Understanding global pedestrian behaviour in 565 cities with dashcam videos on YouTube

MD SHADAB ALAM*, Eindhoven University of Technology, The Netherlands
MARIEKE H. MARTENS, Eindhoven University of Technology, The Netherlands and TNO, The Netherlands
OLENA BAZILINSKA, National University of "Kyiv-Mohyla Academy", Ukraine
PAVLO BAZILINSKYY, Eindhoven University of Technology, The Netherlands

The interactions between future cars and pedestrians should be designed to be understandable and safe worldwide. Although 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, Pedestrians in YouTube (PYT), which includes 1562.80 hours of YouTube day and night dashcam footage from 565 cities in 104 countries. The included videos feature continuous urban driving, are at least 10 min long, feature no atypical events and represent everyday conditions, and are from cities with a minimum of 20,000 population. We detected pedestrian movements, focussing on the speed and the pedestrian crossing decision time during road crossings based on the bounding boxes given by YOLOv11x. The results revealed statistically significant variations in pedestrian behaviour influenced by socioeconomic and environmental factors such as Gross Metropolitan Product, traffic-related mortality, Gini coefficient, traffic index, and literacy. The dataset is publicly available to encourage further research on global pedestrian behaviour.

Additional Key Words and Phrases: Human Behaviour, Pedestrian Safety, Cross-City Evaluation, Computer Vision, Dataset

ACM Reference Format:
Md Shadab Alam, Marieke H. Martens, Olena Bazilinska, and Pavlo Bazilinskyy. 2025. Understanding global pedestrian behaviour in 565 cities with dashcam videos on YouTube. 1, 1 (March 2025), 17 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Advances in the field of automated driving (AD) demonstrated by companies such as Waymo[1], Tesla[2], and Baidu[3] have spurred increased efforts among researchers and industry stakeholders to collect comprehensive datasets and analyse traffic behaviour on a global scale [64]. Data collection for AD research primarily relies on the acquisition of expensive hardware [32, 71], which often requires considerable time and resources. For instance, Dingus et al. [20] conducted a large-scale study that collected 43,000 hours of driving data from 100 instrumented vehicles operating across the United States for 12 months. Similarly, the study by Neale et al. [49] collected approximately 2,000,000 vehicle miles, comprising nearly 43,000 hours of data, involving 241 primary and secondary drivers, and spanning a 12 to 13-month data collection period across 100 different vehicles in the USA. Another notable example is the Shanghai Naturalistic Driving Study (SH-NDS) conducted by Zhu et al. [74], which was a joint effort between Tongji University, General Motors, and the Virginia Tech Transportation Institute. This study involved a three-year data collection procedure

[1] https://waymo.com/
[2] https://www.tesla.com/autopilot
[3] https://www.apollo.auto/apollo-self-driving

Authors' Contact Information: Md Shadab Alam, m.s.alam@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands; Marieke H. Martens, m.h.martens@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands and TNO, Helmond, The Netherlands; Olena Bazilinska, olena.bazilinska@ukma.edu.ua, National University of "Kyiv-Mohyla Academy", Kyiv, Ukraine; Pavlo Bazilinskyy, p.bazilinskyy@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands.

2025. Manuscript submitted to ACM

that commenced in December 2012 and ended in December 2015. Driving data were collected daily from 60 drivers in Shanghai, China, accumulating a total of 161,055 km travelled during the study period. However, such studies are often geographically restricted, limiting their ability to capture diverse cross-cultural variations in driving behaviour.

1.1 Pedestrian behaviour

Pedestrians and modern cars share roads in urban environments. In 2019, pedestrians accounted for 20% of road fatalities in the European Union [22]. These fatalities often occur because the driver is unable to understand the intentions of the pedestrian [61], the time of the day [53], the complexity of the traffic patterns and the higher density of the population [15], etc. Cultural differences between drivers and pedestrians can sometimes lead to traffic fatalities, especially in cases where the driver is operating a rental car in a foreign country [54, 62, 70]. Variations in traffic regulations, such as differing signing policies, often contribute to this problem. For example, Summala [62] studied the behaviour of American drivers navigating European roads, where uncontrolled intersections operate under different sign policies and priority rules. The study found that American drivers initially exhibited riskier crossing behaviour, likely influenced by their familiarity with US traffic rules.

Walking constitutes a significant portion of urban traffic. Even in highly motorised modern urban transportation systems in developed countries, like the US, pedestrian traffic still plays an important role, especially in short- to medium-distance travel and in the process of connecting different modes of transportation. According to the Nationwide Personal Transportation Survey (NHTS) conducted in 2019, 34% of the total trips in Switzerland were made on foot and 11% in the United States [11]. In particular, 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) [68] indicates 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 behaviour that can be difficult to anticipate. This variability is influenced by numerous factors, including individual decision-making, environmental context, cultural norms, and social interactions [19]. As a result, accurately modelling pedestrian behaviour in traffic scenarios presents a considerable challenge. Pedestrians can make sudden stops, change directions, or take unpredictable actions, such as crossing outside designated crosswalks or interacting with their surroundings in ways that are difficult to quantify. This unpredictable behaviour complicates efforts to develop advanced driver assistance systems (ADAS) and automated driving systems (ADS) technology that can interact effectively and safely with pedestrians on the road.

Studies by Bazilinskyy et al. [5], Oudshoorn et al. [51], Medury et al. [44], Saffo et al. [60] and Alam et al. [1] leverage crowdsourcing experiments to analyse pedestrian 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. [50], Tram et al. [65], Zhao et al. [72] and Bazilinskyy et al. [6], faces challenges due to hardware constraints that limit their applicability across different global contexts. In addition, these studies often focus on a specific set of scenarios predefined by the researchers, neglecting other variables such as diverse environmental conditions, time of day, 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.

1.2 Datasets for pedestrian behaviour

Soon, future traffic will be mixed, comprised of manually-driven (MDVs) and automated vehicles (AVs), as well as vulnerable road users (VRUs) [7, 45]. The algorithms deciding on the behaviour of AVs are being designed by computer-human interaction experts in both industry and academia today [2, 36]. 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.

Pivotal datasets such as KITTI [25], NuScenes [12], One Thousand and One Hours [42], Caltech Pedestrian detection benchmark [21], Pedestrian Intention Estimation (PIE) [56], A2d2 [26], Waymo Open Dataset [63], ApolloScape Auto [67], Cityscapes [16], A*3D dataset [55], Argoverse [13] and others have become standard benchmarks that support a variety of tasks in computer vision (CV) and AD research. Studies such as Oxley et al. [52] and Rasouli et al. [57] used videos to investigate the intentions of pedestrians in traffic. Oxley et al. [52] discovered that crossing decisions were primarily based on vehicle distance and less on vehicle arrival time. Rasouli et al. [57] trained a neural network for the classification of pedestrians looking to cross or walking. Similarly, Mordan et al. [48] used the JAAD dataset [58] to detect 32 attributes for a pedestrian. Meanwhile, Vajgl et al. [66], and Mauri et al. [43] used KITTI datasets to train their algorithm to detect cars, pedestrians, and cyclists in traffic with distance estimation.

On-road studies that involve instrumented vehicles for the collection of traffic data often limit their scope to one or a few specific cities or countries due to: (1) high costs [17], (2) the complexity of the method [10], (3) the use of specialised hardware and closed source software [29], and (4) difficulties with obtaining ethics approval for cross-country research [27].

Another critical limitation of the existing datasets is the scarcity of research on regional differences in pedestrian behaviour, particularly at the city level across the globe. For example, NuScenes [12], which has 1000 20-second scenes, features scenes from Singapore and Boston, MA, USA, while Argoverse [13], with 300 km of mapped road lanes, focusses on Miami, FL, USA, and Pittsburgh, PA, USA. The datasets such as the Waymo Open Dataset [63], which contains 1,050 scenes each spanning 20 s, attempt to cover parts of the USA, namely San Francisco, CA, Detroit, CO, Seattle, WA, Phoenix, AZ, Los Angeles, CA, and Mountain View, CA. Similarly, the D^2-dataset [14], which contains a total of 11,211 driving videos that equate to approximately one hundred hours, covers five cities in China. The KITTI dataset contains scenes only from Karlsruhe, Germany.

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 entails insights into traffic behaviour in different cities worldwide. We achieve this goal by collecting 1562.80 hours of day and nighttime dashcam footage from 387 cities from the video hosting platform YouTube (https://www.youtube.com). The dataset was used to analyse 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: Gross Metropolitan Product (GMP) per capita, rate of traffic-related mortality, Gini coefficient, traffic index, average height, and 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 the absolute majority placed on YouTube. Sharing "special" dashcam traffic videos is a phenomenon in which people exchange them for amusement. A notable example is a Telegram group BadShofer (https://t.me/s/badshofer) with ~43,600 users and ~52,500 shared videos as of 4 February 2025. Most of such videos are short and contain isolated footage of abnormal events, such as accidents and unexpected behaviour of traffic participants. 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 ~803,000 followers and 818 uploaded videos, as of 4 February 2025). These videos frequently generate an Autonomous Sensory Meridian Response (ASMR) in viewers, a tingling sensation induced by specific auditory or visual stimuli [4, 39, 40]. Additionally, such lengthy and monotonous driving footage is commonly used by viewers as relaxing "background noise", simultaneously providing income for the authors through continuous playbacks [30]. As these videos are hosted on YouTube, they are under the fair use on YouTube[4], which allows their use for research. To populate the dataset with YouTube videos, we define the following inclusion criteria:

- **C1**: A continuous focus on urban settings, excluding clips that feature highways, rural routes or parking areas.
- **C2**: A minimum duration of 10 minutes to ensure adequate coverage of urban environments.
- **C3**: Avoidance of atypical events, such as accidents or special events, that do not represent everyday conditions.
- **C4**: A population threshold of at least 20,000 for the cities of interest to ensure sufficient pedestrian interactions. However, in cases where the capital city or the largest city of a country has a population of less than 50,000, it should still be included in the dataset to ensure comprehensive geographical representation, especially in smaller or island nations (e.g., Liechtenstein).

In this study, we collected dashcam videos from YouTube, complying with the licencing conditions. The authors included videos that passed the C2 criterion from the results of search queries on YouTube. The queries were as follows: "*dashcam video in [city]*", "*driving videos in [city]*", "*dashcam videos in cities*", "*driving videos in [city]*", and "*dashcam video driving [country]*". One video could feature multiple cities, out of which those that satisfy criterium C4, were added. If a part of the video does not meet criteria C1 and C3, the part was removed from the study. Furthermore, if the video contains scenes from both daylight and nighttime, the video was split into two based on street lights (i.e., nighttime driving started when the street lights went on). The search for videos took place between 3 February 2024, and 19 February 2025. A total of 2,757 videos were included from 565 cities in 104 countries. See Figure 1 for the overview of included countries and cities. They represent 79.23% of the global population. Of these videos, 2396 featured daytime driving, 335 showed nighttime driving, and 26 included footage from both daytime and night. The combined length of the collected footage is 1527 hours, 58 minutes, and 59 seconds. The videos were downloaded from YouTube using the pytubefix (https://pytubefix.readthedocs.io) and yt-dlp (https://github.com/yt-dlp/yt-dlp) libraries with a resolution of 1280 x 720 px. Supplementary material (section 4.1) contains the codebase used in this work.

The videos in the dataset were then analysed with the You Only Look Once (YOLO) algorithm [59], specifically YOLOv11 [34]. It processes images in a single pass, predicting bounding boxes and class probabilities simultaneously.

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

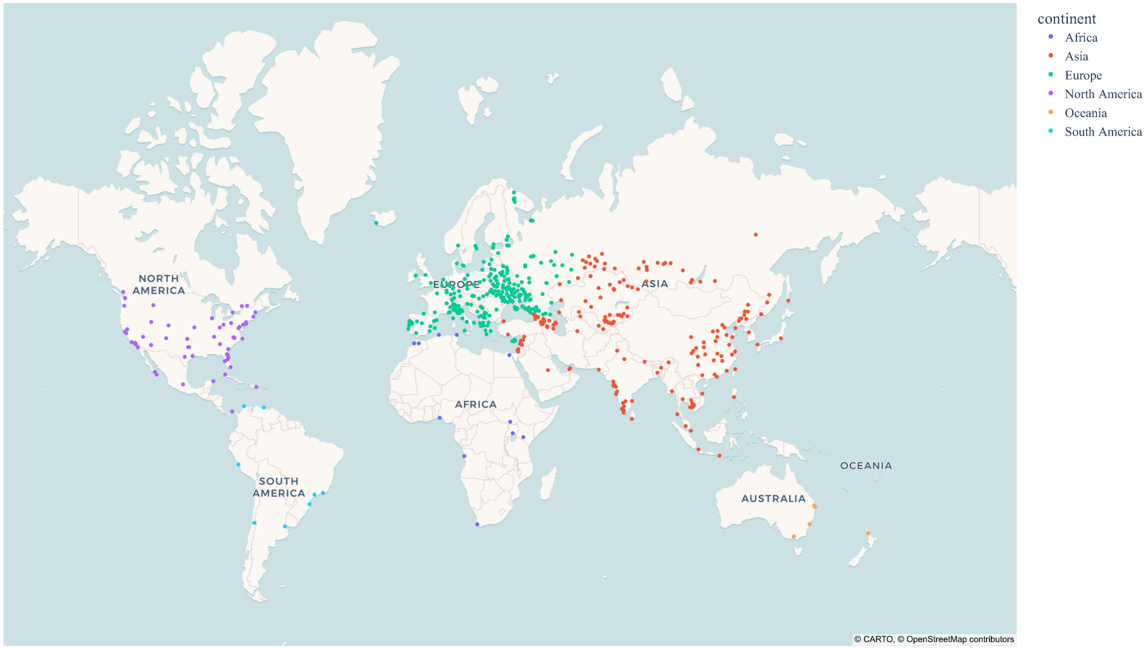

Fig. 1. 565 cities from 104 countries from which videos were included in the dataset. Colour of markers indicate the continent.

It can detect objects from 80 different classes and provides the width (W), height (H), and centre coordinates of each bounding box (X-centre, Y-centre). 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=2,432,518;), bicycles (N=159,542), motorcycles (N=939,522), cars (N=4,734,629), buses (N=218,510), trucks (N=664,139), traffic lights (N=346,781) and stop signs (N = 25,521) with a minimum of 70% confidence. Each detected object was assigned a unique ID, which was tracked across frames to analyse movement patterns. However, these unique IDs were not transferred when videos transitioned from day to night or vice versa, as the tracking was reset at these transitions.

To enhance the analysis, we incorporated various statistical data for each city. These parameters include:

(1) **Population Data**: We collect both national and city-level population data, sourced from https://restcountries.com/.
(2) **Road Traffic Mortality**: The number of traffic-related deaths per 100,000 inhabitants, obtained from https://data.worldbank.org/indicator/SH.STA.TRAF.P5.
(3) **Income Inequality (Gini Coefficient)**: Gini coefficient values, reflecting economic disparity, were sourced from https://restcountries.com/.
(4) **Gross Metropolitan Product (GMP)**: For cities in the United States, GMP data was recovered from https://www.bea.gov/, while data for other cities were sourced from https://data-explorer.oecd.org/.
(5) **Literacy Rate**: Data on literacy levels in different regions were obtained from https://data.worldbank.org/indicator/SE.ADT.LITR.ZS.

(6) **National Average Height**: The average height per country was referenced from https://www.kaggle.com/datasets/dilaraozcerit/averageheightdata.
(7) **Traffic Index**: Congestion and traffic conditions in each city were measured using the https://www.tomtom.com/traffic-index/.

Further details, including the list of cities included with their corresponding YouTube video IDs and incorporated statistical data, can be found in the supplementary material section 4.1.

2.2 Detection of pedestrian crossing

The acquired dataset allows us to investigate pedestrian behaviour on the city level (and automatically on the level of countries and cultures). Detecting pedestrian crossings accurately is essential for understanding and modelling pedestrian behaviour in urban settings. The algorithm detailed in Algorithm 1 employs this detection strategy to systematically identify and count pedestrian crossings.

Algorithm 1 Detection of pedestrian crossing.

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.
Initialise 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 $crossedIdsList$.

The algorithm defines a pedestrian crossing when their tracked path moves from less than 0.45 to more than 0.55 of the frame width, or vice versa. It filters YOLO-detected pedestrians, tracks their movement using unique IDs, and checks if their X-coordinates meet the crossing criterion. If so, their ID is stored. This method ensures adaptability to different traffic orientations and road structures, enabling cross-city comparisons and further behavioural analysis.

2.3 Calculation of pedestrian road crossing speed

Algorithm 2 calculates the speed at which pedestrians cross a road. 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-metre conversion, is processed. The algorithm iterates through each group of pedestrians, identified by a unique ID, and calculates the mean height of each individual. Using the mean height and average height, it computes a pixel-per-metre (ppm) scaling factor, which allows the conversion of X-coordinate movement (in pixels) into real-world distance (in metres). The time taken by each pedestrian to cross the road 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 1.42 m/s are treated as outliers and excluded from the results [9, 37, 47], to eliminate people on a skateboard or a bicycle. The algorithm generates a list of speed values for pedestrians who have successfully crossed the road.

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-centre and Y-centre), 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 speed values.
Initialise an empty list: speedResults ← []
function TIME_TO_CROSS(dataframe, crossingData)
    Initialise an empty dictionary: var ← {}
    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 ← difference between the index of the minimum and maximum X-coordinates:
            time[crossingData] ← count/fps of the video
        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 ←
dataframe.get_group(id)
        Calculate the mean pedestrian height (in pixels) across all frames: meanHeight ←
groupedWithId['Heightoftheboundingbox'].mean()
        Determine the minimum X-coordinate: minXCenter, and the maximum X-coordinate: maxXCenter
        Calculate the pixel-per-meter ratio: ppm ← meanHeight/heightofperson
        Compute the real-world distance crossed: distance ← (maxXCenter − minXCenter)/ppm
        Retrieve the time taken by the pedestrian to cross: time ← groupedWithId['Time'].max() −
groupedWithId['Time'].min()
        Compute the pedestrian's speed: speed ← distance/time
        if speed ≤ 1.42 then                                    ▷ Filter out unrealistic speeds [24]
            Append the calculated speed to speedResults
        end if
    end for
    return speedResults
end function
```

2.4 Calculation of pedestrian crossing decision time

We developed a method to measure the decision time of pedestrians as they approach and begin to cross the road, as detailed in Algorithm 3. This analysis utilises video data tagged and processed through the You Only Look Once (YOLOv11) object detection system, specifically focussing on individual pedestrians identified by their unique *'YOLO id'*. The algorithm calculates the time it takes a pedestrian to decide to cross a road based on their movement data. The algorithm determines the time taken for a pedestrian to decide to cross based on their movement data. It first isolates data for each pedestrian based on their unique ID and processes their movements individually. For each pedestrian, the algorithm calculates their mean height to set a margin of 10% to detect consistent movement. 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 X coordinate data of the pedestrian in steps, checking whether the pedestrian remains within the defined margin for 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.

Algorithm 3 Calculation of pedestrian crossing decision time.

```
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 = fps/10 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
```

3 Results

3.1 Pedestrian crossing decision time and pedestrian road crossing speed

Figure 2 shows the distribution of the speed of pedestrian crossing the road and the time it takes for pedestrian to start crossing the road. The average pedestrian speed during the day across the cities analysed is 0.82 m/s (SD = 0.14). At night, the average speed is 0.80 m/s (SD = 0.16). The city with the highest daytime speed is Coimbra, Portugal, at 1.19 m/s, while Chongqing, China, records the lowest speed during the day at 0.29 m/s. At night, Monaco has the highest speed at 1.13 m/s and Shanghai, China, the lowest at 0.23 m/s.

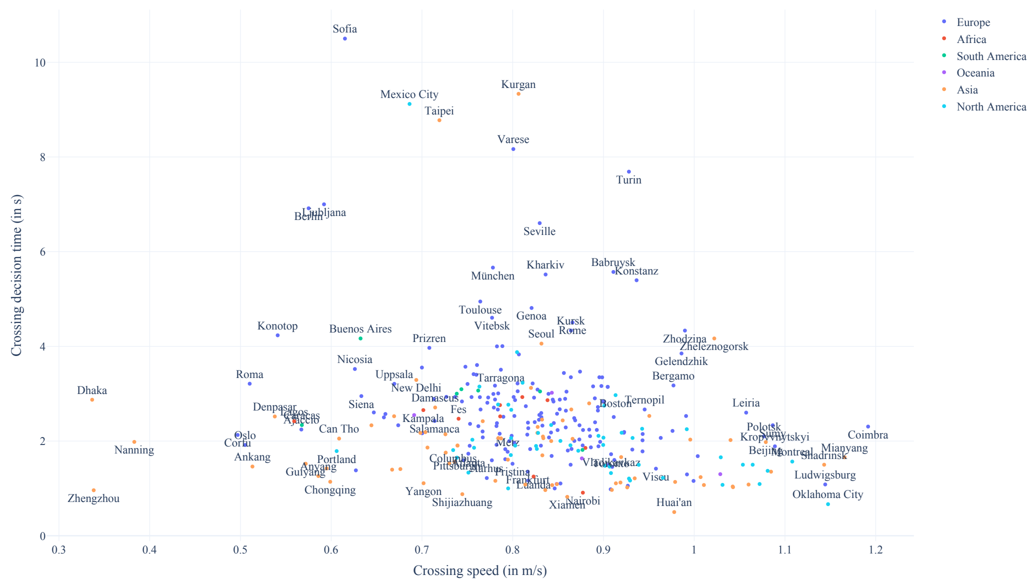

Fig. 2. Speed of crossing of pedestrian over time for pedestrian to cross the road. Each marker represents one of the 565 cities in the dataset, with colour depicting the continent.

The average time it takes pedestrians to cross the road during the day is 0.31 s (SD = 0.13). 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, Bulgaria, at 1.05 s, while the shortest time is in Huaián, China, at 0.10 s. At night, the longest time to begin crossing is observed in Mexico City, Mexico, at 1.61 s, while Monaco has the shortest time at 0.14 s.

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

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

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

Figure 3 shows the dependency between the amount of time pedestrians take to start crossing the road and traffic mortality. 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.

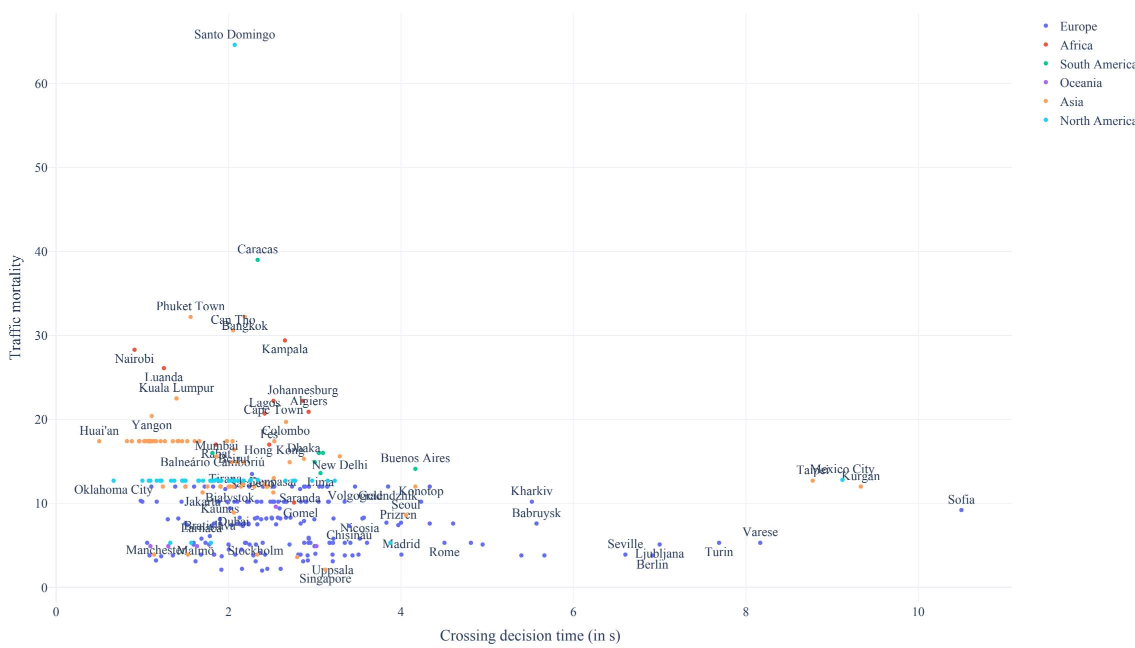

Fig. 3. Time taken by pedestrians to start crossing the road as a function of traffic-related mortality (per 100,000 of population). Each point represents a city, where the size of the point corresponds to the GMP per capita.

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.

3.3 Correlation matrix

The correlation matrix (Figure 4) illustrates the strength and direction of relationships between various parameters analysed in this study. The crossing decision time negatively correlates with pedestrian crossing speed (r = -0.29), indicating that pedestrians who take longer to initiate crossing typically walk slower. A negative correlation is observed between traffic-related mortality and pedestrian crossing decision time (r = -0.22), suggesting that cities with higher traffic mortality rates tend to have shorter pedestrian hesitation times.

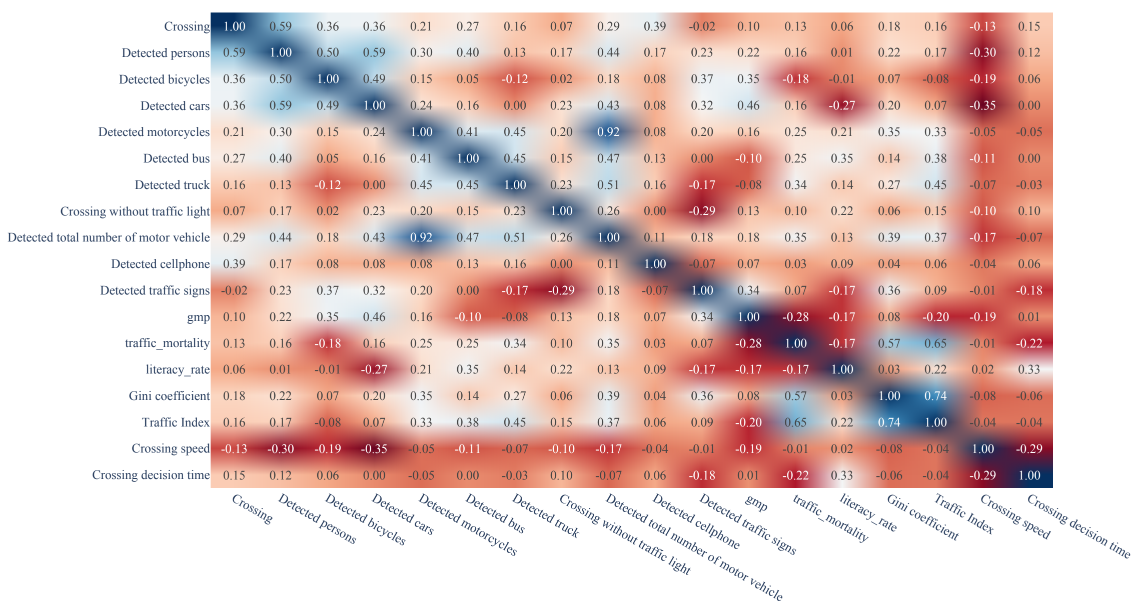

Fig. 4. Correlation matrix illustrating the relationships between pedestrian behaviour, including pedestrian crossing speed, decision time, socioeconomic indicators, and traffic-related factors.

There is also a positive correlation between literacy rates and pedestrian crossing decision time (r = 0.33), highlighting that higher literacy rates are associated with longer pedestrian decision-making times. The Gross Metropolitan Product (GMP) shows a weak negative correlation with crossing speed (r = -0.19), suggesting wealthier cities may have slightly higher pedestrian speed. Pedestrian crossing decision time shows negligible correlations with the GMP (r = 0.01) and a very weak negative correlation with the traffic index (r = -0.04).

Finally, there is a weak negative correlation between pedestrian speed and the Gini coefficient (r = -0.08), indicating that cities with lower traffic mortality could have marginally higher income equality. Pedestrian crossing speed also weakly correlates negatively with the number of detected vehicles (r = -0.17), suggesting that pedestrian speeds decrease slightly as the number of vehicles increases.

4 Discussion

The present study introduces the "PYT" dataset to explore pedestrian behaviour across the world, contributing to the vehicle-VRU interaction research community by examining differences specifically at the city level. By focussing on the analysis at the city level, this study captures the uniqueness inherent in the behaviours of pedestrians that arise from localised conditions. The analysis of this dataset reveals substantial variability in the time required to initiate the crossing and the speed of the pedestrian during road crossings. These city-specific variations are influenced by local traffic regulations, urban infrastructure, and prevailing cultural norms, highlighting the complexity and distinctiveness of pedestrian behaviour in different urban contexts.

The use of the YOLOv11 algorithm helped identify significant differences in pedestrian crossing behaviour between cities with varying socioeconomic conditions, such as traffic-related mortality rates, GMP, 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 Düsseldorf, Germany and Paris, France, 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, Mexico, and Taipei, Taiwan 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, China and Tokyo, Japan, demonstrate consistently high walking speeds 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, Bangladesh, and Can Tho, Vietnam, show slower walking speeds, which may be related to challenges such as inadequate pedestrian infrastructure or increased safety concerns at night. The significant differences between day and night pedestrian behaviour, particularly in cities like Mexico City, Mexico and Hong Kong, China, suggest that local traffic patterns, lighting conditions, and cultural factors influence pedestrian speed.

The analysis of pedestrian crossing decision time reveals further variability between cities. In cities like Warsaw, Poland and Sofia, Bulgaria, longer waiting times during the day suggest higher traffic volumes or a more cautious pedestrian approach. However, 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, Japan and Shanghai, China, exhibit moderate walking speeds and shorter start times, indicating the presence of organised pedestrian traffic systems. In contrast, cities with less developed infrastructure, such as Dhaka, Bangladesh and Yangon, Myanmar, exhibit slower speeds and shorter 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 can 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 advances in traffic systems, such as automated 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 focussing on enhancing pedestrian infrastructure and addressing socioeconomic disparities that impact traffic behaviour.

4.1 Limitations and future work

Although the PYT data set provides substantial information on pedestrian behaviour globally, there are inherent limitations associated with using YouTube videos for data collection. In particular, reliance on video footage limits the depth of analysis to the characteristics of the bounding box provided by YOLOv11, restricting the evaluation of detailed attributes such as exact pedestrian actions, precise vehicle speeds, or lane characteristics. Furthermore, the lack of information on the specific seasons and exact dates of the videos limits the exploration of seasonal or temporal variations in pedestrian behaviour.

Future research should extend the analysis to include more comprehensive behavioural assessments, such as jaywalking patterns [69, 73], cycling dynamics [18, 28], and pedestrian follower behaviours [41]. Extending the focus to other vulnerable road users (VRUs), particularly cyclists, could further enrich the dataset and broaden the understanding of mixed traffic interactions globally [46].

Future research should also address these limitations by using more advanced CV models capable of detecting a wider range of attributes within video footage. For example, 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 intentions to cross the road, time to cross predictions, current behaviour such as direction and walking posture, and appearance factors like clothing and accessories [48]. In addition, vehicle-related attributes such as lane detection [23], speed estimation [38], and distance calculations [66] between objects could also be incorporated.

In addition, a more comprehensive national analysis [31] would allow a deeper understanding of regional behavioural variations and facilitate more targeted policy making decisions [8]. Developing an entirely automated framework capable of fetching, preparing, and analysing video data autonomously would significantly enhance the scalability of such research. This approach could offer valuable tools for policymakers and urban planners to make informed decisions about infrastructure development [33], traffic safety standards [35], and transportation policies [3], ultimately contributing to safer and more efficient urban environments worldwide.

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://www.kaggle.com/datasets/anonymousauthor123/pedestrian-in-youtubepyt.

CRediT authorship contribution statement

4.2 Authors' contributions

Conceptualisation: Md Shadab Alam, Pavlo Bazilinskyy; Methodology: Md Shadab Alam, Pavlo Bazilinskyy; Data Curation: Md Shadab Alam, Olena Bazilinska, Pavlo Bazilinskyy; Formal analysis and investigation: Md Shadab Alam; Writing - original draft preparation: Md Shadab Alam Writing - review and editing: Marieke H. Martens, Pavlo Bazilinskyy; Resources: Pavlo Bazilinskyy; Software: Md Shadab Alam, Pavlo Bazilinskyy; Investigation: Md Shadab Alam; Supervision: Marieke H. Martens, Pavlo Bazilinskyy;

Statements & Declarations

Conflict of interest/Competing interests

The authors have no relevant financial or non-financial interests to disclose

Ethics approval

This study does not require any ethics approval.

Consent to participate

Not applicable

Consent to publication

Not applicable

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). doi: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). doi:10.48550/arXiv.2310.14938

[3] Saeed Asadi Bagloee, Madjid Tavana, Mohsen Asadi, and Tracey Oliver. 2016. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. *Journal of modern transportation* 24 (2016), 284–303. doi:10.1007/s40534-016-0117-3

[4] Emma L Barratt, Charles Spence, and Nick J Davis. 2017. Sensory determinants of the autonomous sensory meridian response (ASMR): understanding the triggers. *PeerJ* 5 (2017), e3846. doi:10.7717/peerj.3846

[5] 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. doi:10.1016/j.trf.2019.10.013

[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. doi:10.1016/j.trf.2019.05.008

[8] Michael GH Bell. 2006. Policy issues for the future intelligent road transport infrastructure. In *IEE Proceedings-Intelligent Transport Systems*, Vol. 153. IET, 147–155. doi:10.1049/ip-its:200600

[9] Raymond C Browning, Emily A Baker, Jessica A Herron, and Rodger Kram. 2006. Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of applied physiology* 100, 2 (2006), 390–398. doi:10.1152/japplphysiol.00767.2005

[10] A.K. Budhkar, G. Asaithambi, A.K. Maurya, and S.S. Arkatkar. 2022. Emerging Traffic Data Collection Practices Under Mixed Traffic Conditions: Challenges and Solutions. In *Transportation Research in India*, A.K. Maurya, L.D. Vanajakshi, S.S. Arkatkar, and P.K. Sahu (Eds.). Springer, Singapore, 101–129. doi:10.1007/978-981-16-9636-7_6

[11] Ralph Buehler and John Pucher. 2024. The challenge of measuring walk trips in travel surveys: problems of undercounting and incomparability among countries and over time. 7 pages. doi:10.1080/01441647.2024.2319415

[12] 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. doi:10.48550/arXiv.1903.11027

[13] 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. doi:10.1109/CVPR.2019.00895

[14] Zhengping Che, Guangyu Li, Tracy Li, Bo Jiang, Xuefeng Shi, Xinsheng Zhang, Ying Lu, Guobin Wu, Yan Liu, and Jieping Ye. 2019. D²-city: a large-scale dashcam video dataset of diverse traffic scenarios. *arXiv preprint arXiv:1904.01975* (2019). doi:10.48550/arXiv.1904.01975

[15] European Commission. 2021. *Pedestrians - Road Safety Thematic Report*. Technical Report. European Commission. https://road-safety.transport.ec.europa.eu/system/files/2021-07/road_safety_thematic_report_pedestrians_tc_final.pdf Accessed: 2024-01-19.

[16] 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. doi:10.48550/arXiv.1604.01685

[17] StreetLight Data. 2023. 3 Challenges with Transportation Data Collection—and How to Solve Them. https://www.streetlightdata.com/3-challenges-with-transportation-data-collection-and-how-to-solve-them/ Accessed: 2025-01-19.

[18] Tim De Ceunynck, Gert Jan Wijlhuizen, Aslak Fyhri, Regine Gerike, Dagmar Köhler, Alice Ciccone, Atze Dijkstra, Emmanuelle Dupont, and Mario Cools. 2023. Behavioural profiling of cycling and walking in nine European cities. *Sustainability* 15, 3 (2023), 1996. doi:10.3390/su15031996

[19] 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. doi:10.1111/1540-6040.00052

[20] 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 ....

[21] 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. doi:10.1109/TPAMI.2011.155

[22] 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.

[23] 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. doi:10.48550/arXiv.2203.02431

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

[25] 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. doi:10.1109/CVPR.2012.6248074

[26] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, et al. 2020. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320* (2020). doi:10.48550/arXiv.2004.06320

[27] Stinne Glasdam, Katharina Ó Cathaoir, and Sigrid Stjernswärd. 2024. Balancing Different Legal and Ethical Requirements in the Construction of Informed Consents in Qualitative International Collaborative Research Across Continents-Reflections from a Scandinavian Perspective. *Journal of Academic Ethics* (2024), 1–14. doi:10.1007/s10805-024-09536-5

[28] Rahul Goel, Anna Goodman, Rachel Aldred, Ryota Nakamura, Lambed Tatah, Leandro Martin Totaro Garcia, Belen Zapata-Diomedi, Thiago Herick de Sa, Geetam Tiwari, Audrey de Nazelle, et al. 2022. Cycling behaviour in 17 countries across 6 continents: levels of cycling, who cycles, for what purpose, and how far? *Transport reviews* 42, 1 (2022), 58–81. doi:10.1080/01441647.2021.1915898

[29] Shivam Gupta, Albert Hamzin, and Auriol Degbelo. 2018. A low-cost open hardware system for collecting traffic data using Wi-Fi signal strength. *Sensors* 18, 11 (2018), 3623. doi:10.3390/s18113623

[30] Bo Han. 2020. How do YouTubers make money? A lesson learned from the most subscribed YouTuber channels. *International Journal of Business Information Systems* 33, 1 (2020), 132–143. doi:10.1504/IJBIS.2020.10026504

[31] Lorena Hell, Janis Sprenger, Matthias Klusch, Yoshiyuki Kobayashi, and Christian Müller. 2021. Pedestrian behavior in japan and germany: A review. In *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1529–1536. doi:10.1109/IV48863.2021.9575809

[32] Rasheed Hussain and Sherali Zeadally. 2019. Autonomous Cars: Research Results, Issues, and Future Challenges. *IEEE Communications Surveys Tutorials* 21, 2 (2019), 1275–1313. doi:10.1109/COMST.2018.2869360

[33] Eva Ivanova and Jana Masarova. 2013. Importance of road infrastructure in the economic development and competitiveness. *Economics and management* 18, 2 (2013), 263–274. doi:.Importanceofroadinfrastructureintheeconomicdevelopmentandcompetitiveness.

[34] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. *YOLO by Ultralytics*. https://github.com/ultralytics/ultralytics

[35] Md Nasim Khan and Subasish Das. 2024. Advancing traffic safety through the safe system approach: A systematic review. *Accident Analysis & Prevention* 199 (2024), 107518. doi:10.1016/j.aap.2024.107518

[36] 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. doi:10.1109/ROBOT.1986.1087444

[37] Robert V Levine and Ara Norenzayan. 1999. The pace of life in 31 countries. *Journal of cross-cultural psychology* 30, 2 (1999), 178–205. doi:10.1177/0022022199030002003

[38] 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. doi:10.1155/2021/1577614

[39] Tobias Lohaus, Sara Yüksekdag, Silja Bellingrath, and Patrizia Thoma. 2023. The effects of Autonomous Sensory Meridian Response (ASMR) videos versus walking tour videos on ASMR experience, positive affect and state relaxation. *Plos one* 18, 1 (2023), e0277990. doi:10.1016/j.jad.2021.12.015

[40] German Lopez. 2015. *ASMR, explained: why millions of people are watching YouTube videos of someone whispering*. https://www.vox.com/2015/7/15/8965393/asmr-video-youtube-autonomous-sensory-meridian-response

[41] Wei Lv, Zhiming Fang, Xiaoge Wei, Weiguo Song, and Xuan Liu. 2013. Experiment and modelling for pedestrian following behavior using velocity-headway relation. *Procedia Engineering* 62 (2013), 525–531. doi:10.1016/j.proeng.2013.08.096

[42] 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). doi:10.48550/arXiv.2106.11037

[43] 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. doi:10.3390/jimaging7080145

[44] Aditya Medury, Offer Grembek, Anastasia Loukaitou-Sideris, and Kevan Shafizadeh. 2019. Investigating the underreporting of pedestrian and bicycle crashes in and around university campuses- a crowdsourcing approach. *Accident Analysis & Prevention* 130 (2019), 99–107. doi:10.1016/j.aap.2017.08.014

[45] 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. doi:10.1080/15472450.2017.1291351

[46] 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. doi:10.1080/15487733.2020.1756188

[47] Betty J Mohler, William B Thompson, Sarah H Creem-Regehr, Herbert L Pick, and William H Warren. 2007. Visual flow influences gait transition speed and preferred walking speed. *Experimental brain research* 181 (2007), 221–228. doi:10.1007/s00221-007-0917-0

[48] 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). doi:10.1109/TITS.2021.3107587

[49] Vicki L Neale, Thomas A Dingus, Sheila G Klauer, Jeremy Sudweeks, and Michael Goodman. 2005. An overview of the 100-car naturalistic study and findings. *National Highway Traffic Safety Administration, Paper* 5 (2005), 0400.

[50] 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. doi:10.1016/j.trf.2021.10.017

[51] Max Oudshoorn, Joost de Winter, Pavlo Bazilinskyy, and Dimitra Dodou. 2021. Bio-inspired intent communication for automated vehicles. *Transportation research part F: traffic psychology and behaviour* 80 (2021), 127–140. doi:10.1016/j.trf.2021.03.021

[52] 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. doi:10.1016/j.aap.2005.04.017

[53] Gregory C Patek and Todd G Thoma. 2013. Commentary: pedestrian fatalities—a problem on the rise. *Annals of emergency medicine* 62, 6 (2013), 613–615. doi:10.1016/j.annemergmed.2013.09.019

[54] Eleni Petridou, Nick Dessypris, Alkistis Skalkidou, and Dimitrios Trichopoulos. 1999. Are traffic injuries disproportionally more common among tourists in Greece? Struggling with incomplete data. *Accident Analysis & Prevention* 31, 6 (1999), 611–615. doi:10.1016/S0001-4575(99)00017-2

[55] 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. doi:10.48550/arXiv.1909.07541

[56] 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. doi:10.1109/ICCV.2019.00636

[57] 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. doi:10.1109/IVS.2017.7995730

[58] 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. doi:10.1109/ICCVW.2017.33

[59] 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. doi:10.48550/arXiv.1506.02640

[60] 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. doi:10.1145/3334480.3382829

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

[62] Heikki Summala. 1998. American drivers in Europe: Different signing policy may cause safety problems at uncontrolled intersections. *Accident Analysis & Prevention* 30, 2 (1998), 285–289. doi:10.1016/S0001-4575(97)00083-3

[63] 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. doi:10.48550/arXiv.1912.04838

[64] Government Technology. 2025. Waymo's Driverless Taxis Monitor San Francisco Traffic. https://www.govtech.com/transportation/waymos-driverless-taxis-monitor-san-francisco-traffic Accessed: 2025-01-19.

[65] Thi Minh Tran Tram, Callum Parker, Xinyan Yu, Debargha Dey, Marieke H. Martens, Pavlo Bazilinskyy, and Martin Tomitsch. 2024. Evaluating autonomous vehicle external communication using a multi-pedestrian VR simulator. *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies* 8 (2024), 2. doi:10.1145/3678506

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

[67] 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). doi:10.1109/TPAMI.2019.2926463

[68] 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.

[69] SQ Xie, SC Wong, Tsz Man Ng, and William HK Lam. 2017. Pedestrian crossing behavior at signalized crosswalks. *Journal of transportation engineering, Part A: Systems* 143, 8 (2017), 04017036. doi:10.1061/JTEPBS.0000055

[70] 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. doi:10.1016/j.jsr.2007.01.014

[71] 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. doi:10.1109/ACCESS.2020.2983149

[72] Xiaoyuan Zhao, Xiaomeng Li, Andry Rakotonirainy, Samira Bourgeois-Bougrine, Zishuo Zhu, and Patricia Delhomme. 2024. Crossing roads in a social context: how behaviors of others shape pedestrian interaction with automated vehicles. *Transportation research part F: traffic psychology and behaviour* 102 (2024), 88–106. doi:10.1016/j.trf.2024.02.008

[73] Yinan Zheng, Thomas Chase, Lily Elefteriadou, Bastian Schroeder, and Virginia P Sisiopiku. 2015. Modeling vehicle–pedestrian interactions outside of crosswalks. *Simulation Modelling Practice and Theory* 59 (2015), 89–101. doi:10.1016/j.simpat.2015.08.005

[74] Meixin Zhu, Xuesong Wang, Andrew Tarko, et al. 2018. Modeling car-following behavior on urban expressways in Shanghai: A naturalistic driving study. *Transportation research part C: emerging technologies* 93 (2018), 425–445. doi:10.1016/j.trc.2018.06.009