PCICF: A Pedestrian Crossing Identification and Classification Framework

Junyi Gu
junyi.gu@gu.se
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden

Beatriz Cabrero-Daniel
beatriz.cabrero-daniel@gu.se
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden

Ali Nouri
ali.nouri@volvocars.se
Volvo Cars,
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden

Lydia Armini
lydia.armini@gu.se
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden

Christian Berger
christian.berger@gu.se
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden

Abstract

We have recently observed the commercial roll-out of robotaxis in various countries, including the USA and Germany. They are deployed within an operational design domain (ODD) on specific routes and environmental conditions, and are subject to continuous monitoring to regain control in safety-critical traffic situations. Since ODDs typically cover urban areas, robotaxis must reliably detect and interact with vulnerable road users (VRUs) such as pedestrians, bicyclists, and e-scooter riders. To better handle such varied traffic situations, end-to-end AI, which directly computes vehicle control actions from multi-modal sensor data instead of only for perception, is on the rise. High quality data is needed to systematically train and evaluate such systems within their ODD. In this work, we propose PCICF, a framework to systematically identify and classify VRU situations to support ODD's incident analysis. We base our work on the existing synthetic dataset SMIRK, and enhance it by extending its single-pedestrian-only design into the *MoreSMIRK* dataset, a structured dictionary of multi-pedestrian crossing situations constructed systematically. We then use space-filling curves (SFCs) to transform multi-dimensional features of scenarios into characteristic patterns, which we match with corresponding entries in MoreSMIRK. We evaluate PCICF with the large real-world dataset PIE, which contains more than 150 manually annotated pedestrian crossing videos. We show that PCICF can successfully identify and classify complex pedestrian crossing situations, even when groups of pedestrians merge or split during their crossing. By leveraging computationally efficient components like SFCs, PCICF has also potential to be used onboard of robotaxis for out-of-distribution (OOD) detection, for example. We share an open-source replication package for PCICF, including its algorithms, the complete MoreSMIRK dataset and dictionary, and our experiment results, available at https://github.com/Claud1234/PCICF.

Keywords

Automotive, Dataset, Object Detection, Space-Filling Curve, Pedestrian Crossing Classification

1 Introduction

The development of information technology has changed people's opinions about vehicles, which are no longer simply mechatronic systems but software-intensive innovation platforms compared to their predecessors. Today, a commercial passenger car could contain more than 100 million lines of code, compared with only 50,000 lines decades ago [10], depicting the challenge the automotive industry faces in maintaining system quality. Meanwhile, the implementation of AI and ML in the automotive industry has been extensively studied in recent years, which has led to two challenges for modern vehicles: First, the need for large computational power, and second, the rapid increase in onboard data to feed AI/ML software.

1.1 Problem Domain & Motivation

Traffic event identification and classification from such growing data is an important research direction in autonomous driving (AD) and traffic safety. Investigating the most safety-related mishaps highlights the importance of detecting and predicting pedestrian behavior, as they are among the VRUs. For instance, a pedestrian was severely injured in a recent mishap involving a robotaxi, which led to the revocation of its AD deployment and testing permit [21]. Avoiding collisions with pedestrians typically requires the highest Automotive Safety Integrity Level (ASIL) due to the high severity of potential accidents [20]. The implementation of these systems becomes even more challenging because pedestrian behavior and appearance are less predictable compared with those of other road users. For example, the vehicle may encounter pedestrians even on roads such as highways, where they are not permitted. Hence,

the development, verification, and validation of such systems require special attention. For instance, the decomposition of safety requirements with a high ASIL (i.e., ASIL D) is prescribed by the standard, which then requires diversity in the implementations of two parallel software components performing the same task. Moreover, recent research is exploring how AI/ML-based software can directly calculate vehicle control parameters from multi-modal sensor data. However, this trend is posing important challenges to be addressed: On the one hand, the development of AI/ML algorithms relies on growingly complex neural network (NN) architectures, and on the other hand, the development and systematic assessment of such AI/ML-based software has become dramatically dependent on large-scale datasets, which need to cover sufficiently diverse traffic situations to be effective. We focus on two research directions in this work: a) the identification and detailed classification of VRUs when they interact with vehicles, and b) the qualification of datasets for AL/ML-based software to assess their potential to be used for specific safety-related traffic events. Our motivation is to propose and evaluate an alternative to typical NN models that rely on complex architectures and vast computational resources to extract traffic events from large-scale non-annotated datasets.

1.2 Research Goal & Research Questions

Safely detecting and classifying pedestrian crossing situations in video data is critical for the robustness of perception stacks in autonomous vehicles (AVs). However, pedestrian crossing identification and classification are more than simple yes/no warnings, but also provide information about scene details such as the number of pedestrians, crossing directions, and pedestrian behaviors [25].

We pursue the research goal in this work to conceptualize and prototype a framework that processes camera images directly to provide a detailed description of the pedestrian crossing event. We first construct *MoreSMIRK*, a systematic dictionary of synthetic pedestrian crossing events. For identifying and classifying real world traffic situations, we use the broadly adopted YOLO network for preliminary object detection, and then leverage on SFC to compute a domain-specific representation of filtered pedestrian crossing tracks; this representation can be thought of as a 'fingerprint' of a specific crossing event having the shape of a bar code (cf. Fig. 3(b)). These fingerprints enable us to compare the similarity to find corresponding ones in our dictionary MoreSMIRK to conclude the semantic classification of pedestrian crossing events. We structure our research along the following three research questions that build on top of each other:

RQ-1: What are the design decisions to create a synthetic dataset for event classification?

RQ-2: What are the design decisions for a traffic event identification and classification framework?

RQ-3: What is the performance of the proposed framework?

1.3 Contributions & Scope

We present a framework for multi-pedestrian crossing identification and classification. We summarize our contributions as follows:

- We introduce a framework consisting of pedestrian extraction, dimensionality reduction, and crossing analysis, which incorporate NNs, SFCs, and a systematically constructed crossing event dictionary as essential algorithms, respectively.
- We extend the public synthetic dataset SMIRK, which focuses on single-pedestrian cases only, to multi-pedestrian scenarios and systematically construct the MoreSMIRK dataset as a benchmark to retrieve crossing events from real-world datasets.
- We demonstrate and evaluate the capability and prospect of our framework in identifying and classifying pedestrian crossings in real-world scenarios. Moreover, we demonstrate the framework's potential to efficiently validate and annotate large-scale datasets for AI/ML-based software.

Our system delivers two primary advantages: a) Our novel combination of deep learning (DL) and SFCs for event identification offers computational and storage efficiency while maintaining domain-specific information, which allows us to deploy the framework in practical applications and consider even large-scale AD data for other traffic events; b) the MoreSMIRK dataset has the flexibility to define even complex pedestrian crossing scenarios and hence, the benchmark dictionary, which is based on the MoreSMIRK dataset, is versatile and applicable for other traffic event detections.

The remainder of the paper is structured as follows. Section 2 reviews the pedestrian safety-related regulations, datasets, and detection approaches. Section 3 presents the design and detailed architecture of PCICF. Specifically, we provide the construction principles of the MoreSMIRK dataset, which is one of our work's essential contributions. Section 4 presents the methodology used to evaluate PCICF. Section 5 reports on the results from the evaluation experiments and provides a discussion. Finally, a conclusion with outlook for future work is provided in Section 7.

2 Related Work

This section covers the state of the art of pedestrian identification and crossing classification: subsection 2.1 describes safety-related regulations, subsection 2.2 moves on to discuss existing image datasets that can be used to train AD and Advanced Driver Assistance Systems (ADAS) functions, and subsection 2.3 presents strategies to deal with large scale datasets to ensure computational efficiency in edge devices such as public road vehicles.

2.1 Pedestrian Safety, a Requirement for AD

All safety-related software systems in AVs shall avoid causing unreasonable risks to other road users, including VRUs, especially pedestrians [21]. Hence, different types of VRUs, such as pedestrians, cyclists, or e-scooter riders, shall be considered when designing, developing, and testing perception systems in both AD and ADAS. Rigorous software safety assurance processes are prescribed by standards and regulations. For instance, ISO 21448 (Safety of the Intended Functionality), also known as SOTIF, provides requirements and recommendations for the design, verification, and validation of safety-related systems [15]. SOTIF emphasizes the importance of unknown hazardous scenarios and the need to identify and mitigate them. This is particularly challenging for pedestrians, as specifying all combinations is impossible. For example, as described in the SOTIF, an unknown hazardous scenario might involve a person on
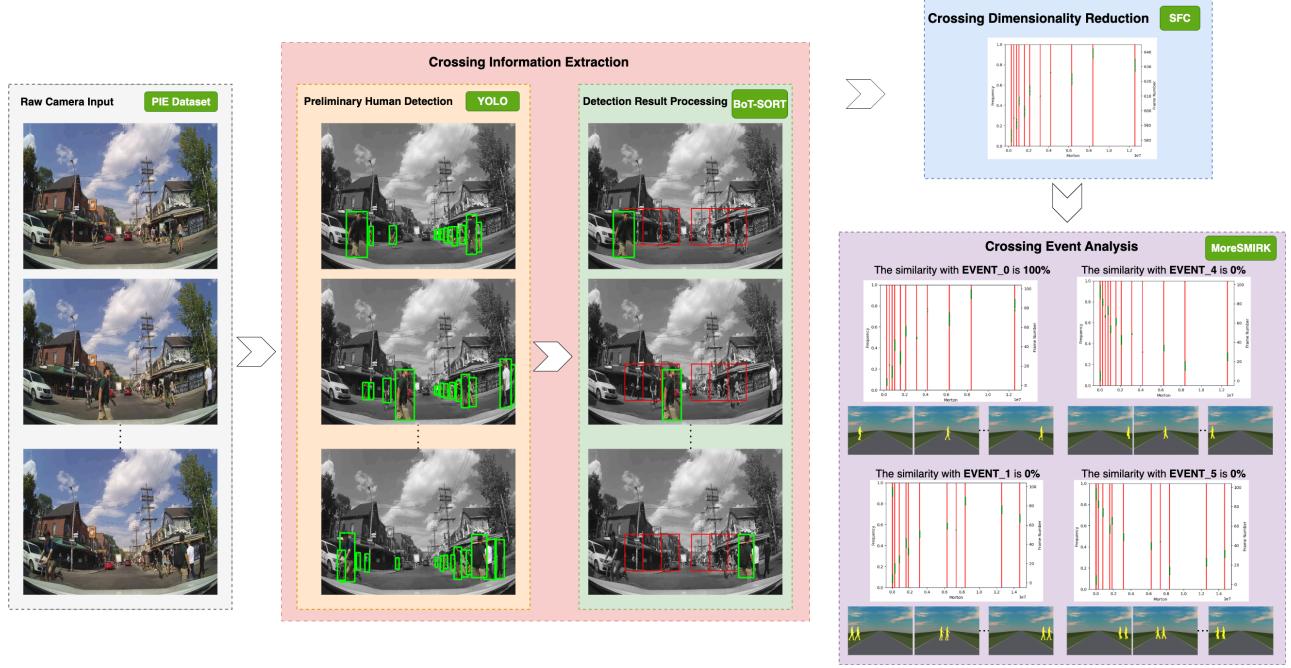
Figure 1: Overall workflow of PCICF: The raw camera input is a sequence from the PIE dataset, and the dashed rectangles represent the modules detailed in Section 3. Dark-green boxes at the top indicate the datasets and algorithms used in each module. Finally, four similarity checks in crossing event analysis out of the 104 entries from MoreSMIRK are shown to obtain semantic descriptions for a crossing event.

a skateboard being ignored by detection algorithms because their speed exceeds the typical pedestrian walking speeds observed in the training data [15]. More detailed considerations, such as diversity and sufficiency (i.e., size, age, pose) in training and testing datasets, are mentioned in ISO 8800 (Safety and Artificial Intelligence) [16]. Complementary guidelines are provided by the European Commission's High-Level Expert Group on Artificial Intelligence: Since AVs interact with humans, a number of high-level requirements, not only during operation (i.e., transparency), but also on the data for training of the underlying AI/ML (i.e., diverse) need to be considered during development and for deployment [12].

2.2 Pedestrian Image Datasets for AD training

To train AI models for pedestrian identification and crossing classification, researchers often turn to already existing datasets, some of which are openly available. Depending on the nature of the model and the task, different types of data are used. For instance, the Zenseact Open Dataset (ZOD) offers video frames from a high-resolution RGB front-looking camera, together with data from a LiDAR sensor [2], and covers a number of geographical areas. The large real-world dataset PIE, captured by York University in Canada, also contains more than 150 manually annotated pedestrian crossing videos. Another well-known example is the Waymo Open Dataset, which contains recordings from five cameras in different directions side-by-side to other sensor readings and segmentation information, which is useful for scene understanding [19].

While the usefulness of these datasets has been largely demonstrated, and they often include annotations such as bounding-boxes for other traffic agents, they do not reflect well the importance of focusing on vulnerable road users, such as pedestrians. On the other hand, STCrowd is a large-scale multi-modal dataset that focuses on 3D pedestrian perception in challenging crowded scenarios [11].

Synthetic data can also be used to train safety-critical AI models. Frameworks for pedestrian simulation exist [24], even though not all focus on visual realism, which might hinder their use in verification processes [9]. Some synthetic datasets focused on AD-related scenarios: For instance, SMIRK is an automatic emergency braking (AEB) system to protect pedestrians developed following the process defined in SOTIF [15, 24] and has been used to conduct a systematic safety analysis of an AEB system using the AM-LAS methodology [6]; in addition, SMIRK is also the underlying synthetic dataset focusing on single pedestrian situation used to evaluate the AEB. Data used to train the SMIRK system is publicly available [24]; our work will build on it and augment it with common pedestrian motion patterns as described in Sec. 3.

2.3 Real-Time Image Analysis in Edge Devices

As more edge cases and malfunctions in the real world are discovered, datasets for training and evaluating AVs continue to grow. Moreover, real-time analysis of onboard sensor data (i.e., camera feeds) is needed to enable live monitoring of surroundings of AVs and ensure compliance with pedestrian safety rules and regulations. The large sets of data as well as the need to process them

in real-time, poses challenges for the software architecture and on computationally efficient processing.

Properly addressing these challenges onboard of vehicles is growingly important, where computational resources for AI systems are constrained, and connectivity might not be reliable. We propose SFC [3] to obtain characteristic patterns in pedestrian crossings. SFCs are mappings from high-dimensional spaces to a single-dimensional space that have practical applications in fields such as database indexing due to their ability to preserve spatial relationships between data points [3]. Given the computational efficiency of SFCs, they have the potential even to be used onboard of vehicles, as demonstrated by Berger et al. in 2023 [3, 4].

3 PCICF: A Pedestrian Crossing Identification and Classification Framework

We propose PCICF, as shown in Fig. 1, as an end-to-end framework for identifying and classifying pedestrian crossing events. Such events can be characterized as (a) spanning over a certain amount of time (i.e., consecutive input frames in our case), and (b) the crossing event itself needs to fulfill the semantic properties as starting either on the left or right hand side of the road, and ending on the opposite side, (i.e., crossing the field of view during the selected frame sequence). In PCICF, camera images are fed as input to extract individual pedestrian tracks (labeled as 'PIE Dataset' in Fig. 1, cf. Section 4 for evaluation details), which are transformed to event-specific fingerprints that intuitively take the form of barcodes with the help of SFCs (in Fig. 1). We match these barcodes with those from our systematically constructed dictionary, MoreSMIRK, to obtain detailed explanations of the particular pedestrian crossing events (labeled as 'MoreSMIRK' in Fig. 1). In the following, we describe the relevant modules of PCICF in detail.

3.1 Constructing the MoreSMIRK Dataset

We use a look-up dictionary of known patterns for pedestrian crossing events. This dictionary forms the core of PCICF when analyzing a dataset of interest to obtain semantically plausible explanations for potential pedestrian crossing events. We based the systematic construction of MoreSMIRK on the original SMIRK dataset [24], which contains 4,928 varying single pedestrian crossing configurations, along with their respective semantic segmentation and corresponding labels. The SMIRK dataset was created using the ESI Pro SiVIC simulator to support the systematic analysis of an automotive safety function that initiates automatic emergency braking (AEB) for a vehicle in the event of an unexpected situation where a pedestrian is crossing the road in front of a vehicle [6]. The dataset contains various synthetic pedestrians, including both male and female individuals in different visual appearances (business and casual), as well as a child.

However, the current SMIRK dataset is not sufficiently representative of what a vehicle could potentially face when approaching an inner-city intersection with traffic lights or crosswalks, as it only covers single pedestrian crossing events. Thus, we extend SMIRK into MoreSMIRK to improve the representativeness of pedestrian crossing situations, which are potentially present in real-world scenarios. Our extensions systematically add more complex pedestrian crossing configurations that were constructed according to two

properties: a) initial location, and b) pedestrian grouping configuration, as shown in Fig. 2. The initial location indicates where pedestrians start to cross, either from left to right or from right to left. We include systematically varying pedestrian configurations of up to three individuals, clearly separated and following each other when crossing a road. We choose this generating pattern because three people following each other with some spacing in between would almost occupy half of the street to cross, and when two groups with the same configuration would cross from both sides, the entire area in front of the ego vehicle is nearly fully covered. All crossing events were generated as image sequences of 100 frames, with the first pedestrian entering the scene in frame 1 and the last pedestrian leaving in frame 100.
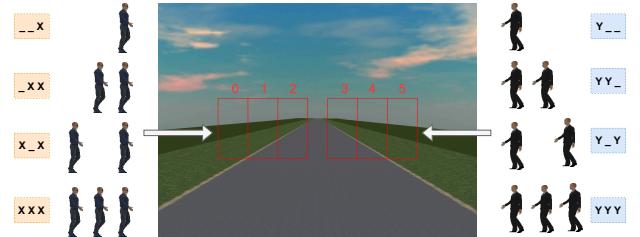

Figure 2: The illustration of configuration principles to generate the MoreSMIRK dataset. The red boxes indicate the locations of regions of interest (RoI) in the dataset.

As illustrated in Fig. 2, X is the pedestrian crossing from left to right, and Y represents the one crossing from right to left. Based on the four pedestrian group configurations, there are eight basic crossing patterns for single-directional pedestrian crossings (four from each side). For two-directional crossings, there are $4 * 4 = 16$ unique crossing patterns when pedestrians from two sides are moving synchronously. We include an additional and optional offset Φ from 0 to 5, corresponding to the number of six regions of interest (RoI) in front of the ego vehicle (cf. red rectangle boxes in Fig. 2), to delay the start of the crossings for the pedestrians X from the left side. For instance, offset $\Phi = 1$ means that when pedestrian X on the left reaches the RoI grid 0, the right pedestrian Y is already located at RoI grid 4, i.e., has entered the street. We result in $16 * 6 = 96$ crossing patterns when following this principle; in total, we provide $8 + 96 = 104$ crossing patterns in the MoreSMIRK dataset, including the pedestrian configurations without facing pedestrians starting on the other side, respectively. Please note that mirroring the 96 two-directional sequences s is equivalent to applying the offset Φ to the pedestrian Y on the right side. We also complement the generated scenarios with semantic ground truth annotations for all sequences in MoreSMIRK. Table 1 shows a part of pedestrian crossing configurations. The complete MoreSMIRK dataset is publicly available via AI Sweden[1].

3.2 Crossing Information Extraction

PCICF consists of three individual parts: (a) preliminary pedestrian detection, (b) intermediate detection result processing, and (c)

---

[1]https://www.ai.se/en/ai-labs/technology-infrastructure/datasets/smirk-and-moresmirk-datasets#moresmirk

Table 1: Part of pedestrian crossing event configurations in the MoreSMIRK Dataset. The X and Y represent the pedestrians crossing from left to right and right to left, respectively.

Events (E)	Left Pedestrians (L)	Offset (Φ)	Right Pedestrians (R)
0	_ _ X	N/A	_ _ _
1	_ X X	N/A	_ _ _
2	X _ X	N/A	_ _ _
3	X X X	N/A	_ _ _
...	...	...	...
8	_ _ X	0	Y _ _
9	_ X X	0	Y _ _
10	X _ X	0	Y _ _
11	X X X	0	Y _ _
...	...	...	...
100	_ _ X	5	Y Y Y
101	_ X X	5	Y Y Y
102	X _ X	5	Y Y Y
103	X X X	5	Y Y Y

crossing event identification and classification. We use YOLO for preliminary pedestrian detection [17] and then apply tracking and filtering to obtain relevant detections as potential candidates for pedestrian crossing sequences. Next, we transform these sequences into their corresponding single-dimensional representations by leveraging an SFC [3] to obtain their characteristic fingerprints. An SFC is a curve exhibiting a repetitive pattern, which passes through every point in a multi-dimensional data space to transform it into its corresponding single-dimensional representation while, depending on the choice of the recursive pattern, preserving the original data-space's properties, such as locality between two points. When observing such single-dimensional representations over time, we obtain characteristic stripe patterns as shown in Fig. 3 that we can match to corresponding entries within our look-up dictionary MoreSMIRK to query for matching pedestrian crossings.

For the first part, preliminary pedestrian detection, we use YOLO detections in the form of bounding-boxes to filter pedestrian behavior of interest. YOLO is one of the most popular choices for general-purpose object detection applications because it has been optimized to balance speed and accuracy. However, one current shortcoming of YOLO is its relatively weak performance at separating groups of objects or densely-packed objects.

Next, for intermediate detection result processing, we extract pedestrian crossings from YOLO's detections by selecting, tracking, and filtering bounding-boxes in order to identify pedestrians' bounding-boxes within a valid crossing sequence, which spans from one side to the other side of the road. The valid crossing sequences are marked as a RoI grid, shown as the red boxes in Fig. 2, serving as a reference for bounding-box tracking and filtering. The definition of the RoI grid is based on domain-knowledge as the most relevant pedestrian crossing events are likely to occur in this region [8]. We exclude the image's edges from the RoI grid to reduce the influence of factors such as image distortion and vehicle motion (i.e., pitching and rolling) on pedestrian detection.

We utilize the BoT-SORT [1, 5] algorithm to track pedestrians, which combines a baseline model (SBS-S50) [18] from the open-source object identification library [14] with a Kalman Filter to achieve two-tier bounding-box tracking. Moreover, BoT-SORT proposes a camera motion compensation module to address the lack of sensor-related information, which is well-suited for our objective to process raw camera data directly. We choose the sparse optical flow [7] as the backbone for feature tracking and then estimate the background motion transformation from one frame to the next.

We filter potential pedestrian crossing events by selecting pedestrian crossing sequences that span continuously from one side of the road to the other. We apply the following two filtering criteria to the tracking results: (a) minimum and maximum x coordinates of the pedestrian bounding-box's central point must fall in the left and right pre-defined RoI clusters (red boxes in Fig. 2), and (b) the horizontal displacement of a pedestrian bounding-box has to be bigger than the specific threshold (half of the whole RoI region in our experiments) that we determined heuristically. We illustrate the process of pedestrian crossing information extraction as described above in Algorithm 1. Specifically, D represents the pedestrian's bounding-box information, such as the (x, y) coordinates, width, and height. The BoT-SORT configuration $conf$ contains the parameters for the first- and second-tier matching thresholds, the new tracking initiation threshold, and the tracking buffer frames.

Algorithm 1: Pedestrian Crossing Information Extraction.

Input: A full input image sequence S; BoT-SORT configuration $conf$; bounding-box coordinates of pre-defined RoI regions R_n; valid-pedestrian-crossing threshold τ.

Output: The bounding-boxes of valid crossing event E.

```
/* The detection and tracking of bounding-boxes */
```
1: **for** frame f_n in S **do**
2:     $D_n \leftarrow YOLO(f_n)$
3:     $uniqueID \leftarrow BoT\text{-}SORT(D_n, conf)$
4: **end for**
5: $D_{track} \leftarrow append(argwhere(D_n, uniqueID))$
```
/* The filtering of bounding-boxes */
```
6: **if** $min(x \in D_{track}) < R_{n/2} < max(x \in D_{track})$ **then**
7:     $D_{1st-criteria} \leftarrow D_{track}$
8: **end if**
9: **for** d in $D_{1st-criterion}$ **do**
10:     **if** $(max(x \in d) - min(x \in d)) > \tau$ **then**
11:         $D_{2nd-criterion} \leftarrow d$
12:     **end if**
13: **end for**
14: $E \leftarrow D_{2nd-criterion}$

3.3 Crossing Dimensionality Reduction

The previous processing step in PCICF provides us with semantically plausible candidates for the crossing event. However, the fact that pedestrians may merge or split into groups when crossing is a challenge for the classification. We need to further process such candidates to meet our research goal of not only knowing that we face a pedestrian crossing but also obtaining further descriptive information, such as type of pedestrian crossing. Hence, we
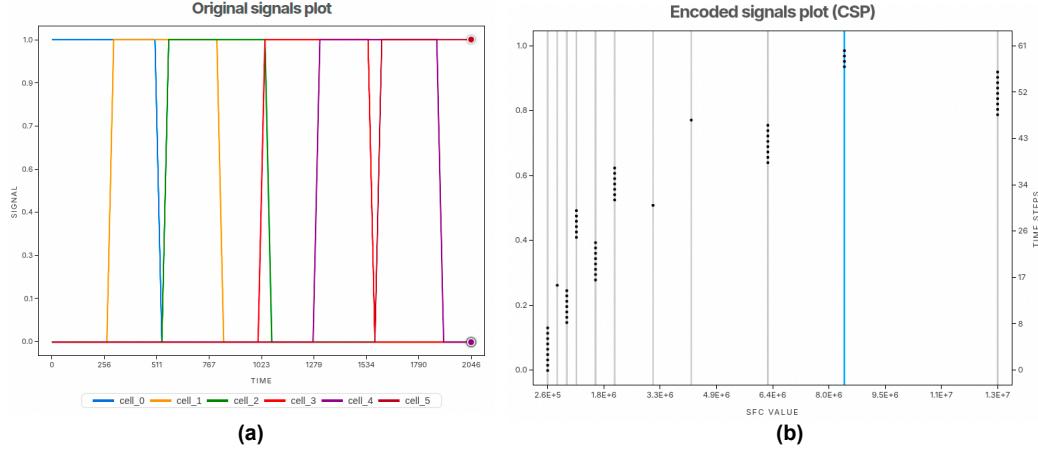
Figure 3: Data dimensionality reduction made with our AutoSFC tool[2] (same PIE dataset sequence as in Fig. 1): (a) shows the activation of the six RoIs (i.e., red boxes in Fig. 2), and (b) depicts the corresponding single-dimensional representation of the 6D-RoIs over time; the vertical stripes represent the crossing-specific fingerprint to be matched within MoreSMIRK.

need to match results from the previous processing step with the systematically constructed events in MoreSMIRK.

We leverage the Z-order SFC [3] to obtain fingerprints of crossing events that we use to look up potentially matching candidates in MoreSMIRK. We calculate fingerprints by transforming the bounding-box information across an entire sequence into a single-dimensional numerical representation that encodes the pedestrian's temporal and spatial occurrences. We first compute the intersection-over-union (IoU) between the valid crossing pedestrian's bounding-box and pre-defined RoI grids (the green and red rectangular boxes in Fig. 1), and then, we feed the IoU values to the Z-order SFC as input to calculate the fingerprints. Please note that overlapping pedestrians' bounding-boxes are only considered once in the process, and we apply the floor-ceiling threshold in this step to further reduce the SFC's computational complexity. For the entire input sequence, the SFC algorithm produces a single-dimensional numerical representation of the activation of each box in the RoI grid, in terms of temporal and spatial perspectives.

Fig. 3 visualizes the SFC transformation process with our tool AutoSFC[2]: Fig. 3(a) plots the original, multi-dimensional data from the previous pedestrian crossing extraction step in PCICF, and Fig. 3(b) depicts the corresponding, single-dimensional representation after applying the Z-order SFC over time. The X-axis in Fig. 3(a) is the time of a given sequence, and the Y-axis is the normalized IoU between the bounding-boxes and RoI grids. The six pre-defined RoI grids are represented by different colors, and their plots show the activation order of all boxes in the RoI grid for a given pedestrian crossing sequence. The X and Y axes in 3(b) are the Morton code and frame ID of the given sequence after dimensionality reduction, respectively. The SFC transformation represents spatio-temporal information of the activation of boxes in the RoI grid in a single-dimensional representation, which is traceable to the original signals.

3.4 Crossing Event Analysis

The final module in PCICF is to analyze the resulting fingerprints to obtain descriptions for the crossing events. By using SFCs to calculate such fingerprints, the identification and classification of pedestrian crossing events over time becomes the exploration of matching fingerprints within MoreSMIRK. The fingerprints describe spatiotemporal events in a multi-dimensional data-space as characteristic stripe patterns (CSPs) that emerge on a single-dimensional spectrum, intuitively referred to as barcodes; we use these CSPs to identify and classify pedestrian crossing events.

First, we need to transform the 104 pedestrian crossing sequences from MoreSMIRK via SFC into their corresponding CSPs to obtain the barcodes for the different crossing events contained in our dictionary. Please note that this step needs to be repeated only when the content of MoreSMIRK changes to retrieve the updated CSPs, and otherwise, it is a one-time generation step. We conduct this transformation by feeding MoreSMIRK into PCICF; the synthetic pedestrian and fixed-view background in the MoreSMIRK dataset reduces false detections in our crossing information extraction step. Furthermore, the steady kinetic characteristics of the MoreSMIRK dataset's pedestrians, such as speed and start/stop positions, ensure that the dimensionality reduction module can accurately determine the CSPs of each crossing event.

When analyzing the crossing event in a sequence, the final step in PCICF is to look up potentially matching spatiotemporal CSPs within MoreSMIRK. The similarity between the testing sequence's CSP and the one from MoreSMIRK reflects the tendency of a crossing sequence towards the one contained in our systematically constructed dictionary. In real-traffic scenarios, a pedestrian crossing sequence often exhibits multiple characteristics simultaneously. For instance, a group of occluded pedestrians starts crossing with different walking speeds, and hence, they split into several clusters during or at the end of the crossing. For such crossing sequences,

---

[2]https://beatrizcabdan.github.io/AutoSFC/

the number of pedestrians is dynamic from the ego vehicle's perspective. A novel contribution from PCICF is that, by systematically constructing varying pedestrian crossing configurations in MoreSMIRK and using fingerprints from SFC-transformed, multidimensional data, we can even identify and classify such ambiguous real-world pedestrian crossing sequences, as we report in our experiments (cf. Sec. 5). The output from the crossing event analysis step provides similarities of the testing sequence with all crossing events defined in MoreSMIRK. Furthermore, these similarities also represent the temporal duration of the testing sequence that falls in each pedestrian crossing classification.

4 Methodology for Evaluating PCICF

After introducing the design rationale behind PCICF, we outline its evaluation methodology based on the real-world dataset PIE [22].

4.1 The PIE Dataset

While PCICF contains MoreSMIRK as a systematically constructed dataset with synthetically generated pedestrian crossing events, we utilize the real-world PIE dataset [22] for evaluation. PIE is an openly available, large-scale dataset focusing on pedestrians in urban traffic scenes collected by York University, Canada. The video in PIE was captured with a wide-angle camera covering the entire pedestrian crossing area, and the dataset provides continuous, pedestrian-dense sequences as MP4 videos. The PIE dataset claimed to have addressed issues in other popular pedestrian-related datasets, such as KITTI [13] and JAAD [23], regarding the number of samples and short discontinuous chunks.

However, as the original purpose of the PIE dataset is to support pedestrian intention analysis, we reorganized and re-labeled it to make it applicable to classify pedestrian crossing events. We manually labeled the ground truth for pedestrian crossing events based on a similar principle that is used for constructing the MoreSMIRK dataset, i.e., start/end location of pedestrians, crossing direction, and group formation, if relevant; thereby, applying an objective labeling principle.

We provide a total of 21,327 camera frames spanning 158 sequences, classified into seven sub-categories based on the number of pedestrians and their crossing directions. Moreover, for complex pedestrian crossing sequences involving multiple pedestrians and directions, we provide detailed textual descriptions to indicate critical information, such as 'four pedestrians in three clusters' and 'two pedestrians with gap from left to right, one pedestrian from right to left'. Table 2 presents an excerpt of the detailed properties of the PIE dataset that we used for the evaluation of PCICF.

4.2 Experimental Setup for Evaluating PCICF

We conduct all experiments on a portable computer powered by an Intel(R) Core(TM) i5-9400 6-core CPU and an NVIDIA GTX 1660 Ti GPU. The parameter configurations of PCICF are crucial for achieving efficient and accurate classification. Please note that the framework's domain-specific parameters, such as RoI location, can be adjusted based on the application context. In our experiments, the input image size is 640×480, and each RoI box is 120×60.

Table 2: The properties of part of the PIE dataset used in this work (Ped, L, and R indicate pedestrian, left, and right, respectively).

Subset	Number of frames	Number of sequences
Single Ped L to R	6649	55
Single Ped R to L	6769	56
Multiple follow-up Peds L to R	1211	9
Multiple follow-up Peds R to L	1404	11
Multiple no-follow-up Peds L to R	1859	10
Multiple no-follow-up Peds R to L	1520	9
Multiple Peds both directions	1915	8

The thresholds for the two-tier bounding-box tracking in the BoT-SORT algorithm are 0.25 and 0.1, respectively. For the bounding-box filtering module, the valid-pedestrian-crossing threshold τ, representing the displacement of the pixels to be regarded as a valid crossing, is 180. We share all parameters from the experimental setup in our open-source GitHub repository.

The metrics used to indicate classification results are the similarities of the Morton codes for a sequence from the PIE dataset and the corresponding sequences from the MoreSMIRK-based dictionary. The numerical values and sequential order of Morton codes represent spatial and temporal information, which are two critical characteristics when computing the similarity percentages. We feed all 158 sequences from PIE into PCICF and report on the similarity between the manual annotations and the results after matching the fingerprints from PIE with those in MoreSMIRK.

4.3 Threats to Validity

Our work consists of two parts: (a) a systematically constructed dataset containing the dictionary of pedestrian crossing events, and (b) algorithms and prototypical implementation to process input frames, extract, and match potential pedestrian crossing sequences with the aforementioned crossing event dictionary. Potential threats may originate from these two aspects of PCICF.

The pedestrian crossing event dictionary is produced based on the MoreSMIRK dataset for precision and minimal interference, but this may also potentially result in a limited number of pedestrian event configurations. For instance, the size and moving speed of synthetic pedestrians in the MoreSMIRK dataset are fixed. However, the distance between pedestrians and the ego vehicle varies in the real world, leading to cases where multiple pedestrians are treated as a single pedestrian (i.e., when they are far away from the vehicle or within the same RoI grid), and vice versa. Therefore, the classification for multiple non-following pedestrians can be biased towards single or follow-up events. The typical examples are sequence 161, 342, and 350 in Table 6. The similarity of biased events is more dominant than the ground truth (90% against 63%, 80% against 54%, and 90% against 63%, respectively). Furthermore, and similar to the pedestrian intention aspect of the original PIE dataset, annotating pedestrian crossing events is a subjective topic to some extent, especially for complex pedestrian crossing sequences, which can be interpreted into different event classes. We follow the principle of the PIE dataset's original pedestrian intention annotation to
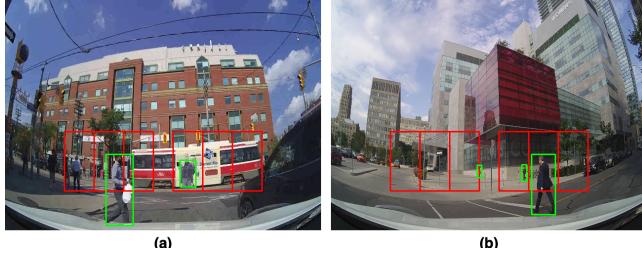
Figure 4: Misclassifications of single-directional crossing for single pedestrian: The ground truth for (a) and (b) are '_ _ X; N/A; _ _ _' and '_ _ _; N/A; Y _ _', respectively.

manually label its sequences with neutral descriptions of pedestrian crossing events.

For the prototypical implementation, the choice of algorithms and related parameters may affect the accuracy and efficiency of classification. For instance, YOLO's weak performance on some pedestrian crossings in Tables 3 and 4 contributes to its failure to classify sequences in which pedestrians are far away from the ego vehicle. To cope with this limitation, the parameters for pedestrian tracking and filtering in PCICF would need to be calibrated and fine-tuned. Moreover, the tracking algorithm is the bottleneck for the classification of complex multiple-pedestrian both-directional crossings due to the heavy conclusions and background objects. These domain-specific nuances pose challenges for generalizability.

5 Results

This section describes the experimental results and analysis of applying PCICF to identify and classify pedestrian crossing sequences. We utilize PCICF to analyze a real-world dataset and then compare the similarity between of those results with the systematically constructed pedestrian crossing dictionary, MoreSMIRK. We present here the basic information on environmental and parameter setup, along with some insights from the experimental results.

As shown in Table 2, the majority of the PIE dataset consists of single pedestrian crossings, which is partially due to the real traffic at locations where the dataset was collected. For the scenarios of single pedestrian crossing from left to right and from right to left, corresponding to event 0 and event 4 in the MoreSMIRK dataset, respectively, our PCICF achieves accuracies of 85% and 80% to precisely map the PIE dataset sequences to corresponding events in MoreSMIRK. The primary reason for false classifications is that pedestrians at a distance are not correctly filtered or detected. Fig. 4 shows two typical false detections for single pedestrian crossing sequences. Fig. 4 (a) is the single pedestrian crossing from left to right. The human printed in the billboard on the tram is falsely detected; thus, PCICF returns the higher similarity for event 8 ('_ _ X; N/A; Y _ _ ' in Table 1) at 66% than the ground truth event 0 ('_ _ X; N/A; _ _ _') at 54%. Fig. 4 (b) is the sequence that crosses from right to left. The far-away pedestrian detection results in the similarity for event 5 ('_ _ _; N/A; Y Y _') at 70% against the ground truth event 4 ('_ _ _; N/A; Y _ _'), which is at 54%.

Table 3 and Table 4 show PCICF's classification results for sequences of multiple follow-up pedestrians crossing from left to

Table 3: Follow-up pedestrians (peds) crossing from left to right. PCICF results show the similarity of the event.

Seq. ID	Start-End Frame	PCICF Results	Manual Annotation
003	598-690	E1: 70% E3: 54%	two peds
012	160-230	E1: 80% E3: 63%	two peds merge to one
021	139-200	E0: 54% E1: 70%	two peds
034	463-602	N/A	far away, challenging
036	232-479	E0: 63% E1: 50%	two peds
037	434-637	E0: 54 %	two peds merge to one
276	22-87	E0: 63% E1: 70%	two peds
292	281-400	E0: 72% E1: 80%	two peds merge to one
430	192-538	E0: 90% E1: 80%	two peds merge to one then split to two

right, and right to left, respectively. We invoke a series of criteria to filter PCICF's results for multi-pedestrian single-directional crossing events: (a) only consider the similarity over 50%; (b) discard other results if the similarity of one event is 100%; (c) pick the first two events with the highest similarity percentages. Most of the follow-up crossing sequences involve only two pedestrians, PCICF returns the highest similarity with event 1 ('_ X X; N/A; _ _ _') and event 5 ('_ _ _; N/A; Y Y _'), which corresponds to their crossing event ground truth. However, as discussed in Section 4.1, real-world pedestrian crossing sequences can exhibit characteristics of multiple events simultaneously, i.e., pedestrians split or merge during the crossing from the ego vehicle's visual view. For these sequences with dynamic events (sequence 037 and 430 in Table 3, or 008 and 009 in Table 4), PCICF can indicate a relatively high similarity with another event. For sequences containing more than two pedestrians, PCICF classifies them into events that are more consistent with real-world situations. In Table 4, sequence 045 has four pedestrians, PCICF shows a dominating 81% similarity of event 7 ('_ _ _; N/A; Y Y Y'); the multiple pedestrians in sequence 128 are from far to close in vehicle's ego view, such change equals to the occupation of RoI grids from one to two, thus, PCICF returns a 54% similarity of event 4 ('_ _ _; N/A; Y _ _') and 50% of event 5 ('_ _ _; N/A; Y Y _').

Table 5 and Table 6 present the classification results for multiple pedestrians who do not follow each other to cross in a single direction. As mentioned in Section 3.1, our crossing event classification only includes up to three follow-up pedestrians from one side, as it would occupy almost half of the ego vehicle's front-facing field-of-view. Thus, the sequence that has more than three pedestrians, for instance, sequence 217 in Table 5, shows a dominating similarity of 72% to event 3 ('X X X; N/A; _ _ _'). For the manual annotation of multiple no-follow pedestrians crossing, a significant drawback is that the gap between pedestrians is represented in a subjective and relative tone. For instance, sequence 555 in Table 5, the human annotator describes the gap with the word 'large'

Table 4: Follow-up pedestrians (peds) crossing from right to left. PCICF results show the similarity of the event.

Seq. ID	Start-End Frame	PCICF Results	Manual Annotation
008	571-697	E4: 81% E5: 70%	one ped splits to two
009	50-210	E4: 81% E5: 70%	one ped splits to two
011	621-731	E5: 100%	two peds merge to one
045	343-410	E5: 50% E7: 81%	four peds
128	101-232	E4: 54% E5: 50%	four peds, far to close
280	280-419	N/A	far away, challenging
282	1-166	E4: 72% E5: 80%	two peds merge to one
391	73-175	E5: 100%	two peds
391	176-271	E5: 100%	two peds
546	440-590	E4: 72% E5: 70%	two peds

Table 5: No-follow-up pedestrians (peds) crossing from left to right. PCICF results show the similarity of the event.

Seq. ID	Start-End Frame	PCICF Results	Manual Annotation
012	710-790	E1: 90% E3: 63%	three peds, two follow-up, one with gap
048	1-274	E0: 54% E1: 50%	two peds start with large gap, merge to one at the end
161	655-871	E3:54%	two peds with gap
217	1-150	E2: 58% E3: 72%	four peds
276	22-87	E0: 63% E1: 70%	two peds start with large gap, merge to one at the end
292	1-280	E0: 81% E1: 50%	three peds in different speeds
313	1-101	E1: 50% E3: 63%	three peds
419	230-584	E2: 82% E3: 81%	three peds clusters with gap
445	501-700	E2: 76% E3: 81%	two peds clusters with gap
555	1-144	E0: 45 % E2: 47%	two peds with large gap

Table 6: No-follow-up pedestrians (peds) crossing from right to left. PCICF results show the similarity of the event.

Seq. ID	Start-End Frame	PCICF Results	Manual Annotation
024	49-247	E5: 70% E7: 81%	two peds start with gap, merge to one at the end
029	456-599	E40: 70% E56: 80%	two peds with large gap, half crossing
116	129-247	E5: 100%	four peds in three clusters
161	497-654	E5: 90% E7: 63%	two peds with gap
292	656-807	E5: 60% E6: 58%	three peds in two clusters with gap
342	564-720	E5: 80% E7:54%	two peds with gap
350	520-795	E5: 90% E7: 63%	two peds with gap
531	33-181	E7: 72%	two peds with gap
543	522-687	E93: 70% E103: 78%	four peds in two clusters with large gap

Table 7: Multiple pedestrians (peds) crossing both directions. L and R indicate left and right, respectively. PCICF results show the similarity of the event.

Seq. ID	Start-End Frame	PCICF Results	Manual Annotation
025	551-800	E11: 31%	four peds L to R one ped R to L
031	1-253	E72: 80%	four peds L to R one ped R to L
377	451-671	E88: 70%	one ped L to R two peds R to L
377	91-357	E73: 30%	two peds L to R (gap) three peds R to L (gap)
409	567-822	E74: 31%	two peds L to R (gap) one ped R to L
418	404-650	E23: 35%	one ped L to R three ped clusters R to L
537	1-240	E40: 23%	one ped L to R (offset) two peds R to L (large gap)
540	96-276	E92: 30%	one pedestrian L to R one pedestrian R to L (offset)

relative to other similar sequences. In contrast, PCICF reports the almost equal similarities (45% and 47%) of event 0 ('_ _ X; N/A; _ _ _') and event 2 ('X _ X; N/A; _ _ _'), which indicate the gap is large enough that the sequence can be regarded as two single-pedestrian crossing sequences. The major false detections occur in sequences with abnormal crossing patterns, i.e., sequence 029 in Table 6, where pedestrians cross half of the road with a large gap, and sequence 543, where four pedestrians have irregularly large gaps. PCICF returns the classification of both-directional crossing

with high offset Φ. The reason for such false classifications is that pedestrian tracking is incorrect due to occlusion and speed changes.

Table 7 shows PCICF's classification results for complex multiple pedestrians crossing in irregular patterns from both directions, which is the most compelling real-world case. Please note that these sequences usually occur at intersections or in urban areas with busy traffic. We notice that factors such as background objects and occlusion pose significant challenges for PCICF's pedestrian detection and tracking modules, which result in relatively low similarities and misclassifications of offset Φ. Instead of setting multiple criteria as the single-directional crossing scenarios, we report the event classification with the highest similarity. For some sequences with

simpler scenarios, PCICF returns a trustworthy classification. For instance, sequence 025 and 540's classification results are event 11 ('X X X; N/A; Y _ _') and event 92 ('_ _ X; 5; Y Y _'), which generally correspond to the ground truth from human annotators.

6 Discussion

Grounded on the aforementioned results, this section addresses the research questions. The first research question focused on the design decisions to create a synthetic dataset for pedestrian crossing classification. First and foremost, MoreSMIRK's strength lies in being a synthetic dataset that has been systematically constructed and can be systematically augmented. Since the focus is on pedestrian crossing scenarios, any crossing direction and group configuration is guaranteed to be covered, as opposed to previous attempts to gather video data, such as PIE [22] or Waymo [19], which rely on chance to capture as many pedestrian motion patterns as possible.

> **Answering RQ-1**: The design of a synthetic dataset for event classification should be based on a systematic exploration of the event space, and should be scalable. The goal is to allow for the augmentation with unique pedestrian crossing events, with specific crossing directions or group configurations.

It is important to note that while MoreSMIRK focuses on pedestrian motion, these principles apply to classifiers for any event type: the systematic construction of the event dictionary allows synthetic datasets to be extensible and adaptable to new configurations.

In turn, the second research question concerned the design decisions required to develop a traffic event identification and classification framework. To do that, the proposed framework must cover the full process from sensory data (e.g., from dashboard cameras) to an end-user-readable classification result. Given the many processing steps needed in such pipelines:

> **Answering RQ-2**: Traffic event identification and classification frameworks should rely on a modular and scalable processing pipeline with interchangeable and extendable components.

For instance, in the proposed PCICF, the YOLO component could be replaced with another method or algorithm if a specific performance level was required for the preliminary pedestrian detection. Similarly, the specific SFC-based algorithm for dimensionality reduction could be replaced w another technique [3].

Finally, the third research question focuses on analyzing the performance of the proposed framework. In that regard, the goal of the proposed MoreSMIRK dataset was to mimic real-world pedestrian crossing scenarios that are challenging in terms of motion direction or group size and pattern. Each systematically constructed and specific motion pattern was then encoded into the MoreSMIRK dictionary and made available to PCICF. Based on the results presented in section 5, we address the research question:

> **Answering RQ-3**: PCICF identifies single-directional crossings with high accuracy, especially single and two-follow-up pedestrians cases. However, complex multiple pedestrian two-directional crossing configurations pose challenges for PCICF.

As shown in section 5, the detections and classifications mostly coincide with the manual annotations of the pedestrian crossings in

the case of single-directional pedestrian crossings. However, as seen in Table 7, more complex pedestrian crossing events have a chance for misclassification. Such events often involve heavy occlusions and background objects, however, they are also challenging to other state-of-the-art algorithms for object detection and tracking. While such algorithms use rather end-to-end AI/ML components, PCICF's focus is on computational efficiency. A possible extension of the PCICF pipeline could be to include a module for pedestrian crossing intention detection [25]. This could serve as a prior for PCICF to improve the accuracy of the classifications.

7 Conclusions and Future Work

In urban scenarios, ADAS and AD systems must safely interact with pedestrians crossing the road in various configurations, i.e., groups and speeds. Standards such as ISO 26262 and ISO 21448 guide the development and assessment of these safety-critical systems, whose robustness must be assessed systematically.

This work proposes a framework, called PCICF, to identify and classify pedestrian crossings. PCICF incorporates several state-of-the-art modules into its internal pipeline, including the well-known YOLO, whose detections are fed into our algorithm to semantically filter potential pedestrian crossing matches. The pipeline also leverages SFCs to reduce pedestrian crossing sequences into their corresponding single-dimensional representations while preserving their spatiotemporal information; this allows the creation of fingerprints that visually take the shape of barcodes [4], which serve as keys in a dictionary of pedestrian crossing events called MoreSMIRK. The MoreSMIRK dataset, which is systematically constructed and publicly available, is used in the last module of the PCICF pipeline to obtain the final classification.

To evaluate the proposed PCICF, we rely on the large-scale, real-world dataset PIE, which contains multiple pedestrian crossing examples. In this dataset, PCICF achieves a maximum accuracy of 85% in scenarios with dominating unique classifications, such as uni-directional crossings of a single pedestrian. On the other hand, for scenarios with ambiguous group patterns, PCICF does not provide a unique match, but instead reports the likelihoods for each crossing event type as defined in MoreSMIRK. Thanks to this, PCICF can also identify sub-patterns (i.e., multiple single-pedestrian crossings) within such ambiguous crossing configurations.

It is important to note that while the accuracy of PCICF, specially in challenging scenarios, might not surpass other state-of-the-art solutions, the computational efficiency of the individual algorithms in PCICF, like SFC, make it usable onboard of AVs, which typically have limited computational resource for OOD detection, for example (i.e., checking whether a configuration is not covered in the look-up dictionary). Moreover, given that the look-up dictionary is constructed systematically, as described in subsection 3.1, more configurations and motion patterns could be added to the look-up dictionary to expand the OOD detector.

As previously discussed, complex two-directional multiple pedestrian scenarios, common in crowded intersections, still pose challenges for PCICF. Thus, refining the pedestrian crossing event dictionary by systematically augmenting the range of pedestrian motion patterns, and improving the pedestrian tracking will be the focus of future work. For example, additional vulnerable road user types,

such as e-scooter riders and people with disabilities, should be included in future editions of MoreSMIRK and PCICF to increase the inclusiveness of technology in our transportation systems. These improvements are planned be added to the open-source PCICF on GitHub, and to MoreSMIRK, hosted at AI Sweden, to foster further important research around VRUs.

Acknowledgments

Disclaimer

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of Volvo Cars.

References

[1] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. 2022. BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *arXiv preprint arXiv:2206.14651* (2022).

[2] Mina Alibeigi, William Ljungbergh, Adam Tonderski, Georg Hess, Adam Lilja, Carl Lindstrom, Daria Motorniuk, Junsheng Fu, Jenny Widahl, and Christoffer Petersson. 2023. Zenseact Open Dataset: A large-scale and diverse multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

[3] Michael Bader. 2012. *Space-filling curves: an introduction with applications in scientific computing*. Vol. 9. Springer Science & Business Media.

[4] Christian Berger, Beatriz Cabrero-Daniel, M. Cagri Kaya, Maryam Esmaeili Darestani, and Hannah Shiels. 2023. Systematic Evaluation of Applying Space-Filling Curves to Automotive Maneuver Detection. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. 2932–2938. doi:10.1109/ITSC57777.2023.10422366

[5] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. 2016. Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP)*. Ieee, 3464–3468.

[6] Markus Borg, Jens Henriksson, Kasper Socha, Olof Lennartsson, Elias Sonnsjö Lönegren, Thanh Bui, Piotr Tomaszewski, Sankar Raman Sathyamoorthy, Sebastian Brink, and Mahshid Helali Moghadam. 2023. Ergo, SMIRK is safe: a safety case for a machine learning component in a pedestrian automatic emergency brake system. *Software quality journal* 31, 2 (2023), 335–403.

[7] J.-Y. Bouguet. 1999. Pyramidal implementation of the lucas kanade feature tracker. https://api.semanticscholar.org/CorpusID:9350588

[8] Tayssir Bouraffa, Elias Kjellberg Carlson, Erik Wessman, Ali Nourir, Pierre Lamart, and Christian Berger. 2024. Comparing Optical Flow and Deep Learning to Enable Computationally Efficient Traffic Event Detection with Space-Filling Curves. In *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 199–206.

[9] Beatriz Cabrero-Daniel, Ahmed Yasser Abdelkarim, and Axel Broberg. 2024. Digital Twins for Early Verification and Validation of Autonomous Driving Features: Open-source Tools and Standard Formats. In *2024 IEEE Intelligent Vehicles Symposium (IV)*. 2477–2482. doi:10.1109/IV55156.2024.10588808

[10] Robert N. Charette. 2009. This Car Runs on Code - IEEE Spectrum. https://spectrum.ieee.org/this-car-runs-on-code

[11] Peishan Cong, Xinge Zhu, Feng Qiao, Yiming Ren, Xidong Peng, Yuenan Hou, Lan Xu, Ruigang Yang, Dinesh Manocha, and Yuexin Ma. 2022. STCrowd: A multimodal dataset for pedestrian perception in crowded scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19608–19617.

[12] European Commission, Directorate-General for Communications Networks, Content and Technology. 2021. *EUR-Lex - 52021PC0206 - EN - EUR-Lex*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[13] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 3354–3361.

[14] Lingxiao He, Xingyu Liao, Wu Liu, Xinchen Liu, Peng Cheng, and Tao Mei. 2023. Fastreid: A pytorch toolbox for general instance re-identification. In *Proceedings of the 31st ACM International Conference on Multimedia*. 9664–9667.

[15] ISO 21448 2022. *ISO 21448, Road vehicles — Safety of the intended functionality*. Standard. International Organization for Standardization.

[16] ISO/PAS 8800:2024 2024. *ISO/PAS 8800:2024 Road vehicles — Safety and artificial intelligence*. Standard. International Organization for Standardization.

[17] Glenn Jocher. 2020. *Ultralytics YOLOv5*. doi:10.5281/zenodo.3908559

[18] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. 2019. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[19] Jieru Mei, Alex Zihao Zhu, Xinchen Yan, Hang Yan, Siyuan Qiao, Yukun Zhu, Liang-Chieh Chen, Henrik Kretzschmar, and Dragomir Anguelov. 2022. Waymo Open Dataset: Panoramic Video Panoptic Segmentation. doi:10.48550/ARXIV.2206.07704

[20] Ali Nouri, Christian Berger, and Fredrik Törner. 2023. On STPA for Distributed Development of Safe Autonomous Driving: An Interview Study. In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 5–12.

[21] Ali Nouri, Beatriz Cabrero-Daniel, Fredrik Törner, and Christian Berger. 2025. The DevSafeOps dilemma: A systematic literature review on rapidity in safe autonomous driving development and operation. *Journal of Systems and Software* 230 (2025), 112555. doi:10.1016/j.jss.2025.112555

[22] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John K Tsotsos. 2019. Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*. 6262–6271.

[23] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. In *Proceedings of the IEEE international conference on computer vision workshops*. 206–213.

[24] Kasper Socha, Markus Borg, and Jens Henriksson. 2022. SMIRK: A Machine Learning-Based Pedestrian Automatic Emergency Braking System with a Complete Safety Case. *Software Impacts* 13 (2022). doi:10.1016/j.simpa.2022.100352

[25] Chi Zhang and Christian Berger. 2023. Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review. *IEEE Transactions on Intelligent Transportation Systems* 24, 10 (2023), 10279–10301. doi:10.1109/TITS.2023.3281393