data

MDPI

Data Descriptor

SIT-PET: Long-Term Multimodal Traffic Trajectory Data with PET-Based Interaction Events at a Signalized Intersection

Markus Steinmaßl *, Karl Rehrl and Timo Vornberger

Salzburg Research Forschungsgesellschaft m.b.H., Jakob Haringer Straße 5/3, 5020 Salzburg, Austria
* Correspondence: markus.steinmassl@salzburgresearch.at

Abstract

In this paper, we present a curated dataset derived from continuous multi-object tracking observations over a two-year period from a signalized urban intersection in Salzburg, Austria. The dataset includes time-resolved trajectories of multimodal road users, post-processed object attributes, movement relations, and Post-Encroachment Time values computed for a fixed set of eight predefined multimodal traffic conflict scenarios. Moreover, traffic signal data are included and can be used as contextual information. A temporal six-month subset is published via Zenodo including usage examples written in python. The full dataset can be provided on request. Potential applications include traffic safety analysis, behavioral modeling, method development for interaction detection, and educational use in data-driven traffic research.

Dataset: 10.5281/zenodo.18234283.

Dataset License: CC BY 4.0

Keywords: intersection analysis; Post-Encroachment Time (PET); road user interactions; signalized intersection; surrogate safety measures; traffic safety; traffic signal states; traffic trajectories; urban traffic

1. Summary

High-resolution trajectory data derived from video- or LiDAR-based multi-object tracking (MOT) systems enable detailed analyses of traffic behavior and safety [1]. In particular, surrogate safety measures (SSMs) such as the Post-Encroachment Time (PET) [2] provide a means to study conflict situations without relying on rare and often incomplete crash records [3]. Despite their relevance, publicly available datasets that combine long-term MOT observations with PET-based interactions and rich contextual information remain scarce.

This data paper describes a curated dataset derived from continuous MOT observations at a signalized urban intersection over a multi-year period. The dataset comprises time-resolved trajectories of individual road users, post-processed object attributes and movement relations, PET values computed for a fixed set of eight predefined conflict scenarios, and traffic signal data as contextual information.

The data were collected during a four-year research initiative that aimed at the development of techniques and methods that practitioners could apply in order to utilize MOT data for traffic analysis. A small four-hour dataset [4] from a different intersection, acquired with the same MOT system, has been published and used for research on data

Check for updates

Academic Editors: Diego Gabriel Rossit and Pedro Moreno-Bernal

Received: 23 February 2026
Revised: 17 March 2026
Accepted: 18 March 2026
Published: 25 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

quality evaluation techniques [5]. A one-week subset of the present dataset has been used for research on the reliability of surrogate safety assessment with a special focus on the influence of data quality [6]. Finally, the present multi-year dataset was curated for research on factors affecting traffic safety by means of PET. The outcomes of this research will be published in a forthcoming master's thesis.

The dataset is published via Zenodo, an open-access research data repository operated by CERN that provides persistent DOIs for research outputs, to ensure long-term availability, versioning, and citability. The dataset is accompanied by this data paper. By releasing the data in a structured and well-documented form, this work aims to facilitate reuse for a broad range of applications, including traffic safety research, behavioral modeling, and method development for surrogate safety measures and interaction detection algorithms. Beyond its original use case, the dataset may also serve educational purposes and support reproducible research in the growing field of data-driven traffic analysis.

2. Data Description

This section describes the structure, content, and organization of the dataset of which a temporal subset is published via Zenodo. The full dataset can be provided on request. The dataset is derived from MOT observations at a single signalized urban intersection and focuses on traffic interactions characterized by the PET.

2.1. Overview of the Dataset

The dataset consists of multiple interrelated data tables that represent different stages of processing, from raw object trajectories to PET-based interaction events. Traffic light signal states at the intersection are included as contextual information, along with auxiliary information and figures to make the data generation process comprehensible and reproducible. Data tables are provided as CSV files compressed with gzip. Missing values are represented as empty entries. The main data products are:

- Tracks: Time-resolved trajectories of individual road users detected by the MOT system;
- Scenarios: Pre-defined interaction scenarios for which PET is calculated;
- Objects: Post-processed object-level attributes and movement relations of scenario-relevant objects;
- PET events: All computed PET values for predefined conflict scenarios;
- Signal states: Traffic light signal states at the intersection.

The dataset was collected at an urban, signalized intersection in Salzburg, Austria, located at 13.0640°, 47.8096°, and spans the period from 1 August 2023 to 31 August 2025. The data description in the following sections is based on the whole period, while only a temporal subset has been published on Zenodo, covering 1 January 2025 to 30 June 2025.

2.2. Tracks

Tracks constitute the most fine-grained data level and represent trajectories of individual traffic participants observed within the intersection area. Each track corresponds to a single object and consists of time-stamped positional and kinematic attributes. Tracks are stored as daily files, each containing all trajectories whose first timestamp occurs on the respective day. Each record includes:

- A unique object identifier;
- A timestamp;
- Spatial coordinates in WGS84;
- Object class, dimensions, orientation, and speed;
- Detection quality indicators.
- The column description of the *tracks* table is provided in Table 1.

Table 1. Tracks: column description.

Column	Type	Description
object_id	int	Identifier of the tracked object
timestamp_ms	int	Timestamp in milliseconds since Unix epoch (UTC)
det_points_count	int	Number of LiDAR points associated with the object (detection quality indicator)
lon_deg	float	Object position—Longitude (WGS84, degrees)
lat_deg	float	Object position—Latitude (WGS84, degrees)
heading_deg	float	Object heading (degrees clockwise from North)
speed_ms	float	Object speed (meters per second)
length_m	float	Object length (meters)
width_m	float	Object width (meters)
height_m	float	Object height (meters)
tracking_status	string	Object tracking status reported by the MOT system: VALIDATING = Checking validity in the early stage of tracking; INVALIDATING = Short term prediction when tracking is lost in VALIDATING status; TRACKING = Stable tracking (recommended value to use for tracking, ignore the rest); DRIFTING = Short term prediction when tracking is lost in TRACKING status
object_class	string	Object class as reported by the MOT system (vehicle, two-wheeler, pedestrian, misc)

Tracks have been filtered on those trajectories that intersect the central area of the intersection, depicted in red in Figure 1. The temporal resolution and spatial accuracy are determined by the underlying MOT system which is described in Section 3. Due to the real-time nature of the tracking system, object dimensions and class labels may vary over the lifetime of a track.

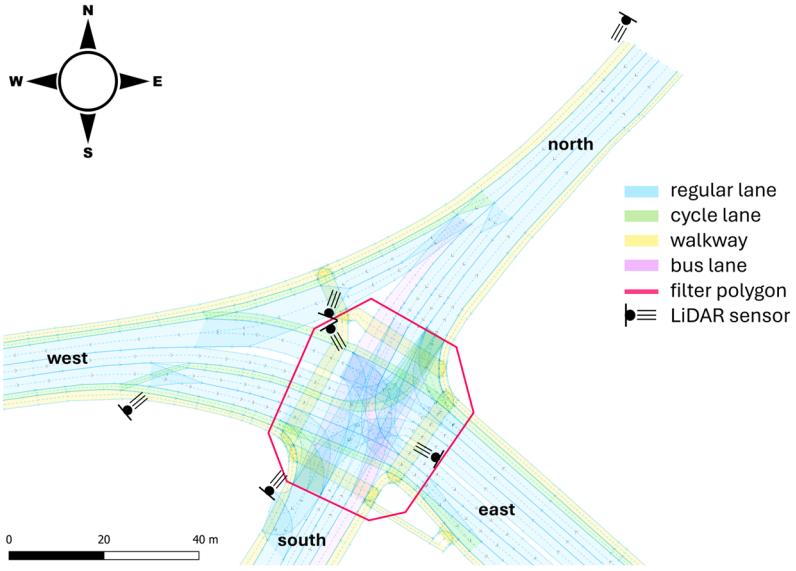


Figure 1. Intersection layout.

2.3. Scenarios

The *scenarios* table describes a fixed set of eight predefined interaction scenarios for which PET-based safety assessment is meaningful. Each scenario defines a pair of compatible movement relations and an implicit priority structure between the involved road users relevant to PET calculation. Conflict scenarios serve as the conceptual basis for PET computation and for structuring interaction-level analyses. Each PET event record is associated with exactly one scenario. The table's column description is provided in Table 2.

Table 2. Scenarios—column description.

Column	Type	Description
scenario_id	string	Unique identifier of the PET scenario
encroaching_movement_category	string	Movement category of the encroaching road user
encroaching_conflict_leg	string	Logical intersection leg associated with the encroaching movement
encroaching_turning_movement	string	Turning movement of the encroaching road user (empty if not applicable)
priority_movement_category	string	Movement category of the priority road user
priority_conflict_leg	string	Logical intersection leg associated with the priority movement
priority_turning_movement	string	Turning movement of the priority road user (empty if not applicable)
description	string	Human-readable description of the scenario

2.4. Objects

To support conflict-based analyses, the subset of scenario-relevant objects is post-processed. For these objects, dynamic attributes are consolidated into static estimates (e.g., object length, width, and height), and a stable object class is assigned. Each object is associated with a movement relation describing its dominant path through the intersection. Movement relations are defined based on spatial geofencing logic and object trajectories and are limited to those required for the predefined set of conflict scenarios. Objects whose movements do not match any of the defined relations are excluded from this data product. The *objects* table's column description is provided in Table 3.

Table 3. Objects—column description.

Column	Type	Description
object_id	int	Object identifier (references tracks.object_id)
object_day	string	Calendar day on which the object was first observed, expressed as an ISO 8601 date (YYYY-MM-DD, UTC)
object_class_static	string	Stabilized object class
length_m_static	float	Static object length (meters)
width_m_static	float	Static object width (meters)
height_m_static	float	Static object height (meters)
movement_category	string	High-level category describing the type of movement performed by the road user at the intersection
conflict_leg	string	Logical intersection leg associated with the movement, defined according to the intersection layout
turning_movement	string	Turning movement performed by the road user at the intersection (empty if not applicable)

2.5. PET Events

The *pet_events* table contains all computed Post-Encroachment Time values for the defined conflict scenarios. For each potential interaction, the dataset records:

- Identifiers of the encroaching and priority objects;
- Entry and exit times of both objects into the conflict area;
- The resulting PET value;
- The spatial location of the conflict point;
- The associated conflict scenario.

PET values are computed directly from the observed entry and exit times without aggressive post-filtering. As a result, the table may include large PET values corresponding to temporally distant interactions as well as negative PET values. These characteristics are intentional and allow users to apply their own filtering strategies depending on the

application. A detailed description of the calculation method and further explanation can be found in Section 3. The *pet_events* table's column description is provided in Table 4.

Table 4. PET events—column description.

Column	Type	Description
event_id	int	Unique event identifier
scenario_id	string	Interaction scenario identifier
encroaching_object_id	int	Object of the encroaching movement relation
priority_object_id	int	Object of the priority movement relation
ts_enter_encroaching_ms	int	Entry time of encroaching object (ms since epoch)
ts_leave_encroaching_ms	int	Exit time of encroaching object (ms since epoch)
ts_enter_priority_ms	int	Entry time of priority object (ms since epoch)
ts_leave_priority_ms	int	Exit time of priority object (ms since epoch)
encroachment_duration_s	float	Duration of encroachment (seconds)
pet_s	float	Post-Encroachment Time (seconds)
conflict_lon_deg	float	Conflict point longitude (WGS84, degrees)
conflict_lat_deg	float	Conflict point latitude (WGS84, degrees)

2.6. Signal Data

The *signal_states* table contains traffic light data at the intersection intended as contextual information to PET event analysis. These data are available for almost the entire dataset time span. Unfortunately, signal state data contain some artifacts and errors (described in Section 3) and need to be handled with caution. The table's column description is provided in Table 5.

A *signal_locations* table is included in the published dataset that contains the locations of signal groups as Well-Known-Text point geometries in WGS84 coordinates referenced via signal_group_id. These positions are also visualized in Figure 2. Additionally, a convenience mapping between signal groups and scenarios (*scenario_signal_mapping*) is included that provides a representative signal group for each scenario's encroaching and priority movement relations.

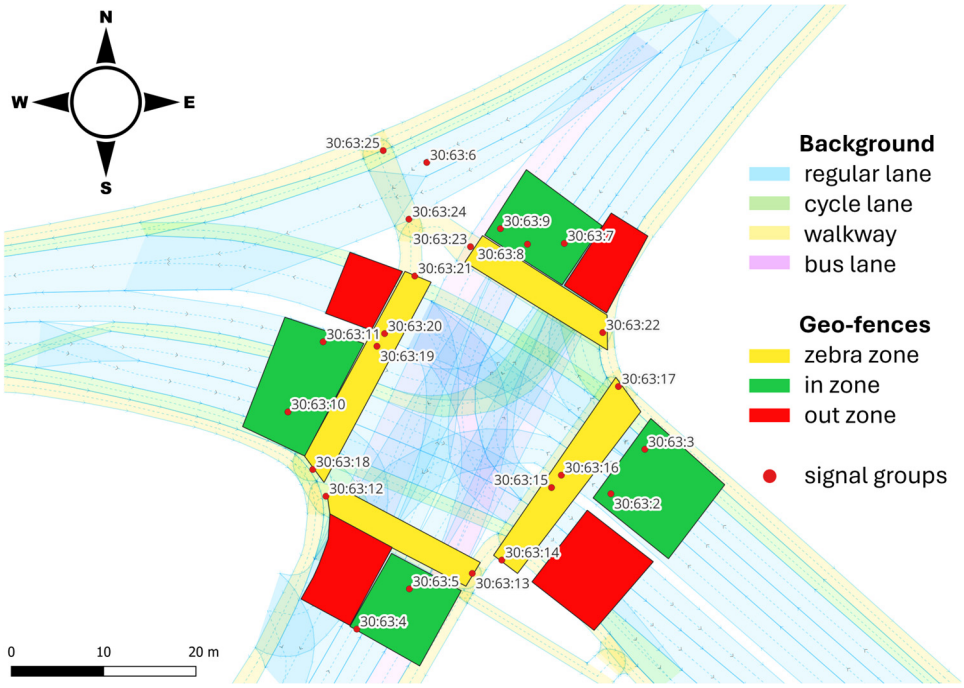


Figure 2. Location of signal groups and geo-fences for determining objects' movement relations.

Table 5. Signal states—column description.

Column	Type	Description
signal_group_id	string	Identifier of the signal group
start_timestamp_ms	int	Start time of the signal state (ms since epoch)
end_timestamp_ms	int	End time of the signal state (ms since epoch)
signal_state	string	Human-readable signal state (green, yellow, red, red-yellow)

2.7. Auxiliary Files

In addition to the main data products, auxiliary files for documentation and reproducibility are included in the dataset:

- Geofence used for track selection (see Figure 1);
- Geofences used for determining the movement relation of objects and consequently their relevance for interaction scenarios (see Figure 2).

These figures are included in the dataset as well.

2.8. Entity Relationships

The dataset consists of multiple logical tables linked through object identifiers, scenario identifiers, signal group identifiers, and timestamps.

- Trajectory records in *tracks* are associated with individual road users via *object_id*.
- Static object attributes in *objects* reference the same road users via *object_id*:
  - *tracks.object_id* → *objects.object_id*
- PET-based interaction events in *pet_events* reference two road users:
  - *pet_events.encroaching_object_id* → *objects.object_id*
  - *pet_events.priority_object_id* → *objects.object_id*
- Each PET event is associated with exactly one predefined interaction scenario:
  - *pet_events.scenario_id* → *scenarios.scenario_id*
- Scenario definitions in *scenarios* describe combinations of object-level movement attributes that determine whether PET is computed for a given pair of road users.
- Objects are therefore indirectly related to scenarios through their movement attributes and their participation in PET events.
- A convenience mapping between scenarios and traffic signal groups is provided:
  - *scenario_signal_mapping.scenario_id* → *scenarios.scenario_id*
  - *scenario_signal_mapping.encroaching_signal_group_id* → *signal_states.signal_group_id*
  - *scenario_signal_mapping.priority_signal_group_id* → *signal_states.signal_group_id*
- The *scenario_signal_mapping* table enables straightforward temporal association of PET events with traffic signal states based on the scenario context.
- For scenarios involving pedestrians, multiple signal groups may correspond to priority movements; the mapping table provides a representative signal group with aligned phase timing.
- *pet_events* can be temporally linked to *signal_states* using timestamps.

All relationships are based on identifiers and time references; no tables are pre-joined in the published dataset.

3. Methods

3.1. MOT System

The system used to acquire the movement data of traffic participants consists of six HESAI Pandar XT 32 LiDAR sensors (Hesai Technology Co., Ltd., Shanghai, China)

mounted at a height of 4.5 m at the intersection. The location of the sensors can be seen in Figure 1. The acquired point-clouds are processed directly at the intersection using the SENSR perception software by Seoul Robotics (Version 2.3.4). Object-level data are transmitted via MQTT to a central service, which assembles the iteratively received object information into trajectories based on the object identifier provided by the perception software and assigns a persistent identifier to each trajectory (*object_id*). The system operates at an intended detection frequency of 10 Hz.

3.2. Preparation of Tracks Data and Quality Notes

For the published dataset, tracks are filtered to those that intersect the central area of the intersection (see Figure 1). This achieves a trade-off between manageable data quantity and inclusion of traffic context. Tracks that do not correspond to any of the movement relations relevant to a defined scenario can therefore be included and could be used to analyze the traffic context within the central area of the intersection. Outside the central area, trajectory data might be incomplete since data from objects that did not cross the central area at all are excluded. Additionally, tracking of objects can be interrupted (e.g., due to severe occlusion), such that one real object is represented as multiple objects in the original data (*object_id*) and fragments outside the central area will be missing in the published data due to the filter.

The intended detection interval of the MOT system is 100 ms. Every object detected within one processing cycle is assigned the same timestamp. About 89.5% of sampling intervals are in the range of 95–105 ms. Occasionally, a cycle is skipped, resulting in 4.9% in the range of 195–205 ms. Additionally, 2.4% are off by 5–15 ms from 100 ms and 0.3% are off by 5–15 ms from 200 ms. Consequently, 2.9% are off by more than 15 ms from 100 ms and 200 ms.

Between 7 April 2024 and 29 April 2024, the MOT system did not work properly and data from that period are mostly missing. On 12 February 2023, the system probably was disturbed by heavy snowfall such that it detected exceptionally many different objects, 20% of which were not classifiable.

The MOT system occasionally provides fragmented detections, i.e., a single road user is detected as multiple objects. In this case, one of the fragments is usually dominant and tracked reliably with a consistent *object_id* while the other fragments are typically only tracked for a few frames. When applying geo-fencing with two or more different intersection zones—for example, the method we used to determine movement relations of objects (see Section 3.3)—such short, fragmented tracks are automatically filtered out. Larger vehicles seem to be more prone to fragmented detection with the given MOT system, e.g., buses due to reflection issues because of their large windows or semi-trucks due to tractor and trailer being detected as separate objects.

Moreover, a road user might not be detected at all for a short time. In that case, the MOT system estimates the movement of the object (e.g., tracking_status = DRIFTING) to assign the correct *object_id* once the object is detected again. If this association fails, the movement of one real road user might be represented with two different *object_id* values in the dataset.

Since the system operates live, some attributes of objects that are constant in reality are estimated in every detection cycle and consequently vary within the same *object_id*. This applies to *object_class*, *object_length_m*, *object_width_m*, and *object_height_m*. Moreover, the speed estimate that the system provides is affected by a temporal lag. To produce a stable speed time series without heavy fluctuations, a live estimation cannot immediately react to a strong change in the calculated speed value (which could be a measurement error)

but needs to wait for more measurements to make sure that the object actually changed its speed. An example is given in [6].

Reference [6] uses the same MOT system at the same intersection and shows that the data quality issues described above do not affect the reliability of safety assessment based on PET calculated from these data.

Since the MOT system was operated as a commercial black-box system and raw sensor data are not available, no manually annotated ground truth or quantitative accuracy metrics (e.g., position or classification errors) can be provided for the trajectory data.

3.3. Preparation of Objects Data and Notes on Movement Relations

A static object class attribute is created per *object_id* by considering both the count of *object_class* values (e.g., "vehicle") N and the mean of the attribute *det_points_count* per *object_class* Q. Both indicators are combined in a weighted sum to derive a final score S per object class, and the object class with the highest score is selected as the static object class.

$$S = a * N + (1 - a) * Q, \quad (1)$$

This approach balances two concepts. First, if an object is often estimated to be of a certain class, this class probably is the true class of the object. And second, the class estimation is more accurate when an object's bounding box corresponds to many LiDAR points. The weighting factor has been empirically determined and finally set to a = 0.08.

The score serves only as a ranking heuristic for candidate classes. The weighting factor was empirically chosen to balance the contributions of the frequency of class assignments and the detection quality indicator. Since the score is only used for ranking candidate classes, explicit normalization of the indicators was not required.

The movement relation of objects corresponding to the defined scenarios is determined based on their static object class and their movement through the intersection. Figure 2 shows the geo-fencing zones that have been used to determine the movement relations of scenario-relevant objects and the convention for naming intersection legs. For vehicles and two-wheelers, an in-zone and an out-zone are defined which trajectories need to intersect to be considered for a certain movement relation in one of the defined scenarios. To mitigate the risk of false positive assignments of movement relations in case of untypical trajectories, further conditions have been formulated for each relation by using every geo-fence. For every geo-fence, a must-intersect or must-not-intersect criterion has been formulated. For example, cars approaching from the north leg and turning left need to intersect four geo-fences (north in, zebra north, zebra east, east out) and must not intersect any of the other geo-fences. For scenarios involving pedestrians, only one geo-fence covering the zebra crossing is used because pedestrian movements vary more and are less aligned with the dedicated infrastructure, e.g., pedestrians might short-cut the entry or exit of a zebra crossing. Note that the zones depicted in Figure 2 have been aligned with aerial images and therefore do not perfectly align with the schematic map used as reference in the figure.

Static object dimensions are calculated based on the 10% of each object_id with the highest *det_points_count* values as the weighted average of *object_length_m*, *object_width_m*, and *object_height_m*, using *det_points_count* as weights.

3.4. Notes on Scenario Selection

The eight defined scenarios have been selected based on suitability for PET-based safety assessment. They are schematically visualized in Figure 3. These are scenarios where road users of one relation (encroaching) need to yield to road users of another relation (priority). If road users of the encroaching relation traverse the conflict area, i.e., the area shared between both relations, just before a priority road user, their interaction

was potentially unsafe. At this signalized intersection, permissive left-turns exist from the northern and from the southern approach, i.e., left-turning vehicles need to yield to oncoming traffic as both receive a green signal concurrently (1A, 1B). Interactions between vehicles and vulnerable road users have been considered in vehicle–pedestrian and vehicle–bicycle scenarios corresponding to permissive turns: right-turning vehicles approaching from west or east need to yield to cyclists from the same approach that want to go straight (3A, 3B); left-turning vehicles from north and south and right-turning vehicles from east and west need to yield to pedestrians crossing the outbound leg (2A–D).

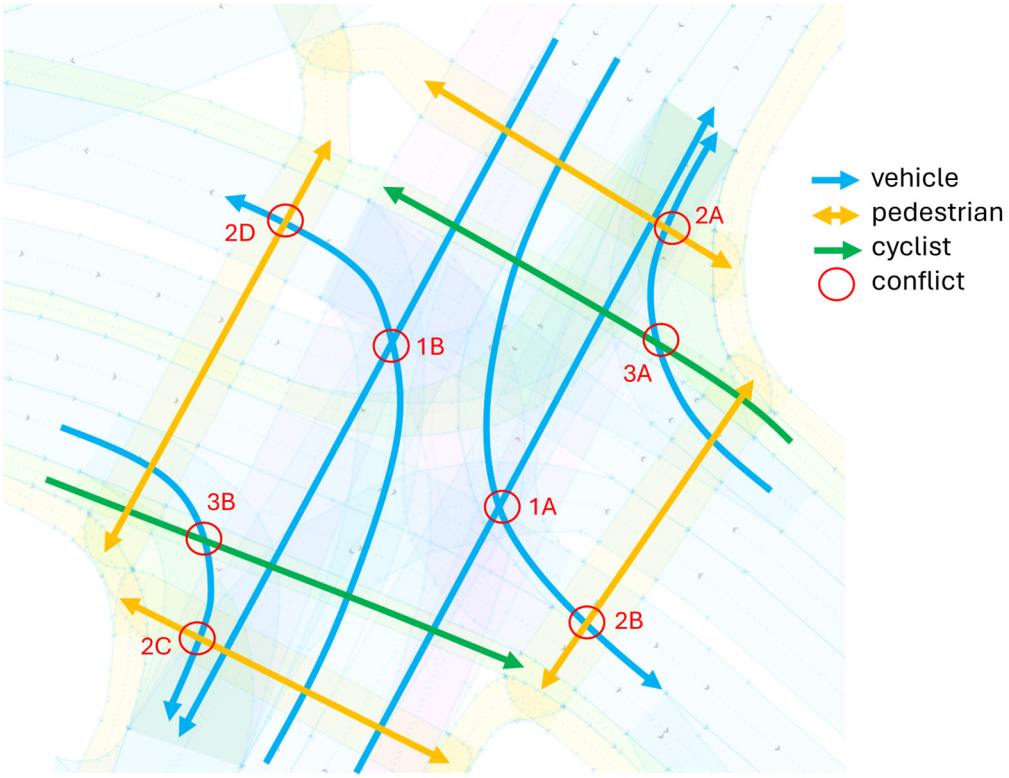


Figure 3. Schematic visualization of the eight interaction scenarios.

3.5. PET Calculation and Quality Notes

PET [2] is suitable for investigating interactions in which a road user encroaches the path of a priority road user that has the right of way. It is defined as the time from the encroaching road user leaving the shared space (conflict area) to the priority road user entering this space. The algorithm that we used to calculate PET values determines this conflict area dynamically for each interaction instead of using a predefined area. The same algorithm was used in [6].

First, we identify the pairs of objects (*object_id*) that we want to calculate PET for. To determine interaction pairs, we join objects of the encroaching movement relation to objects of the priority movement relation based on their minimum and maximum timestamps, i.e., the first and last time they were detected by the MOT system anywhere at the intersection. In particular, the following two join conditions are applied:

- encroaching max. timestamp + 5 s $\geq$ priority min. timestamp;
- encroaching min. timestamp $\leq$ priority max. timestamp.
- The 5 s tolerance ensures that all PETs up to 5 s can be calculated from the pairs considered.

The following steps are then conducted for each identified pair:

- Determine the positions where the paths of both objects intersect as conflict points. If there are no conflict points, there is no PET for the pair.
- Determine a conflict area per conflict point:
  - For both objects, find the record with the nearest position to the conflict point;
  - For both objects, create a rectangle around the conflict point based on the object's heading, length, and width from the nearest record;
  - The intersecting area of these two rectangles defines the conflict area.
- For each record of both objects, reconstruct the 2D bounding box of the objects based on their position, heading, length, and width resulting in a trajectory of rectangles.
- For both objects and every conflict area, determine the timestamps corresponding to the first and last bounding box that intersect the conflict area and calculate the PET and the encroachment duration.

Due to the rather inclusive definition of interaction pairs, negative and large PET values exist. We include these in our dataset and leave further filtering to the user. However, interactions with PET values above 5 s might be missing in the dataset, due to the definition of interaction pairs (see the join conditions above).

Further usage notes:

- Multiple PET events can exist for the same pair of objects. These might be artifacts requiring special attention.
- One object can encroach multiple priority objects in the data. Depending on the scenario, it might be adequate to only consider the first event per encroaching object. One vehicle can encroach multiple pedestrians on the crosswalk. Arguably, one left-turning vehicle cannot encroach multiple opposing vehicles on a single lane but rather only the first of them.
- One priority object can be encroached by multiple objects. Arguably, this is meaningful in every defined scenario.

The PET calculation algorithm has been evaluated for several scenarios and on several data subsets. In particular, negative, small positive, and very large values have been investigated thoroughly. The algorithm calculates PET values from the given trajectory data as intended. In general, PET values could be affected by artifacts of the given trajectory data. Due to the lack of a ground truth that could be used for comparison, the resulting values have not been directly compared to a "true" PET value. But as described in Section 3.2, PET has shown good robustness against the quality issues present in the given data [6].

3.6. Preparation of Signal State Data and Quality Notes

Signal data were received as SPAT message content [7] in an XML format via MQTT and collected in a relational database, where for each signal group the first record of each new signal state is stored. For publication, these tabular data were cleaned further with the following logic:

- Signal state duration and signal state end timestamp are derived from the timestamp of the next record (=changed state) within a signal group.
- The collecting system occasionally receives an unknown state. Sometimes, known states and unknown states alternate with high frequency in the raw data. Therefore, we removed any unknown states that lasted for not more than two seconds.
- Consecutive records of the same state (now possible due to the filter in the previous step) are fused into a single record of that state.

Unknown states lasting longer than two seconds are not treated prior to publication. The true signal state in these cases might be estimated by considering the state of other

signal groups at that moment or by considering the general pattern of signal phases derived from historic data.

Signal state data are affected by missing data. For the following days, no data are available:

- 1 August 2023 to 2 August 2023;
- 31 August 2023 to 4 September 2023;
- 18 January 2024;
- 21 January 2024;
- 30 January 2024;
- 1 February 2024;
- 23 February 2024 to 25 February 2024;
- 17 March 2024;
- 16 June 2024;
- 13 July 2024;
- 22 August 2024 to 31 August 2024;
- 5 June 2025 to 11 June 2025;
- 17 July 2025;
- 16 August 2025 to 17 August 2025;
- 28 August 2025 to 31 August 2025.
- Other days might be partially affected.

During our own work with these signal data, we found indication that signal timings might be rather imprecise. The duration of yellow and red-yellow signals, which should be of constant duration, vary. Mainly, the duration sometimes seems to be too long or too short by roughly full-second steps, but variance around full-second marks is also evident. Likely, the start timestamps of green and red signals are affected by this imprecision as well. Any application of the data needs to consider signal timing imprecision.

The signal group 30:63:9 is a special bus signal corresponding to public transport prioritization of the northern approach. Almost all signal states of this group in the dataset are either “green” or “unknown”.

4. User Notes

Usage examples are provided with the published dataset as python scripts.

Author Contributions: Conceptualization, M.S., T.V. and K.R.; methodology, M.S. and T.V.; software, M.S. and T.V.; validation, M.S., T.V. and K.R.; formal analysis, M.S. and T.V.; investigation, M.S. and T.V.; data curation, M.S. and T.V.; writing—original draft preparation, M.S.; writing—review and editing, M.S., T.V. and K.R.; visualization, M.S.; supervision, K.R.; project administration, M.S. and K.R.; funding acquisition, M.S. and K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Austrian Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie, grant number GZ 2021-0.641.557.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The full dataset will be made available by the authors on request. A 6-month temporal subset is openly available in Zenodo at https://doi.org/10.5281/zenodo.18234283.

Acknowledgments: The authors gratefully thank ALP.Lab GmbH for their valuable support throughout the project and for their constructive collaboration. We further thank the municipality of Salzburg and GEVAS software GmbH for providing access to the traffic signal state data. During the preparation of this manuscript/study, the authors used ChatGPT 5.2 for the purposes of text generation and

grammar correction. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: All authors were employed by the company Salzburg Research Forschungsgesellschaft m.b.H., a non-profit research organization. The funders stated in the Funding statement above had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MOT	Multi-object tracking
PET	Post-encroachment time
SSM	Surrogate safety measure

References

1. Abdel-Aty, M.; Wang, Z.; Zheng, O.; Abdelraouf, A. Advances and Applications of Computer Vision Techniques in Vehicle Trajectory Generation and Surrogate Traffic Safety Indicators. *Accid. Anal. Prev.* **2023**, *191*, 107191. [CrossRef] [PubMed]
2. Allen, B.L.; Shin, B.T.; Cooper, P.J. Analysis of Traffic Conflicts and Collisions. In *Transportation Research Record*; Transportation Research Board: Washington, DC, USA, 1978.
3. Johnsson, C.; Laureshyn, A.; De Ceunynck, T. In Search of Surrogate Safety Indicators for Vulnerable Road Users: A Review of Surrogate Safety Indicators. *Transp. Rev.* **2018**, *38*, 765–785. [CrossRef]
4. Steinmaßl, M.; Beeking, M. Salzburg Intersection Trajectory Dataset 2023. Available online: https://osf.io/rdzxp/ (accessed on 17 March 2026).
5. Steinmaßl, M.; Rehrl, K. Evaluating Traffic Trajectories from Stationary Multiple Object Tracking Systems. *J. Locat. Based Serv.* **2025**, *19*, 257–286. [CrossRef]
6. Steinmaßl, M.; Beeking, M.; Troth, N.; Rehrl, K. A Framework for Reliable Traffic Surrogate Safety Assessment Based on Multi-Object Tracking Data. *Traffic Saf. Res.* **2025**, *9*, e000123. [CrossRef]
7. *TS 102 894-2*; Intelligent Transport Systems (ITS); Users and Applications Requirements; Part 2: Applications and Facilities Layer Common Data Dictionary. European Telecommunications Standards Institute (ETSI): Sophia Antipolis, France, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.