



OPEN

DATA DESCRIPTOR

# A Fine-Grained Lightweight Urban Signalized-Intersection Dataset of Dense Conflict Trajectories

Yiyang Chen<sup>1,2</sup>, Zhigang Wu<sup>1,2</sup>, Guohong Zheng<sup>1,2</sup>, Xuesong Wu<sup>1,2,3</sup>, Liwen Xu<sup>1,2</sup>, Haoyuan Tang<sup>1,2</sup>, Zhaocheng He<sup>1,2,3</sup>✉ & Haipeng Zeng<sup>1,2</sup>

The trajectory data of traffic participants (TPs) is a fundamental resource for evaluating traffic conditions and optimizing policies, especially at urban intersections. Although data acquisition using drones is efficient, existing datasets still have limitations in scene representativeness, information richness, and data fidelity. This study introduces FLUID, comprising a fine-grained trajectory dataset capturing dense conflicts at typical urban signalized intersections, and a lightweight, full-pipeline framework for drone-based trajectory processing. FLUID covers three distinct intersection types, with approximately 5 hours of recording time and featuring over 20,000 TPs across 8 categories. Notably, the dataset records an average of 2.8 vehicle conflicts per minute across all scenes, with roughly 15% of all recorded motor vehicles directly involved in these conflicts. FLUID provides comprehensive data, including trajectories, traffic signals, maps, and raw videos. Comparison with the DataFromSky platform and ground-truth measurements validates its high spatio-temporal accuracy. Through detailed classification of motor vehicle conflicts and violations, FLUID reveals diverse interactive behaviors, demonstrating its value for human preference mining, traffic behavior modeling, and autonomous driving research.

## Background & Summary

Trajectory datasets of traffic participants (TPs) are fundamental for advancing Intelligent Transportation Systems. Modern data acquisition techniques provide these datasets with unprecedented granularity, enabling the empirical observation of microscopic interactions<sup>1</sup>. Consequently, they support a range of research areas, including microscopic simulation<sup>2</sup>, behavioral modeling<sup>3</sup>, and trajectory generation<sup>4</sup>. This level of detail is especially critical at urban intersections, which are dense with complex traffic conflicts. Specifically, the insights gained from this data—covering yielding strategies, violation patterns, and near-collision events<sup>5</sup>—are essential for designing effective control strategies<sup>6</sup>, developing safety interventions, and testing autonomous driving systems. Therefore, high-resolution intersection trajectory data is invaluable for improving urban traffic operations and proactive safety.

Methods for observing traffic participant (TP) mobility and interactions are primarily categorized into ego-centric, roadside, and aerial perspectives<sup>7</sup>. Ego-centric approaches, foundational to autonomous driving datasets like KITTI<sup>8</sup>, Apolloscape<sup>9</sup>, Waymo<sup>10</sup>, and nuPlan<sup>11</sup>, equip vehicles with multi-modal sensors to capture surrounding interactions<sup>12</sup>, but are inherently limited by occlusions and a restricted field of view that prevent a complete scene understanding. Roadside methods, employing either single-sensor (e.g., NGSIM<sup>13</sup>, Zen Traffic Data<sup>14</sup>, I-24 MOTION<sup>15</sup>, TJRD<sup>16</sup>) or multi-sensor systems (e.g., DLR-UT<sup>17</sup>), offer long-term observation but suffer from limited spatial coverage and potential error accumulation<sup>18</sup>. A critical drawback for both ground-based methods is that visible hardware can alter driver behavior, compromising data naturalness<sup>19</sup>. In contrast, aerial observation—typically using drones for their stability and cost-effectiveness—naturally overcomes these limitations by providing a complete, bird's-eye view of the entire traffic scene and all simultaneous interactions within it. Crucially, high-altitude operation ensures the drone remains virtually unnoticed, thus preserving the naturalness of TP behavior. These advantages have spurred the development of numerous drone-based datasets for

<sup>1</sup>School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, Guangdong, China. <sup>2</sup>Guangdong Provincial Key Laboratory of Intelligent Transportation Systems, Shenzhen, 518107, Guangdong, China. <sup>3</sup>Pengcheng Laboratory, Shenzhen, 518055, Guangdong, China. ✉e-mail: [hezch@mail.sysu.edu.cn](mailto:hezch@mail.sysu.edu.cn)

highways (e.g., highD<sup>20</sup>, exiD<sup>21</sup>, CQSkyEyeX<sup>22</sup>, MiTra<sup>23</sup>), urban junctions (e.g., inD<sup>24</sup>, SIND<sup>25</sup>, Hohhot-HDI<sup>26</sup>, Songdo Traffic<sup>27</sup>, Round<sup>28</sup>), and mixed urban environments (e.g., INTERACTION<sup>29</sup>, pNEUMA<sup>30</sup>, CitySim<sup>31</sup>).

Among the various traffic scenarios, urban intersections are particularly critical due to the significant efficiency losses and safety risks associated with the interrupted traffic flow. In the domain of intersection datasets, our review focuses on three key aspects: the captured scenes, the provided information, and the quality of trajectory data. Prominent existing datasets typically cover a limited number of scenarios (e.g., 3-4 intersections)<sup>24,25,29,31</sup>, but generally provide comprehensive information, including precise object dimensions and detailed scene elements. In contrast, the HDI dataset<sup>26</sup>, while featuring diverse trajectory patterns across more intersections, lacks the aforementioned details. This discrepancy precludes a direct and fair comparison. Therefore, our subsequent analysis evaluates the datasets presented in Table 1 against these three aspects, with separate considerations for motor vehicles (MVs) and vulnerable road users (VRUs, including pedestrians and two-wheelers).

The first aspect is the captured scenes, analyzed from both static (intersection type) and dynamic (traffic flow) perspectives. From a static perspective, intersections are classified by their geometries and channelization—such as three-way/four-way intersections (TI/FI), and four-way with dedicated right-turn lanes (FIDRT)—and by their control methods. The latter include unsignalized (e.g., uncontrolled/right-before-left, all-way-stop, priority-controlled) and signalized types, which feature permissive or protected phases for conflicting directions<sup>32</sup>. Different characteristics of intersections directly affect the speed and angle of TPs when crossing them, which is closely related to the situation of traffic conflicts. However, Fig. 1 shows that existing drone datasets seldom focus on intersections with dedicated channelization and usually provide incomplete coverage of diverse control strategies. From a dynamic perspective, we evaluate traffic flow using three key metrics: TP arrival rates, MV conflict ratio, and the number of associated MV per conflict ( $N_{CMVCP}$ ), where conflicts are identified using time-based surrogate safety measures (SSMs). A comparative analysis reveals several limitations in existing datasets, such as low or imbalanced participant counts (e.g., a low VRU proportion in INTERACTION and a low overall TP count in inD) and consistently low MV conflict ratio across several major datasets. Furthermore, the pNEUMA dataset's high shooting altitude leads to decimeter-level target positioning accuracy, making it unsuitable for conflict analysis. In contrast, our proposed FLUID dataset shows clear advantages, featuring a higher overall TP arrival rate and a higher MV conflict ratio of over 15%. Its  $N_{CMVCP}$  metric, comparable to that of SIND, further underscores its strength in capturing dense and interconnected traffic conflicts.

Beyond the scenes themselves, the richness and standardization of the provided information present further challenges. This encompasses both object attributes and behaviors. Large-scale datasets such as pNEUMA and Songdo Traffic lack structured maps, which constrains spatial analysis and simulation of traffic behavior. For TPs from aerial perspectives, the inherent lack of detail hinders both VRU detection and MV classification. Consequently, CitySim omits VRUs entirely, INTERACTION lacks a classification for them, and even in class-annotated datasets like SIND, misclassifications are common. Furthermore, comprehensive behavioral annotations are rare. While analyzing the spatial distribution of traffic conflicts, they seldom provide individual conflict attributes (e.g., object types, conflict types), and the conflicts themselves are often temporally sparse (see Fig. 2). Intent information, crucial for understanding driving behavior, is another neglected area, with only the SIND dataset offering preliminary annotations for turning and violation intentions.

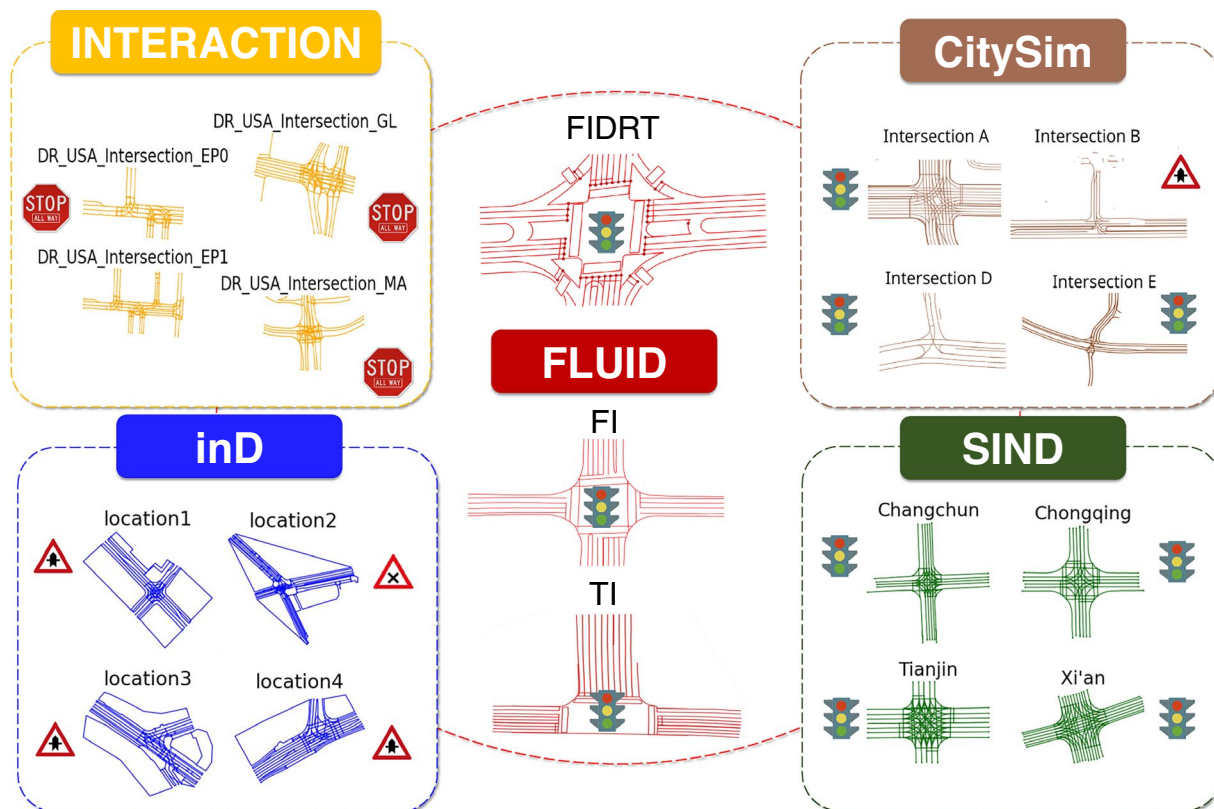
The ultimate value hinges on data quality. It presents challenges related to both the final data outputs and the methodological transparency of the generation process. Regarding the final outputs, quantitative assessments of spatio-temporal accuracy are scarce. Most works vaguely mention manual annotation ratios and typically only release polished, final-version trajectories without the corresponding raw data (e.g., videos). This practice prevents users from independently verifying data fidelity or assessing potential error accumulation from over-processing. Methodological transparency is also a widespread issue. Implementation details for detection algorithms are often sparse (e.g., the application of U-Net<sup>24</sup>, Mask R-CNN<sup>31</sup>, YOLOv5<sup>25</sup>) or undisclosed, and descriptions of tracking algorithms are almost universally absent. Furthermore, commercial platforms like DataFromSky<sup>33</sup> or GoodVision<sup>34</sup> are not viable alternatives for many researchers due to barriers such as high costs, accuracy issues, and stringent eligibility requirements. Collectively, the lack of transparency and access to raw data severely undermines the reproducibility of existing datasets.

To address the aforementioned challenges, we introduce FLUID, a new fine-grained trajectory dataset for urban intersections. We conducted 14 flight campaigns at three carefully selected signalized intersections in Xuancheng, Anhui, China, while simultaneously recording their traffic signal states. This process yielded a dataset rich in diverse traffic behaviors, generated via a clear and lightweight pipeline. FLUID is characterized by the following key features:

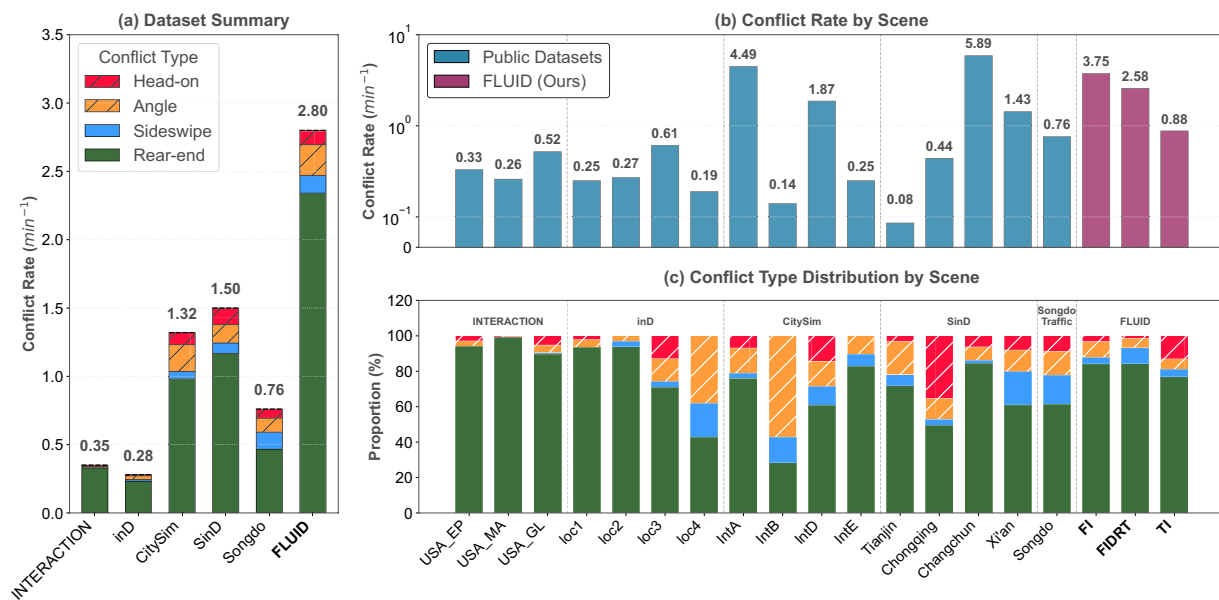
- **Scene Representativeness:** FLUID features three distinct types of signalized intersections, chosen to cover a range of common traffic conflict types. The high arrival rates of TPs and a significant proportion of conflict-involved vehicles result in a dataset characterized by dense and frequent conflict scenarios.
- **Information Richness:** The dataset provides detailed attributes for multiple classes of TPs. It is further supplemented with synchronized traffic signal states, road maps, and fine-grained annotations of traffic conflicts and behavioral intentions (e.g., turning maneuvers and traffic violations).
- **Data Fidelity:** The spatio-temporal accuracy of the trajectory data was validated against the DataFromSky platform and ground-truth measurements from the RTK-GNSS device. Furthermore, we provide a comprehensive description of our entire data acquisition, processing, and fusion pipeline. This provides a basis for assessing the dataset's reliability and extending the methodology to new scenarios.

Dataset	Year	SCENE		INFORMATION										DATA QUALITY	
		Intersection Type		Traffic Flow			Object Info				Behavior Info			Validation	
		Geometries and channelization	Control methods	MV arrival rate ( $min^{-1}$ )	VRU arrival rate ( $min^{-1}$ )	Conflict MV ratio	Assoc. MV ( $\leq 10$ m)	Structured Map	MV types	VRU types	Conflict annot.	Intention annot.	Speed	Position	
INTERACTION	2019	FI, TI	U (AWS)	34.29	1.82	1.80%	1.46	✓	car	pedestrian / bicycle	—	—	—	—	
inD	2019	FI, TI	U (RBL/prio.)	13.59	9.11	3.72%	1.25	✓	car	pedestrian, bicycle	—	—	—	—	
pNEUMA	2020	Multiple	Unknown	7.08	3.61	—	—	—	Car/Taxi, Medium/Heavy Vehicle, Bus	Motorcycle	—	—	—	—	
CitySim	2022	FI, TI	S (prot./perm.) / U (prio.)	27.69	—	8.14%	1.36	✓	car	—	—	—	—	—	
SIND	2022	FI	S (perm.)	21.41	11.82	11.78%	1.74	✓	car, truck, bus	pedestrian, bicycle, motorcycle	—	✓	—	—	
Songdo Traffic	2024	Multiple	S (prot.)	2.03	0.06	3.12%	1.28	—	car/van, truck, bus	motorcycle	—	—	✓	✓	
FLUID (ours)	2025	FI, TI, FIDRT	S (prot./perm.)	31.10	37.29	15.14%	1.66	✓	car, tricycle, van, truck, bus, trailer	pedestrian, moped	✓	✓	✓	✓	

**Table 1.** Summary of intersection drone datasets. **Notes:** • **INTERACTION dataset/pNEUMA dataset/Songdo Traffic dataset:** We use the publicly available multi-agent prediction version of INTERACTION dataset, which features pre-extracted intersection scenarios. For pNEUMA and Songdo Traffic, they do not have structured road network data, so we summarize the types of intersections by observing their sample images. Additionally, the statistical results of pNEUMA were calculated based on 60 major intersections. • **FI/TI/FIDRT/Multiple:** Abbreviations for intersection geometry and channelization types, as defined in the text. *Multiple* indicates that these types are included, and it also covers other abnormal intersections. • **Signalized/Unsignalized intersection.** • **RBL/AWS/prio./prot./permi.:** Strategies for conflict directions, including Right-Before-Left (RBL), All-Way-Stop (AWS), Priority-controlled (prio.), Protected (prot.), and Permissive (permi.). • **Arrival rate:** The average number of new MVs/VRUs arriving per minute during the recording period. • **MV Conflict ratio:** The percentage of total MVs involved in at least one traffic conflict. A detailed calculation is provided in the Methods section. • **Assoc. MV ( $N_{MVCP}$ ):** Associated Motor Vehicles. For each conflict pair (CP), this metric is the total count of other conflict-involved MVs located within a 10-meter radius of either of the two vehicles in the pair at the moment of conflict, explained in Fig. 12. • **annot.:** Annotation. Indicates whether the corresponding information is explicitly labeled in the dataset.



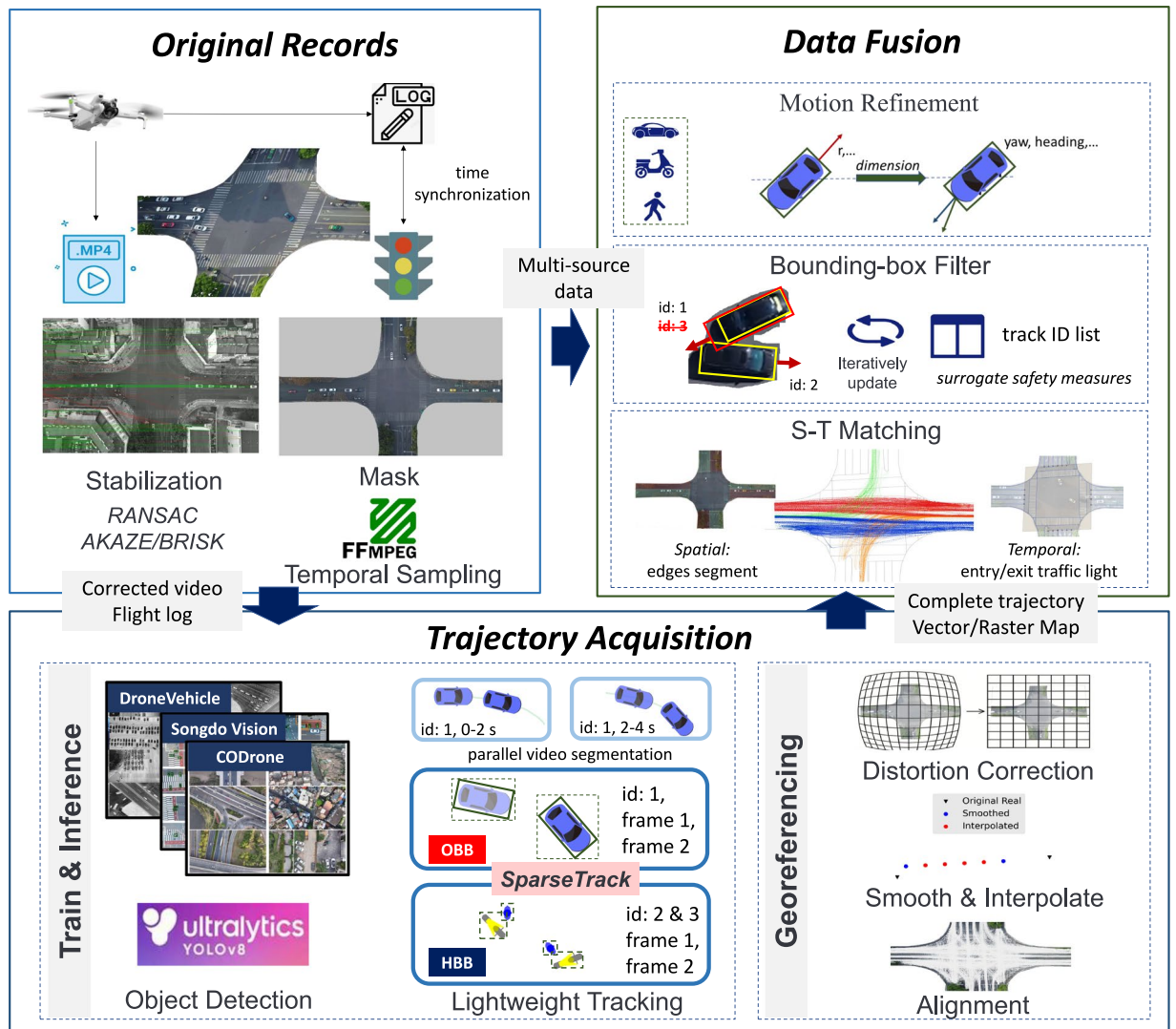
**Fig. 1** Comparative view of intersection networks from drone-based datasets.



**Fig. 2** Comparison of conflict types and frequencies across drone-based intersection datasets. Notes: Conflicts are classified by vehicle yaw angle difference; rates shown are conflicts per minute.

### Methods

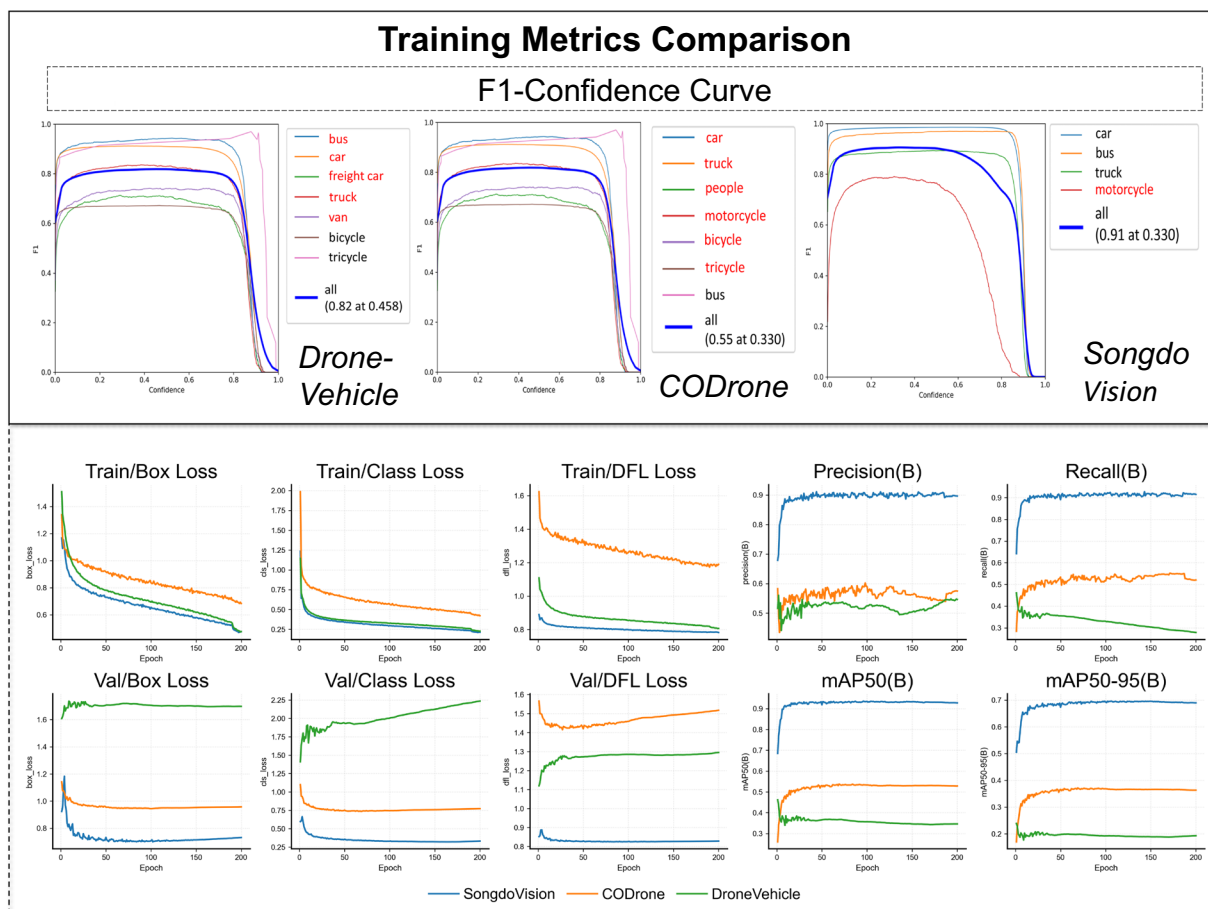
To obtain high-quality and fine-grained annotated trajectory results from raw collected data, we propose the construction and quality enhancement framework shown in Fig. 3, which is divided into three parts: raw recording, trajectory acquisition, and data fusion.



**Fig. 3** Road map of the construction and quality enhancement of FLUID.

**Original Records.** *Raw Data.* The raw data was collected by a three-person team: one operator for the drone, and two observers who recorded traffic signal phases using ground-based cameras. We used a DJI Mini 3 drone to capture high-definition video at 4K resolution ( $3840 \times 2160$  pixels) and approximately 30 FPS (29.97 FPS). Due to regulations, the maximum altitude for drones is 120 meters, which is sufficient for capturing microscopic traffic behavior. The drone maintained a consistent flight altitude of 100 ~ 120 meters (100 ~ 105 m for the FI scenario to capture finer VRU details; 120 m for FIDRT and TI). The positional drift of this type of drone during stationary shooting does not exceed 1.5 meters, with a maximum recording duration per flight of up to 30 minutes. The precise flight altitude for each session can be retrieved from the drone's flight logs. To synchronize the timestamps, the drone and ground cameras were aligned to a unified mobile phone clock with second-level precision, using the coordinated movement of a ground marker as a temporal reference.

*Video Pre-processing.* Due to the drone's lightweight design, the raw footage was susceptible to wind-induced instability and required stabilization. We implemented a two-stage stabilization pipeline based on the open-source tools developed by Fonod *et al.*<sup>35</sup>. The lower level employed the feature detector based on video quality: AKAZE<sup>36</sup> is employed when sharpness remains consistent, leveraging its robust scale-space construction via non-linear diffusion filtering; BRISK<sup>37</sup> is preferred for videos with significant clarity variations. Both offer a favorable balance of speed and accuracy compared to alternatives like SIFT or ORB<sup>38</sup>. The upper level then utilized a RANSAC algorithm with block matching and motion compensation to robustly estimate inter-frame motion parameters while rejecting false matches from dynamic objects. This process effectively eliminated significant jitter from the footage. Moreover, we developed a masking program to define a Region of Interest (ROI) that strictly encompassed the intersection area. This step served to both anonymize areas outside the road network and reduce the computational load for subsequent processing. The FIDRT scenario was exempted from masking to preserve the full range of VRU activities. Finally, the videos were downsampled to 10 FPS. This frame



**Fig. 4** Training result metrics of three detection models (the *advantage categories* of the model are marked in red).

rate is sufficient to capture the decision-making time range of human TPs and is also conducive to detecting the normal motion of VRUs<sup>25</sup>. The primary output of this stage is the set of pre-processed videos.

**Trajectory Acquisition.** The foundation of our analysis lies in the extraction of TP trajectories from videos, which involves identifying track points and associating them with specific individuals in the video.

**Object Detection.** Detection was accomplished using the YOLOv8 architecture<sup>39</sup>, selected for its native support for both horizontal/oriented bounding boxes (HBB/OBB) and its efficiency as a single-stage detector. We prioritized an OBB representation as it provides a tighter and more accurate encapsulation of object boundaries and orientation from our aerial perspective, which is critical for the subsequent analysis of kinematic parameters and TP behaviors. To overcome single-dataset limitations and achieve higher detection accuracy across diverse object classes, we employ a multi-detector ensemble strategy. This strategy involves training the same YOLOv8 architecture independently on three distinct datasets, resulting in three sets of specialised model weights, each optimised to leverage the unique strengths of its training data.

The three specialised detectors were trained on the following datasets:

- **DroneVehicle\_ Revised:** Our custom-curated dataset, which augments the DroneVehicle benchmark<sup>40</sup> with manually-annotated samples of underrepresented classes (e.g., tricycles from FLUID) and additional vehicle types (e.g., motorcycles from the VETRA dataset<sup>41</sup>) to enhance its robustness.
- **CODrone**<sup>42</sup>: A recent high-quality, **4K-resolution** OBB detection benchmark, used to ensure state-of-the-art performance on common object categories, especially small objects such as pedestrians and two-wheelers.
- **Songdo Vision**<sup>35</sup>: A third detector was trained on this dataset to enhance the detection of VRUs (e.g., motorcycles). Although notable for its extensive and precise HBB annotations, its HBB detections were subsequently converted to the OBB format to serve as a supplement to the primary OBB detectors.

For each of the three training processes, the respective dataset was re-partitioned into training, validation, and test subsets using a 70%/20%/10% ratio, and each model instance was trained for 200 epochs. Subsequently, the final detections were generated through a category-based fusion of the outputs from these specialised detectors, as illustrated in Fig. 4.

To evaluate the performance of each detector, we adopt the standard metrics from the YOLOv8 framework, including bounding box loss, classification loss, and Distribution Focal Loss (DFL) for both training and validation phases. The detection performance is quantified by Precision, Recall, and mean Average Precision (mAP) at Intersection over Union (IoU) thresholds of 0.5 and 0.5:0.95. Detailed definitions of each metric are provided in the supplementary materials<sup>43,44</sup>. F1-Confidence curve is a curve that shows the variation of F1-score as Confidence gradually increases. The F1-score exceeded 0.8 for all categories, indicating a well-balanced performance between precision and recall and confirming the robustness of the detection model.

In this process, we identified the object categories where each detector outperforms the others (termed its *advantage categories*). The final detection output for any given frame was constructed by integrating only the detections for these designated advantage categories from their respective specialist detector. This ensemble approach ensures a comprehensive and precise set of object detections.

**Lightweight Tracking.** Once object detection was complete, we employed SparseTrack<sup>45</sup> to link the detections over time and form continuous trajectories. This algorithm was chosen for its effectiveness in dense and complex scenes, as it uses only intersection over union (IoU) for matching. As an enhancement to the widely-used ByteTrack<sup>46</sup>, SparseTrack introduces pseudo-depth estimation and deep cascade matching (DCM), ensuring robustness against occlusions and in mixed-traffic scenarios with VRUs. The pseudo-depth ( $d_p$ ) is defined as:

$$L_p = H - y_p \quad (1)$$

The pseudo-depth of a target is determined by its distance from the camera. This value, denoted as  $d_p$  (where a larger value implies a greater distance), is calculated using the image height  $H$  and the y-coordinate of the bounding box's bottom-center point,  $y_p$ , within the image's pixel coordinate system. Next, the Deep Cascade Matching (DCM) algorithm refines the association process for confirmed tracks by assigning matching priorities. This multi-level strategy employs a cost matrix that combines cosine and Mahalanobis distances alongside the standard IoU score. The final data association is then resolved using the Hungarian algorithm.

For efficient inference on large video files, we employed a stream-based reading and parallel preprocessing pipeline to enable lightweight data loading and mitigate the risk of out-of-memory (OOM) errors.

**Georeferencing.** To georeference the pixel-based trajectories, we performed camera calibration and lens distortion correction. Initial intrinsic and extrinsic parameters were sourced from the image metadata and the drone's flight logs. The distortion coefficients were then refined via a least-squares optimization. This process minimized the discrepancy between theoretical distances, calculated using the Ground Sampling Distance (GSD) from a constant flight altitude, and measured pixel distances in the pre-stabilized video feed. These coefficients account for both radial distortion ( $k_1, k_2, k_3$ ) and tangential distortion ( $p_1, p_2$ ):

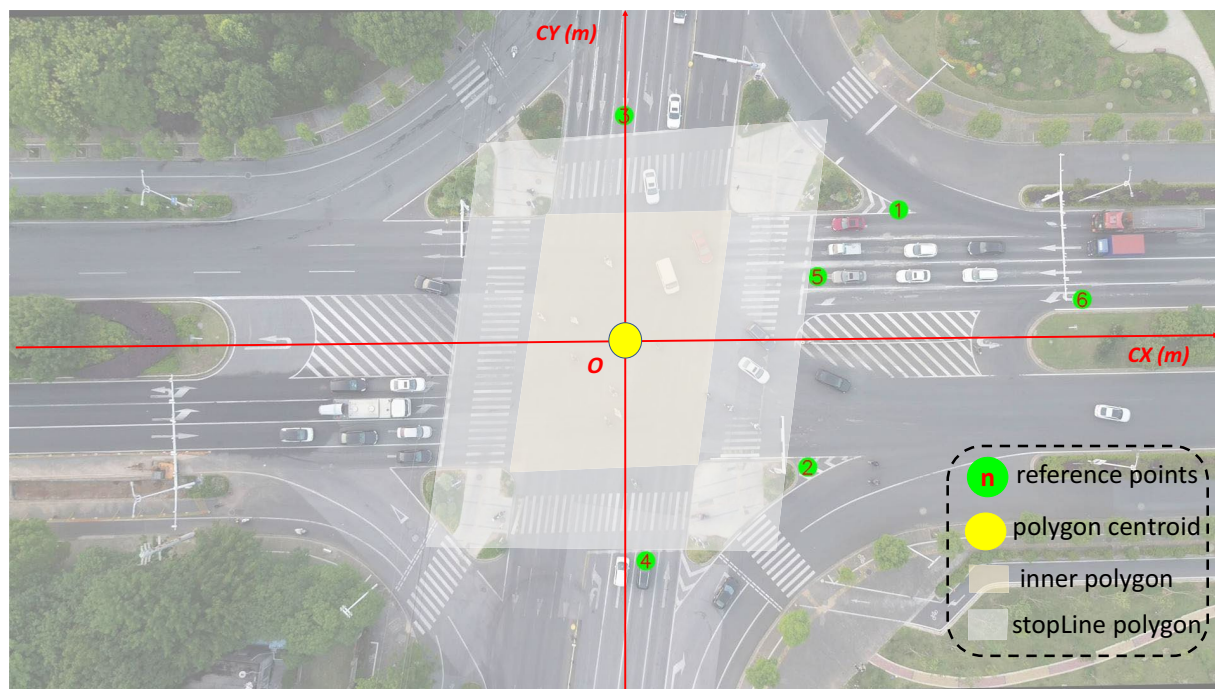
$$\text{Dist} = [k_1, k_2, p_1, p_2, k_3] \quad (2)$$

Given that the distortion function is modeled as a polynomial, the corrected coordinates ( $x_{dist}, y_{dist}$ ) can be calculated from  $(x, y)$  and the coefficient Dist, where  $(x, y)$  is the normalized coordinate computed from the pixel position using the camera intrinsic parameters.

Following distortion correction, we obtained a set of refined pixel coordinates for each trajectory. These trajectories, however, contained jitter and missing frames resulting from residual stabilization errors, occlusions, or indistinct object features. A multi-stage process was implemented in the pixel domain to refine these trajectories, focusing on interpolation and smoothing:

- Savitzky-Golay (S-G) Filter: The S-G filter with a dynamic window size was applied to the pixel coordinates of the bounding box vertices and their orientation angles. This procedure mitigates high-frequency jitter in the raw detections.
- Kinematic Interpolation: Missing data points in the trajectories were interpolated. The positions were estimated by calculating the linear and angular velocities from adjacent valid data points in the pixel coordinate system. This kinematic method was supplemented by linear and nearest-neighbor interpolation as fallback routines. All interpolated points were explicitly flagged.
- Rauch-Tung-Striebel (RTS) Smooth: An RTS-smoother was applied to the complete trajectories (containing both original and interpolated points). It is based on a Constant Velocity (CV) motion model. From the resulting smoothed velocity profile, acceleration was computed using the central difference method.

Upon completion of the pixel-domain processing, the trajectories were georeferenced to a local Cartesian coordinate system (in meters), as illustrated in Fig. 5. This was achieved by first computing a homography matrix that mapped pixel coordinates to WGS84 geodetic coordinates. The matrix was calibrated using several ground control points, whose positions were precisely measured with an RTK-GNSS device. These geodetic coordinates were then projected into the final local system via a Universal Transverse Mercator (UTM) projection. The origin of this local system is the centroid of the intersection's inner polygon, which is derived from a larger stopLine polygon by excluding the crosswalk areas. The stopLine polygon itself is the area bounded by the stop lines and is used for subsequent violation analysis.



**Fig. 5** Illustration of local geographic coordinate system construction.

**Data Fusion.** This process is applied to complete trajectories, aiming to fuse tracks of diverse TP types from multiple sources and align them through spatio-temporal matching.

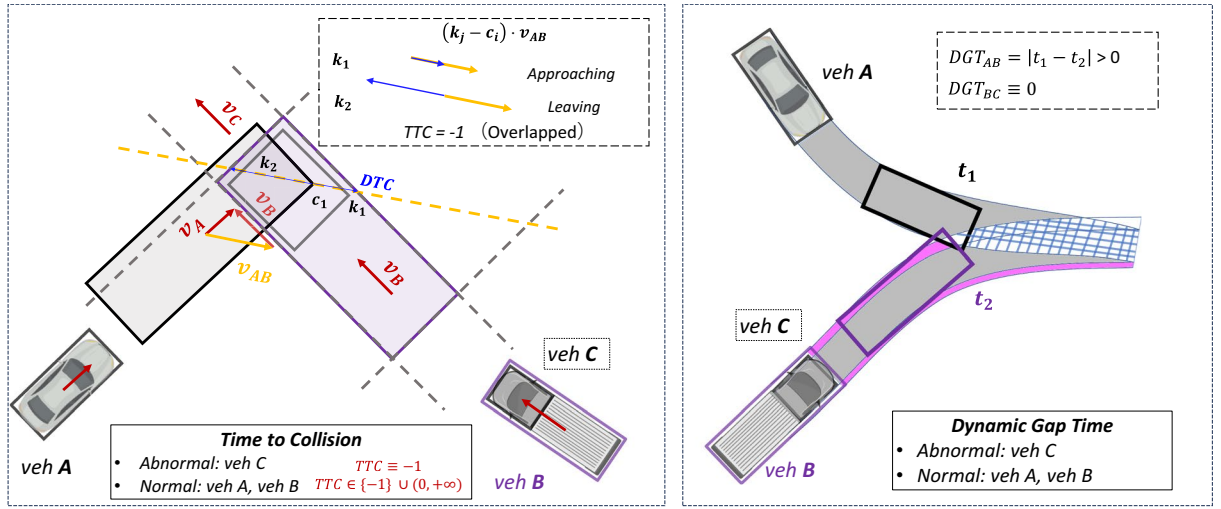
**Motion Refinement.** A stable representation for each target's physical dimensions was established by searching the median width and height (converted to meters) from its complete set of detections. Subsequently, we performed kinematic correction by computing and refining two orientation angles: heading (direction of motion) and yaw (TPs' longitudinal orientation). The raw yaw sequence was first stabilized via a bidirectional method, which traverses the sequence to identify stable values and replace intermittent outliers, mitigating abrupt changes. Concurrently, the heading angle, computed from the target's displacement vector, was used as a reference to correct anomalous yaw values that deviated significantly from the direction of motion.

**Bounding-box Filter.** Erroneous and redundant bounding boxes were then filtered in a two-stage process. First, a heuristic pass removed trajectories with a short duration, minimal displacement, or an average confidence below 0.5, targeting false positives like static objects and shadows. The second stage resolved persistent overlaps on single objects (i.e., dual detections caused by classification ambiguity). Although Surrogate Safety Measures (SSMs) are typically used to analyze traffic conflicts<sup>47</sup>, their inherent ability to quantify spatio-temporal proximity provides a new perspective for lightweight trajectory post-processing. We perform scene-wide removal of redundant bounding boxes using Two-dimensional SSMs (2D-SSMs)—specifically **Time-to-Collision** (TTC) and **Dynamic Gap Time** (DGT)—which distinguish persistent redundancy from transient overlaps. As shown in Fig. 6, vehicle A and B represent valid, distinct targets, whereas vehicle C is abnormal.

As depicted in the left panel of Fig. 6, we consider a scenario where at least one of two objects, vehicle A and vehicle B, is in motion. Building upon the open-source work of Jiao *et al.*<sup>48</sup>, we compute TTC to assist in determining potential overlaps between the bounding boxes of these two relatively moving objects.

Let  $v_{AB}$  be the relative velocity vector between vehicle A and vehicle B. For each corner point  $c_i$  of vehicle A's bounding box, we define  $k_j$  as the intersection point of the line originating from  $c_i$  with direction  $v_{AB}$  and the line segments forming the bounding box of vehicle B. If such an intersection point  $k_j$  exists, the vector  $(k_j - c_i)$  represents the relative displacement from the corner point  $c_i$  to the edge of vehicle B's bounding box. To determine whether an edge is approaching or receding from a corner point, we compute the scalar product of this displacement vector with the relative velocity vector,  $v_{AB}$ . As illustrated in the figure, for the intersection point  $k_1$ , the scalar product  $v_{AB} \cdot (k_1 - c_i)$  yields a positive value, indicating that the corresponding edge of B is approaching corner  $c_i$ . Conversely, for  $k_2$ , the product  $v_{AB} \cdot (k_2 - c_i)$  is negative, signifying that the edge is receding from  $c_i$ . We use an indicator function  $I(c_i)$  to denote this relationship, where  $I(c_i) = +1$  for an approaching edge and  $I(c_i) = -1$  for a receding one.

Let  $d_{i \rightarrow j}$  denote the minimum distance from corner  $c_i$  to the bounding box of vehicle B along the direction of the relative velocity vector  $v_{AB}$ . It is calculated as:



**Fig. 6** Filtering abnormal vehicles based on 2D-SSMs calculations (left: TTC, right: DGT).

$$d_{i \rightarrow j} = \begin{cases} \|k_j - c_i\|, & (k_j - c_i) \cdot v_{AB} \geq 0 \\ \text{inf}, & (k_j - c_i) \cdot v_{AB} < 0 \text{ or } k_j \text{ does not exist} \end{cases} \quad (3)$$

By iterating over all corner points  $c_i$  of vehicle A, we define the Distance-to-Collision (DTC) as the minimum magnitude among all possible distances  $d_{i \rightarrow j}$ . The DTC between vehicles A and B at the current time step is thus given by:

$$\begin{aligned} DTC_{A \rightarrow B} &= \min D_{i \rightarrow j} \\ DTC &= \min\{DTC_{A \rightarrow B}, DTC_{B \rightarrow A}\} \end{aligned} \quad (4)$$

Then, TTC can be calculated from the DTC as follows:

$$TTC = \begin{cases} -1, & \text{if } \left(\sum_{c \in C_i} I_+(c) > 0\right) \wedge \left(\sum_{c \in C_i} I_-(c) > 0\right) \\ \frac{\min\{\|k_c - c\| \cdot I_+(c)\}}{\|v_{ij}\|}, & \text{if } \sum_{c \in C_i} I_+(c) > 0 \text{ and } \sum_{c \in C_i} I_-(c) = 0 \\ \infty, & \text{if } \sum_{c \in C_i} I_+(c) = 0 \end{cases} \quad (5)$$

The decision rule for TTC is as follows: a value of  $-1$ , which indicates that parts of the two objects are simultaneously approaching and receding, signifies a bounding box overlap. In such cases, the trajectory with the shorter duration is identified as redundant and is removed.

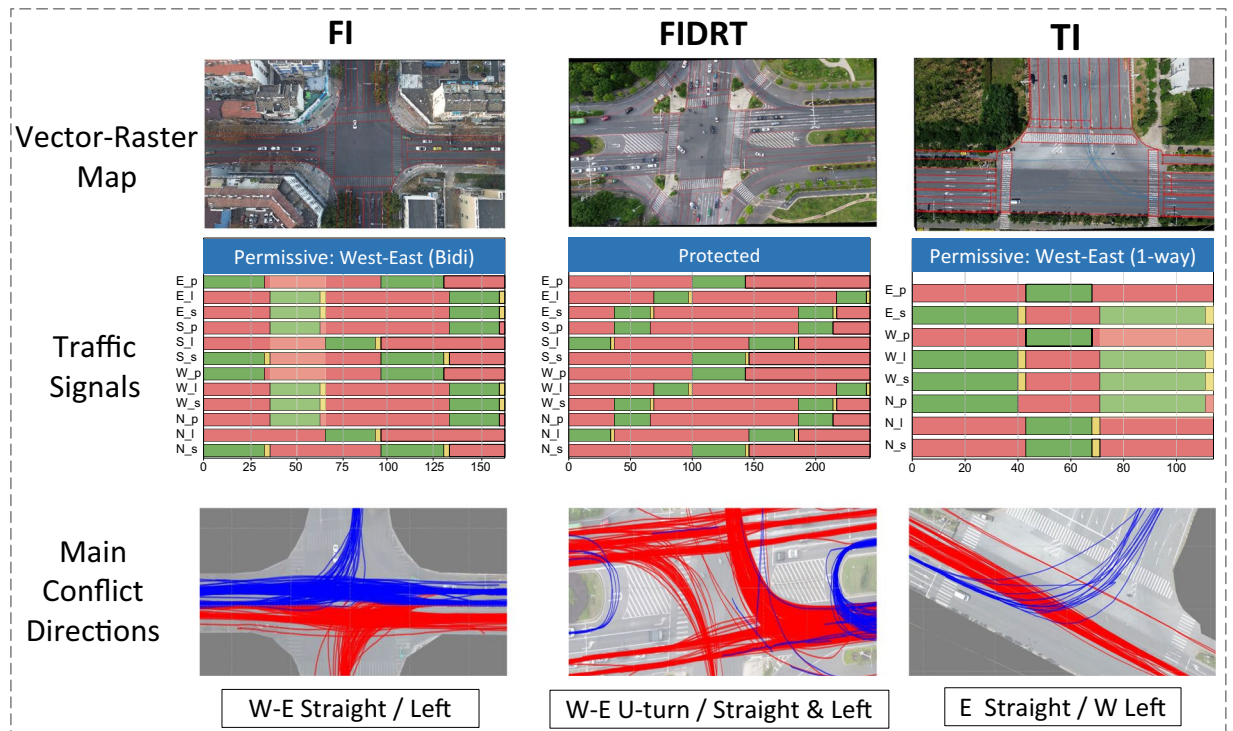
However, the TTC-based method is insufficient for cases where an erroneous bounding box moves in close parallel with the true target, maintaining a near-constant relative velocity (i.e., they are relatively static). To address this limitation, we introduce DGT, as illustrated in the right panel of Fig. 6.

The most common time-based SSM, Post-Encroachment Time (PET), is ill-suited for this task because it requires a clear *exit time* from a conflict zone, which is ambiguous when considering full bounding boxes. We therefore turned to the concept of Gap Time (GT), which only considers the time difference between the two vehicles entering the conflict zone<sup>49</sup>. Based on this, we define DGT. We generalize the conflict point to a conflict area—the intersection of the two bounding boxes’ swept areas, detected via the Separating Axis Theorem (SAT)—and calculate DGT as the time difference between the moments each vehicle first enters this area. A DGT that remains zero for a sustained duration signifies a persistent overlap, prompting the removal of the shorter trajectory.

By applying this sequential filtering process (first TTC, then DGT), we effectively screened out the majority of overlapping and erroneous bounding boxes, retaining only the validated TP trajectories.

**S-T Matching.** The final step in our data fusion pipeline is the precise spatio-temporal matching of the validated TP trajectories.

Spatial matching was achieved by integrating the trajectory data with geographic information. For each aerial video, a georeferenced TIFF image with a local Cartesian coordinate system was generated. Then, the intersection’s road layout was semantically segmented. We partitioned each road edge into its constituent entry and exit



**Fig. 7** Scene information recorded in FLUID.

lane groups. By jointly matching a trajectory's position and its refined yaw angle against these directional layers, we were able to assign a specific turning movement (e.g., left-turn, straight, right-turn) to each TP.

In the temporal matching stage, we focused on time-series data. Each trajectory was synchronized with intersection entry/exit time and the corresponding traffic signal status. This temporal integration allowed for the fusion of supplementary information, such as traffic violations (e.g., red-light running).

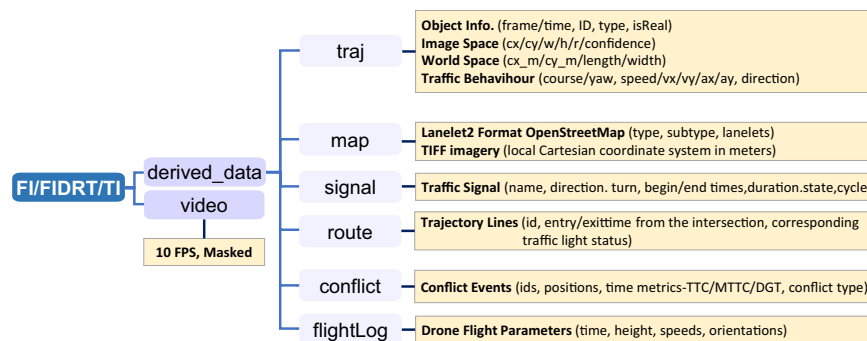
### Data Records

The full dataset is available through the figshare repository<sup>50</sup>. Figure 7 shows all the intersections, which are located in the central urban area. For privacy protection, the specific names and locations of these three intersections have been anonymized. Data was collected at these sites during clear daytime hours on selected days in January and May 2025. After an anomaly screening process, a total of approximately 5 hours of raw video footage was obtained. An overview of these three intersections is as follows:

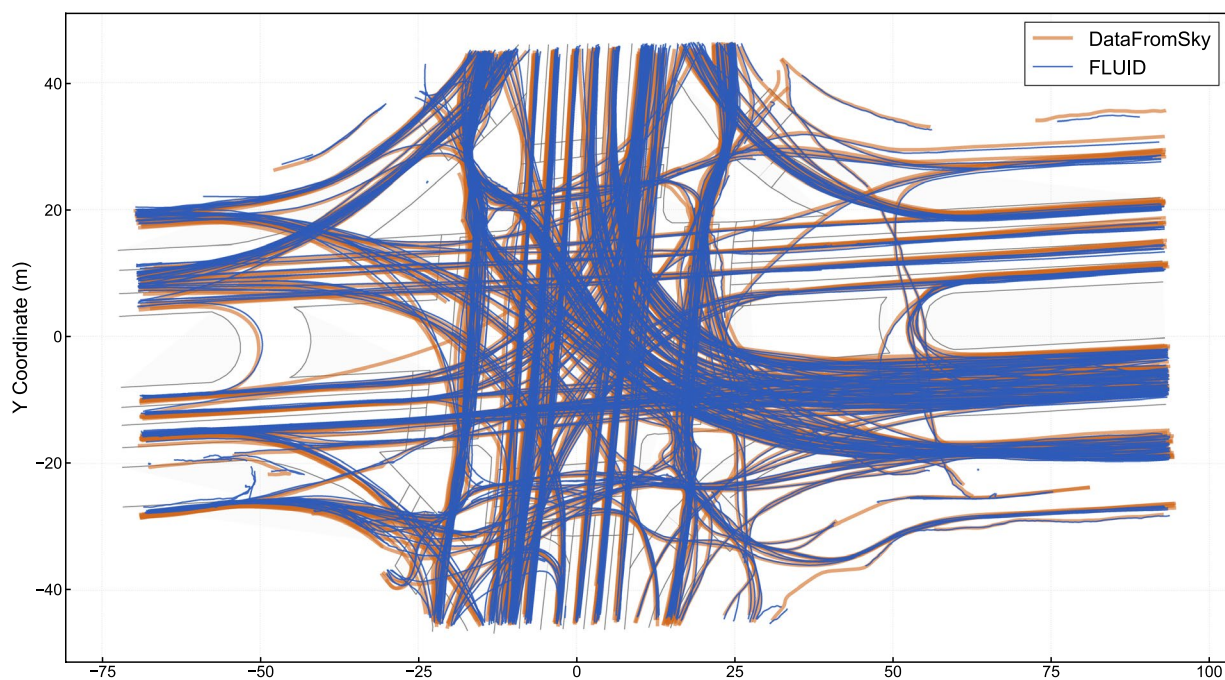
- **FI** (Four-way Intersection): This intersection is governed by a three-phase signal control. It is characterized by high volumes of MVs and VRUs, leading to direct conflicts. Furthermore, conflicts arise from the concurrent release of through and left-turning traffic in the east-west direction.
- **FIDRT** (Four-way Intersection with Dedicated Right-turn Lanes): Operating on a four-phase signal, this intersection has no direct internal conflict points but exhibits high traffic density. Conflicts are present externally at U-turn spots where paths cross with other traffic streams.
- **TI** (T-Intersection): This three-way intersection is managed by a two-phase signal. It serves as a valuable site for observing frequent conflicts between opposing through and left-turning vehicles, particularly in low-volume traffic scenarios.

The formats of files in the FLUID dataset are: MPEG-4 Part 14 (**MP4**), Comma-Separated Values (**CSV**), Tagged Image File Format (**TIFF**), and OpenStreetMap (**OSM**). Since there are too many fields, their meanings are explained in the Markdown document **README.md**. As the structure shown in Fig. 8, the following components are available:

- Privacy-preserved videos (*video*): Provided in MP4 format at 10 FPS.
- Signal timings (*signal*): Manually annotated signal control data in CSV format, sampled per second.
- Flight log during videos (*flightLog*): Drone flight posture recording in CSV format.
- Maps (*map*): Georeferenced TIFF images and vector maps largely compatible with the Lanelet2 (OSM) format<sup>51</sup>.
- Processed trajectories (*traj/route/conflict*): Offered in CSV format (the annotations for conflicts and violations belong to different files than the original tracks).



**Fig. 8** Structure of files of FLUID.



**Fig. 9** Position comparison (the coordinate of trajectories).

It is worth noting that FLUID offers not only a unified processing framework, but also high-quality trajectory extraction results, which act as a solid baseline and can be further enhanced by the research community.

### Technical Validation

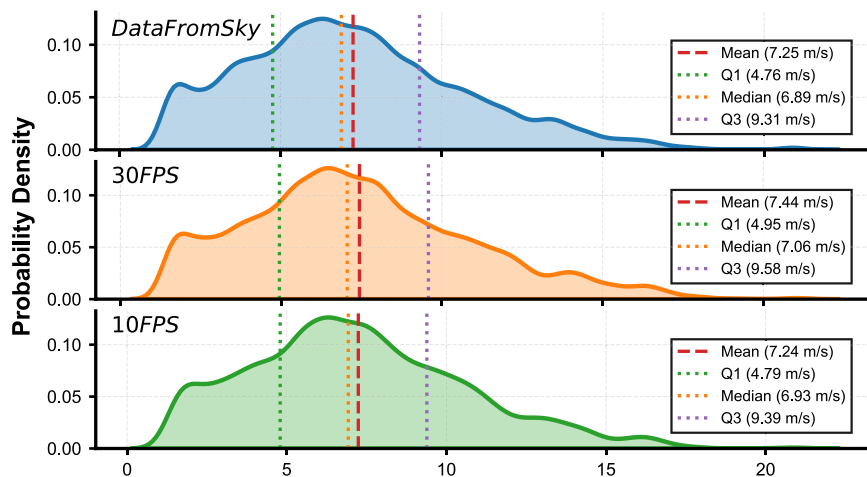
The validation of the FLUID dataset is **three-fold**. First, we assess the effectiveness of our data processing pipeline by benchmarking its results against the DataFromSky platform and ground truth data. Second, we establish the significance of our chosen scenes by contrasting their conflict profiles with those of other public datasets. Finally, we confirm the dataset's richness by demonstrating that each of the three scenes features a unique distribution of conflicts and violations, capturing a wide spectrum of behaviors.

**Trajectory Accuracy.** Previous datasets have rarely benchmarked their results against alternative processing methods or conducted systematic accuracy validation using supplementary data sources. Acknowledging that many prominent trajectory datasets—such as pNEUMA<sup>30</sup>, MAGIC<sup>52</sup>, and Mitra<sup>23</sup>—rely on the DataFromSky (DFS)<sup>33</sup> platform, we selected DFS as a benchmark to validate our FLUID framework's effectiveness. Furthermore, inspired by vehicle kinematics studies that often leverage ground-based positioning for validation<sup>53</sup>, we equipped a test vehicle with an RTS-GNSS device to establish a high-precision ground truth trajectory.

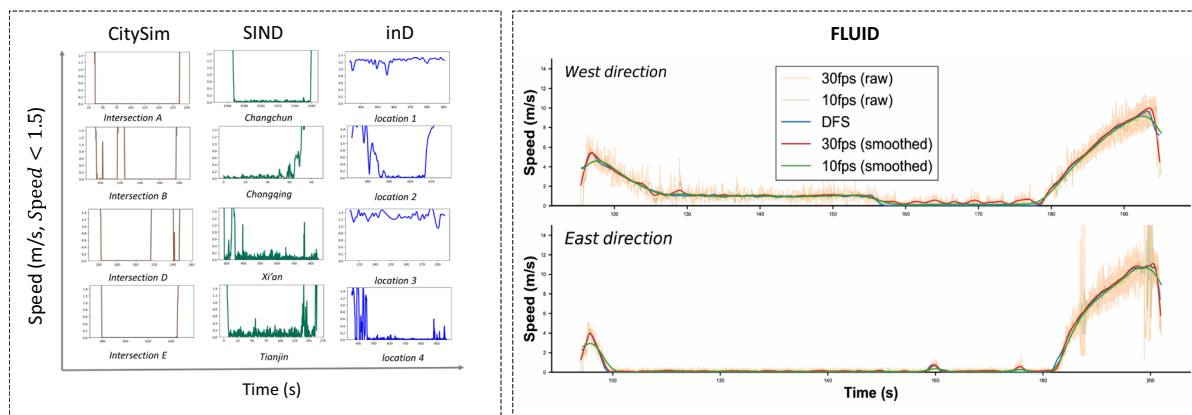
For this validation, we collected ground control points for georeferencing, as shown in Fig. 5. We utilized the 5-minute video analysis offered by the DFS free tier, processing the first five minutes of footage from the FIDRT scene recorded on May 26, 2025. This DFS-generated trajectory set serves as a baseline. We then compared it against the trajectories extracted by our FLUID framework from two video sources: the original  $\sim 30$  FPS footage and a downsampled 10 FPS version. Manual object counts were used as the ground-truth for quantity

Type	Source			Error (%)
	FLUID	DFS	Ground Truth	
MV	288	267	281	+2.5/−5.0
VRU	206	184	196	+5.1/−6.1
Total	494	451	477	+3.6/−5.4

**Table 2.** Comparison of MV and VRU counts across different sources.



**Fig. 10** Speed distribution of MVs of different methods.



**Fig. 11** Comparison of processed results of speed.

assessment. This comparative analysis focuses on three aspects: the accuracy of MV/VRU position and count, overall speed distribution, and individual speed profiles.

**Position and Count.** Figure 9 presents the aligned trajectories positions. For consistency, all object classes were grouped into MV and VRU categories. The spatial comparison reveals that FLUID-extracted trajectories closely correspond with the DFS output. This positional accuracy is further corroborated by the RTS-GNSS ground truth: the Hausdorff distance between the reference and extracted trajectories for the test vehicle is 0 ~ 0.97m, with errors under 0.3m on straight segments. Table 2 reveals object counts. DFS exhibited a 5 ~ 6% miss rate against the ground-truth, a consequence of its policy to discard stationary and short trajectories. While this may improve tracking precision, it sacrifices recall. In contrast, FLUID achieved near-zero missed detections and a low ID switch rate of 2 ~ 5%, demonstrating performance comparable or superior to the DFS benchmark in comprehensive object detection.

**Speed Distribution.** Figure 10 presents the overall MV speed distributions from DFS and FLUID. A notable finding is that the speed profile from the 10 FPS video appears more stable. We hypothesize that the lower frame rate mitigates detection jitter from bounding boxes, suggesting that processing at very high frame rates may, counterintuitively, complicate the velocity post-processing stage.

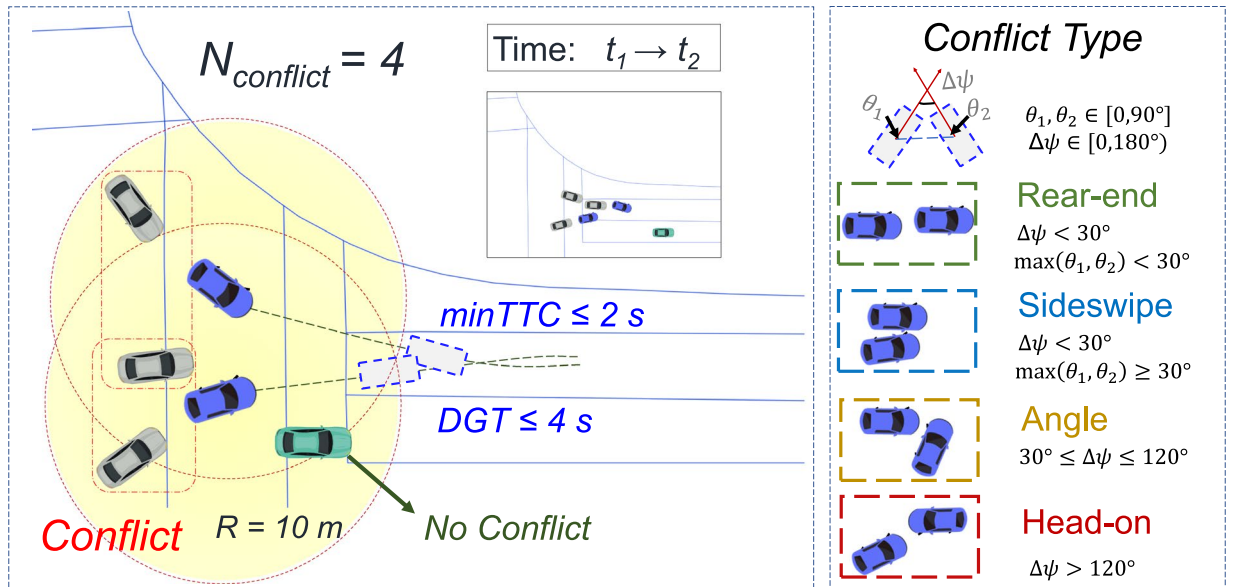


Fig. 12 Traffic conflicts identification and classification.

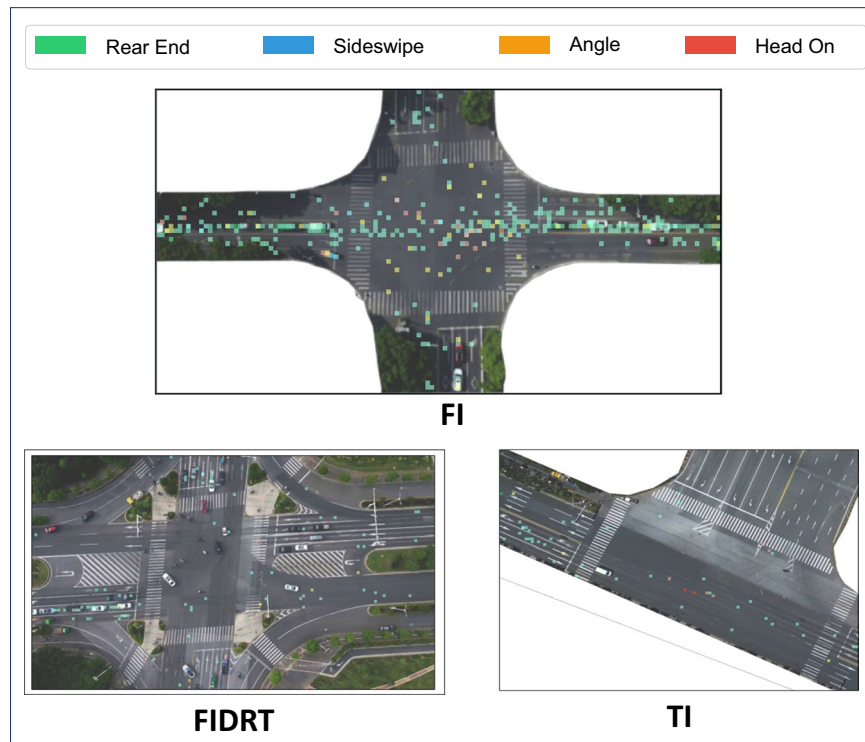
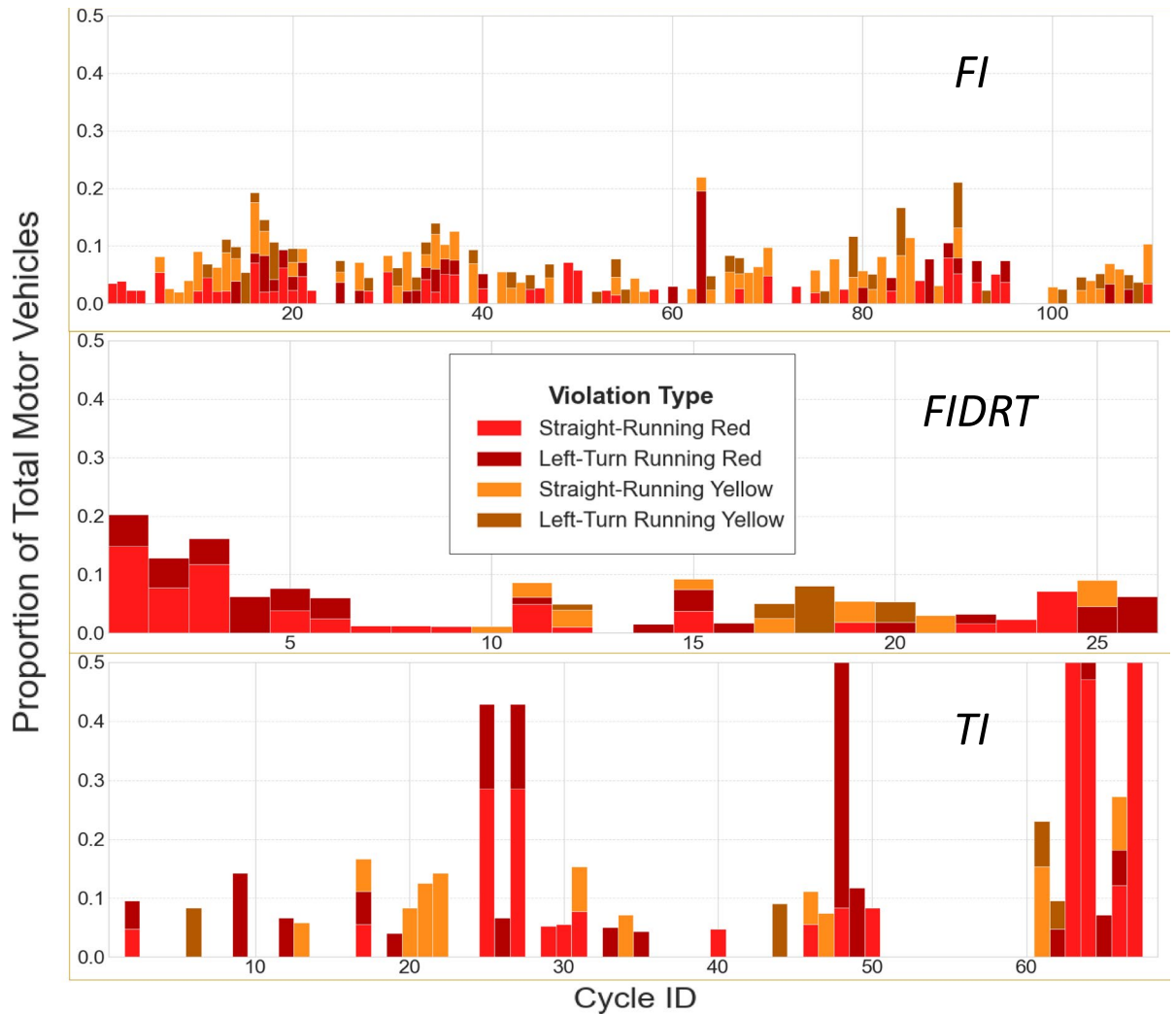


Fig. 13 Heatmaps of spatial density for different conflict types.

**Individual Speed Profiles.** This observation is reinforced by examining individual speed profiles (the two MVs with the longest trajectories in the FLUID scenario). In contrast, the left panel of Fig. 11 illustrates that existing datasets employ inconsistent smoothing strategies. For instance, CitySim thresholds all near-zero speeds to zero, whereas SIND and inD apply this only to persistently stationary objects, leaving transient stops untreated. The variation in smoothing granularities across scenes—as seen in datasets like SIND and inD—is substantial and could lead to analytical biases. Conversely, FLUID enhances transparency by providing both raw and smoothed data (Fig. 11, right). The analysis again demonstrates that the 10 FPS processing yields accurate, stable speed profiles that align well with the DFS results, thus validating our methodological choices.

**Scene Significance.** A central application of the FLUID dataset is traffic conflict analysis. Defined as hazardous traffic interactions<sup>54</sup>, conflicts serve as a proxy for collisions, and their link to accident risk can be quantified



**Fig. 14** Per-cycle violation rates for straight and left-turn movements, calculated as a proportion of total MVs.

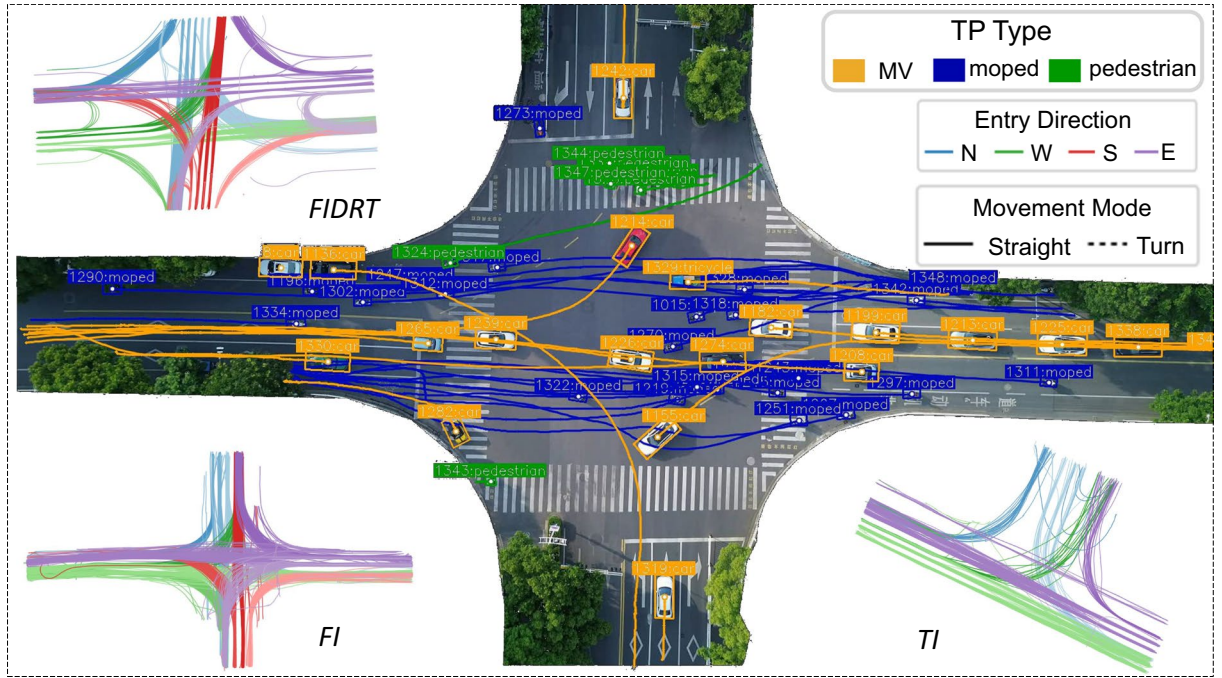
using Surrogate Safety Measures (SSMs)<sup>49</sup>. Due to the lack of standardized SSM thresholds across different scenarios, we developed a new conflict quantification process based on a comprehensive literature review. This process includes conflict extraction, classification, and the identification of associated TPs.

To maintain a consistent comparative basis, our analysis is confined to MVs. We selected Time-to-Collision (TTC) as the primary predictive SSM, given its robustness after standardizing the temporal resolution of velocity data across datasets. Potential conflicts are initially identified using the minimum TTC (minTTC) observed between any two trajectories<sup>49</sup>. Subsequently, to improve precision, we introduce our Dynamic Gap Time (DGT) method for post-validation, which effectively filters out kinematically plausible but physically impossible conflicts (e.g., those separated by infrastructure). Drawing from a review of established practices<sup>55</sup>, we define a conflict event using the thresholds:  $0s \leq DGT \leq 4.0s$ ,  $0s \leq TTC \leq 2.0s$ .

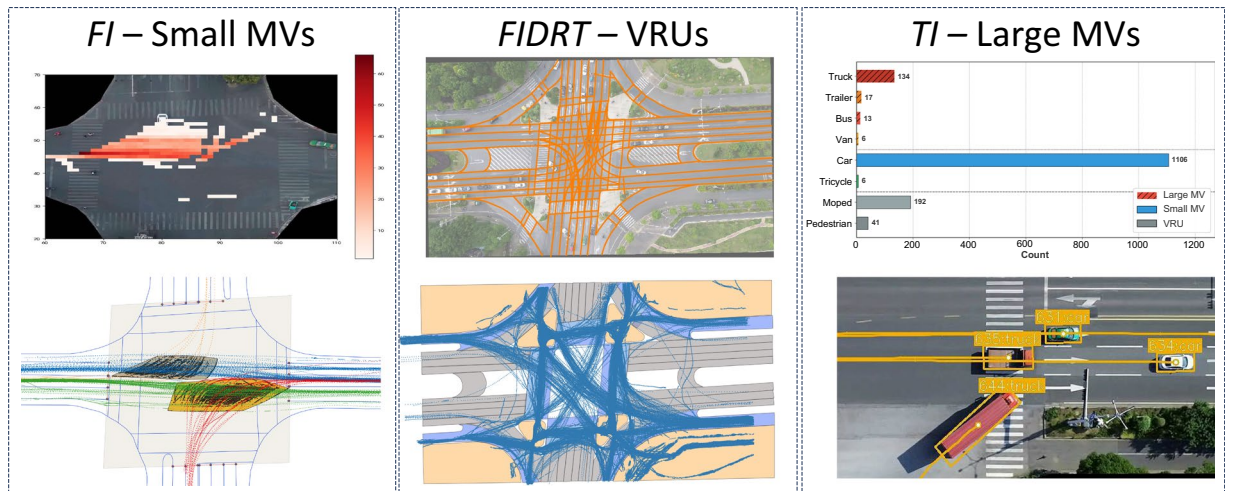
The conflict angle,  $\Delta\psi$ , is a critical determinant of potential accident severity. While recent research widely adopts angle-based classification for conflicts and collisions, existing methodologies often suffer from ambiguous criteria<sup>56,57</sup>, overly complex rules requiring bounding box overlap analysis<sup>58</sup>, or incomplete taxonomies<sup>59</sup>. To address these limitations, we adopt the comprehensive definition<sup>60</sup>, categorizing conflicts into four distinct types: rear-end, sideswipe, angle, and head-on, as depicted in Fig. 12.

Figure 12 illustrates the process of identifying associated conflicting objects. At instant  $t_1$ , the two blue vehicles constitute a primary conflict pair. All other vehicles within a 10-meter radius of this pair—the two gray vehicles—are designated as *associated objects*, which will subsequently engage in new conflicts at a later time ( $t_2$ ). The green vehicle, being outside this radius, is excluded. Associated objects can be repeatedly identified if they are proximate to multiple conflict TPs.

Table 1 shows that FLUID's conflict quantification demonstrates clear advantages over other datasets. A key differentiator is the traffic composition surrounding conflict events. Our analysis reveals that VRUs constitute 35.4% of all agents within a 10-meter radius of a conflict pair in FLUID. This proportion is substantially higher than that in SIND (7.2%), inD (23.7%), and INTERACTION (4.4%), indicating that FLUID provides a unique environment for studying MV interactions in the presence of VRUs, which impacts driver decision-making.



**Fig. 15** Trajectory visualization and turn annotation result.



**Fig. 16** Unique application prospects of FLUID.

**Behavioral Richness.** In FLUID, conflict and violation annotations exhibit rich spatiotemporal behavioral characteristics. The  $1m \times 1m$  grid is used for discrete division, enabling finer conflict identification. Subsequently, the previously classified traffic types are clustered according to the grid. Figure 13 showed that different scenarios exhibit diverse traffic conflicts type-density distributions. Figure 14 shows that the violation rates for each signal cycle (non-consecutive videos, 110 for FI, 26 for FIDRT, and 68 for TI) vary. The calculated violation rates may be higher during cycles with lower traffic volume.

**Usage Notes**

Benefiting from the fine-grained details of our dataset, we can categorize distinct traffic behavior patterns through turn labels, as illustrated in Fig. 15. Beyond basic traffic flow analysis, the high precision of FLUID facilitates multi-domain research, including human preference mining, traffic behavior modeling, and autonomous driving. Figure 16 showcases three representative cases that highlight the dataset’s unique potential in these specialized research contexts.

**Analysis of Passing and Yielding Behaviors.** Leveraging the standardized geometric layout, high density of passenger vehicles, and integrated traffic signal data in the FI scenario, FLUID provides a robust foundation

for analyzing complex interaction behaviors. By calculating the convex hull of conflicting trajectories, we can accurately delineate conflict zones and analyze the temporal sequences of vehicles entering and exiting both these zones and the intersection boundaries. This methodology enables a refined quantification of passing and yielding dynamics for specific maneuvers. For instance, among the through-left interaction events identified during green phases across the FI dataset, we recorded 502 instances of left-turn yielding, 415 instances of through-moving yielding, and 5 instances where no clear yielding behavior was discernible (Fig. 16, Left).

**Spatiotemporal Violation Analysis of VRUs.** While previous research<sup>61</sup> (conducted on a scenario similar to FIDRT) was often constrained by limited tracking precision and necessitated grid-based spatial partitioning, the FLUID dataset provides individual-level trajectories of VRUs integrated with precise Lanelet2 semantic maps. Moreover, the significantly higher VRU arrival rate in FLUID, compared to existing benchmarks, allows for a more granular analysis of spatial occupancy. These features facilitate a deeper understanding of VRUs' spatiotemporal violation intentions and movement patterns (Fig. 16, Middle).

**Optimization of Large Vehicle Detection.** In the TI scenario, large vehicles constitute over 15% of the traffic, where their diverse dimensions present substantial challenges for accurate detection. By providing both raw data and bounding box labels, we offer significant optimization potential for large-scale target detection in complex environments. Ultimately, this provides a contribution of similar value to pNEUMA Vision<sup>62</sup> for multi-object tracking and detection research (Fig. 16, Right).

In addition to these unique capabilities, our dataset also serves as a high-quality resource for broader applications in traffic engineering, such as:

- *Driving Decision Modeling:* Analyze the impact of interconnected information on driving decisions and traffic operations in interactive dilemmas involving two vehicles encountering conflicting directions at an intersection<sup>63</sup>.
- *Intersection Operation Quantification:* Evaluate the safety and efficiency characteristics of different intersection control strategies<sup>26</sup>.
- *Conflict Correlation Analysis:* Quantify conflict severity using SSMs, and analyze the relationship between this severity and other kinematic parameters<sup>64</sup>.
- *Trajectory Generation:* Learn from human mobility patterns to generate human-like and socially-inspired behaviors and movements<sup>65</sup>.

## Data availability

The full dataset is available at figshare<sup>50</sup>. The DOI to data record is <https://doi.org/10.6084/m9.figshare.29974954>.

## Code availability

Regarding raw data processing, the object detection and tracking code itself is based on the open-source YOLOv8 and SparseTrack. Therefore, the entire code is not repeated. The methods for processing, verification and visualization of the FLUID dataset is publicly available on GitHub: <https://github.com/sysu19351014/FLUID>.

Received: 1 September 2025; Accepted: 19 March 2026;

Published online: 30 March 2026

## References

1. Jiang, X. *et al.* A naturalistic trajectory dataset with dense interaction for autonomous driving. *Scientific Data* **12**, 1084, <https://doi.org/10.1038/s41597-025-05344-7> (2025).
2. Chen, D., Zhu, M., Yang, H., Wang, X. & Wang, Y. Data-driven traffic simulation: A comprehensive review. *IEEE Transactions on Intelligent Vehicles* **9**, 4730–4748, <https://doi.org/10.1109/tiv.2024.3367919> (2024).
3. Rowan, D., He, H., Hui, F., Yasir, A. & Mohammed, Q. A systematic review of machine learning-based microscopic traffic flow models and simulations. *Communications in Transportation Research* **5**, 100164, <https://doi.org/10.1016/j.commtr.2025.100164> (2025).
4. Choi, S., Jin, Z., Ham, S. W., Kim, J. & Sun, L. A gentle introduction and tutorial on deep generative models in transportation research. *Transportation Research Part C: Emerging Technologies* **176**, 105145, <https://doi.org/10.1016/j.trc.2025.105145> (2025).
5. Gore, N., Chauhan, R., Easa, S. & Arkatkar, S. Traffic conflict assessment using macroscopic traffic flow variables: A novel framework for real-time applications. *Accident Analysis & Prevention* **185**, 107020, <https://doi.org/10.1016/j.aap.2023.107020> (2023).
6. Sarkar, M., Kweon, O., Kim, B.-I., Choi, D. G. & Kim, D. Y. Synergizing autonomous and traditional vehicles: a systematic review of advances and challenges in traffic flow management with signalized intersections. *IEEE Transactions on Intelligent Transportation Systems* <https://doi.org/10.1109/tits.2024.3450471> (2024).
7. Wang, Y. *et al.* City-scale holographic traffic flow data based on vehicular trajectory resampling. *Scientific Data* **10**, 57, <https://doi.org/10.1038/s41597-022-01850-0> (2023).
8. Geiger, A., Lenz, P., Stiller, C. & Urtasun, R. Vision meets robotics: The kitti dataset. *The international journal of robotics research* **32**, 1231–1237, <https://doi.org/10.1177/0278364913491297> (2013).
9. Huang, X. *et al.* The apollo scape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 954–960, <https://doi.org/10.1109/CVPRW.2018.00141> (2018).
10. Sun, P. *et al.* Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2446–2454, <https://doi.org/10.1109/CVPR42600.2020.00252> (2020).
11. Karnchanachari, N. *et al.* Towards learning-based planning: The nuPlan benchmark for real-world autonomous driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 629–636, <https://doi.org/10.1109/ICRA57147.2024.10610077> (IEEE, 2024).
12. Wang, J. *et al.* Realtime wide-area vehicle trajectory tracking using millimeter-wave radar sensors and the open tjrd ts dataset. *International Journal of Transportation Science and Technology* **12**, 273–290, <https://doi.org/10.1016/j.ijst.2022.02.006> (2023).

13. Punzo, V., Borzacchiello, M. T. & Ciuffo, B. On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (ngsim) program data. *Transportation Research Part C: Emerging Technologies* **19**, 1243–1262, <https://doi.org/10.1016/j.trc.2010.12.007> (2011).
14. Seo, T. *et al.* Evaluation of large-scale complete vehicle trajectories dataset on two kilometers highway segment for one hour duration: Zen traffic data. In *2020 International Symposium on Transportation Data and Modelling*. <https://limos.engin.umich.edu/istdm2021/wp-content/uploads/sites/2/2021/05/ISTDM-2021-Extended-Abstract-0019.pdf> (2020).
15. Gloudemans, D. *et al.* I-24 motion: An instrument for freeway traffic science. *Transportation Research Part C: Emerging Technologies* **155**, 104311, <https://doi.org/10.1016/j.trc.2023.104311> (2023).
16. Wang, J., Fu, T. & Shangquan, Q. Wide-area vehicle trajectory data based on advanced tracking and trajectory splicing technologies: Potentials in transportation research. *Accident Analysis & Prevention* **186**, 107044, <https://doi.org/10.1016/j.aap.2023.107044> (2023).
17. Schickanz, C. *et al.* The dlr urban traffic dataset (dlr-ut): A comprehensive traffic dataset from an urban research intersection. In *2025 IEEE Intelligent Vehicles Symposium (IV)*, 398–405, <https://doi.org/10.1109/IV64158.2025.11097400> (IEEE, 2025).
18. Coifman, B. & Li, L. A critical evaluation of the next generation simulation (ngsim) vehicle trajectory dataset. *Transportation Research Part B: Methodological* **105**, 362–377, <https://doi.org/10.1016/j.trb.2017.09.018> (2017).
19. Wang, Z. *et al.* A comprehensive review of drone-based autonomous driving datasets: Methodology, taxonomy, and prospects. *IEEE Transactions on Intelligent Transportation Systems*, <https://doi.org/10.1109/tits.2025.3571726> (2025).
20. 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. In *2018 21st international conference on intelligent transportation systems (ITSC)*, 2118–2125, <https://doi.org/10.1109/ITSC.2018.8569552> (IEEE, 2018).
21. Moers, T. *et al.* The exid dataset: A real-world trajectory dataset of highly interactive highway scenarios in germany. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, 958–964, <https://doi.org/10.1109/IV51971.2022.9827305> (IEEE, 2022).
22. Xu, J., Pan, C., Dai, Z. & Zhang, H. Cqskyeyex: a drone dataset of vehicle trajectory on chinese expressways. In *International Conference on Green Intelligent Transportation System and Safety*, 463–479, [https://doi.org/10.1007/978-981-97-3052-0\\_33](https://doi.org/10.1007/978-981-97-3052-0_33) (Springer, 2022).
23. Chaudhari, A. A., Treiber, M. & Okhrin, O. Mitra: A drone-based trajectory data for an all-traffic-state inclusive freeway with ramps. *Scientific Data* **12**, 1174, <https://doi.org/10.1038/s41597-025-05472-0> (2025).
24. Bock, J. *et al.* The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, 1929–1934, <https://doi.org/10.1109/IV47402.2020.9304839> (IEEE, 2020).
25. Xu, Y. *et al.* Sind: A drone dataset at signalized intersection in china. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2471–2478, <https://doi.org/10.1109/ITSC55140.2022.9921959> (IEEE, 2022).
26. Pu, Q. *et al.* Drone data analytics for measuring traffic metrics at intersections in high-density areas. *Transportation Research Record* **0**, <https://doi.org/10.1177/03611981241311566> (2025).
27. Fonod, R., Cho, H., Yeo, H. & Geroliminis, N. Songdo traffic: High accuracy georeferenced vehicle trajectories from a large-scale study in a smart city <https://doi.org/10.5281/zenodo.13828383> (2025).
28. Krajewski, R., Moers, T., Bock, J., Vater, L. & Eckstein, L. The round dataset: A drone dataset of road user trajectories at roundabouts in germany. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6, <https://doi.org/10.1109/ITSC45102.2020.9294728> (IEEE, 2020).
29. Zhan, W. *et al.* Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*, <https://doi.org/10.48550/arXiv.1910.03088> (2019).
30. Barmounakis, E. & Geroliminis, N. On the new era of urban traffic monitoring with massive drone data: The pneuma large-scale field experiment. *Transportation research part C: emerging technologies* **111**, 50–71, <https://doi.org/10.1016/j.trc.2019.11.023> (2020).
31. Zheng, O. *et al.* Citysim: A drone-based vehicle trajectory dataset for safety-oriented research and digital twins. *Transportation research record* **2678**, 606–621, <https://doi.org/10.1177/03611981231185768> (2024).
32. Chandler, B. E. *et al.* Signalized intersections informational guide. Tech. Rep., United States. Department of Transportation. Federal Highway Administration. [https://rosap.ntl.bts.gov/view/dot/42546/dot\\_42546\\_DS1.pdf](https://rosap.ntl.bts.gov/view/dot/42546/dot_42546_DS1.pdf) (2013).
33. DataFromSky. TrafficSurvey. <https://datafromsky.com/trafficsurvey/>.
34. GoodVision. Traffic data collection. <http://goodvisionlive.com/solutions/traffic-data-collection/>.
35. Fonod, R., Cho, H., Yeo, H. & Geroliminis, N. Advanced computer vision for extracting georeferenced vehicle trajectories from drone imagery. *Transportation Research Part C: Emerging Technologies* **178**, 105205, <https://doi.org/10.1016/j.trc.2025.105205> (2025).
36. Pablo Alcantarilla (Georgia Institute of Technology), A. B., Jesus Nuevo (TrueVision Solutions AU). Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *Proceedings of the British Machine Vision Conference*. <https://doi.org/10.5244/C.27.13> (BMVA Press, 2013).
37. Leutenegger, S., Chli, M. & Siegwart, R. Y. Brisk: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*, 2548–2555, <https://doi.org/10.1109/ICCV.2011.6126542> (Ieee, 2011).
38. Tareen, S. A. K. & Saleem, Z. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In *2018 International conference on computing, mathematics and engineering technologies (iCoMET)*, 1–10, <https://doi.org/10.1109/ICOMET.2018.8346440> (IEEE, 2018).
39. Jocher, G., Qiu, J. & Chaurasia, A. Ultralytics yolo. <https://ultralytics.com> (2023).
40. Sun, Y., Cao, B., Zhu, P. & Hu, Q. Drone-based rgb-infrared cross-modality vehicle detection via uncertainty-aware learning. *IEEE Transactions on Circuits and Systems for Video Technology* **32**, 6700–6713, <https://doi.org/10.1109/tcsvt.2022.3168279> (2022).
41. Hellekes, J., Mühlhaus, M., Bahmanyar, R., Azimi, S. M. & Kurz, F. Vetra: A dataset for vehicle tracking in aerial imagery—new challenges for multi-object tracking. In *European Conference on Computer Vision*, 52–70, [https://doi.org/10.1007/978-3-031-73013-9\\_4](https://doi.org/10.1007/978-3-031-73013-9_4) (Springer, 2024).
42. Ye, K. *et al.* More clear, more flexible, more precise: A comprehensive oriented object detection benchmark for uav. *arXiv preprint arXiv:2504.20032* <https://doi.org/10.48550/arXiv.2504.20032> (2025).
43. Zou, Z., Chen, K., Shi, Z., Guo, Y. & Ye, J. Object detection in 20 years: A survey. *Proceedings of the IEEE* **111**, 257–276, <https://doi.org/10.1109/JPROC.2023.3238524> (2023).
44. Li, X. *et al.* Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Advances in neural information processing systems* **33**, 21002–21012 (2020).
45. Liu, Z., Wang, X., Wang, C., Liu, W. & Bai, X. Sparsetrack: Multi-object tracking by performing scene decomposition based on pseudo-depth. *IEEE Transactions on Circuits and Systems for Video Technology* <https://doi.org/10.1109/tcsvt.2024.3524670> (2025).
46. Zhang, Y. *et al.* Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*, 1–21, [https://doi.org/10.1007/978-3-031-20047-2\\_1](https://doi.org/10.1007/978-3-031-20047-2_1) (Springer, 2022).
47. Arun, A., Haque, M. M., Washington, S., Sayed, T. & Mannering, F. A systematic review of traffic conflict-based safety measures with a focus on application context. *Analytic methods in accident research* **32**, 100185, <https://doi.org/10.1016/j.amar.2021.100185> (2021).
48. Jiao, Y. A fast calculation of two-dimensional Time-to-Collision. <https://github.com/Yiru-Jiao/Two-Dimensional-Time-To-Collision> (2023).
49. Wang, C., Xie, Y., Huang, H. & Liu, P. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis & Prevention* **157**, 106157, <https://doi.org/10.1016/j.aap.2021.106157> (2021).
50. Chen, Y. *et al.* FLUID: A Fine-Grained Lightweight Urban Signalized-Intersection Dataset of Dense Conflict Trajectories <https://doi.org/10.6084/m9.figshare.29974954.v1> (2025).
51. Poggenhans, F. *et al.* Lanelet2: A high-definition map framework for the future of automated driving. In *2018 21st international conference on intelligent transportation systems (ITSC)*, 1672–1679, <https://doi.org/10.1109/ITSC.2018.8569929> (IEEE, 2018).

52. Ma, W., Zhong, H., Wang, L., Jiang, L. & Abdel-Aty, M. Magic dataset: Multiple conditions unmanned aerial vehicle group-based high-fidelity comprehensive vehicle trajectory dataset. *Transportation research record* **2676**, 793–805, <https://doi.org/10.1177/03611981211070549> (2022).
53. Mi, T., Takács, D., Liu, H. & Orosz, G. Capturing the true bounding boxes: vehicle kinematic data extraction using unmanned aerial vehicles. *Journal of intelligent transportation systems* 1–13, <https://doi.org/10.1080/15472450.2024.2341395> (2024).
54. Klebelsberg, D. Derzeitiger sand der verhaltensanalyse des kraftfahrens. zrbait und leitung. *Abtl. Arbeitswissenschaft soziale betriebspraxis* **18**, 33–37 (1964).
55. Wang, X. *et al.* Conflict extraction and characteristics analysis at signalized intersections using trajectory data. *International Journal of Transportation Science and Technology* <https://doi.org/10.2139/ssrn.4518329> (2025).
56. Park, N., Park, J., Joo, Y.-J. & Abdel-Aty, M. Micro-level hotspot identification at intersections using traffic conflict analysis. *Accident Analysis & Prevention* **220**, 108167, <https://doi.org/10.2139/ssrn.5217411> (2025).
57. Anisha, A. M., Abdel-Aty, M., Abdelraouf, A., Islam, Z. & Zheng, O. Automated vehicle to vehicle conflict analysis at signalized intersections by camera and lidar sensor fusion. *Transportation research record* **2677**, 117–132, <https://doi.org/10.1177/03611981221128806> (2023).
58. Wu, Y., Abdel-Aty, M., Zheng, O., Cai, Q. & Zhang, S. Automated safety diagnosis based on unmanned aerial vehicle video and deep learning algorithm. *Transportation research record* **2674**, 350–359, <https://doi.org/10.1177/0361198120925808> (2020).
59. Zhang, S. & Sze, N. Real-time conflict risk at signalized intersection using drone video: A random parameters logit model with heterogeneity in means and variances. *Accident Analysis & Prevention* **207**, 107739, <https://doi.org/10.1016/j.aap.2024.107739> (2024).
60. Feng, S. *et al.* Dense reinforcement learning for safety validation of autonomous vehicles. *Nature* **615**, 620–627, <https://doi.org/10.1038/s41586-023-05732-2> (2023).
61. Xu, H., He, Z., Chen, Y., Wu, Z. & Zhu, Y. Behavior recognition of non-motorized transport at intersections using dual-channel grid model based on disordered trajectory point data. *Physica A: Statistical Mechanics and its Applications* **650**, 129994, <https://doi.org/10.1016/j.physa.2024.129994> (2024).
62. Kim, S., Anagnostopoulos, G., Barmponakis, E. & Geroliminis, N. Visual extensions and anomaly detection in the pneumonia experiment with a swarm of drones. *Transportation Research Part C: Emerging Technologies* **147**, 103966, <https://doi.org/10.1016/j.trc.2022.103966> (2023).
63. Yang, M. *et al.* Does information provision always enable drivers to make better decisions?—a study on decision-making dilemmas at uncontrolled intersections. *Transportation Research Part F: Traffic Psychology and Behaviour* **109**, 320–335, <https://doi.org/10.1016/j.trf.2024.12.018> (2025).
64. Shen, J. & Jin, W. Analysis of the factors influencing the severity of straight-left conflicts under left turn permit phase. In *International Conference on Smart Transportation and City Engineering (STCE 2023)*, vol. 13018, 80–87, <https://doi.org/10.1117/12.3024121> (SPIE, 2024).
65. Wang, S., Ni, Y., Miao, C., Sun, J. & Sun, J. A multiagent social interaction model for autonomous vehicle testing. *Communications in Transportation Research* **5**, 100183, <https://doi.org/10.1016/j.commtr.2025.100183> (2025).

## Acknowledgements

We are grateful to the Joint Research and Development Laboratory of Smart Policing in Xuancheng Public Security (a joint initiative between Xuancheng Public Security Bureau and Sun Yat-sen University) for their essential contribution to this work, specifically in facilitating the selection of the investigation's spatiotemporal scope and data acquisition. We thank Junfeng Li and Kaiyuan Yang for their assistance in vector map processing, as well as the anonymous undergraduate students from the School of Intelligent System Engineering at SYSU for their help with data integration. The research was also supported by the National Key Research and Development Program of China (No. 2023YFB4301900), the Shenzhen Science and Technology Program (No. JCYJ20240813151445059), the National Natural Science Foundation of China (No. U21B2090), and the Science and Technology Planning Project of Guangdong Province (No. 2023B1212060029).

## Author contributions

Conceptualization and Methodology: Y.C., Z.W., G.Z., H.Z.; Investigation: Y.C., Z.W., L.X.; Data Curation and Formal Analysis: Y.C., G.Z.; Validation and Visualization: Y.C., L.X., H.T., Z.W.; Writing - Original Draft Preparation: Y.C.; Writing - Review & Editing: Z.W., X.W., Z.H.; Project Administration and Funding Acquisition: H.Z., Z.H. All authors have reviewed and agreed to the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Z.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026