Evaluation of Pedestrian Behaviour in 157 Cities with 285 Hours of Dashcam Footage from YouTube

Md Shadab Alam*
m.s.alam@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Marieke Martens
m.h.martens@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands
TNO
Helmond, The Netherlands

Pavlo Bazilinskyy
p.bazilinskyy@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Аннотация

Interaction between future cars and pedestrians should be designed to be understandable and safe globally. While previous research has studied vehicle-pedestrian interactions within specific cities or countries, this study offers a more scalable and robust approach by examining pedestrian behaviour worldwide. We present a dataset, "PYT which includes 285 hours of day and night dashcam YouTube footage from 157 cities and 59 countries. We detected pedestrian movements, focusing on the speed and the pedestrian crossing decision time during road crossings based on the bounding boxes given by YOLO. Videos were carefully selected based on specific criteria to ensure urban settings and adequate pedestrian interactions. Results revealed statistically significant cross-cultural variations in pedestrian behaviour influenced by socioeconomic and environmental factors such as Gross Metropolitan Product (GMP), traffic-related mortality and literacy. The dataset is publicly available to encourage further research into global pedestrian behaviour.

Keywords

Human Behaviour, Pedestrian Safety, Cross-Country Evaluation, Computer Vision, Dataset

ACM Reference Format:
Md Shadab Alam, Marieke Martens, and Pavlo Bazilinskyy. 2024. Evaluation of Pedestrian Behaviour in 157 Cities with 285 Hours of Dashcam Footage from YouTube. In . ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Most studies in the domain of Human Factors share processed results and insights, with no data published in open access. In recent years, there has been some improvement in the accessibility of datasets for automated driving (AD) applications. Data collection for AD research mostly assumes the acquisition of expensive hardware, [49], which is often unreliable and requires considerable time and resources. For example, Dingus et al. [14] collected 43,000 hours of data with 100 instrumented vehicles driving in the USA for 12 months. Such studies are mostly geographically constrained, limiting the potential to capture diverse, cross-cultural variations in behaviour.

Conference'17, July 2017, Washington, DC, USA
2024. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1.1 Pedestrian Behaviour

Pedestrians and modern cars share roads in urban environments. In 2019, vulnerable road users (VRUs, such as pedestrians and cyclists) accounted for 29% of road fatalities in the EU [16]. These fatalities often occur due to the driver not being able to understand the intentions of the pedestrian [43]. Sometimes, such fatalities also occur as the driver and the pedestrian belong to different cultures, e.g., in the case of the driver operating a rental car in a foreign country [48].

Walking constitutes a significant portion of urban traffic. Particularly in densely populated cities in the developing world, where walking remains a primary mode of transportation. A study by LUTP (Leaders in Urban Transport Planning) [47] indicate that in major Indian cities, between 25% and 50% of trips are made entirely on foot, while in major African cities, this figure reaches around 50%. Even in terms of distance travelled, walking accounts for over 50% of all trips in countries like Tanzania. Unlike vehicles, which follow more predictable patterns dictated by traffic rules and road designs, pedestrians exhibit a wide range of behaviours that can be difficult to anticipate. This variability is influenced by numerous factors, including individual decision-making, environmental context, cultural norms, and social interactions [13]. As a result, accurately modelling pedestrian behaviour in traffic scenarios presents a considerable challenge. Pedestrians may make sudden stops, change directions, or engage in unpredictable actions, such as crossing outside designated crosswalks or interacting with their surroundings in ways that are hard to quantify. Such unpredictable behaviour complicates efforts to develop advanced driver-assistance systems (ADAS) and AD technology that can effectively and safely interact with pedestrians on the road.

Studies by Bazilinskyy et al. [3–5], Saffo et al. [42] and Alam et al. [1] leverage crowdsourcing experiments to analyse human decision-making across different cultures. However, these studies often lack realism due to their controlled experimental conditions. Similarly, research utilising Virtual Reality (VR) and driving simulators, such as the studies by Onkhar et al. [33] and Bazilinskyy et al. [6], faces challenges due to hardware constraints that limit their applicability across different global contexts. Moreover, these studies often focus on a specific set of scenarios predefined by the researchers, neglecting other variables such as diverse environmental conditions, the time of day, the age of the pedestrian, and various unexpected or spontaneous events that can significantly influence pedestrian behaviour and safety outcomes. This gap underscores the need for more comprehensive research methodologies that consider a broader range of real-world conditions and pedestrian demographics.

Pedestrian hesitation [21] at crossings is a significant indicator of both perceived and actual safety risks present in urban traffic environments. Hesitation often occurs when pedestrians intend to cross the road but pause due to traffic conditions, uncertainty, or other inhibitory factors. These moments of hesitation are crucial for understanding pedestrian behaviour and the effectiveness of road safety measures. Harkey et al. [20] found that pedestrian hesitation at roundabouts is common, particularly at exit legs where uncertainty about vehicle yielding increases. Similarly, Jay et al. [21] studied pedestrian hesitation when crossing illegally at a red light. They found that hesitation often results in pedestrians either abandoning the crossing or accelerating to avoid an oncoming vehicle. The study highlights that hesitation increases the risk of accidents, particularly when pedestrians misjudge the situation or follow social cues without verifying traffic conditions, leading to unsafe crossing decisions.

Understanding the speed at which pedestrians cross roads is critical for assessing pedestrian behaviour, traffic flow, and road safety [18]. Crossing speed can indicate the level of comfort or urgency pedestrians feel in urban environments, and it is a key factor in designing pedestrian-friendly infrastructure. This approach provides quantitative insights into pedestrian crossing speed under various urban conditions, including differing levels of traffic congestion, time of day, and pedestrian demographics. Such data is invaluable for urban planners and traffic engineers who aim to enhance road safety and optimise pedestrian traffic management systems. By analysing crossing speeds, we can better understand pedestrian preferences and behaviours, ultimately contributing to more effective and safer urban environments.

Another variable is the time it takes for a pedestrian to decide to cross the road, which is a critical measure that reflects both individual behaviour and broader traffic environment conditions. To the best of our knowledge, no prior studies have measured this variable either at a local or global level. Crossing decision time can indicate the perceived safety, the level of caution, and the overall efficiency of crossing locations. This metric is pivotal for evaluating the impact of traffic signals, signage, and other urban infrastructure on pedestrian behaviour. By quantitatively assessing crossing decision times, we gain valuable insights into how different variables, such as traffic density, time of day, and pedestrian density, influence pedestrian decisions at crossings. This information is crucial for urban and traffic planners seeking to improve pedestrian safety and optimise traffic flow, thereby enhancing the overall efficiency and safety of urban transport systems.

1.2 Datasets for Analysing Pedestrian Behaviour

Soon, future traffic will be mixed, comprised of manually-driven (MDVs) and automated vehicles (AVs), as well as VRUs [7, 30]. The algorithms deciding on the behaviour of AVs are being designed by computer-human interaction experts in both industry and academia today [2, 24]. Often, decisions are made based on the verification done with a limited base of potential users comprised of individuals from a single city/country/culture. However, such algorithms must be scalable to all cultures, and it is important to understand how pedestrians behave cross-culturally.

The field of computer vision (CV) began in the 1960s, initially focusing on the challenge of dividing, defining, and identifying backgrounds and objects within a scene [37]. However, a significant breakthrough occurred in 2012 with the introduction of AlexNet [23], a deep neural network that dramatically improved image recognition accuracy by substantially reducing error rates. With these advancements in technology, researchers started employing vehicle-mounted cameras to gather extensive video data from driving in urban environments. Pivotal datasets such as KITTI [19], NuScenes [9], One Thousand and One Hours [28], Caltech Pedestrian detection benchmark [15], Pedestrian Intention Estimation (PIE) [38], Waymo Open Dataset [44], ApolloScape Auto [46], Cityscapes [12], A*3D dataset [35], Argoverse [10] and others have become standard benchmarks that support a variety of tasks in CV and AD research. Studies such as Oxley et al. [34] and Rasouli et al. [39] used videos to investigate the intentions of pedestrians in traffic. Oxley et al. discovered that crossing decisions were primarily based on vehicle distance and less on the time of arrival of the vehicle. Rasouli et al. trained a neural network for the classification of pedestrians who were looking to cross or walking. Similarly, Mordan et al. [32] used the JAAD dataset [40] to detect 32 attributes for a pedestrian. Meanwhile, Vajgl et al. [45], and Mauri et al. [29] used KITTI datasets to train their algorithm to detect cars, pedestrians and cyclists in traffic with distance estimation.

On-road studies involving instrumented vehicles for the collection of traffic data often limit their scope to one or few specific cities or countries due to (1) high costs, (2) the complexity of the method, (3) the use of specialised hardware and closed-source software, and (4) difficulties with obtaining the ethics approval for cross-country research. Another critical limitation of the existing datasets is the scarcity of research on the cultural and regional differences in pedestrian behaviour, particularly at the city level across the globe. For example, NuScenes features scenes from Singapore and Boston, while Argoverse focuses on Miami and Pittsburgh. Datasets such as Waymo Open Dataset attempt to cover some parts of the USA, namely San Francisco, Detroit, Seattle, Phoenix, Los Angeles and Mountain View. Similarly, D^2-dataset [11] encompasses five cities from China, whereas datasets like KITTI only contain scenes from Karlsruhe (Germany). Bazilinskyy et al. [5] compared risk perception by showing dashcam video to participants across the world through crowdsourcing, while Bellone et al. [8] conducted a user experience of automated public transport in the cities around the Baltic Sea namely Finland, Estonia, Norway and Gdansk. While these studies begin to address regional and contextual differences, they often remain focused on limited geographical areas or specific aspects of the issue.

1.3 Aim of study

As shown above, the present literature lacks an understanding of how pedestrians behave in different cities (and countries). The present study aims to provide a dataset **PYT** ("Pedestrians in YouTube") that will entail cross-cultural insights into traffic behaviour across different cities worldwide. We achieve this goal by collecting 285 hours and 57 minutes of day and nighttime dashcam footage from 157 cities from the video hosting platform YouTube (https://www.youtube.com). The dataset was used to analyze pedestrian crossing speed and time to initiate crossing without the use of any pre-trained neural networks, relying solely on object detection in each video frame. Furthermore, the study's objective is to explore the difference in traffic behaviour and compare traffic behaviour and city-specific parameters, namely: (1) Gross Metropolitan Product (GMP) per capita, (2) rate of traffic-related mortality, and (3) level of literacy.

2 Method

2.1 Dataset

Numerous videos of driving footage from the perspective of the dashcam are available on the Internet, with an absolute majority placed on YouTube. Sharing "special"dashcam traffic videos is a phenomenon where people exchange them for amusement. A notable example is a Telegram group BadShofer (https://t.me/s/badshofer) with ~43,800 users and ~48,700 shared videos as of September 1, 2024. Most such videos are short and contain isolated footage of abnormal events like accidents and unexpected behaviour of participants in traffic. The angle of the camera is often not constant throughout such videos. For this study, we collected dashcam footage, which is representative of regular driving situations. Numerous users on YouTube share relatively long (often longer than 40 min) dashcam videos made with professional recording equipment (e.g., https://www.youtube.com/@jutah with ~787,000 followers and 781 uploaded videos, as of September 1, 2024); likely as they can be used as "background noise"by viewers, generating income for the authors. As these videos are hosted on YouTube, they are under the Fair use on YouTube[1], which allows their use for research. The research was approved by the Human Research Ethics Committee of the Eindhoven University of Technology. To populate the dataset with videos from YouTube, we defined the following inclusion criteria:

- **C1**: A continuous focus on urban settings, excluding clips featuring highways or rural routes.
- **C2**: A minimum video duration of 10 minutes to ensure adequate coverage of urban environments.
- **C3**: Avoidance of atypical events, such as accidents or special events, which do not represent everyday conditions.
- **C4**: A population threshold of at least 50,000 for the cities of interest to guarantee sufficient pedestrian interactions.

It is important to note that YouTube's license allows the use of content for research under its Fair use policy. In this study, we gathered dashcam videos from YouTube, adhering to these licensing conditions.

The authors included videos that passed the C2 criterion from the results of search queries on YouTube, where cities were input individually. The queries were as follows: "*dashcam video in [city name]* "*driving videos in [city name]* "*dashcam videos in cities*"and some queries in the regional language like "*відео з реєстратора*"("video from dashcam"in Ukrainian) and "*видео с регистратора*"("video from dashcam"in Russian). If a part of the video does not satisfy the criteria C2, the part was removed from the study. Furthermore, if the video contains scenes from both daylight and nighttime, the video was split in two based on the street lights. The search for videos took place between February 3, 2024, and April 15, 2024. Each video selected in this study went through visual inspection by the authors to verify whether criteria C1, C3 and C4 were satisfied. A total of 338 videos from 157 cities across 59 countries were included. See Figure 1 for the overview of included countries. They represent 76.67% of the global population. Of these videos, 267 featured daytime driving, 62 showed nighttime driving, and 4 included footage from both day and night. The combined length of collected footage was 285 h 58 min. The videos were downloaded from YouTube using the pytube library [2] in resolution 1280 x 720 px. Supplementary material contains the codebase used in this work.

The videos in the dataset were then analysed with the You Only Look Once (YOLO) algorithm [41], specifically YOLOv8 [22]. YOLO is free to use and is renowned for its speed and accuracy in real-time object detection. It processes images in a single pass, predicting bounding boxes and class probabilities simultaneously. It can detect objects from 80 different classes and provides the width (W), height (H), and centre coordinates of each bounding box (X-center, Y-center). The output from YOLO is normalised between 0 and 1, representing the relative position and size of objects within the frame. We processed the videos with YOLO to extract detailed annotations of persons (N = 464,805), bicycles (N = 30,459), motorcycles (N = 73,409), cars (N = 946,916), buses (N = 41,408), trucks (N = 100,105), traffic lights (N = 178,906) and stop signs (N = 5,918) at 30 frames per second with minimum of 70% of confidence. Each detected object was assigned a unique ID, which was tracked across frames to analyse movement patterns.

We also incorporated additional statistical data for each city: (1) the population of the country[3], (2) road traffic mortality (deaths per 100,000 population)[4], (3) population of the city [5], (4) GMP per city[6], where the GMP of 37 cities were not available, so we took the GDP of the country from World Bank[7] and divided it from the population of the country (GDP per capita).(5) literacy rate[8], and (6) the average height of the nation[9]. See supplementary material for the list of included cities with the YouTube IDs, included videos, and the aforementioned statistical data.

2.2 Pedestrian Crossing Detection

The acquired dataset allows us to investigate the behaviour of pedestrians on the city (and automatically country and culture) level. Detecting pedestrian crossings accurately is essential for understanding and modelling pedestrian behaviour in urban settings.

[1] https://support.google.com/youtube/answer/9783148?hl=en&sjid=15816069292466875886-EU

[2] https://pytube.io/en/latest/

[3] https://data.worldbank.org/indicator/SP.POP.TOTL; last accessed on June 20, 2024

[4] https://extranet.who.int/roadsafety/death-on-the-roads/#deaths; last accessed on June 20, 2024

[5] https://en.wikipedia.org/wiki/List_of_cities_by_GDP; last accessed on June 20, 2024

[6] https://en.wikipedia.org/wiki/List_of_cities_by_GDP; last accessed on June 20, 2024

[7] https://data.worldbank.org/indicator/NY.GDP.MKTP.CD; last accessed on June 20, 2024

[8] https://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate; last accessed on June 21, 2024

[9] https://en.wikipedia.org/wiki/Average_human_height_by_country; last accessed on June 21, 2024

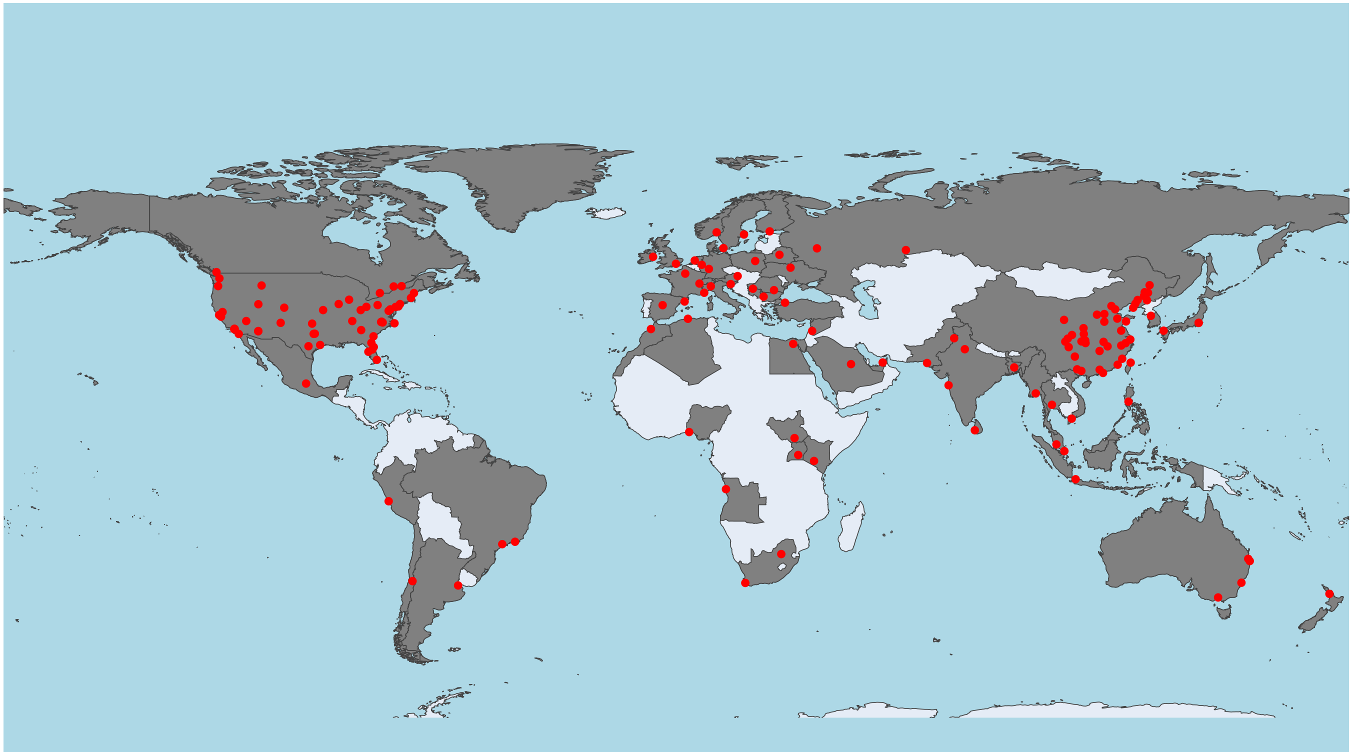

Рис. 1: Geographic distribution of cities included in the study. Red points indicate cities from which dashcam videos were collected, while grey-shaded regions represent countries included in the dataset.

The algorithm detailed in Algorithm 1 employs this detection strategy to systematically identify and count pedestrian crossings.

The criterion for a pedestrian crossing is defined by the pedestrian's entry and exit points on the video frame. A pedestrian is considered to have crossed the road if their tracked path crosses from less than 0.45 to greater than 0.55 of the frame's width, or vice versa. This accounts for both directions of crossing, thereby capturing a comprehensive view of pedestrian movement patterns. Importantly, this criterion is designed to account for different traffic orientations, whether right-sided or left-sided, by focusing on relative frame positions rather than specific road configurations. The algorithm filters the video data to focus only on those instances where the *'YOLO id'* matches specific pedestrians, groups the data by unique identifiers, and then checks whether these identified groups cross the predefined screen width boundaries.

Algorithm 1 Identification of Pedestrian Crossings.

Input: CSV file containing tracking data of individual objects, including their unique ID and position (X and Y coordinates in the frame).
Output: List of IDs of individuals who crossed the pedestrian area
Filter the dataset to include only entries corresponding to persons.
Initialize an empty list: $crossedIdsList \leftarrow []$.
Group the filtered dataset by $Unique\ Id$, such that all entries corresponding to a single individual across multiple frames are grouped together and named as $groupedData$.
for each $group$ in $groupedData$ **do**
    Extract the X-coordinate values of the individual across frames: $xValues \leftarrow group[X - center].values$.
    if $min(xValues) \leq 0.45$ and $max(xValues) \geq 0.55$ **then**
        The individual is classified as having crossed the pedestrian area.
        Add the individual's ID (i.e., $group.name$) to $crossedIdsList$.
    end if
end for
return $crossedIdsList$.

2.3 Calculation of Pedestrian Speed of Crossing a Road

Algorithm 2 calculates the speed of crossing a road by pedestrians. The algorithm calculates the speed of pedestrians crossing a road using frame-by-frame tracking data and removes outlier speeds to ensure accuracy.

First, the input data, consisting of video-specific metadata, a list of pedestrian IDs who crossed the road, and an average person's height in the country for pixel-to-meter conversion, is processed. The algorithm iterates through each group of pedestrians, identified by a

Algorithm 2 Calculation of Pedestrian Road Crossing Speed

Input:
- *dataframe*: DataFrame containing the tracking data of objects, including their unique ID, coordinates of the centre (X-center and Y-center), size of the bounding box (height and width) and time information.
- *crossingData*: List of pedestrian IDs who crossed the road, as determined by the pedestrian crossing identification algorithm.
- *heightofperson*: Real-world reference length used to calculate pixel-per-meter (ppm) scaling (refer to 2.1).

Output: A list of valid pedestrian speeds.
Initialise an empty list: $speedResults \leftarrow []$
function time_to_cross(dataframe, crossingData)
  Initialise an empty dictionary: $var \leftarrow \{\}$
  for each *crossingData* in *dataframe* **do**
    for each *id* in *crossingData* **do**
      Find the minimum and maximum X-coordinate for the pedestrian: x_min and x_max
      Find the index of the minimum and maximum X-coordinates.
      $count \leftarrow$ difference between the index of the minimum and maximum X-coordinates:
      $time[crossingData] \leftarrow count/30$ ▷ 30 frames per second
    end for
  end for
  return *time*
end function
function Calculate speed to cross(dataframe, crossingData, length, time)
  for each *id* in *crossingData* **do**
    Extract the data for the pedestrian with ID *id* from *dataframe*: $groupedWithId \leftarrow dataframe.get\_group(id)$
    Calculate the mean pedestrian height (in pixels) across all frames: $meanHeight \leftarrow groupedWithId['Heightoftheboundingbox'].mean()$
    Determine the minimum X-coordinate: $minXCenter$, and the maximum X-coordinate: $maxXCenter$
    Calculate the pixel-per-meter ratio: $ppm \leftarrow meanHeight/heightofperson$
    Compute the real-world distance crossed: $distance \leftarrow (maxXCenter - minXCenter)/ppm$
    Retrieve the time taken by the pedestrian to cross: $time \leftarrow groupedWithId['Time'].max() - groupedWithId['Time'].min()$
    Compute the pedestrian's speed: $speed \leftarrow distance/time$
    if $speed \leq 1.2$ **then** ▷ Filter out unrealistic speeds [18]
      Append the calculated speed to *speedResults*
    end if
  end for
  return *speedResults*
end function

unique ID, and calculates the mean height of each individual. Using the mean height and average height, it computes a pixel-per-meter (ppm) scaling factor, which allows the conversion of X-coordinate movement (in pixels) into real-world distance (in meters). The time taken by each pedestrian to cross is calculated from the difference between the maximum and minimum timestamps of their movement. The pedestrian's speed is then determined by dividing the real-world distance by the time taken. Speeds greater than 1.2 m/s are treated as outliers and excluded from the results (taken from [18]) to eliminate people who are on skateboard or cycle. The algorithm outputs a list of valid speeds for pedestrians who successfully crossed the road.

2.4 Quantifying Pedestrian Crossing Decision Time

In our study, we have developed a method to measure the decision time for pedestrians as they approach and begin to cross the road, detailed in Algorithm 3. This analysis utilises video data tagged and processed through the You Only Look Once (YOLO) object detection system, focusing specifically on individual pedestrians identified by their unique *'YOLO id'*.

The algorithm calculates the time a pedestrian takes to decide to start crossing a road based on their movement data. It first filters the data for the specified pedestrian ID, then groups it by the pedestrian's unique ID to process each pedestrian's movements individually. For each pedestrian, the algorithm calculates their mean height to set a margin of 10 % for detecting consistent movement. It identifies the initial X-coordinate and determines the crossing direction (left-to-right or right-to-left) based on the pedestrian's starting position. The algorithm then iterates through the pedestrian's X-coordinate data in steps, checking whether the pedestrian remains within the defined margin for thirty consecutive frames (1 second). Once consistent movement is detected, the time taken to initiate crossing is recorded. The result is stored in a dictionary, where the keys are pedestrian IDs, and the values are the respective decision times for starting the crossing.

3 Results

3.1 Time taken by pedestrians to start crossing and speed of crossing

Figure 2 shows the distribution of the speed of pedestrian crossing the road and time taken for the pedestrian to start crossing the road as described in subsection 2.3 and subsection 2.4.

The average pedestrian speed during the day across the cities analysed is 0.77 m/s, with a standard deviation of 0.13 m/s. At night, the average speed is 0.81 m/s, with a standard deviation of 0.14 m/s. The city with the highest daytime speed is Hong Kong at 1.19 m/s, while Dhaka records the lowest speed during the day at 0.27 m/s. At night, Monaco has the highest speed at 1.13 m/s, and Moscow the lowest at 0.23 m/s.

The average time it takes pedestrians to start crossing the road during the day is 0.31 s, with a standard deviation of 0.13 s. At night, the average crossing decision time is 0.34 s, with a standard deviation of 0.29 s. The longest time to start crossing during the day is in Sofia, at 1.05 s, while the shortest time is in Huaián, at

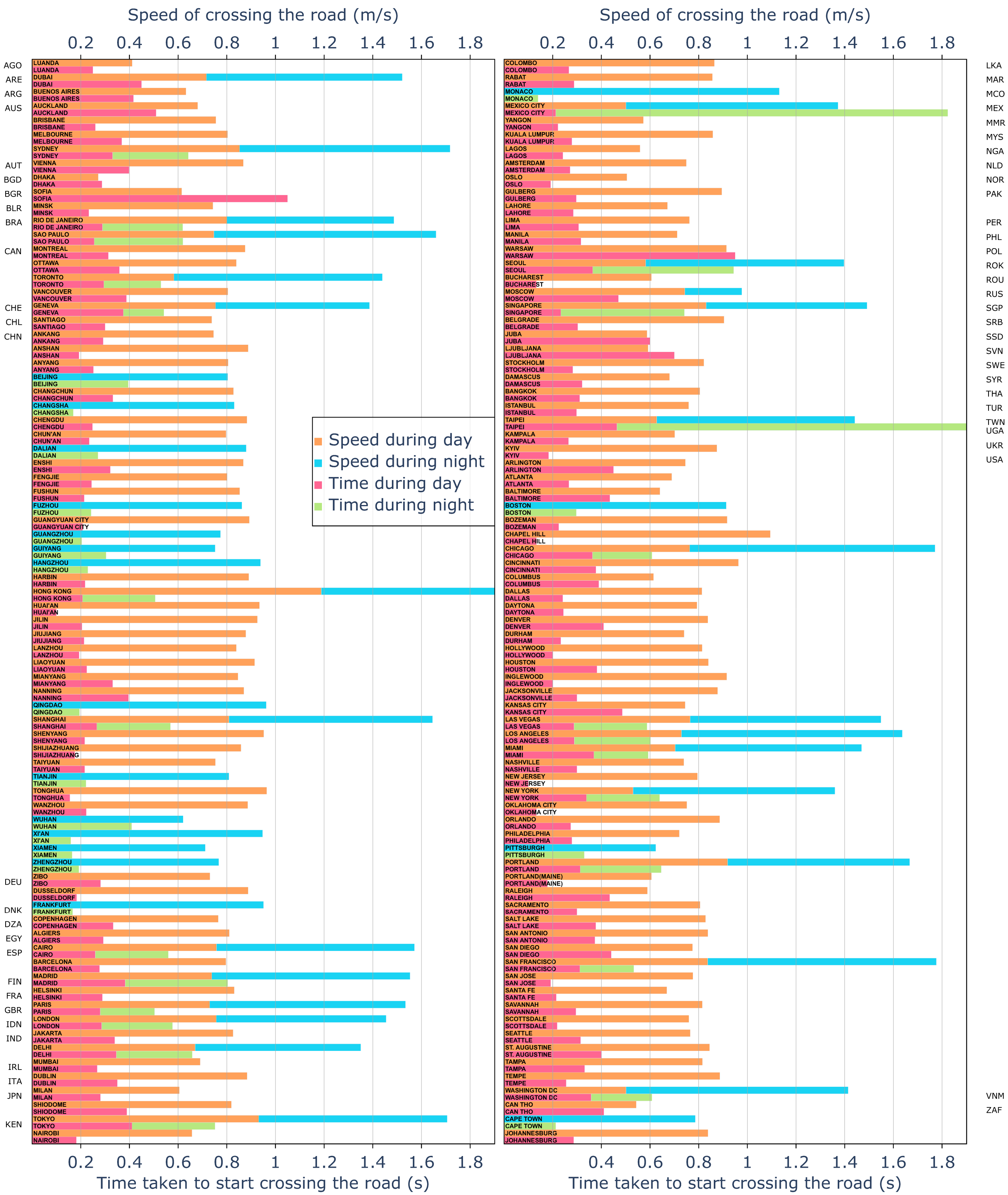

Рис. 2: Comparison of time taken by pedestrians to start crossing and speed of crossing the road across multiple cities, highlighting both day and night scenarios.

Algorithm 3 Calculation of time a pedestrian takes to decide to start crossing a road.

```
function Calculate Crossing Decision Time(dataframe, crossingdata)
    Initialise dictionary: decisionTimes ←
    Group by pedestrian Unique Id: groupedData ← personData.groupby('UniqueId')
    for each (uniqueId) in groupedData do
        Extract X-coordinates: xValues ← groupData['X − center'].values
        Calculate mean pedestrian height across all frames: meanHeight
        Set initial X-coordinate: initialX
        Set margin for movement: margin ← 0.1 × meanHeight
        Initialise counters: consecutiveFrame ← 0, flag ← 0
        Determine crossing direction:
        if initialX < 0.5 then
            Set direction to left-to-right: direction ← 1
        else
            Set direction to right-to-left: direction ← −1
        end if
        for i ← 0 to len(xValues) − 10 step 10 do
            if xValues[i] − margin × direction ≤ xValues[i + 10] ≤ xValues[i] + margin × direction then
                consecutiveFrame ← consecutiveFrame + 1
                if consecutiveFrame = 3 then
                    Set flag: flag ← 1
                end if
            else if flag = 1 then
                Store crossing decision time: decisionTimes[uniqueId] ← consecutiveFrame
                break
            else
                Reset consecutive frame counter: consecutiveFrame ← 0
            end if
        end for
    end for
    return decisionTimes
end function
```

0.10 s. At night, the longest time to start crossing is observed in Mexico City at 1.61s, while Monaco has the shortest time at 0.14s.

In cities such as Delhi (0.01 m/s), Sydney (0.01 m/s), Cairo (0.02 m/s), and Shanghai (0.02 m/s), pedestrian speeds exhibit minimal variation between day and night. In contrast, Moscow (0.51 m/s) and Hong Kong (0.48 m/s) display significantly greater differences in pedestrian speeds during these periods. Furthermore, the time taken by pedestrians to start crossing shows the largest differences in Moscow (0.51 s) and Hong Kong (0.48 s), whereas London (0.00 s) and Las Vegas (0.01 s) demonstrate negligible differences. Additionally, regional patterns indicate that Asian cities tend to have faster speeds and start crossing the road sooner compared to North American and European cities, where speeds are generally slower.

3.2 Time taken by pedestrian to start crossing as a function of literacy rate

Cities with higher literacy rates generally exhibit quicker crossing decision times. For example, Dusseldorf (99% literacy) has a crossing decision time of 0.18s, Paris (99% literacy) has a crossing decision time of 0.28 s during daytime and 0.22 s at night, and Bucharest (99% literacy) has a crossing decision time of 0.13 s. Other cities with high literacy rates, like Ottawa (99% literacy) and Taipei (86% literacy), also show relatively short crossing decision times, at 0.35 s and 0.33 s, respectively, in the daytime and at night respectively. Some cities with high literacy rates show longer crossing decision times, especially at night. Mexico City (95% literacy) has a crossing decision time of 1.61 s at night, and Taipei (98.7% literacy) has a crossing decision time of 1.52 s at night. Lower literacy cities like Juba (35% literacy) have longer crossing decision times during the day, with 0.6 s. Other cities with lower literacy, such as Lagos (62% literacy), exhibit a crossing decision time of 0.24 s.

3.3 Time taken by pedestrian to start crossing as a function of traffic-related mortality

The Figure 4 shows that cities like Geneva (traffic mortality rate: 2.7 per 100,000, crossing decision time in night time: 0.16) and Oslo (traffic mortality rate: 2.7 per 100,000, crossing decision time: 0.19 s in daylight) have both low traffic mortality rates and relatively short crossing decision times. Similarly, Frankfurt has a traffic mortality rate of 4.1 per 100,000 and a crossing decision time of 0.16 s, while Amsterdam has a traffic mortality rate of 3.8 per 100,000 and a crossing decision time of 0.27 s. These cities exhibit low values in both mortality rates and crossing decision times.

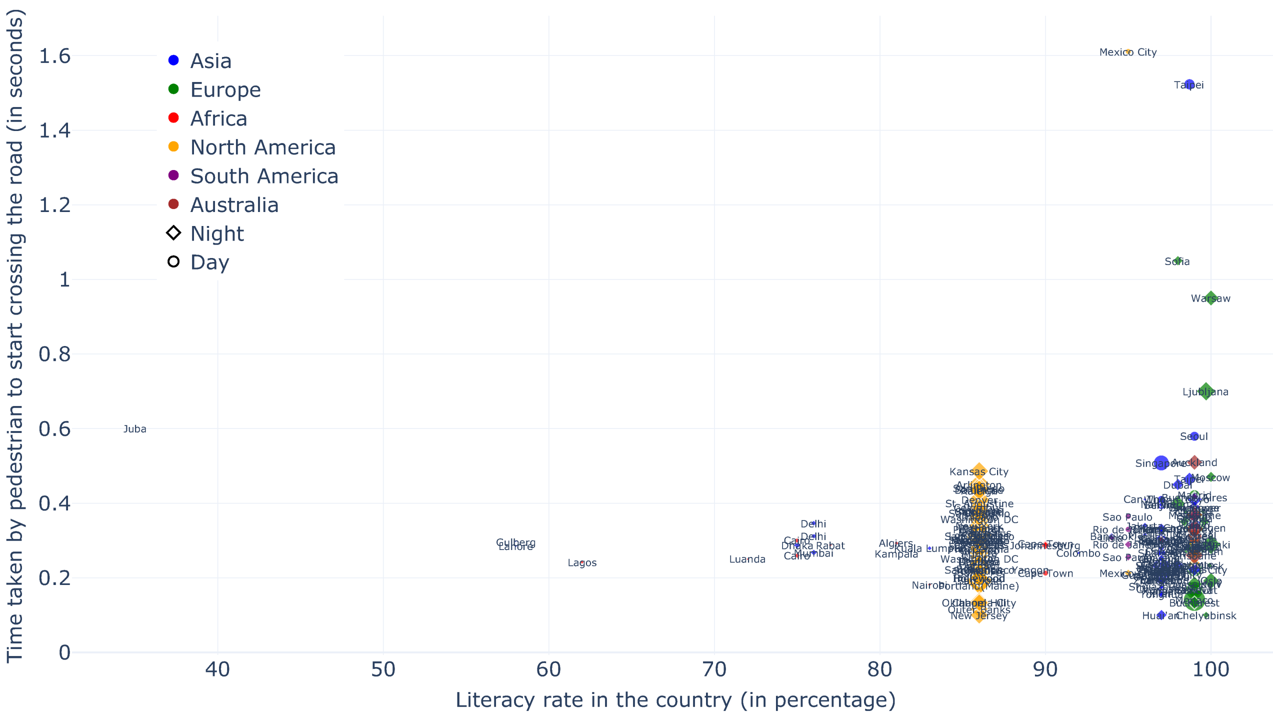

Рис. 3: Time taken by pedestrians to start crossing the road as a function of literacy percentage of the nation. Each point represents a city, with the size of the point corresponding to GMP per capita.

Cross-city Pedestrian Behaviour as a Function of traffic-related mortality

In contrast, cities with higher traffic mortality rates generally have longer crossing decision times. Moscow has a traffic mortality rate of 18.0 per 100,000 and a crossing decision time of 0.47 s, while Juba shows a traffic mortality rate of 25.7 per 100,000 and a crossing decision time of 0.60 s. Can Tho reports a traffic mortality rate of 26.4 per 100,000 and a crossing decision time of 0.41 s, while Dubai has a traffic mortality rate of 18.1 per 100,000 and a crossing decision time of 0.45 s. These cities have higher mortality rates alongside longer or moderate crossing decision times.

4 Discussion

The present study introduces the "PYT"dataset to explore pedestrian behaviour across globe, contributing significantly to the Car-VRU Interaction research community by offering cross-cultural insights into traffic behaviour. The analysis of this dataset reveals substantial variability in the time required to initiate crossing and walking speed while crossing the road. These variations are influenced by local traffic regulations, urban infrastructure, and cultural norms, underscoring the complexity of pedestrian behaviour in different regions.

The use of the YOLOv8 algorithm helped the identification of significant differences in pedestrian crossing behaviours across cities with varying socio-economic conditions, such as traffic-related mortality rates, GMP per capita, and literacy levels. The study also highlights a notable correlation between literacy rates and pedestrian crossing behaviours. In cities with near-perfect literacy rates, such as Dusseldorf and Paris, pedestrians tend to exhibit faster decision-making, resulting in shorter crossing initiation times. This trend suggests that higher literacy levels may contribute to better awareness of traffic rules and quicker reactions in traffic environments. However, outliers such as Mexico City and Taipei exhibit prolonged crossing decision times at night, despite high literacy rates, indicating that additional factors, such as infrastructure quality and nighttime lighting conditions, play critical roles in pedestrian safety.

When comparing pedestrian speeds, cities with advanced infrastructure, such as Hong Kong and Tokyo, demonstrate consistently high walking speeds both during the day and night. These results suggest that well-designed urban environments promote pedestrian confidence and efficiency in movement. In contrast, cities such as Dhaka and Can Tho show slower walking speeds, which may be linked to challenges such as inadequate pedestrian infrastructure or heightened safety concerns at nighttime hours. The significant differences between day and night pedestrian behaviours, particularly in cities like Mexico City and Hong Kong, suggest that local traffic patterns, lighting conditions, and cultural factors influence pedestrian speed. Faster speeds at night in these cities may reflect a tendency to navigate more quickly in lower traffic conditions during off-peak hours.

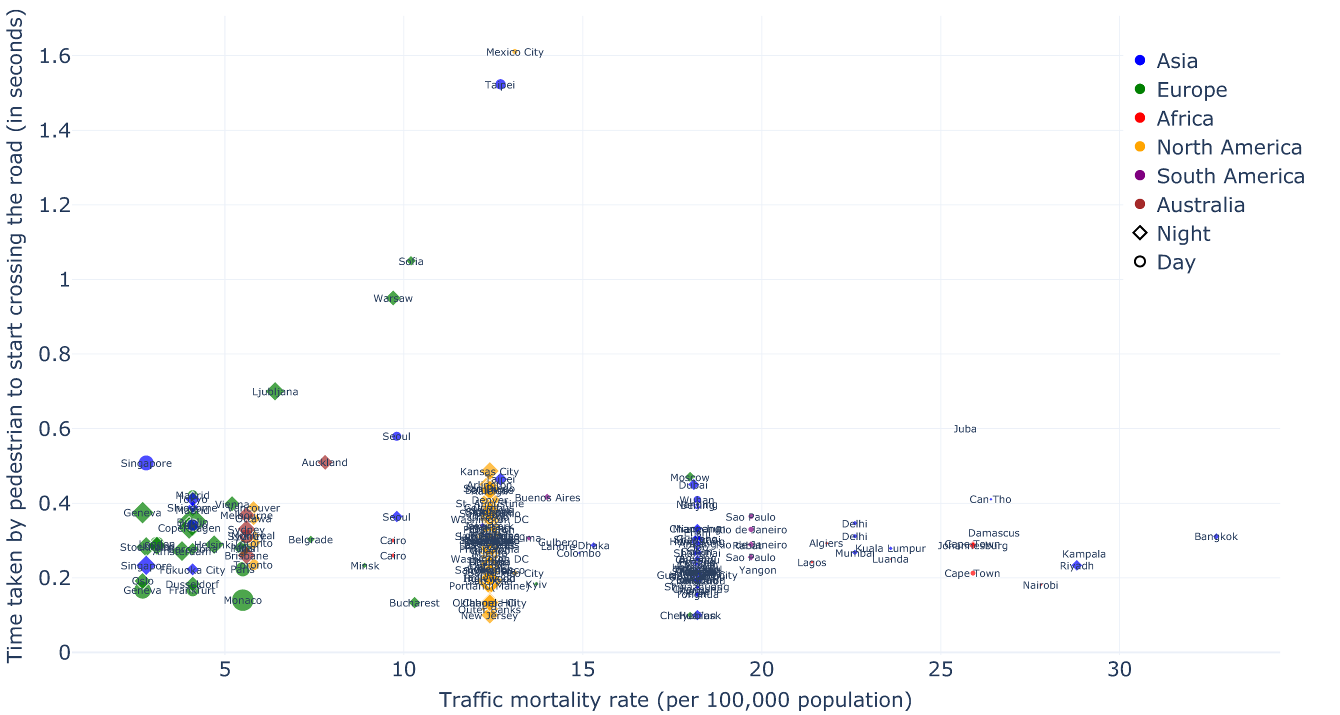

Рис. 4: Time taken by pedestrians to start crossing the road as a function of traffic-related mortality(per 100,000 population). Each point represents a city, where the size of the point corresponds to the GMP per capita.

The analysis of pedestrian crossing decision time reveals further variability across cities. In cities like Warsaw and Sofia, longer waiting times during the day suggest either higher traffic volumes or a more cautious pedestrian approach. On the other hand, cities such as Geneva and Monaco exhibit shorter waiting times, indicating more efficient pedestrian crossings, likely supported by favourable traffic regulations and well-planned urban infrastructure. Cities with advanced urban planning, including Tokyo and Shanghai, exhibit moderate walking speeds and shorter start times, pointing to the presence of organized pedestrian traffic systems. Conversely, cities with less developed infrastructure, such as Dhaka and Yangon, exhibit slower speeds and lower crossing decision times, highlighting the challenges faced in these regions.

An important observation from the study is the relationship between pedestrian crossing behaviour and traffic-related mortality rates. Cities with low mortality rates generally exhibit shorter pedestrian crossing decision times, while cities with higher mortality rates show longer hesitation times. This correlation suggests that the level of traffic safety directly influences pedestrian confidence and decision-making processes. European cities tend to cluster at the lower end of mortality rates, indicating more consistent safety standards, whereas cities in Africa and Asia display greater variability in pedestrian behaviour, reflecting disparities in infrastructure and safety measures. The relationship between crossing decision times and mortality rates presents an interesting pattern, with a gap in the mid-to-high range of mortality rates and crossing decision times. This indicates that cities with moderately high mortality rates may still implement safety measures that prevent excessive pedestrian hesitation, while cities with extremely high mortality rates struggle with inadequate infrastructure, leading to significant delays in crossing decisions.

The findings also suggest that while technological advancements in traffic systems, such as autonomous vehicles, are important, socio-economic and environmental factors play a more immediate role in shaping pedestrian behaviour. Observable patterns of time and speed adjustments are more reflective of the local socio-cultural context than of technological innovations. Thus, interventions aimed at improving pedestrian safety should be tailored to address these contextual factors rather than relying solely on high-tech solutions. The insights gained from this research provide a foundation for developing targeted strategies to improve road safety globally by focusing on enhancing pedestrian infrastructure and addressing socio-economic disparities that impact traffic behaviour.

5 Limitations and Future Work

It is important to acknowledge the limitations inherent in relying solely on videos from YouTube for data collection. Particularly, the accessibility of videos from certain regions, notably those in Africa, proved challenging due to socioeconomic factors influencing video production and dissemination. Moreover, while the YOLOv8 algorithm facilitated object detection within videos, its scope was limited to bounding box characteristics, precluding detailed analysis of attributes such as vehicle speed or lane length. Furthermore, the videos are also not classified in different seasons as the date is not available for the video.

Future research should address these limitations by utilizing more advanced CV models capable of detecting a broader range of attributes within video footage. For instance, by employing algorithms that can discern detailed features such as vehicle dynamics, lane markings, and specific pedestrian actions, researchers could gain a deeper understanding of traffic behaviour. This includes detecting additional attributes for pedestrians, such as road crossing intentions, time-to-crossing predictions, current behaviours like walking direction and posture, and appearance factors like clothing and accessories [32]. Moreover, vehicle-related attributes such as lane detection [17], speed estimation [27], and distance calculations [45] between objects could also be incorporated. These enhancements would provide a richer, more nuanced dataset, enabling a more comprehensive analysis of how different socio-economic and environmental factors influence pedestrian and cyclist behaviours across diverse urban settings.

Additionally, expanding the geographic scope of data collection to include underrepresented regions will help create a more balanced representation of global traffic patterns. Future studies could also extend these methodologies to other VRUs, such as cyclists, building on previous research but broadening the focus beyond single cities or countries. This approach would offer a more holistic view of global traffic behaviour, ultimately supporting the development of more effective and culturally sensitive traffic management strategies [31] and interventions worldwide.

Future research could also focus on cyclists (another type of VRU). Past papers have investigated how cyclists behave [25, 26, 36], yet these resources also focus on one city/country. This dataset and presented algorithms could be extended to assess the difference in the behaviour of this increase in its share group of participants in traffic.

Supplementary Material

A maintained version of the source code is available at https://github.com/Shaadalam9/pedestrians-in-youtube. The PYT dataset can be found at https://doi.org/10.4121/06e9bb9a-a064-412b-b0f3-9ac5dd62ea16.

References

[1] Md Shadab Alam, Thirumanikandan Subramanian, Marieke Martens, Wolfram Remlinger, and Pavlo Bazilinskyy. 2024. From A to B with ease: User-centric interfaces for shuttle buses. (2024). https://doi.org/10.1145/3641308.3685033

[2] Md Shadab Alam, Sanjeev Kumar Ramkumar Sudha, and Abhilash Somayajula. 2023. AI on the Water: Applying DRL to Autonomous Vessel Navigation. *arXiv preprint arXiv:2310.14938* (2023). https://doi.org/10.48550/arXiv.2310.14938

[3] Pavlo Bazilinskyy, Dimitra Dodou, and Joost De Winter. 2019. Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation research part F: traffic psychology and behaviour* 67 (2019), 175–194. https://doi.org/10.1016/j.trf.2019.10.013

[4] Pavlo Bazilinskyy, Dimitra Dodou, Y. B. Eisma, W . V. Vlakveld, and J. C. F. De Winter. 2022. Blinded windows and empty driver seats: the effects of automated vehicle characteristics on cyclist decision-making. *IET Intelligent Transportation Systems* 17 (2022), 72–84. https://doi.org/10.1049/itr2.12235

[5] P Bazilinskyy, YB Eisma, D Dodou, and Joost CF de Winter. 2020. Risk perception: A study using dashcam videos and participants from different world regions. *Traffic injury prevention* 21, 6 (2020), 347–353. https://doi.org/10.1080/15389588.2020.1762871

[6] Pavlo Bazilinskyy, Lars Kooijman, Dimitra Dodou, and J. C. F. De Winter. 2020. Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. In *Proceedings of Driving Simulation Conference (DSC)*. Antibes, France.

[7] Pavlo Bazilinskyy, Miltos Kyriakidis, Dimitra Dodou, and J. C. F. De Winter. 2019. When will most cars be able to drive fully automatically? Projections of 18,970 survey respondents. *Transportation Research Part F: Traffic Psychology and Behaviour* 64 (2019), 184–195. https://doi.org/10.1016/j.trf.2019.05.008

[8] Mauro Bellone, Azat Ismailogullari, Tommi Kantala, Sami Mäkinen, Ralf-Martin Soe, and Milla Åman Kyyrö. 2021. A cross-country comparison of user experience of public autonomous transport. *European Transport Research Review* 13, 1 (2021), 19. https://doi.org/10.1186/s12544-021-00477-3

[9] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020. Nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11621–11631. https://doi.org/10.48550/arXiv.1903.11027

[10] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. 2019. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8748–8757. https://doi.org/10.1109/CVPR.2019.00895

[11] Zhengping Che, Guangyu Li, Tracy Li, Bo Jiang, Xuefeng Shi, Xinsheng Zhang, Ying Lu, Guobin Wu, Yan Liu, and Jieping Ye. 2019. D^2-city: a large-scale dashcam video dataset of diverse traffic scenarios. *arXiv preprint arXiv:1904.01975* (2019). https://doi.org/10.48550/arXiv.1904.01975

[12] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3213–3223. https://doi.org/10.48550/arXiv.1604.01685

[13] Loren Demerath and David Levinger. 2003. The social qualities of being on foot: A theoretical analysis of pedestrian activity, community, and culture. *City & Community* 2, 3 (2003), 217–237. https://doi.org/10.1111/1540-6040.00052

[14] Thomas A Dingus, Sheila G Klauer, Vicki Lewis Neale, Andy Petersen, Suzanne E Lee, Jeremy Sudweeks, Miguel A Perez, Jonathan Hankey, David Ramsey, Santosh Gupta, et al. 2006. *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment*. Technical Report. United States. Department of Transportation. National Highway Traffic Safety ....

[15] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. 2011. Pedestrian detection: An evaluation of the state of the art. *IEEE transactions on pattern analysis and machine intelligence* 34, 4 (2011), 743–761. https://doi.org/10.1109/TPAMI.2011.155

[16] European Commission. 2022. Road Safety Fact Sheet: Pedestrians. https://road-safety.transport.ec.europa.eu/document/download/aaeb811d-f455-4fb0-8a79-7a373626952d_en?filename=FF_pedestrians_20220209.pdf Accessed: 2024-09-03.

[17] Zhengyang Feng, Shaohua Guo, Xin Tan, Ke Xu, Min Wang, and Lizhuang Ma. 2022. Rethinking efficient lane detection via curve modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17062–17070. https://doi.org/10.48550/arXiv.2203.02431

[18] Kay Fitzpatrick, Marcus A Brewer, and Shawn Turner. 2006. Another look at pedestrian walking speed. *Transportation research record* 1982, 1 (2006), 21–29. https://doi.org/10.1177/0361198106198200104

[19] 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*. 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

[20] David L Harkey and Daniel L Carter. 2006. Observational analysis of pedestrian, bicyclist, and motorist behaviors at roundabouts in the United States. *Transportation research record* 1982, 1 (2006), 155–165. https://doi.org/10.1177/0361198106198200120

[21] Mathilde Jay, Anne Régnier, Anaïs Dasnon, Killian Brunet, and Marie Pelé. 2020. The light is red: Uncertainty behaviours displayed by pedestrians during illegal road crossing. *Accident Analysis & Prevention* 135 (2020), 105369. https://doi.org/10.1016/j.aap.2019.105369

[22] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. [n. d.]. *YOLO by Ultralytics*. https://github.com/ultralytics/ultralytics

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012). https://doi.org/10.1145/3065386

[24] Bruce Krogh and Charles Thorpe. 1986. Integrated path planning and dynamic steering control for autonomous vehicles. In *Proceedings. 1986 IEEE International*

Conference on Robotics and Automation, Vol. 3. IEEE, 1664–1669. https://doi.org/10.1109/ROBOT.1986.1087444

[25] Xiaofei Li, Fabian Flohr, Yue Yang, Hui Xiong, Markus Braun, Shuyue Pan, Keqiang Li, and Dariu M Gavrila. 2016. A new benchmark for vision-based cyclist detection. In *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1028–1033. https://doi.org/10.1109/IVS.2016.7535515

[26] Xiaofei Li, Lingxi Li, Fabian Flohr, Jianqiang Wang, Hui Xiong, Morys Bernhard, Shuyue Pan, Dariu M Gavrila, and Keqiang Li. 2016. A unified framework for concurrent pedestrian and cyclist detection. *IEEE transactions on intelligent transportation systems* 18, 2 (2016), 269–281. https://doi.org/10.1109/TITS.2016.2567418

[27] Cheng-Jian Lin, Shiou-Yun Jeng, and Hong-Wei Lioa. 2021. A Real-Time Vehicle Counting, Speed Estimation, and Classification System Based on Virtual Detection Zone and YOLO. *Mathematical Problems in Engineering* 2021, 1 (2021), 1577614. https://doi.org/10.1155/2021/1577614

[28] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, Jingheng Chen, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, et al. 2021. One million scenes for autonomous driving: Once dataset. *arXiv* (2021). https://doi.org/10.48550/arXiv.2106.11037

[29] Antoine Mauri, Redouane Khemmar, Benoit Decoux, Madjid Haddad, and Rémi Boutteau. 2021. Real-time 3D multi-object detection and localization based on deep learning for road and railway smart mobility. *Journal of imaging* 7, 8 (2021), 145. https://doi.org/10.3390/jimaging7080145

[30] Dimitris Milakis, Bart Van Arem, and Bert Van Wee. 2017. Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of intelligent transportation systems* 21, 4 (2017), 324–348. https://doi.org/10.1080/15472450.2017.1291351

[31] Michael Mögele and Henrike Rau. 2020. Cultivating the "car state": a culturally sensitive analysis of car-centric discourses and mobility cultures in Southern Germany. *Sustainability: Science, Practice and Policy* 16, 1 (2020), 15–28. https://doi.org/10.1080/15487733.2020.1756188

[32] Taylor Mordan, Matthieu Cord, Patrick Pérez, and Alexandre Alahi. 2021. Detecting 32 Pedestrian Attributes for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems (T-ITS)* (2021). https://doi.org/10.1109/TITS.2021.3107587

[33] Vishal Onkhar, Pavlo Bazilinskyy, Dimitra Dodou, and J. C. F. De Winter. 2022. The effect of drivers' eye contact on pedestrians' perceived safety. *Transportation Research Part F: Traffic Psychology and Behaviour* 84 (2022), 194–210. https://doi.org/10.1016/j.trf.2021.10.017

[34] Jennifer A Oxley, Elfriede Ihsen, Brian N Fildes, Judith L Charlton, and Ross H Day. 2005. Crossing roads safely: an experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention* 37, 5 (2005), 962–971. https://doi.org/10.1016/j.aap.2005.04.017

[35] Quang-Hieu Pham, Pierre Sevestre, Ramanpreet Singh Pahwa, Huijing Zhan, Chun Ho Pang, Yuda Chen, Armin Mustafa, Vijay Chandrasekhar, and Jie Lin. 2020. A* 3D dataset: Towards autonomous driving in challenging environments. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2267–2273. https://doi.org/10.48550/arXiv.1909.07541

[36] Ewoud AI Pool, Julian FP Kooij, and Dariu M Gavrila. 2021. Crafted vs learned representations in predictive models—A case study on cyclist path prediction. *IEEE Transactions on Intelligent Vehicles* 6, 4 (2021), 747–759. https://doi.org/10.1109/TIV.2021.3064253

[37] J. Pyfer. 2018. Project MAC. https://www.britannica.com/topic/Project-Mac Accessed: 2024-09-03.

[38] 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. https://doi.org/10.1109/ICCV.2019.00636

[39] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 264–269. https://doi.org/10.1109/IVS.2017.7995730

[40] 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. https://doi.org/10.1109/ICCVW.2017.33

[41] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788. https://doi.org/10.48550/arXiv.1506.02640

[42] David Saffo, Caglar Yildirim, Sara Di Bartolomeo, and Cody Dunne. 2020. Crowdsourcing virtual reality experiments using vrchat. In *Extended abstracts of the 2020 chi conference on human factors in computing systems*. 1–8.

[43] Matus Sucha, Daniel Dostal, and Ralf Risser. 2017. Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention* 102 (2017), 41–50. https://doi.org/10.1016/j.aap.2017.02.018

[44] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. 2020. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2446–2454. https://doi.org/10.48550/arXiv.1912.04838

[45] Marek Vajgl, Petr Hurtik, and Tomáš Nejezchleba. 2022. Dist-yolo: Fast object detection with distance estimation. *Applied sciences* 12, 3 (2022), 1354. https://doi.org/10.3390/app12031354

[46] Peng Wang, Xinyu Huang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. 2019. The apolloscape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019). https://doi.org/10.1109/TPAMI.2019.2926463

[47] World Bank. 2012. The Pedestrian Environment and Non-Motorized Transport (NMT). https://thedocs.worldbank.org/en/doc/3b8c94226d4949640f948f345214befb-0090062024/original/C5-M3-Pedestrian-Enviroment-and-Non-Motorized-Transport-v2-04-2012.pdf Accessed: 2024-09-03.

[48] George Yannis, John Golias, and Eleonora Papadimitriou. 2007. Accident risk of foreign drivers in various road environments. *Journal of safety research* 38, 4 (2007), 471–480. https://doi.org/10.1016/j.jsr.2007.01.014

[49] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. 2020. A Survey of Autonomous Driving: Common Practices and Emerging Technologies. *IEEE Access* 8 (2020), 58443–58469. https://doi.org/10.1109/ACCESS.2020.2983149