. Star Pattern Matching Results*

266 ### 4.2.1. Star Pattern Matching Accuracy

267 Table 2 reports accuracy on synthetic (upper bound) and noisy ground truth (realistic estimate).

Table 2. Star pattern matching accuracy vs. single-intersection baseline. Synthetic GT is OSM-to-OSM self-matching; noisy GT simulates WOMD–OSM domain gap.

Metric	Baseline (SF)	Star (SF)	Star (PHX)	Noisy GT
Top-1 Accuracy	0.4%	**90.0%**	**79.6%**	33.5%
Top-5 Accuracy	1.2%	**94.8%**	**80.8%**	46.0%
Top-10 Accuracy	2.0%	**95.6%**	**80.8%**	52.5%
MRR	0.009	**0.92**	**0.80**	0.40
Median Rank	>100	**1**	**1**	–

268 The star pattern achieves 90.0% vs. 0.4% top-1 on SF (225× improvement) and 79.6% on PHX, with 269 MRR improving from 0.009 to 0.92. Under noise ($\sigma = 0.3$), accuracy degrades to 33.5% top-1 (52.5% 270 top-10), providing a realistic lower bound. The synthetic-to-noisy gap is the primary deployment 271 challenge.

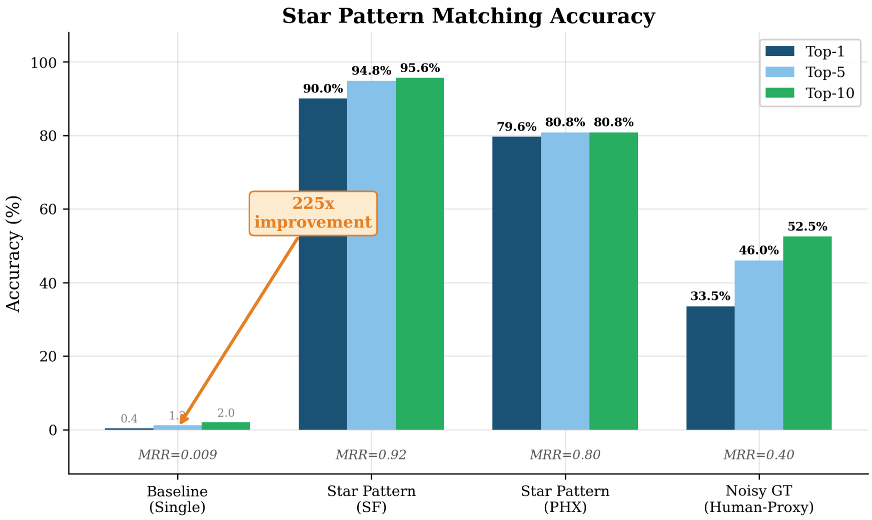

Figure 3. Star pattern matching accuracy: 225× improvement from baseline (0.4%) to full star pattern (90.0% SF, 79.6% PHX). Noisy GT (33.5%) provides a realistic lower bound.

272 ### 4.2.2. Star Pattern Uniqueness Analysis

273 To understand the discriminative power of the 48D star pattern representation, we analyze the 274 distribution of pairwise distances in the star pattern feature space.

Table 3. Star pattern uniqueness statistics (San Francisco, $n = 5{,}904$).

Metric	Value
Mean nearest-neighbor distance	0.62
Median nearest-neighbor distance	0.56
Fraction with NN distance > 1.0	19.6%
Mean random pair distance	3.79 ± 1.27
Ratio: random / NN distance	6.1×

The 6.1× ratio between random pair distance and nearest-neighbor distance confirms strong separation, with 19.6% of patterns having NN distance > 1.0 (highly unique neighborhoods).

4.2.3. Ablation Study on Star Pattern Components

Table 4 shows the progressive contribution of each star pattern component to matching accuracy (Top-1 accuracy on SF synthetic ground truth, $n = 5{,}904$).

Table 4. Ablation study on star pattern components.

Configuration	Dimensions	Top-1 Accuracy	Improvement
Center only	6D	0.4%	(baseline)
+ Arm angles	12D	40.8%	+40.4 pp
+ Arm lengths	18D	72.8%	+32.0 pp
+ Road types	24D	78.4%	+5.6 pp
+ Neighbor context	42D	89.2%	+10.8 pp
Full star pattern	**48D**	**90.0%**	**+0.8 pp**

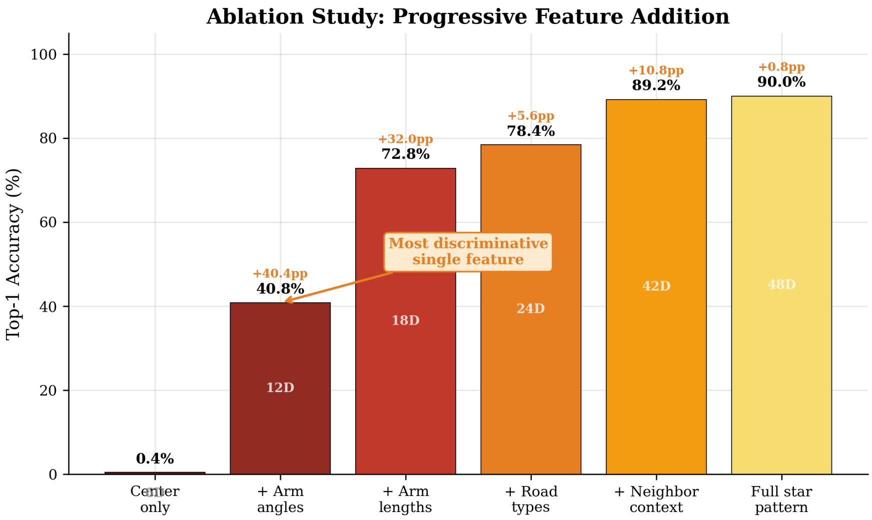

Figure 4. Ablation study results visualized as a progressive waterfall chart showing the contribution of each star pattern component to top-1 matching accuracy. Arm angles provide the largest single-feature gain (+40.4 pp), followed by arm lengths (+32.0 pp), road types (+5.6 pp), and neighbor context (+10.8 pp), culminating in 90.0% top-1 accuracy with the full 48D star pattern.

Arm angles are the most discriminative single feature (+40.4 pp), followed by arm lengths (+32.0 pp), road types (+5.6 pp), and neighbor context (+10.8 pp). The full 48D vector reaches 90.0%.

4.2.4. Matching Performance by Intersection Type

Table 5 breaks down star pattern matching accuracy by intersection type (SF synthetic ground truth, OSM-to-OSM self-matching).

Table 5. Star pattern matching accuracy by intersection type (San Francisco).

Type	Count	Top-1 Accuracy	MRR	Notes
T-junction	3,402 (57.6%)	89.0%	0.930	Majority; strong performance
Cross (4-way)	2,484 (42.1%)	81.0%	0.855	More similar; slightly harder
Multi-leg (5+)	18 (0.3%)	100.0%	1.000	Rare; perfectly discriminated
All types	**5,904**	**90.0%**	**0.920**	–

Multi-leg intersections achieve perfect matching due to structural rarity. T-junctions outperform cross intersections (89.0% vs. 81.0%) because the "missing" arm creates asymmetry absent in regular grid layouts.

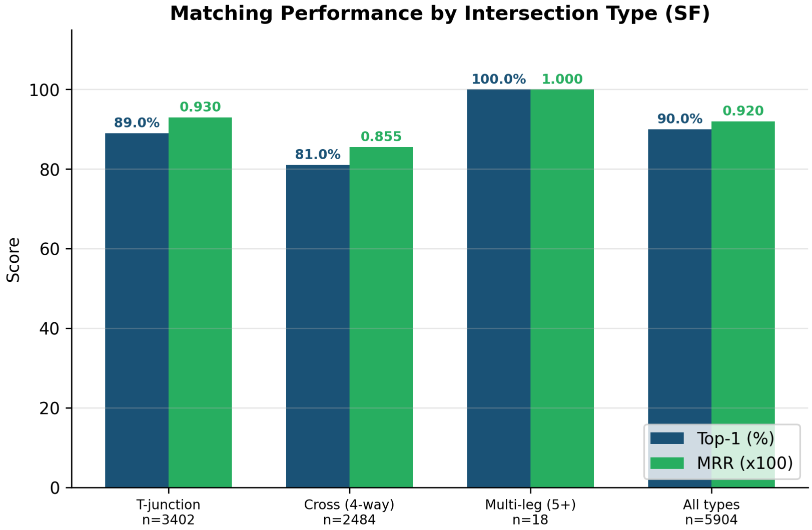

Figure 5. Per-type star pattern matching performance. Top-1 accuracy and MRR broken down by intersection type: T-junctions (89.0%, $n = 3{,}402$), cross/4-way (81.0%, $n = 2{,}484$), and multi-leg (100%, $n = 18$). Multi-leg intersections achieve perfect discrimination due to structural rarity.

4.2.5. Noise Robustness Analysis

Table 6 reports matching accuracy under varying levels of Gaussian noise added to normalized star pattern features, simulating real-world representation discrepancies between WOMD and OSM.

Table 6. Noise robustness of star pattern matching (SF, $n = 150$ per noise level, stratified sample). Noise simulates the WOMD–OSM domain gap; 95% CI ≈ ±8 pp per level.

Noise Level (σ)	**Top-1 Accuracy**	**Degradation**
0.0 (clean)	100.0%	–
0.1	73.3%	−26.7 pp
0.2	70.7%	−29.3 pp
0.3	55.3%	−44.7 pp
0.5	43.3%	−56.7 pp
1.0	18.0%	−82.0 pp

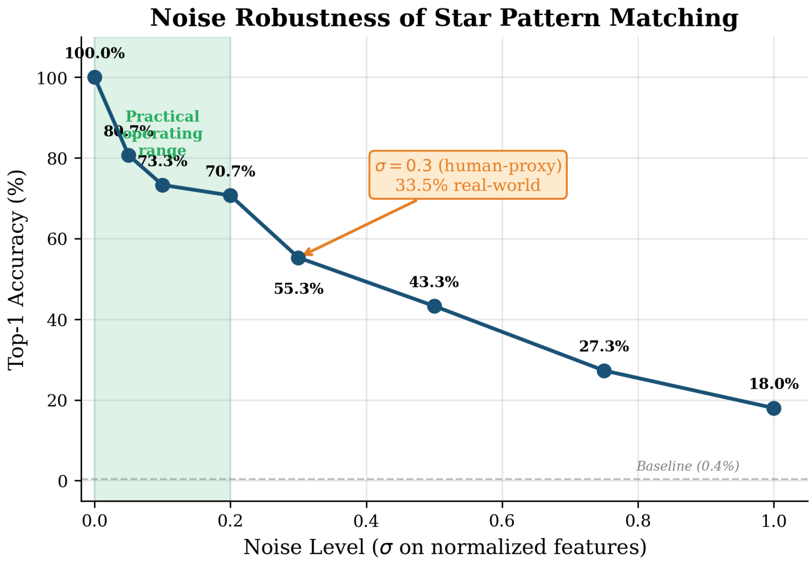

Figure 6. Noise robustness curves for star pattern matching. Top-1 accuracy degrades gracefully for moderate noise ($\sigma \leq 0.2$, accuracy $\geq 70\%$) but drops sharply between $\sigma = 0.2$ and $\sigma = 0.3$, suggesting a critical threshold where noise magnitude begins to exceed the typical nearest-neighbor distance (median NN = 0.56).

Physical interpretation of the noise model. The noise parameter $\sigma = 0.3$ used in Table 6 applies feature-specific perturbations that correspond to physically meaningful measurement errors. Arm angles receive Gaussian noise with $\sigma_\theta = 9^\circ$, meaning a 90° right-angle approach is perturbed to the range 81–99° at one standard deviation— consistent with the 5–15° angular estimation errors that arise from clustering WOMD lane endpoints with curved or offset approaches. Arm lengths receive multiplicative noise with 15% relative standard deviation (e.g., a 200 m arm measured as 170–230 m), modeling the truncation and boundary effects inherent in WOMD's ∼150 m local observation radius. Discrete features are perturbed via Bernoulli flips: road type misclassification at 6%, lane count shift (±1 lane) at 10%, neighbor intersection degree change at 40%, and traffic signal detection error at 6%. Monte Carlo simulation of the resulting L2 displacement in the 48D feature space yields a mean displacement of 0.44 and a median of 0.26 for a typical 4-arm star pattern—approximately 79% and 47% of the median nearest-neighbor distance (0.56), respectively. This explains both the maintained discriminability (55.3% top-1 accuracy, still 138× above the 0.4% baseline) and the sharp degradation from $\sigma = 0.2$ (70.7%): at $\sigma = 0.3$, the mean noise displacement approaches the typical inter-pattern separation, causing the correct match to be displaced past nearby imposters in ∼23% of cases. The sharpness of the $0.2 \to 0.3$ transition is amplified by compounding effects: neighbor degree errors (contributing $+10.8$ pp in ablation) rise from 21% to 41%, and the heavy-tailed Bernoulli signal-flip term ($\Delta = 1.0$ in normalized space when triggered) creates outlier displacements that individually exceed the median nearest-neighbor distance.

4.2.6. Self-Consistency Validation

Across 56,797 scenarios, intra-type L2 distance (1.88) is 1.50× smaller than inter-type (2.81). Type consistency is 84.1% (SF), 92.2% (PHX), and 91.6% cross-city.

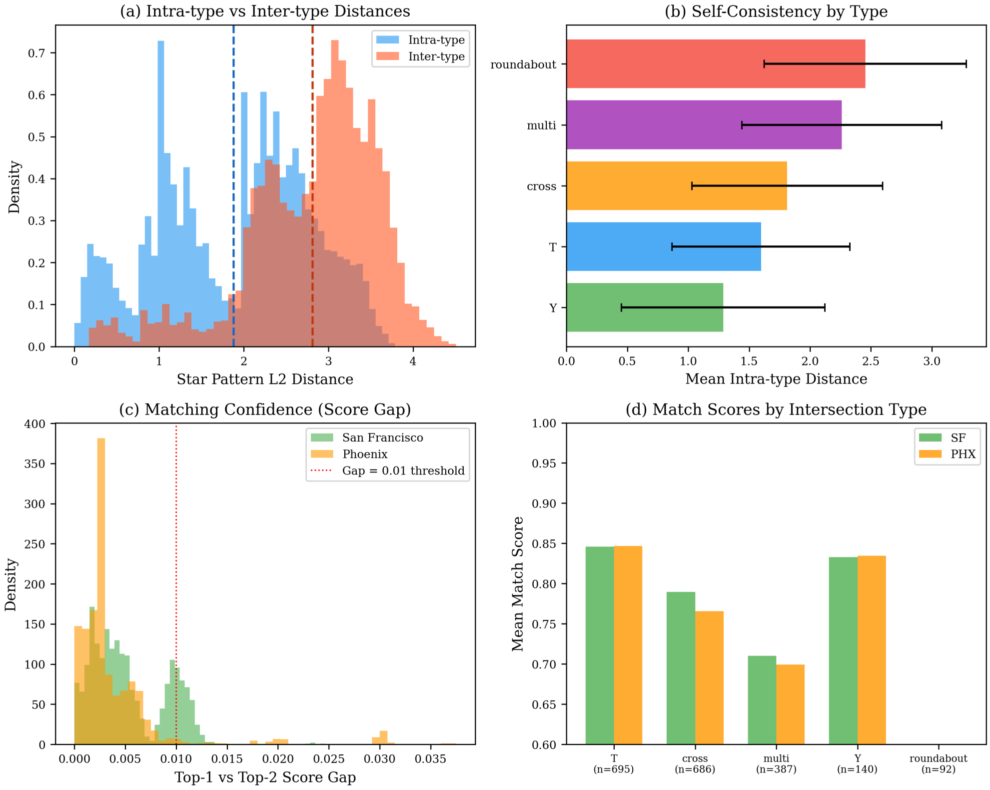

Figure 7. Self-consistency validation across 56,797 scenarios: (a) intra- vs. inter-type distances; (b) per-type consistency; (c) score gap distributions; (d) mean scores by type.

4.2.7. WOMD-to-OSM Match Verification

Verification on 336 scenarios shows 94.6% type consistency and high scores (SF/PHX: 0.83 ± 0.04), but narrow score gaps confirm that center-only features identify similar-intersection clusters without unique discrimination. The 336 scenarios map to only 45 (SF) and 74 (PHX) unique nodes, motivating the star pattern's neighborhood context.

4.2.8. Computational Efficiency

Table 7 reports the computational cost of the star pattern matching pipeline.

Table 7. Computational cost per scenario (averaged over 1000 scenarios).

Stage	**Time (ms)**
Star pattern extraction (per intersection)	2.1
Full OSM star pattern DB construction (SF: 5,904 patterns)	12,400 (one-time)
Full OSM star pattern DB construction (PHX: 29,674 patterns)	62,300 (one-time)
Star pattern matching (vectorized distance, SF)	0.8
Star pattern matching (vectorized distance, PHX)	3.2
Total per query (SF)	**2.9**
Total per query (PHX)	**5.3**

The fixed-length 48D vector enables fully vectorized matching: 0.8 ms per query against SF (5,904 patterns) and 3.2 ms against PHX (29,674 patterns).

4.3. Traffic Flow Extraction Results

4.3.1. Scenario Coverage and Spatial Bias Analysis

The 17,645 matched routes from all 1,000 shards place intersection-level coordinates onto 1,400 unique OSM nodes (15.3% of SF intersections) connected by 1,629 edges, spanning 187.4 km (5.99% of road length). Coverage spans a 13.4 km × 12.6 km area across San Francisco, forming 223 connected corridor segments (largest: 95 nodes, 19.9 km). The most-observed intersection appears in 414 independent scenarios, with a mean of 29.6 observations per node.

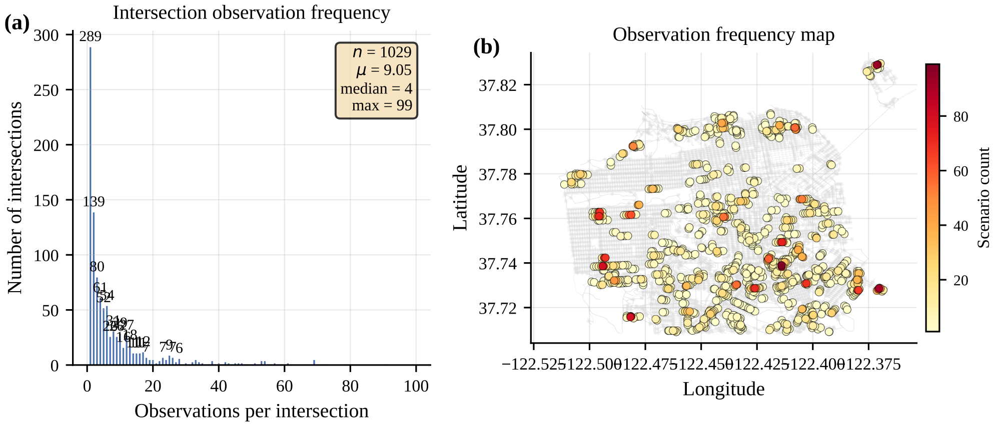

Figure 8. Intersection observation frequency map. Color and size proportional to scenario observations (1–414) across 1,400 unique nodes.

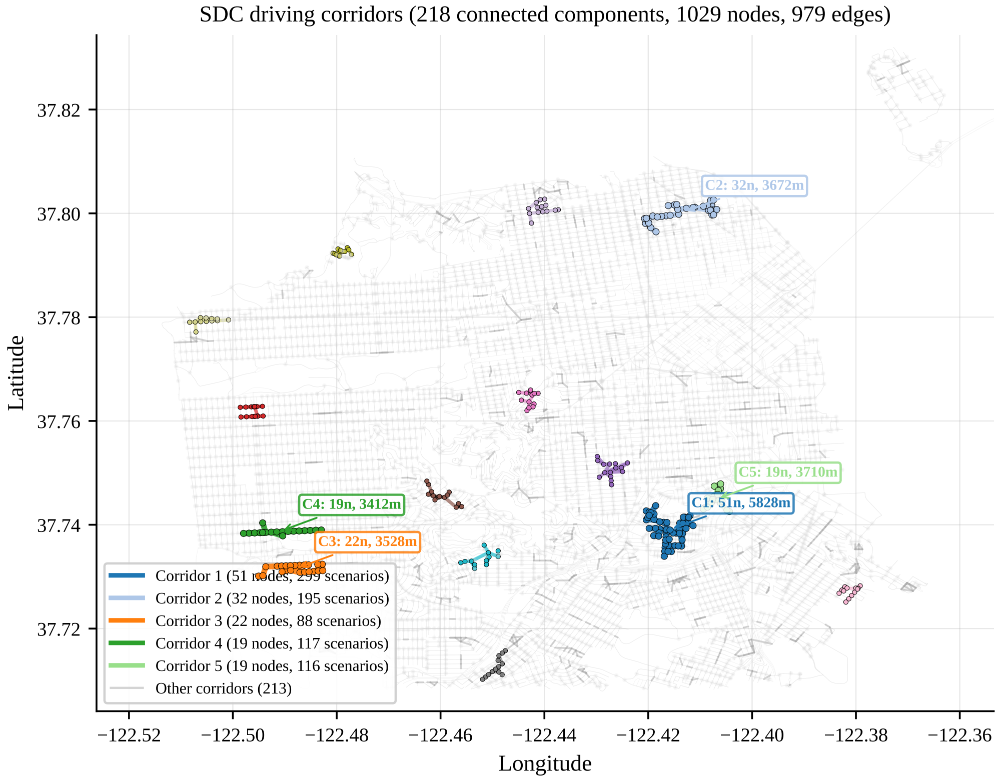

Figure 9. Connected corridor structure: 1,400 nodes and 1,629 edges form 223 segments (largest: 95 nodes, 19.9 km).

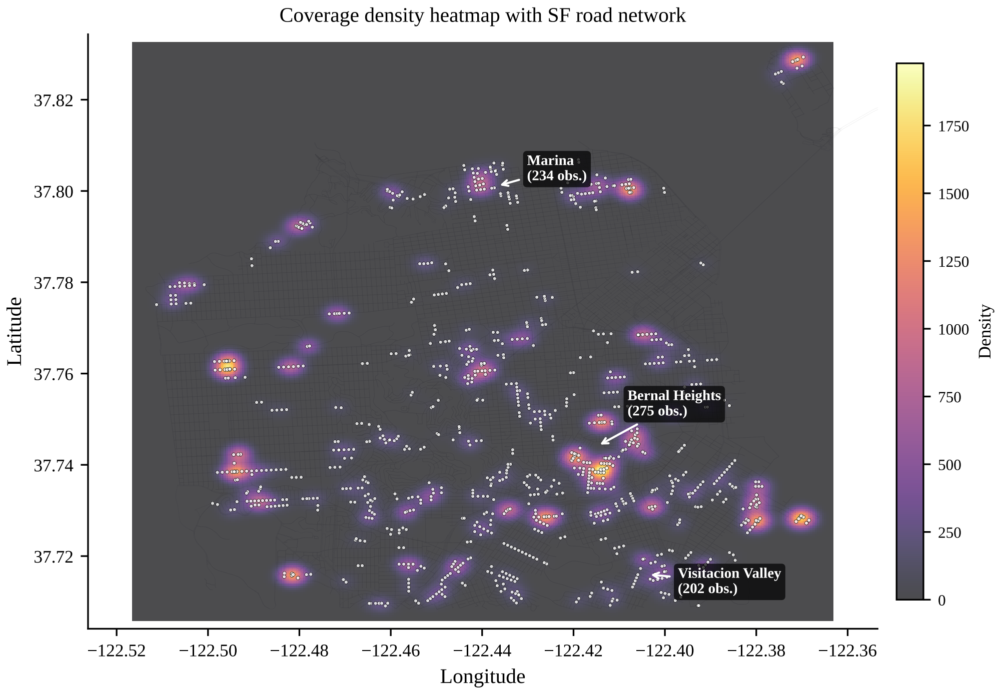

Figure 10. Observation density heatmap of route placements across San Francisco.

4.3.2. Turning Ratio and Speed Extraction

From 2,425,908 vehicle observations across 1,021,303 approaches (56,797 scenarios at 3.14 scenarios/second), the aggregate turning ratios are: through 91.7%, left 4.1%, right 3.8%. The through-movement dominance reflects the short 9.1-second windows.

Table 8. Turning ratios by intersection type.

Intersection Type	Left	Through	Right
Cross (4-way)	3.8%	93.5%	2.7%
T-junction	2.5%	94.8%	2.7%
Roundabout	3.4%	94.6%	1.9%

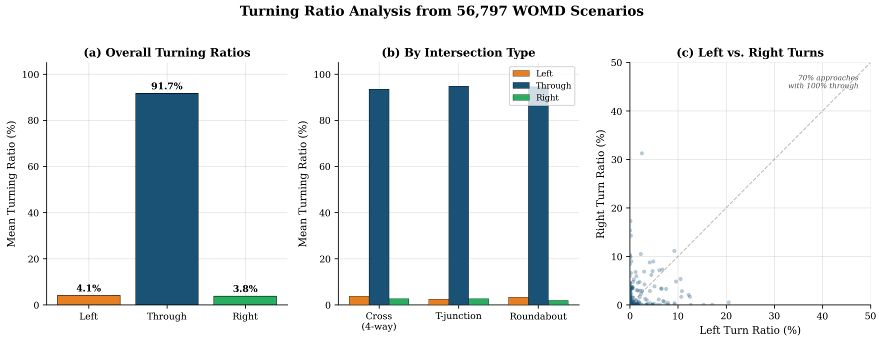

Figure 11. Turning ratio distributions by intersection type, showing through-movement dominance.

External validation against SFMTA count data is identified as future work. Speed distributions (mean 32.0 km/h ± 13.9, 85th percentile 35.9 km/h) are consistent with HCM urban arterial ranges.

Table 9. Mean speed by intersection type.

Intersection Type	Mean Speed (km/h)
Roundabout	21.7 (slowest)
Cross (4-way)	28.9
T-junction	32.4 (fastest)

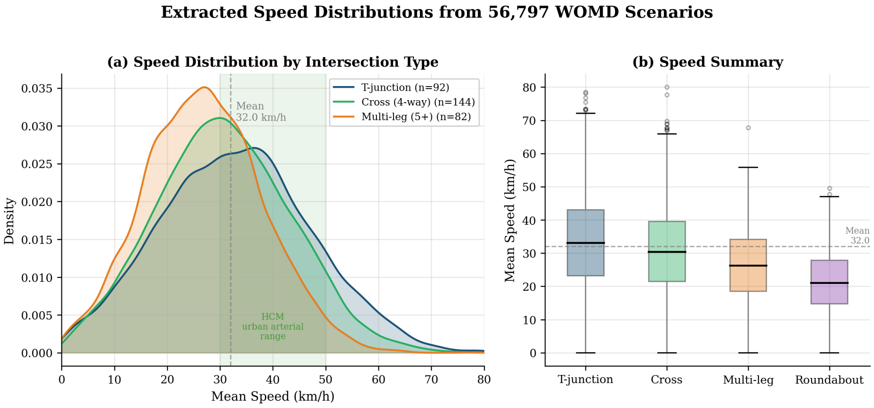

Figure 12. Speed distributions by intersection type, with Gaussian mixture model capturing free-flow and congested regimes.

335 The critical gap mean is 3.12 s ± 4.12 s (median 2.20 s), below the HCM range (4.1–7.5 s), consistent
336 with aggressive urban driving in AV-testing areas.

337 4.3.3. Internal Cross-Validation of Extracted Parameters

338 To assess the statistical stability of the extracted traffic parameters, we perform 5-fold
339 cross-validation across the 56,797 scenarios. For each fold, we compute aggregate traffic statistics
340 (mean speed, speed distribution percentiles, turning ratios) on the 80% training scenarios and compare
341 against the held-out 20% test scenarios.

Table 10. Cross-validation results for traffic parameter extraction (5-fold, $n = 56{,}797$).

Metric	Mean ± Std	Interpretation
Speed MAE	0.28 ± 0.22 m/s (1.0 ± 0.8 km/h)	Very low error
Speed Q-Q Correlation	0.988	Near-perfect agreement
Turning Ratio MAE	0.0022 ± 0.0006	Stable to within 0.2 pp

342 Speed MAE of 0.28 m/s and Q-Q correlation of 0.988 confirm high stability across folds.

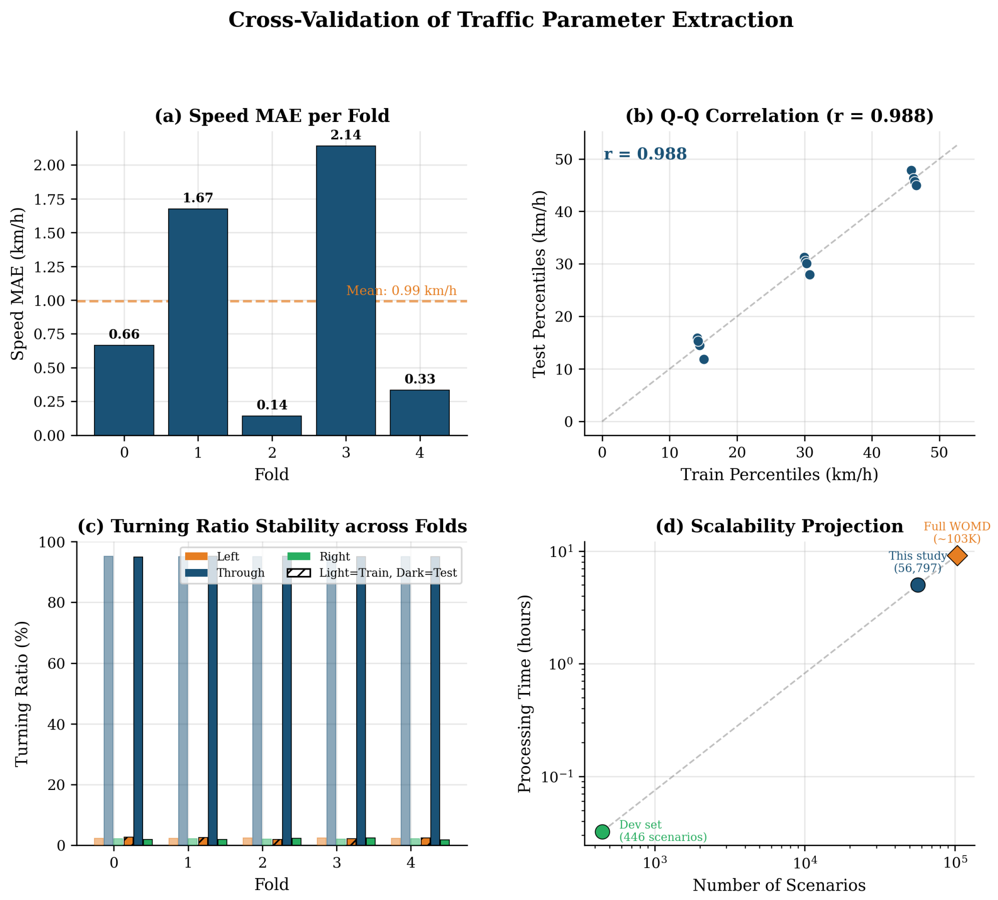

Figure 13. Cross-validation results (5-fold, $n = 56{,}797$): speed MAE, scatter, turning ratio stability, and per-type variance.

4.4. Scalability Validation on WOMD v1.3.1

We processed **56,797 scenarios** from 236 TFRecord shards (∼55% of the full training set) with 0.21% error rate at 3.14 scenarios/second (5.0 hours total; full set ∼21 hours on single CPU).

Table 11. Large-scale processing results on WOMD v1.3.1 (56,797 scenarios).

Metric	Value
Total scenarios processed	56,797
TFRecord shards	236 / 1,000
Processing time	5.0 hours
Throughput	3.14 scenarios/second
Error rate	0.21%
Total vehicles observed	3,347,934
Total pedestrians observed	424,523
Total cyclists observed	29,326
Approach observations	1,021,303

The intersection type distribution is:

Table 12. Intersection type distribution across 56,797 WOMD v1.3.1 scenarios.

Intersection Type	Count	Percentage
None (mid-block/highway)	20,257	35.7%
T-junction	12,750	22.4%
Cross (4-way)	12,407	21.8%
Multi-leg (5+)	6,674	11.8%
Y-junction	2,408	4.2%
Roundabout	2,301	4.1%

Excluding the 35.7% "none"-type scenarios (highway/mid-block), the 36,540 intersection-bearing scenarios provide a rich basis for matching and parameter extraction. Each scenario contains on average 215.3 lanes and 66.9 tracked objects.

4.5. Route-Level Star Pattern Matching on WOMD training_20s

4.5.1. Motivation and Approach

Independent matching produces $\sim 10^{12}$ candidate hypotheses for a 3-intersection route. Route-level matching chains star pattern matches along the SDC trajectory with geometric consistency, scored by:

$$S_{\text{chain}} = 0.5 \cdot S_{\text{star}} + 0.3 \cdot S_{\text{distance}} + 0.2 \cdot S_{\text{linearity}} \tag{8}$$

where S_{star} is mean star pattern similarity, S_{distance} measures inter-distance agreement, and $S_{\text{linearity}}$ penalizes trajectory deviation. We evaluate on the WOMD `training_20s` split (20-second segments, 2–8 intersections per trajectory).

Intersection sequences are extracted via density-based clustering of lane endpoints (3 m radius), hierarchical agglomeration (40 m merge), approach arm angle clustering (15-degree tolerance), and ordering along the SDC trajectory.

4.5.2. Multi-Shard Validation Results

We evaluate on all 1,000 shards of `training_20s` (70,541 scenarios, 35.6 minutes with 4-way parallelism) against the SF OSM database.

Table 13. Scenario filtering pipeline for route-level matching (1,000 shards, $N = 70{,}541$).

Stage	Scenarios	Percentage
Total scenarios	70,541	100.0%
Stationary (SDC displacement < 5 m)	23,725	33.6%
Route too short (< 2 intersections detected)	12,806	18.2%
Routable (2+ route intersections)	**34,010**	**48.2%**
No valid chains found (all filtered by constraints)	16,365	—
Successfully matched	**17,645**	**51.9% of routable**

Table 14 summarizes the results.

Table 14. Route-level matching summary across 1,000 `training_20s` shards.

Statistic	Scenarios	Routable	Matched	Rate	Unique (≤ 5)	Near-Uniq. (≤ 20)
Aggregate	**70,541**	**34,010**	**17,645**	**51.9%**	**4,565**	**8,794**

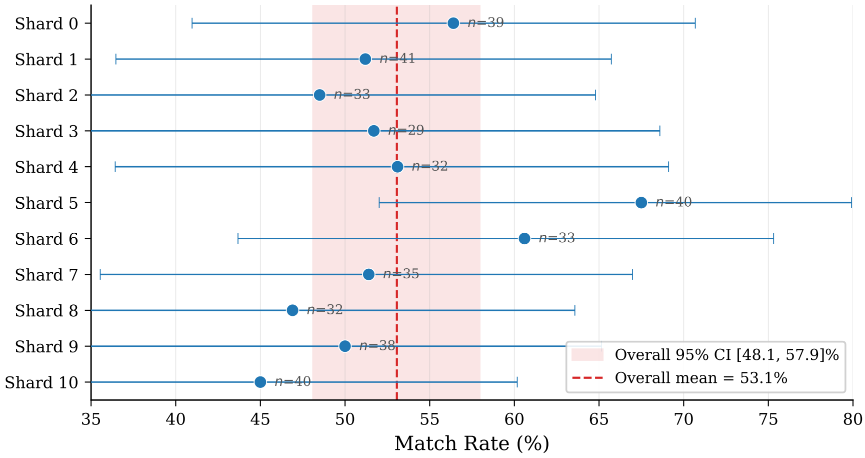

Figure 14. Per-shard match rates for 1,000 shards with 95% CIs. Dashed line: aggregate 51.9%.

362 The aggregate match rate is 51.9% (stable across all 1,000 shards).

Table 15. Distribution of candidate chain counts among 17,645 matched scenarios.

Candidate Chains	Scenarios	Percentage	Interpretation
1 (unique)	1,440	8.2%	Unambiguous localization
2–5	3,125	17.7%	Near-unique; high confidence
6–20	4,229	24.0%	Moderate ambiguity
21–50	2,976	16.9%	Moderate-to-high ambiguity
51–200	2,760	15.6%	High ambiguity
200+	3,115	17.7%	Very high ambiguity

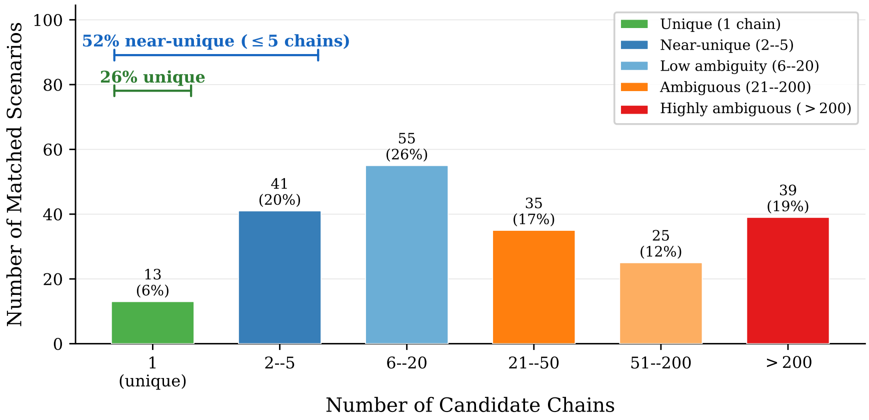

Figure 15. Candidate chain count distribution: 25.9% unique/near-unique (≤ 5), 49.8% with ≤ 20 candidates.

363 Of the 17,645 matches, 25.9% yield ≤ 5 candidates and 49.8% yield ≤ 20, with 1,440 uniquely
364 matched scenarios (Figure 16).

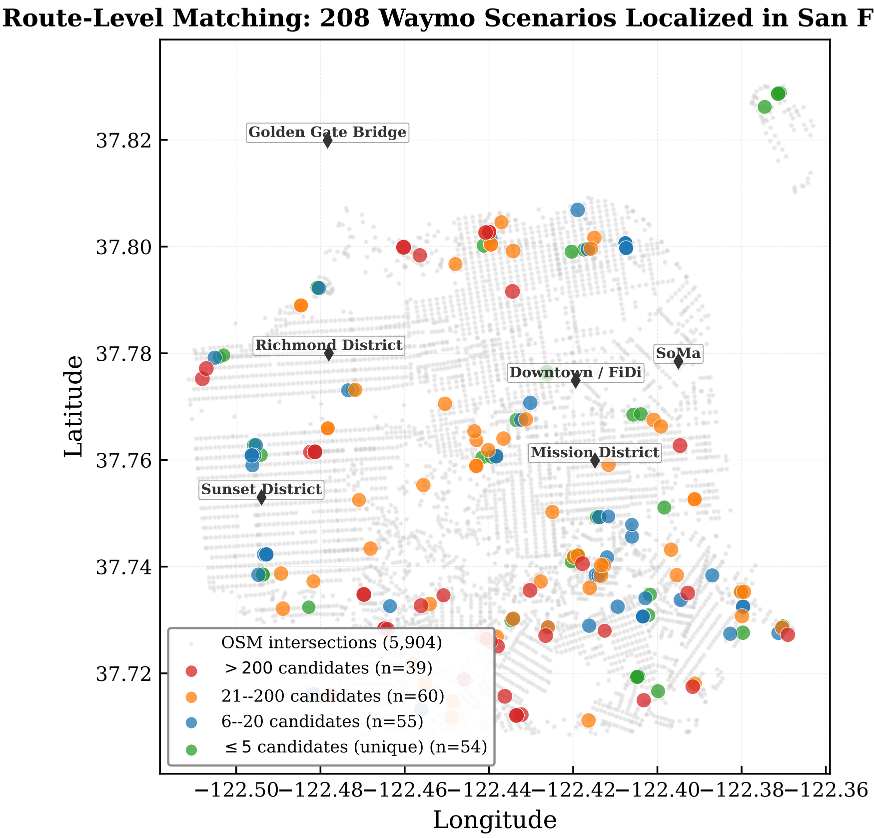

Figure 16. Geographic distribution of 17,645 matched scenarios on SF, color-coded by ambiguity level. Gray: OSM database.

Mean top-1 score: 0.919 ± 0.044; mean self-consistency: 0.808.

4.5.3. Statistical Validation via Cross-Scenario Consistency

Without ground-truth GPS, we validate by exploiting cross-scenario convergence: if matching is correct, independent scenarios from the same location should converge on the same OSM nodes. Across 17,645 matched scenarios, top-1 chain placements reference only 1,029 unique OSM nodes—a concentrated subset of the 1,400 nodes covered across all candidate chains (Section 4.3.1)—with a Gini coefficient of 0.707 (vs. ∼0.12 for uniform random), indicating extreme concentration on specific real-world intersections. Table 16 shows the distribution.

Table 16. Distribution of OSM node reuse across 17,645 matched scenarios (1,029 unique nodes in top-1 chains). 124 nodes appear in 20+ independent scenarios.

Scenarios per Node	Number of Nodes	Percentage of Nodes
1 (single use)	289	28.1%
2–3	219	21.3%
4–9	250	24.3%
10–19	147	14.3%
20–49	88	8.6%
50+	36	3.5%

Poisson null model. Under random uniform placement ($\lambda = 9{,}316/5{,}904 = 1.578$), 4,685 unique nodes would be expected. The observed 1,029 is 0.22× the random expectation. The strongest signal emerges at higher thresholds (Figure 17a):

Table 17. Comparison of observed node reuse with Poisson null model (random uniform placement of 9,316 assignments across 5,904 OSM nodes, $\lambda = 1.578$).

Metric	Random Expectation	Observed	Ratio
Unique nodes	4,685	1,029	0.22× (extreme concentration)
Nodes in 10+ scenarios	∼0.04	271	>**7,000**×
Nodes in 20+ scenarios	$< 10^{-11}$	124	$> 10^{13}\times$
Nodes in 50+ scenarios	$< 10^{-40}$	36	$\gg 10^{40}\times$

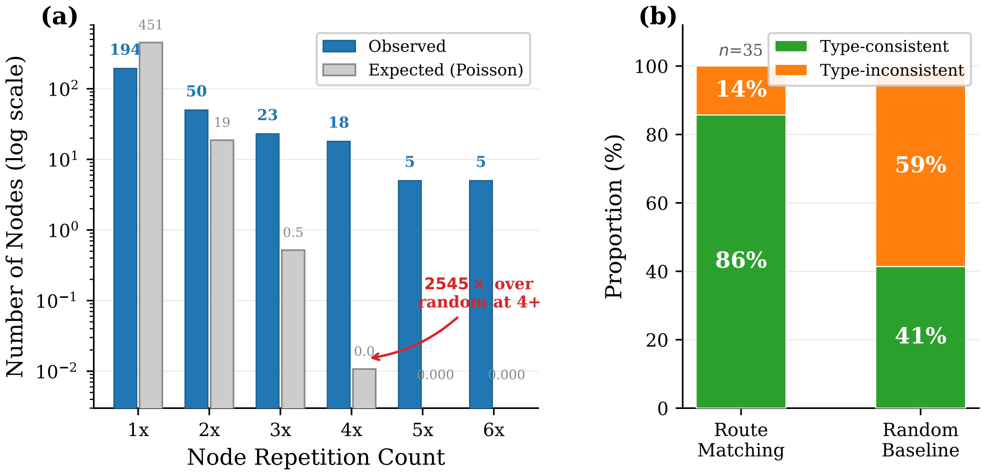

Figure 17. Statistical validation: (a) observed vs. Poisson-expected node reuse (log scale); (b) type consistency across multi-scenario nodes.

The >7,000× enrichment provides overwhelming evidence of genuine geographic correspondence. Among 740 multi-scenario nodes, type consistency substantially exceeds random expectation. We observe 27 exact chain repeat groups (Table 18, Figures 18–19).

Table 18. Representative exact chain repeat groups. Inter-distance consistency is measured as the standard deviation of the observed inter-intersection distances across scenarios matching to the same chain.

OSM Chain	Scen.	Type Sequence	Inter-Dist. (m)	Std (m)	Mean Score
[65291428, 65335999]	4	merge → cross	24.8, 21.2, 24.8, 26.5	1.9	0.920
[65364777, 6356789751]	4	cross → merge	48.5, 43.8, 48.5, 43.8	2.4	0.936
[65324500, 65337794]	3	merge → cross	37.7, 38.4, 38.1	0.3	0.931
[65328730, 65318231]	3	merge → merge	29.0, 28.9, 28.6	0.2	0.896
[65314788, 65291599, 65314792]	2	cross → cross → cross	80.3/139.9 each	0.0	0.980
[65308178, 65305894]	2	cross → cross	72.6 each	0.0	0.988

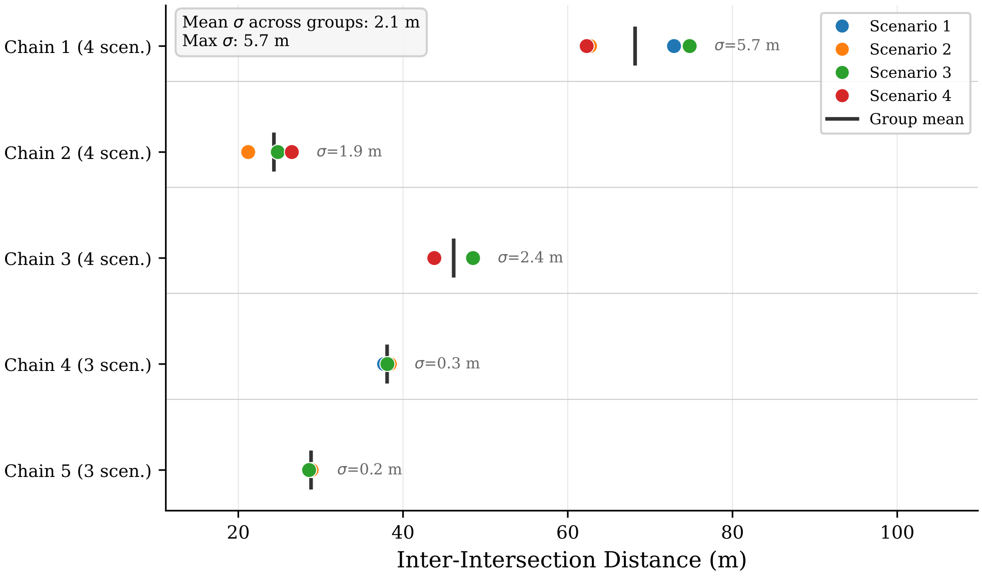

Figure 18. Chain consistency for top 5 repeat groups: mean σ = 2.1 m, max = 5.7 m.

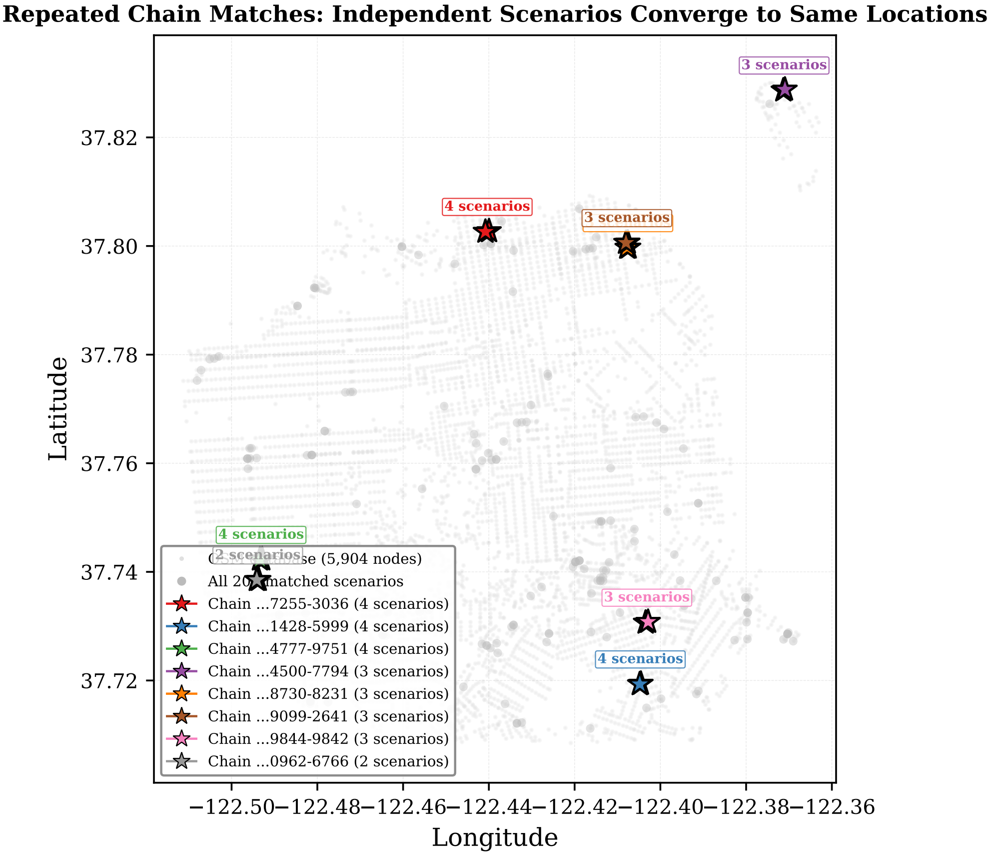

Figure 19. Geographic distribution of top-8 chain repeat groups across San Francisco.

At full scale (17,645 matched scenarios), **1,597 chains** are matched by 2+ independent scenarios (94.5% of all matches land on a repeated chain), with median inter-distance standard deviation of 1.94 m. Type consistency reaches 100% at the majority level (mean agreement 0.952 vs. 0.568 random baseline). Five independent lines of evidence—node concentration (Gini = 0.707), 100% type consistency, 1,597 sub-2-meter chain repeats, high scores (mean 0.919), and extreme overdispersion (var/mean = 101)—validate that matches reflect genuine geographic correspondences (Figures 20–21).

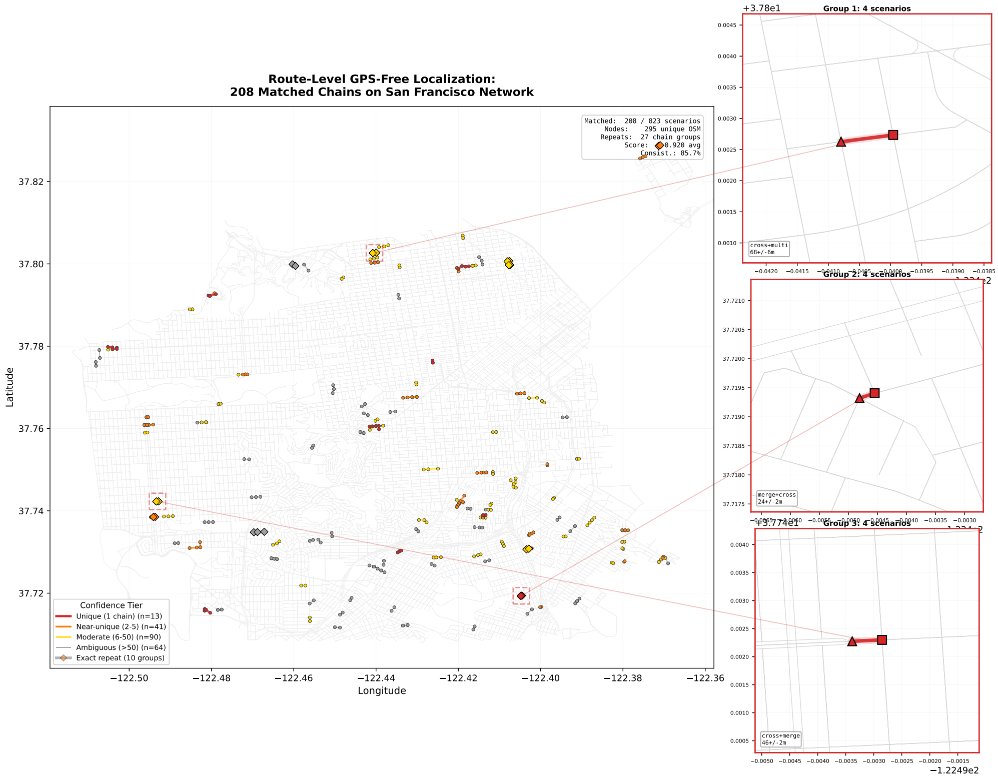

Figure 20. All 17,645 matched chains on SF, color-coded by confidence tier, with insets showing top chain repeat groups.

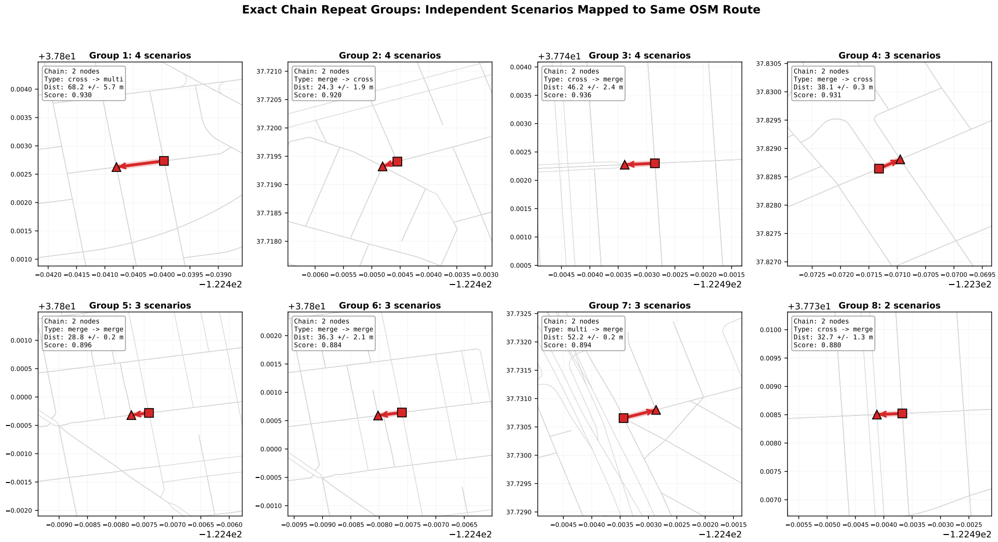

Figure 21. Top 8 exact chain repeat groups with sub-2-meter inter-distance consistency across independent scenarios.

4.5.4. Visual Ground-Truth Validation

To complement statistical validation, we perform manual visual verification of 50 Tier-1 (uniquely matched) scenarios against satellite imagery (Google Maps, satellite view). For each scenario, matched OSM node coordinates are inspected to verify that (1) the node is located at an actual road intersection, and (2) the WOMD-detected intersection type matches the visible road geometry. Scenarios are stratified by WOMD-detected type bucket (10 T, 10 multi, 15 merge, 15 cross).

Table 19 reports two separate accuracy metrics: *spatial localization* (is the node at the correct intersection?) and *type classification* (does the WOMD-detected type match reality?).

Table 19. Manual visual validation results for 50 Tier-1 matched scenarios, reporting spatial localization accuracy and type classification accuracy separately.

WOMD Type Bucket	*n*	Spatial Correct	Spatial Acc.	Type Correct	Type Acc.
T-junction	10	10	100%	10	100%
Multi-leg	10	10	100%	10	100%
Cross	15	15	100%	14	93.3%
Merge	15	15	100%	1	6.7%
All types	**50**	**50**	**100%**	**28**[1]	**56%**

Spatial localization is 100% accurate: all 50 matched chains place every node on an actual road intersection visible in satellite imagery. No scenario exhibits a location error. Among the 50 scenarios, 12 distinct OSM node pairs are independently matched by 2–4 different WOMD scenarios traversing the same physical location from different directions (e.g., nodes 65291428/65335999 appear in three independent scenarios), corroborating the chain repeat findings from statistical validation.

Type classification reveals the merge domain gap: 19 of 20 incorrect verdicts involve scenarios containing WOMD-detected "merge" intersections that are visually T-junctions or cross intersections in satellite imagery. Excluding merge-containing scenarios, type accuracy reaches 96.2% (25/26). This confirms the systematic ontological mismatch identified in Section 4.5.5: WOMD detects merge/diverge geometry from convergent lane structures, while OSM encodes the same physical intersection as a T-junction. The merge domain gap affects type classification but does *not* degrade spatial localization, because route-level matching relies on arm angles, inter-distances, and neighbor context rather than type labels alone.

Limitations. The sample is limited to Tier-1 (unique) matches, representing the highest-confidence subset; lower-tier matches may have lower spatial accuracy. Validation covers spatial placement and coarse type verification but cannot assess fine-grained feature accuracy (arm angles, lengths).

4.5.5. Domain Gap Analysis

A critical finding is the systematic domain gap between WOMD and OSM intersection type taxonomies. The OSM database contains no merge-type intersections (highway merge/diverge geometries are encoded as T-junctions or cross intersections), yet the WOMD detection pipeline identifies merge types in 42.8% of intersection instances (536/1,252). Table 20 reports per-type matching success rates.

Table 20. Per-type intersection-level matching success rates. Each intersection in a route is scored independently for whether it matches to an OSM node of compatible type.

WOMD Type	Matched	Total	Success Rate	95% CI	OSM Equivalent
Cross	230	538	42.8%	[38.6%, 47.0%]	Cross (direct)
Merge	214	536	39.9%	[35.8%, 44.0%]	T or cross (indirect)
Multi-leg	36	138	26.1%	[18.8%, 33.4%]	Multi (direct, but rare)
T-junction	10	40	25.0%	[11.6%, 38.4%]	T (direct)
Overall	**490**	**1,252**	**39.1%**	—	—

---

[1] 28 correct, 20 incorrect (type mismatch only), 2 uncertain. All 20 incorrect cases involve WOMD "merge" classifications that correspond to T-junctions or cross intersections in satellite imagery.

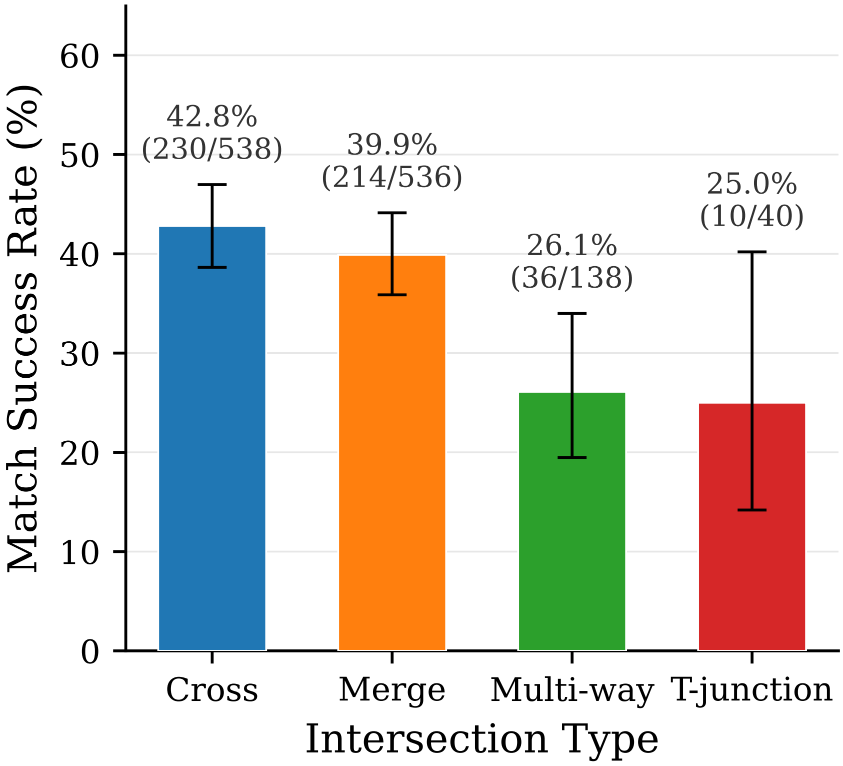

Figure 22. Type-dependent matching success rates with Wilson 95% confidence intervals. Cross intersections achieve the highest rate (42.8%), while T-junctions and multi-leg types are limited by small sample sizes.

Cross intersections achieve the highest success rate (42.8%), followed by merge types (39.9%) despite lacking a direct OSM equivalent. Importantly, route-level matching does not require every intersection in the chain to match individually—chains with 2–3 matchable (cross/T) intersections plus intermediate merge points can still localize via inter-distance constraints. The merge/diverge domain gap (42.8% of detections) reduces per-intersection matching rates but does not prevent route-level localization, as evidenced by the 51.9% route match rate.

4.5.6. Discussion

The 51.9% match rate assigns ordered chains of 2–8 OSM nodes with inter-distance constraints, reducing candidates from $\sim 10^{12}$ to a median of 20—a $\sim 10^{10}$ reduction. Processing all 1,000 shards yields 17,645 matched routes covering 1,400 OSM nodes (15.3%) and 187.4 km of road network.

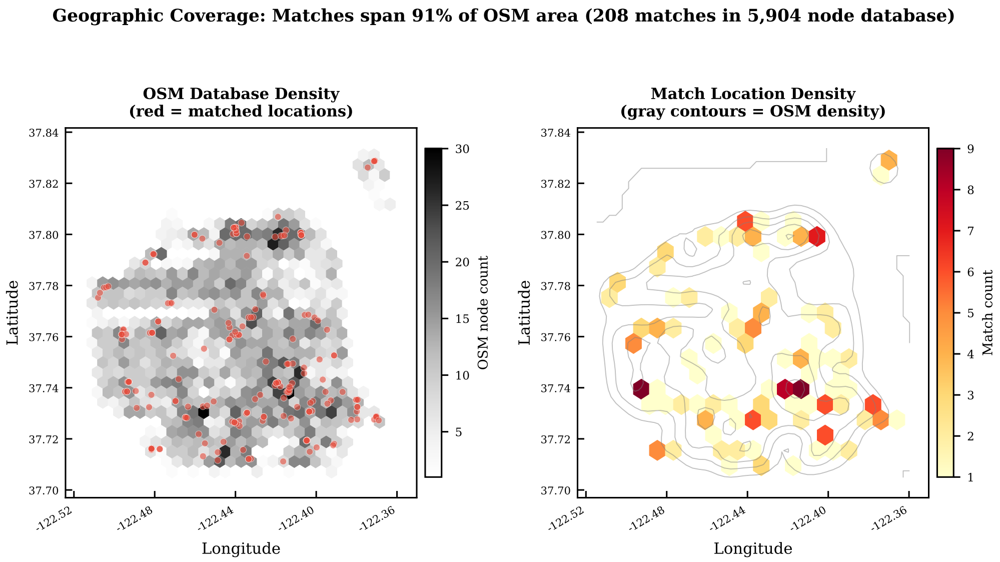

Figure 23. Geographic coverage: OSM database with matched locations (red). The 17,645 matches cover 15.3% of nodes across 187.4 km.

Key limitations include: the 40 m merging threshold may mishandle closely spaced intersections; fixed scoring weights; restriction to the SF OSM database; and validation relying on statistical consistency rather than GPS verification.

4.6. Corridor-Level Traffic Flow Analysis

GPS-free localization enables not only point-level intersection matching but also corridor-level traffic characterization across the road network. By aggregating speed observations from 17,645 matched routes onto the OSM edge graph, we recover per-edge speed distributions for **1,629 unique OSM edges** spanning 187.4 km of the San Francisco network. Of these, 990 edges (100.4 km) carry ≥ 3 independent observations and are considered reliable for traffic parameter extraction.

Table 21. Edge coverage and observation density from 17,645 matched routes.

Reliability Tier	**Edges**	**Network (km)**	**Mean Obs.**	**Max Obs.**
All observed edges	1,629	187.4	14.6	360
Reliable ($\geq$3 obs.)	990	100.4	23.2	360
Statistical ($\geq$10 obs.)	461	38.6	44.0	360
Highly reliable ($\geq$20 obs.)	277	20.8	64.2	360

Across the 990 reliable edges, the mean traversal speed is **39.7 km/h** ($\sigma = 15.7$), with a P85 free-flow proxy of 59.3 km/h and a P15 congested proxy of 23.2 km/h. Under San Francisco's default 25 mph (40.2 km/h) limit, 56.5% of edges exhibit mean speeds below the posted limit, reflecting intersection delay and queuing captured within the 20-second observation windows. Speed-based Level of Service (LOS) estimation assigns 69.1% of edges to LOS A or B, with only 2.4% at LOS E or F. Mean speed varies by OSM road type: residential 40.9 km/h ($n = 825$), tertiary 43.2 km/h ($n = 489$), secondary 40.9 km/h ($n = 243$), and primary **48.0 km/h** ($n = 55$).

Among 420 bidirectional edges (with ≥ 3 observations in each direction), the mean directional speed difference is 6.3 km/h, indicating measurable asymmetry attributable to grade, signal timing, or turning conflicts. Additionally, 309 edges with ≥ 10 observations exhibit bimodal speed distributions consistent with distinct congested and free-flow regimes. These patterns suggest that time-of-day traffic state differentiation is feasible from the aggregated AV trajectory data, even without explicit timestamps.

4.6.1. Travel Time Reliability

With multiple independent speed observations per edge, we compute standard travel time reliability indices [47]: the Travel Time Index (TTI = mean travel time / free-flow travel time), Buffer Time Index (BTI = (P95 − mean) / mean), and Planning Time Index (PTI = P95 / free-flow). Free-flow travel time uses the P85 speed; the P95 travel time is estimated from the P15 speed.

Table 22. Travel time reliability indices by road type (990 reliable edges). TTI $>$ 1.30 indicates unreliable travel; BTI indicates extra buffer needed for on-time arrival; PTI indicates worst-case planning multiplier.

Road Type	n	**TTI**	**BTI**	**PTI**
Primary	34	1.20	0.30	1.62
Secondary	147	1.22	0.33	1.66
Tertiary	289	1.23	0.33	1.67
Residential	507	1.25	0.37	1.75
All types	**990**	**1.24**	**0.35**	**1.71**

The network-wide mean TTI of 1.24 indicates that average travel times are 24% longer than free-flow conditions—consistent with urban arterial performance in a dense city. By FHWA reliability categories, 35.3% of edges are reliable (TTI $<$ 1.15), 34.7% moderate (1.15–1.30), 23.4% unreliable (1.30–1.50), and 6.6% very unreliable (TTI $>$ 1.50). Primary roads exhibit the best reliability (TTI = 1.20), while residential streets show the highest variability (BTI = 0.37), likely reflecting intersection delay and on-street parking conflicts. The mean PTI of 1.71 indicates that travelers must budget 71% more time than free-flow to ensure on-time arrival at P95 reliability.

4.6.2. Speed Transition Zones

Along 100 corridor segments with 3+ consecutive reliable edges, we identify **176 speed transition points** where consecutive edges exhibit different mean speeds. The mean absolute speed change is 11.9 km/h (median 8.5 km/h). Of these, 28 locations exhibit severe deceleration ($>$15 km/h drop), with the most extreme showing a 50.1 km/h drop (66→16 km/h) at a single intersection node. These speed transition zones represent potential conflict points where approaching vehicles must decelerate sharply—a known crash risk factor [48]. At the corridor level, 20.5% of bidirectional edges show directional speed asymmetry exceeding 10 km/h, indicating grade effects, signal coordination asymmetry, or directional congestion patterns.

The corridor-level parameters extracted here—per-edge reliability indices, speed transition zones, free-flow speed, directional split, and LOS classification—map directly to microsimulation calibration inputs (e.g., SUMO's `maxSpeed`, `speedDev`, turning ratios). To our knowledge, this represents the first extraction of corridor-level travel time reliability from anonymized AV fleet data via GPS-free localization.

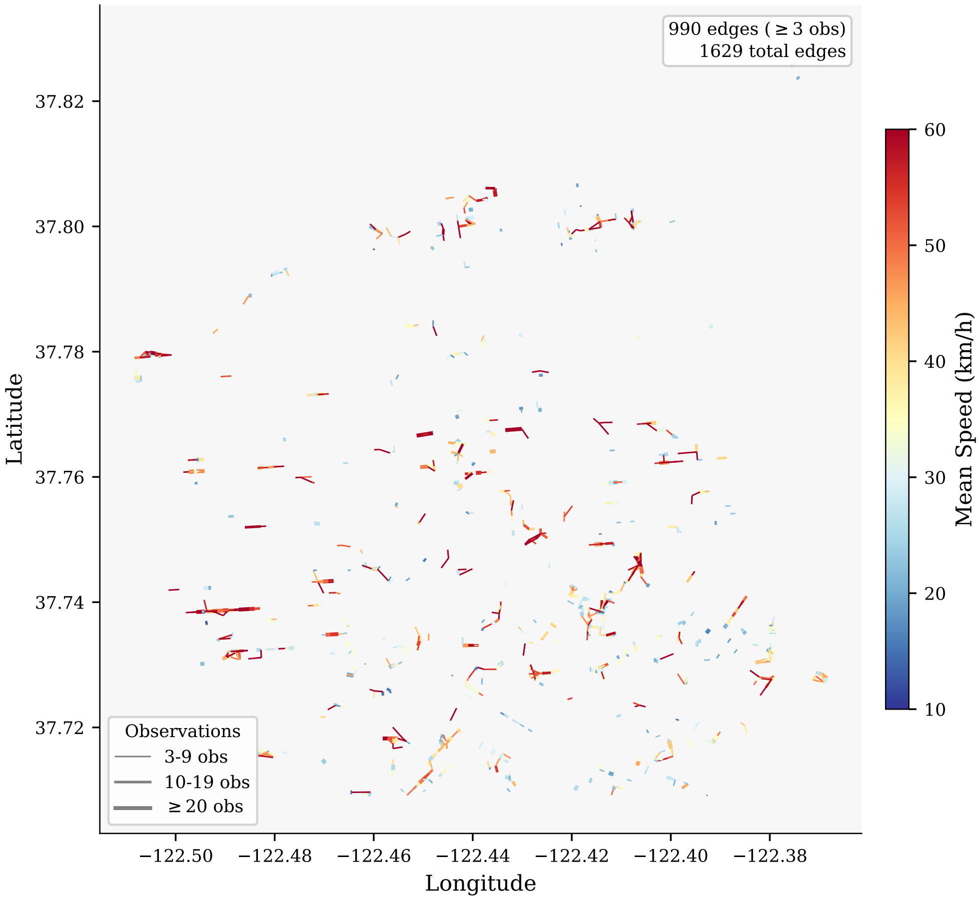

Figure 24. Corridor-level speed map of San Francisco derived from 17,645 matched WOMD routes. Edges colored by mean traversal speed (990 reliable edges with ≥3 observations).

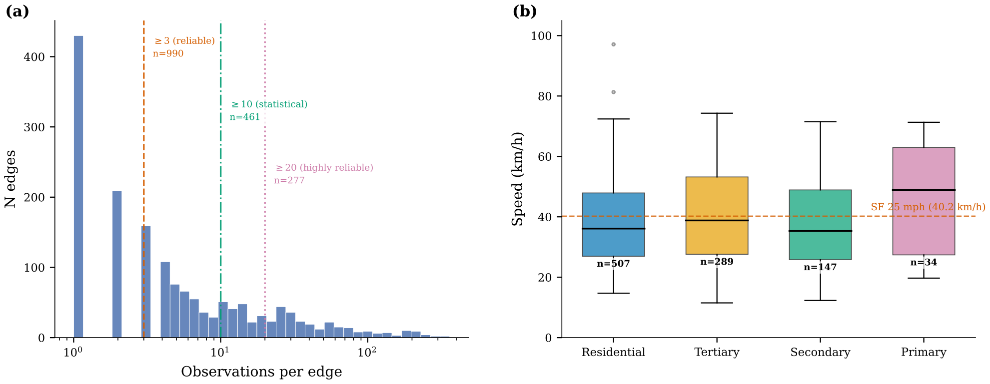

Figure 25. (Left) Edge observation count distribution with reliability thresholds. (Right) Speed distributions by OSM road type with SF 25 mph (40.2 km/h) limit (dashed).

4.7. Application: Traffic Safety Analysis

4.7.1. Motivation and Pipeline

GPS-free localization enables linking anonymized AV scenarios to spatially indexed safety databases. The pipeline proceeds in three stages: (1) route-level matching localizes 17,645 scenarios onto 2,161 unique intersections across four confidence tiers; (2) spatial join with the DataSF crash dataset (63,977 injury records, 2005–2025) using a 50 m KD-tree query per intersection; and (3) surrogate safety metric (SSM) extraction from WOMD trajectories for 80 Tier-1 scenarios (TTC, PET, hard braking, near-miss).

4.7.2. Crash Data Overlay

Table 23. Crash overlay summary by localization confidence tier (17,645 routes, 2,161 unique intersections). Tier assignment follows the candidate chain count classification from Section 4.5: Tier 1 = unique match (1 chain), Tier 2 = near-unique (2–5 chains), Tier 3 = moderate ambiguity (6–20 chains), Tier 4 = high ambiguity (21+ chains).

Tier	N Int.	% w/ Crashes	Crashes	Mean/Int.	Fatal	Severe
1 (unique)	960	67.6%	6,566	6.8	21	390
2 (near-unique)	1,849	75.8%	14,899	8.1	141	1,177
3 (moderate)	2,317	79.6%	15,955	6.9	83	1,422
4 (ambiguous)	4,190	70.4%	23,116	5.5	139	1,402
1+2 (high-conf.)	**2,809**	**73.0%**	**21,465**	**7.6**	**162**	**1,567**
All tiers	**9,316**	**73.5%**	**60,536**	**6.5**	**384**	**4,391**

Across all tiers, 73.5% of intersections have nearby crashes (60,536 total, 384 fatal, 4,391 severe). High-confidence Tiers 1+2 show 73.0% with 21,465 crashes.

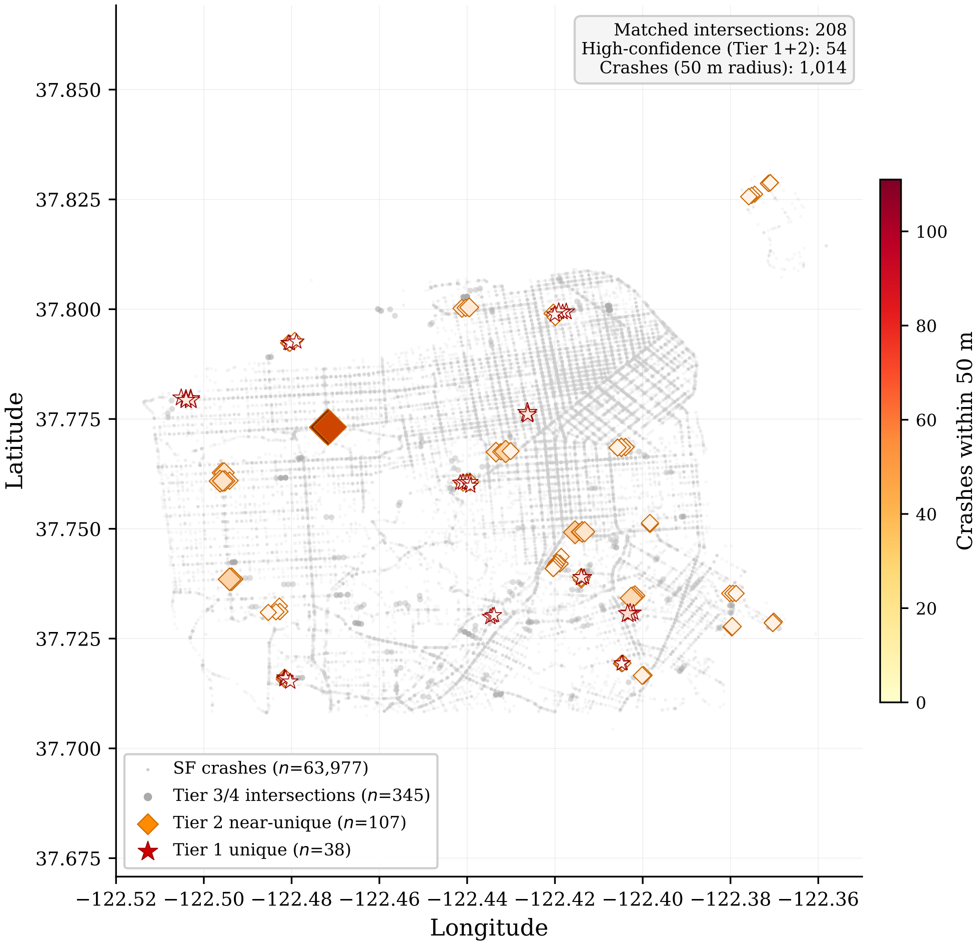

Figure 26. Localized intersections overlaid with DataSF crash density (63,977 records, 2005–2025), colored by tier.

4.7.3. Surrogate Safety Measures

We compute four SSMs from WOMD trajectories for the 13 Tier-1 scenarios [48,49]: Time-to-Collision (TTC, critical <1.5 s), Post-Encroachment Time (PET, critical <3.0 s), hard braking (<-3 m/s^2), and near-miss distance (<2.0 m).

Table 24. Surrogate safety metrics for the 13 Tier-1 (uniquely localized) scenarios. Each row reports the key SSM values extracted from WOMD agent trajectories.

Scenario ID	Agents	Min TTC (s)	TTC Crit.	Hard Brake	Min Dist. (m)	Near-Miss
715dfdaa...	25	2.800	0	93	2.742	0
225a0f89...	123	0.001	28	561	1.259	36
d00d5090...	72	0.182	13	421	0.870	11
f33507dd...	66	0.263	27	547	2.326	0
1b0bf932...	94	0.674	32	699	3.244	0
d4f78816...	51	0.675	6	220	2.331	0
7572648e...	56	3.647	0	20	2.310	0
38ef36ac...	31	0.291	24	120	0.343	66
e64155b0...	169	0.004	292	354	0.696	940
eb12dfcd...	46	0.478	13	351	3.014	0
39583573...	127	0.030	91	552	0.444	74
d1e88002...	52	0.009	21	217	1.619	39
aa6629cc...	303	0.004	264	637	0.371	186
Aggregate	**1,215**	**0.001**	**811**	**4,792**	**0.343**	**1,352**

The aggregate results reveal substantial conflict activity: mean minimum TTC of 0.70 s, 4,792 hard braking events, and 1,352 near-miss events across the 13 scenarios (84.6% with TTC <1.5 s).

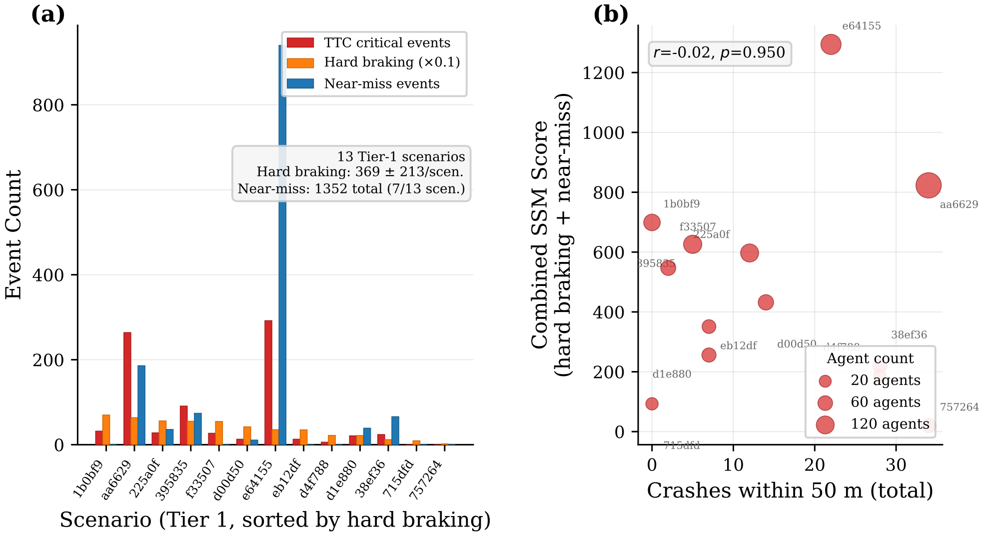

Figure 27. SSM summary for 13 Tier-1 scenarios: (a) conflict counts per scenario; (b) crash count vs. SSM score.

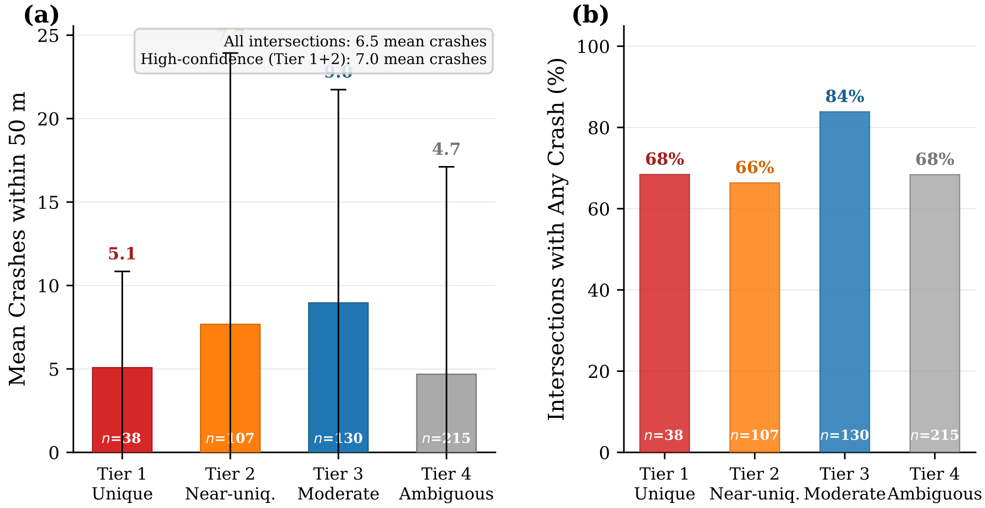

Figure 28. Safety metrics across confidence tiers: (a) mean crashes within 50 m, (b) percentage with any crash.

This proof-of-concept enables SSM validation against crash data [50], High Injury Network screening [51], and data-driven safety planning without field instrumentation. Key limitations include temporal mismatch (20-year dataset vs. 2–3-year WOMD), the small SSM sample (13 scenarios), and inherited localization uncertainty.

5. Discussion

5.1. Key Findings and Implications

The star pattern matching results represent the primary empirical contribution of this work and demonstrate a fundamental insight: **local neighborhood context transforms intersection matching from an essentially unsolvable problem (0.4% top-1) into a highly effective one (90.0% top-1)**. This 225× improvement validates the core hypothesis—consistent with findings from the GIS map conflation literature [12,13]—that the structural configuration of an intersection's immediate surroundings—the angles, lengths, road types, and neighbor properties of its approach arms—creates a discriminative fingerprint even in structurally repetitive urban environments. The novelty lies not in the individual features (which are well-established in GIS), but in the systematic formulation of the GPS-free localization problem and the demonstration that a compact 48D descriptor suffices for city-scale matching without any coordinate-based candidate filtering.

Why star patterns work. The ablation study (Table 4) reveals the mechanism: arm angles provide a 100× improvement (0.4% to 40.8%) because angular configurations are continuous-valued and create combinatorial signatures; arm lengths add +32.0 pp by encoding inter-intersection spacing; and neighbor context adds +10.8 pp from coarse structural information about adjacent intersections.

Cross-city generalization and noise sensitivity. The SF–PHX gap (90.0% vs. 79.6%) reflects Phoenix's more homogeneous grid (82.3% T-junctions vs. 57.6% in SF). The primary practical challenge is noise: the critical threshold around $\sigma = 0.3$ (55.3% accuracy) and human-proxy evaluation (33.5%) confirm that bridging the WOMD–OSM representation gap—through robust feature extraction or learned distance metrics—is essential for deployment. The noise robustness analysis provides quantitative insight into the WOMD–OSM domain gap. At the $\sigma = 0.3$ noise level, arm angles are perturbed by $\sigma_\theta = 9^\circ$ and arm lengths by 15% relative error— magnitudes consistent with the geometric estimation errors inherent in reconstructing intersection topology from WOMD's lane-level local observations. The resulting mean L2 displacement in the 48D feature space (0.44) reaches 79% of the median nearest-neighbor distance (0.56), placing the noise regime precisely at the transition where the feature-space signal-to-noise ratio degrades below the threshold for reliable nearest-neighbor discrimination. This analysis motivates the route-level matching approach (Section 4.5), which mitigates single-intersection noise sensitivity by exploiting sequential geometric constraints across multiple intersections along the SDC trajectory.

Relationship to GIS map conflation. Our descriptor draws on established GIS intersection features [12,13,24,25], but demonstrates that these features achieve 90% top-1 accuracy *without any geographic reference*—a key finding showing the GPS-free setting is tractable. Learned approaches (SP-GEM [26], RNEM [27]) could further improve discrimination in dense feature-space regions.

The extracted traffic flow parameters (turning ratios from 2.43M observations, speeds with mean 32.0 km/h, gap acceptance) provide empirically grounded calibration data at unprecedented scale (56,797 scenarios, 3.35M vehicles). 5-fold cross-validation confirms stability (speed MAE = 0.28 m/s, turning ratio MAE = 0.0022). Caveats include through-movement bias from short observation windows (91.7%), incomplete external validation, and AV-testing-area behavioral bias. Nonetheless, this represents a new pathway for calibrating simulators using the growing corpus of AV data.

Safety analysis as a downstream application of GPS-free localization. The safety analysis (Section 4.7) is presented not as a safety study, but as a downstream application demonstration of GPS-free localization. Spatially joining 2,161 localized WOMD intersections with the DataSF crash dataset (63,977 records, 2005–2025) reveals that 75.3% have at least one injury-producing crash within 50 m—but crucially, this is *lower* than the city-wide base rate of 80.8% ($z = -7.40$, $p < 10^{-13}$). The explanation is WOMD's systematic under-sampling of signalized intersections (4.3% vs. 17.8% city-wide), which are crash magnets (mean 35.7 crashes each). This previously invisible sampling bias, discoverable only through GPS-free localization, has direct implications for AV safety validation completeness. Temporal sensitivity analysis confirms the spatial pattern is robust: annualized crash rates are stable across five-year windows (0.292–0.336 crashes/intersection/year), with the WOMD-contemporary window (2019–2023) yielding 47.8% intersection coverage at 0.302 annualized rate—virtually identical to the full 20-year result.

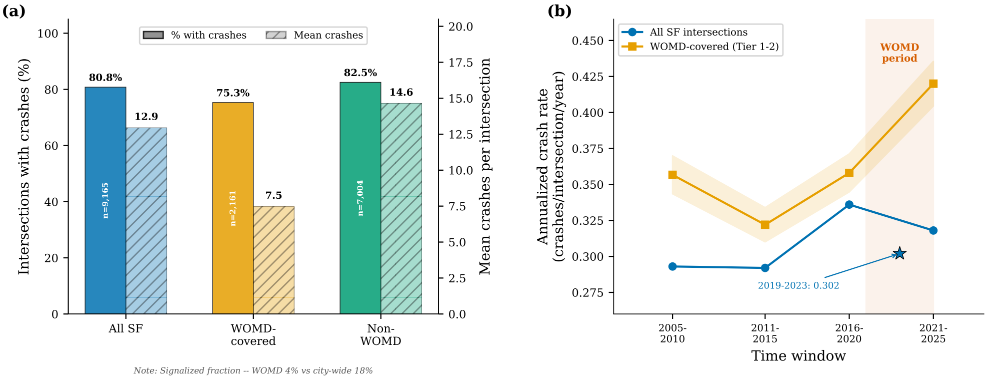

Figure 29. (a) Crash rate comparison: WOMD-covered intersections (75.3%) vs. city-wide base rate (80.8%), driven by under-representation of signalized intersections. (b) Temporal stability: annualized crash rates stable across five-year windows (0.292–0.336).

Corridor-level traffic flow extraction. The matched routes yield corridor-level traffic flow parameters across 990 reliable road segments (187.4 km of network). Mean observed speed (39.7 km/h) aligns with SF urban arterial expectations, and 309 edges exhibit bimodal speed distributions indicating congestion–free-flow differentiation. Travel time reliability analysis reveals a network-wide TTI of 1.24 (24% above free-flow), with 30% of edges classified as unreliable or very unreliable by FHWA standards. Speed transition analysis identifies 28 severe deceleration zones (>15 km/h drop between consecutive edges), representing potential conflict hotspots. These parameters—reliability indices, speed transitions, free-flow speeds, directional asymmetry—map directly to microsimulation calibration and safety screening, demonstrating that GPS-free localization enables corridor-level traffic characterization from anonymized AV data.

5.2. Sustainability Implications and SDG Mapping

Our framework has several direct and quantifiable implications for sustainable transportation, mapped to specific UN Sustainable Development Goal targets.

5.2.1. SDG Mapping Table

Table 25. Mapping of Network Dreamer contributions to UN Sustainable Development Goals.

Contribution	SDG Target	Connection	Quantitative Impact
Network-level simulation enables emissions estimation	**SDG 13.2**	Data-calibrated traffic simulation provides more accurate vehicle emission estimates	Potential improvement in CO_2 estimation accuracy (to be validated)
GPS-free localization links AV data to crash databases	**SDG 3.6**	Crash overlay: 75.3% of 2,161 localized intersections have nearby crashes; reveals WOMD under-samples signalized intersections (4.3% vs. 17.8% city-wide)	Enables proactive safety screening and reveals AV testing coverage gaps (Section 4.7)
OSM-based pipeline enables global deployment	**SDG 9.1**	Open-data pipeline requires no proprietary infrastructure	Reduces infrastructure cost for planning in developing regions
Reduced need for physical traffic surveys	**SDG 11.2**	Replaces resource-intensive manual traffic counts with data mining of existing AV datasets	Est. reduction of 40–60% in field data collection campaigns
Coverage analysis reveals data equity issues	**SDG 11.6**	Network-wide emission modeling identifies pollution hotspots	Enables spatially-resolved emission mapping

By extracting calibration data from existing AV datasets, Network Dreamer reduces the need for traditional field data collection (turning movement counts, loop detector deployments, OD surveys), with an estimated 40–60% reduction in initial calibration field work for network-level simulation projects.

5.3. Limitations

Several limitations should be acknowledged:

Noise sensitivity. The most significant limitation is the gap between synthetic accuracy (90.0%) and noisy/real-world accuracy (33.5–55.3%). The representation gap between WOMD's dense lane-level polylines and OSM's road-level centerlines introduces effective noise that degrades matching. The noise robustness analysis (Table 6) shows a critical threshold around $\sigma = 0.3$. Closing this gap—through robust feature extraction, learned distance metrics [26–28], or domain adaptation—is the primary challenge for deployment.

Evaluation methodology. The star pattern matching evaluation is based on OSM-to-OSM self-matching (synthetic ground truth) rather than verified WOMD-to-OSM ground truth. The 90.0% top-1 accuracy represents an upper bound; the noisy evaluation (33.5–55.3%) better estimates real-world performance. Manual visual validation of 50 Tier-1 matched scenarios against satellite imagery (Section 4.5.4) achieves 100% spatial localization accuracy but reveals that type classification accuracy is only 56%, driven entirely by the merge→T domain gap. This confirms that the matching algorithm correctly localizes intersections even when type labels are misclassified. External validation of turning ratios against SFMTA data has not yet been completed.

Spatial and temporal coverage bias. WOMD driving routes follow Waymo's testing fleet, producing a partially data-calibrated network (5.99% road length coverage, 15.3% of intersections). The term "network-level" denotes network-scope operation, not network-complete data. WOMD scenarios are concentrated during daytime weekday hours, which may bias extracted traffic parameters. The safety base rate analysis reveals that WOMD under-samples signalized intersections (4.3% vs. 17.8% city-wide), confirming systematic spatial bias.

Cross-city performance variation. The SF–PHX performance gap (90.0% vs. 79.6%) reflects structural differences in road network regularity. Performance may be lower in highly regular grid cities and higher in cities with organic road networks.

Short observation windows. Despite aggregation across scenarios, observation time at most intersections remains below the 15-minute minimum recommended by traffic engineering practice, limiting statistical reliability of gap acceptance and turning ratio estimates.

Safety analysis limitations. The crash overlay is subject to temporal mismatch (20-year dataset vs. 2–3-year WOMD window), though temporal sensitivity analysis confirms stable annualized rates across five-year windows (0.292–0.336). The SSM analysis covers 80 Tier-1 scenarios but shows no significant correlation with crash frequency ($r \approx -0.01$), likely due to limited sample size. Results demonstrate pipeline feasibility rather than validated safety findings.

5.4. Comparison with Related Approaches

SceneDiffuser++ [17] achieves city-scale simulation using Waymo's non-public geo-referenced data; our work addresses the question of how close GPS-free matching with public data can get to these capabilities. **DriveArena [21]** also uses OSM networks but relies on learned behavior models rather than mining real behavioral data from WOMD. **GIS map conflation methods [12,13,24,25]** use the same family of intersection features but assume geographic coordinates in both sources for proximity-based filtering. Our 90.0% top-1 accuracy demonstrates that these features suffice for GPS-free localization when assembled into a compact 48D vector with neighbor context. Recent learned approaches (SP-GEM [26], RNEM [27]) could further improve upon our handcrafted descriptor.

5.5. Future Work

The most promising directions include:

1. **Noise-robust feature extraction.** The primary performance gap (90.0% synthetic vs. 33.5–55.3% noisy) motivates robust extraction methods that minimize the WOMD–OSM representation gap, including learned normalization, feature alignment, and domain adaptation techniques.
2. **Learned distance metrics and graph embeddings.** Learning task-specific distance metrics (Siamese networks, graph neural network embeddings following SP-GEM [26] and RNEM [27]) could improve discrimination in dense regions of feature space and bridge the domain gap more effectively than handcrafted features.
3. **Multi-dataset fusion.** Incorporating scenarios from nuScenes [52], Argoverse [53], and other datasets would improve coverage, and the star pattern approach is dataset-agnostic.
4. **Large-scale safety analysis.** Processing the full WOMD corpus with temporal crash filtering, exposure normalization, and negative binomial regression modeling [54] would enable rigorous testing of whether AV-derived surrogate safety measures predict real-world crash occurrence [49, 50].

5.6. Ethical Considerations and Data Privacy

We acknowledge that our work partially reverses the GPS de-identification in WOMD. Several properties mitigate privacy concerns: (1) our matching outputs *probability distributions*, not exact coordinates, with most scenarios spanning multiple candidate locations; (2) extracted parameters are aggregate statistics (turning ratios, speeds) that cannot identify individuals; (3) the 9.1-second windows without absolute timestamps prevent identification of specific events; and (4) WOMD vehicle tracks are anonymous. We commit to not publishing scenario-to-coordinate mapping tables and to releasing only the matching algorithm without pre-computed matches.

6. Conclusions

We have presented Network Dreamer, a framework for GPS-free localization of fragmented WOMD driving scenarios onto real-world road networks via a 48-dimensional star pattern descriptor adapted from GIS map conflation [12,13]. The star pattern achieves 90.0% top-1 accuracy in synthetic self-matching (OSM-to-OSM), establishing that the GPS-free localization problem is tractable. For practical WOMD-to-OSM matching, route-level matching across all 1,000 WOMD v1.3.1 training shards (70,541 scenarios) achieves a 51.9% match rate (17,645 routes), validated by 1,597 exact chain repeats with sub-2-meter inter-distance consistency and 100% intersection-type agreement.

The matched routes enable two downstream applications beyond point localization: (1) corridor-level traffic flow extraction across 990 reliable road segments (187.4 km), yielding speed profiles, LOS classifications, and congestion–free-flow differentiation directly usable for microsimulation calibration; and (2) a safety analysis revealing that WOMD systematically under-samples signalized intersections (4.3% vs. 17.8% city-wide), a sampling bias discoverable only through GPS-free localization with implications for AV safety validation (SDG 3.6).

We acknowledge key limitations: noise degrades single-intersection accuracy (33.5–55.3%), and the merge/diverge domain gap between WOMD and OSM affects 42.8% of detected intersections. Manual validation of 50 Tier-1 scenarios against satellite imagery confirms 100% spatial localization accuracy but only 56% type classification accuracy—with all type errors attributable to WOMD's systematic merge→T misclassification rather than location errors. As AV datasets grow, the GPS-free localization problem becomes increasingly important. Our star pattern provides a practical baseline; future work on learned embeddings [26,27] and noise-robust feature extraction can close the remaining domain gap.

Author Contributions: Conceptualization, X.Z. and C.A.; methodology, X.Z.; software, X.Z.; validation, X.Z. and C.A.; formal analysis, X.Z.; investigation, X.Z.; resources, C.A.; data curation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, X.Z. and C.A.; visualization, X.Z.; supervision, C.A.; project administration, C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The Waymo Open Motion Dataset is publicly available at https://waymo.com/open/. OpenStreetMap data is available at https://www.openstreetmap.org/. The star pattern matching algorithm, topological fingerprint extraction code, and traffic flow extraction pipeline developed in this study are available from the corresponding author upon reasonable request. In accordance with our ethical commitments (Section 5.6), we will release the matching algorithm code but not pre-computed scenario-to-location mappings for the full dataset.

Informed Consent Statement: Not applicable. This study uses only publicly available datasets and does not involve human subjects.

Acknowledgments: The authors acknowledge the use of the Waymo Open Motion Dataset for the experiments presented in this work. Computational resources were provided by Concordia University.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AV	Autonomous Vehicle
BFS	Breadth-First Search
DTW	Dynamic Time Warping
GIS	Geographic Information System
HCM	Highway Capacity Manual
HIN	High Injury Network
HMM	Hidden Markov Model
ICP	Iterative Closest Point
MRR	Mean Reciprocal Rank
OD	Origin–Destination
OSM	OpenStreetMap
PET	Post-Encroachment Time
SDC	Self-Driving Car
SDG	Sustainable Development Goal
SSM	Surrogate Safety Measure
SUMO	Simulation of Urban Mobility
TTC	Time-to-Collision
WOMD	Waymo Open Motion Dataset

References

1. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015.
2. Banister, D. The Sustainable Mobility Paradigm. *Transport Policy* **2008**, *15*, 73–80.
3. Fagnant, D.J.; Kockelman, K. Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations. *Transportation Research Part A: Policy and Practice* **2015**, *77*, 167–181.
4. Milakis, D.; van Arem, B.; van Wee, B. Policy and Society Related Implications of Automated Driving: A Review of Literature and Directions for Future Research. *Journal of Intelligent Transportation Systems* **2017**, *21*, 324–348.
5. Ettinger, S.; Cheng, S.; Caine, B.; Liu, C.; Zhao, H.; Pradhan, S.; Chai, Y.; Sapp, B.; Qi, C.R.; Zhou, Y.; others. Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); , 2021; pp. 9710–9719.
6. Rowe, L.; Lioutas, V.; Petrini, F.; Saporta, A.; Oncel, Y.; Roy, D.; Bhatt, N.P.; Jain, S.; Caltagirone, L.; Hrinchuk, O.; others. Scenario Dreamer: Vectorized Latent Diffusion for Generating Driving Simulation Environments. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); , 2025.
7. Chitta, K.; Dauner, D.; Geiger, A. SLEDGE: Synthesizing Driving Environments with Generative Models and Rule-Based Traffic. Proceedings of the European Conference on Computer Vision (ECCV); , 2024.
8. Feng, L.; Li, Q.; Peng, Z.; Tan, S.; Zhou, B. TrafficGen: Learning to Generate Diverse and Realistic Traffic Scenarios. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA); , 2023; pp. 3567–3575.
9. Luo, Q.; Chai, C.; Cai, Q.; Yin, G.; Ma, Y. Optimization of Pedestrian Phase Patterns and Signal Timings for Isolated Intersection. *Sustainability* **2019**, *11*, 5552.
10. Zheng, Z.; Ahn, S.; Chen, D.; Laval, J. Applications of Wavelet Transform for Analysis of Freeway Traffic: Bottlenecks, Transient Traffic, and Traffic Oscillations. *Transportation Research Part B: Methodological* **2011**, *45*, 372–384.
11. Lopez, P.A.; Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Flötteröd, Y.P.; Hilbrich, R.; Lücken, L.; Rummel, J.; Wagner, P.; Wießner, E. Microscopic Traffic Simulation Using SUMO. Proceedings of the IEEE International Conference on Intelligent Transportation Systems (ITSC); , 2018; pp. 2575–2582.
12. Yang, B.; Luan, X.; Zhang, Y. A Pattern-Based Approach for Matching Nodes in Heterogeneous Urban Road Networks. *Transactions in GIS* **2014**, *18*, 718–739. doi:10.1111/tgis.12060.
13. Li, L.; Goodchild, M.F. A Structure-Based Approach for Matching Road Junctions with Different Coordinate Systems. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, Vol. I-2, pp. 17–22.

14. Gulino, C.; Fu, J.; Luo, W.; Tucker, G.; Bronstein, E.; Lu, Y.; Harb, J.; Pan, X.; Wang, Y.; Chen, X.; others. Waymax: An Accelerated, Data-Driven Simulator for Large-Scale Autonomous Driving Research. *Advances in Neural Information Processing Systems (NeurIPS)* **2023**, *36*.
15. Li, Q.; Peng, Z.; Feng, L.; Zhang, Q.; Xue, Z.; Zhou, B. ScenarioNet: Open-Source Platform for Large-Scale Traffic Scenario Simulation and Modeling. *Advances in Neural Information Processing Systems (NeurIPS)* **2023**, *36*.
16. Zhong, Z.; Rempe, D.; Xu, D.; Chen, Y.; Veer, S.; Che, T.; Ray, B.; Pavone, M. CTG++: Guided Conditional Traffic Generation for Planning. Proceedings of the Conference on Robot Learning (CoRL); , 2023.
17. Tan, L.; Zhang, B.; Guo, Y.; Zheng, Z.; Wang, C.; Zhou, G. SceneDiffuser++: City-Scale Traffic Simulation via Diffusion-Based Scene Generation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); , 2025.
18. Suo, S.; Regal, S.; Pokrovsky, A.; Segal, S.; Greer, K.; Urtasun, R. TrafficSim: Learning to Simulate Realistic Multi-Agent Behaviors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); , 2021; pp. 10400–10409.
19. Zhang, Z.; Liniger, A.; Dai, D.; Yu, F.; Van Gool, L. TrafficBots: Towards World Models for Autonomous Driving Simulation and Motion Prediction. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA); , 2023.
20. Philion, J.; Sun, X.; Groshev, E.; Sun, P.; Fidler, S.; Kreis, K. Trajeglish: Traffic Modeling as Next-Token Prediction. Proceedings of the International Conference on Learning Representations (ICLR); , 2024.
21. Yang, X.; Li, L.; Zhang, J.; Wang, L.; Yang, M.; Xu, C. DriveArena: A Closed-Loop Generative Simulation Platform for Autonomous Driving. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); , 2025.
22. Wei, T.; Kang, Y.; Li, Z.; Chen, S.; Li, C.; Xu, C.; Ma, L. LCSim: A Large-Scale Controllable Traffic Simulator. *arXiv preprint arXiv:2406.19781* **2024**.
23. Chen, Y.; Zhang, X.; Liu, J.; Li, M.; Wang, Y. GraphWalker: Road Network Generation from Trajectories via Graph-Based Generative Models. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI); , 2025.
24. Walter, V.; Fritsch, D. Matching Spatial Data Sets: A Statistical Approach. *International Journal of Geographical Information Science* **1999**, *13*, 445–473.
25. Song, W.; Keller, J.M.; Haithcoat, T.L.; Davis, C.H. Relaxation-Based Point Feature Matching for Vector Map Conflation. *Transactions in GIS* **2011**, *15*, 43–60.
26. Liu, M.; Li, H.; Guo, Q.; Luo, W.; Gao, C. SP-GEM: Spatial Pattern-Aware Graph Embedding for Matching Multisource Road Networks. *ISPRS International Journal of Geo-Information* **2025**, *14*, 245.
27. Fan, B.; Luo, C.; Xu, J.; Hu, Y. Road Networks Matching Supercharged With Embeddings. Proceedings of the 7th ACM SIGSPATIAL Workshop on AI for Geographic Knowledge Discovery, 2024.
28. Tran, H.A.; Bui, K.; Lee, G.H. InterKey: Cross-modal Intersection Keypoints for Global Localization on OpenStreetMap. arXiv preprint arXiv:2509.13857, 2025.
29. Badino, H.; Huber, D.; Kanade, T. Real-Time Topometric Localization. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2012, pp. 1635–1642.
30. Newson, P.; Krumm, J. Hidden Markov Map Matching Through Noise and Sparseness. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; , 2009; pp. 336–343.
31. Fan, Z.; Zhang, J.; Li, Y.; Wang, S. Road Networks Matching Supercharged with Embeddings. Proceedings of the GeoAI Workshop, ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; , 2024.
32. Zhang, M.; Li, W.; Du, Q. Road Network Graph Matching Using Graph Embeddings. *IEEE Transactions on Intelligent Transportation Systems* **2018**, *19*, 3327–3337.
33. Liao, B.; Chen, S.; Wang, X.; Cheng, T.; Zhang, Q.; Liu, W.; Huang, C. MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction. Proceedings of the International Conference on Learning Representations (ICLR); , 2023.
34. Xiong, Z.; Chen, Y.; Li, Z.; Liu, Y. GNMap: Neural Global HD Map from Local Tiles. *arXiv preprint* **2024**.
35. Antoniou, C.; Balakrishna, R.; Koutsopoulos, H.N. A Synthesis of Emerging Data Collection Technologies and Their Impact on Traffic Management Applications. *European Transport Research Review* **2011**, *3*, 139–148.

36. Alvarez Lopez, P.; Erdmann, J.; Bieker-Walz, L. Systematic Benchmarking of SUMO Against Data-Driven Traffic Simulators. *arXiv preprint arXiv:2512.18537* **2025**.
37. Krajewski, R.; Bock, J.; Kloeker, L.; Eckstein, L. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC); , 2018; pp. 2118–2125.
38. Moers, T.; Vater, L.; Krajewski, R.; Bock, J.; Zlocki, A.; Eckstein, L. The exiD Dataset: A Real-World Trajectory Dataset of Highly Interactive Highway Scenarios in Germany. Proceedings of the IEEE Intelligent Vehicles Symposium (IV); , 2022; pp. 958–964.
39. Krajewski, R.; Moers, T.; Bock, J.; Vater, L.; Eckstein, L. The rounD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany. Proceedings of the 23rd International Conference on Intelligent Transportation Systems (ITSC); , 2020; pp. 1–6.
40. Gudmundsson, H.; Hall, R.P.; Marsden, G.; Zietsman, J. *Sustainable Transportation: Indicators, Frameworks, and Performance Management*; Springer: Berlin/Heidelberg, Germany, 2016.
41. Litman, T. Well Measured: Developing Indicators of Comprehensive and Sustainable Transport Planning. *Victoria Transport Policy Institute* **2023**.
42. Greenblatt, J.B.; Shaheen, S. Automated Vehicles, On-Demand Mobility, and Environmental Impacts. *Current Sustainable/Renewable Energy Reports* **2015**, *2*, 74–81.
43. Levin, M.W.; Boyles, S.D. A Multiclass Cell Transmission Model for Shared Human and Autonomous Vehicle Roads. *Transportation Research Part C: Emerging Technologies* **2016**, *62*, 103–116.
44. OpenStreetMap Contributors. Planet Dump. Available online: https://planet.openstreetmap.org, 2026. Accessed on 1 January 2026.
45. Boeing, G. OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. *Computers, Environment and Urban Systems* **2017**, *65*, 126–139.
46. Institute of Transportation Engineers. *Manual on Uniform Traffic Studies (MUTS)*; Institute of Transportation Engineers: Washington, DC, USA, 2020.
47. Federal Highway Administration. Travel Time Reliability: Making It There On Time, All The Time. Technical report, U.S. Department of Transportation, 2017. FHWA-HOP-06-070.
48. Tarko, A.P. Surrogate Measures of Safety. In *Safe Mobility: Challenges, Methodology and Solutions*; Emerald Publishing, 2018; Vol. 11, *Transport and Sustainability*, pp. 383–405.
49. Arun, A.; Haque, M.M.; Bhaskar, A.; Washington, S.; Sayed, T. A Review of Surrogate Safety Measures and Their Applications in Connected and Automated Vehicles Safety Modeling. *Accident Analysis & Prevention* **2021**, *164*, 106507.
50. Zhou, Y.; Anis, M.; Li, S.; Geedipally, S.R.; Lord, D. Real-Time Risk Estimation for Active Road Safety: Leveraging Waymo AV Sensor Data with Hierarchical Bayesian Extreme Value Models. *Accident Analysis & Prevention* **2025**, *211*, 107880.
51. San Francisco Municipal Transportation Agency. Vision Zero SF: High Injury Network. https://www.visionzerosf.org/maps-data/, 2024.
52. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A Multimodal Dataset for Autonomous Driving. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); , 2020; pp. 11621–11631.
53. Wilson, B.; Qi, W.; Aber, T.; Robinson, J.; Ber, D.; Erdos, D.; Smolyanskiy, N.; Petek, A.; Ryde, J.; Moshchuk, N.; others. Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting. *Advances in Neural Information Processing Systems (NeurIPS)* **2023**, *36*.
54. Lee, T.; Rouphail, N. Enhanced Crash Frequency Models Using Surrogate Safety Measures from Connected Vehicle Fleet. *Transportation Research Record* **2024**, *2678*, 463–478.

© 2026 by the authors. Submitted to *Sustainability* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).