applied sciences

MDPI

Article

Road2Synth: Reconstructing Pedestrian–Vehicle Encounters from Video

Paul M. Torrens [1,*], Snigdha Anantharaju [2], Andrew Hwang [2], Ryan Kim [1] and Abhishek Kumar [2]

[1] Department of Computer Science and Engineering and Center for Urban Science + Progress, Tandon School of Engineering, New York University, Brooklyn, NY 10012, USA; kim.ryan@nyu.edu
[2] Center for Urban Science + Progress, Tandon School of Engineering, New York University, Brooklyn, NY 10012, USA; sa9069@nyu.edu (S.A.); yh5975@nyu.edu (A.H.); ak12142@nyu.edu (A.K.)
* Correspondence: torrens@nyu.edu

Abstract

There are many facets of applied science that could benefit from convergent approaches that reach across ethnographic fieldwork, sensing, and simulation in search of unified explanations. This is particularly true in pedestrian and traffic safety science, where empirical traditions for vehicle analysis must often be paired to social science understanding of pedestrian and driver behavior. Alas, these steps are usually pursued in isolation, ceding fidelity of explanation in the gaps in between approaches. There has been growing interest in using video to close these gaps, capitalizing on the prevalence of cameras in fixed traffic infrastructure. This is promising, but fixed infrastructure cameras may miss valuable in-the-moment interactions that occur at eye level as lived experiences of drivers and road-crossers. Here, we introduce Road2Synth, a research pipeline to take cellphone video captured at the roadside and add value to it through a combination of geographic information science, ethnography assisted by deep learning, and reconstructive modeling from extended reality (XR). Road2Synth targets applied analysis of safety in pedestrian–vehicle interactions, supporting statistical modeling from video, as well as establishing the ability to take flagged conflicts from video and replay them from each road user's unique viewpoint. Road2Synth is designed to be easy to use, as a supplement to traditional forms of traffic survey. We demonstrate a proof of concept for studying right-turn conflicts at signalized crossings, proven on fieldwork across 21 sites, 227 videos, 3668 pedestrian tracks, and 735 vehicle tracks in New York City and its outlying suburbs.

Keywords: deep learning; spatiotemporal computer vision; GIS; traffic safety; pedestrian behavior; safety science

Check for updates

Academic Editor: Atsushi Mase

Received: 29 September 2026
Revised: 1 November 2026
Accepted: 3 November 2026
Published: 8 November 2026

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

1. Introduction

In urban science and engineering, there is a rich tradition of engaging in fieldwork to visit, measure, and understand problems, phenomena, and processes [1–7]. In the case of traffic engineering and pedestrian planning and design, fieldwork is traditionally carried out in two modalities that are quite distinct in their approach and information-gathering: by variations of observation and surveying for physical systems [8–10], and by ethnography for behavioral systems [11–18]. Increasingly, the methodological gap between the two has narrowed. Sensors in particular are now routinely used to measure the behavior of people in their natural settings [19–22], while knowledge and ontologies from ethnography are being used by machine learning schemes to build agentic understanding and autonomous measurement [23–25]. In some cases, both traditional approaches and

automated machine intelligence are co-deployed in the field, for example, in streetscape science, where convergence between near-field sensing and computing is leading to real-time understanding of streetscapes as being both smart and sentient [26–32].

Extended Reality (XR) technologies, including augmented reality (AR), mixed reality (MR), and virtual reality (VR), can add value as a third methodology, providing broad flexibility for vicarious fieldwork, hypothesizing about supposed conditions or system states, and evaluating what-if event dynamics. XR has a robust advantage in this regard: it allows for science to be examined in interim spaces that mark up the real world (AR) [33,34], that converge elements of the real world with synthetic counterparts (MR) [35–37], or that replicate would-be variations of reality as simulations and simulacra [38]. For safety science involving pedestrian–vehicle conflicts, this provides a unique synthetic laboratory for experimenting with dangerous situations that are not accessible on the ground [39–41]. Drawing all three—surveying, ethnography, and XR—into explanatory parity could produce new experimental tools across a range of application scenarios with broad theoretical and practical uses. Convergence across these approaches could be of particular benefit in smart cities research [42,43], where the complex dynamics that yield observable scenes of everyday urban life can be ported in synthetic laboratories and enlivened against what-if hypotheses and theses with data from sensing [44].

In this paper, we examine how a unified pipeline that begins with roadside cellphone video and moves through ethnography to statistical modeling and XR reconstruction could be developed and applied to the study of unsafe pedestrian–vehicle interactions. Our main aim is to develop a relatively easy-to-use pipeline that can turn around relatively straightforward analyses of roadways without involving elaborate surveying. As a secondary aim, we explore whether it can be scientifically consistent. To that end, we aim the XR endpoint of the pipeline at virtual geographic environments (VGEs) [45–48]. As a proof of concept, we consider the case of identifying conflicts among pedestrians and vehicles in real intersection scenes across New York City in the United States, where a wide assortment of traffic patterns, passing pedestrians and vehicles, urban designs, and urban geographies can be studied. We report success in these goals, with a pipeline that we term as *Road2Synth*. We show that Road2Synth can capture and explain fleeting encounters between pedestrian and vehicles in roadside crossing applications, framed around the surrogate safety measure, post-encroachment time (PET) [49]. We also show that the captured scenes can be reconstructed as VGEs and used to explore individualized perspectives of pedestrian crossers and drivers.

The contributions of our paper are as follows. We introduce a unified pipeline that accepts point-of-view eye-level, street-level video from naturalistically observed, hyper-local streetscape sites, captured on cellphone cameras. With the assistance of a proactive visual labeling tool, users can ethnographically annotate the scenes captured in video. These annotations then launch a machine learning and computer vision process that automatically builds context by detecting, classifying, and tracking selected road users (vehicles or pedestrians), which the annotator may then edit further in response to that context. We include homography-based ground-plane mapping to geographic information systems (GISs) in the pipeline. This supports the derivation of metric empirical properties from the video for positional and trajectory-level insight, and these metrics can provide situational awareness at multiple scales from pedestrian-to-vehicle encounters all the way up to citywide safety reports that can be checked against accident data. All field data from the pipeline are passed to a statistical engine for determinant analysis and to an XR-based scene graph and rendering environment for immersive analysis. By feeding data to statistical models, we show the pipeline can inform threshold-based safety analysis on dyadic interactions. By passing the data to XR, we show that immersive versions of the recorded and annotated scenes

can be used to construct and replay synthetic individualized pedestrian- and driver-based perspectives "in the moment." Indeed, this XR end phase can be used to edge the analysis closer to the lived experiences of pedestrians and of drivers in ways that the original data may not have been able to capture.

2. Background

A range of virtual immersion technologies can now serve as explorative modalities for simulation-based engineering and science (SBE&S) [50]. Among them, XR models have proven to be particularly useful, owing largely to their ability to present would-be worlds [51] in ways that can be examined vicariously across simulation scenarios. Initially, XR–SBE&S provided a visualization mechanism for simulations, allowing users to visit and peer into dynamic simulated events. But there is now a growing tradition of harnessing XR as the actual scientific substrate for analysis. VR in particular has emerged as a tool for safety science researchers to explore the potential conflicts between human behavior as pedestrians in crossing scenarios and counterpart driver behaviors and traffic dynamics [41,52–59]. Recently, variants of this approach have been introduced for VGEs, where the intent is to establish a direct exchange between the extensible virtuality of modeling and tangible counterparts to the real world as GIS [60,61].

VGEs use a synthetic geography [62,63] as the environment for examining spatial relationships and geographic context in ways that often map one to one to reality [64]. A VGE provides a convenient mathematical translation for synthetic environments, a data infrastructure for organizing and updating information streams, and a visualization environment for assessment and analysis. Traditional forms of VGE are built from static data, usually from cartographies and geometries [65]. Importantly, the additional third and fourth dimensionality of the VGE can "reach back" to two-dimensional cartography and geometry in ways that preserve the structure and ground truth that is captured in the base data, with the implication that added-dimension analysis can be brought to parity with underlying surveys [64]. Alas, including dynamic geography (the fourth dimension) in VGEs has been challenging [48,66], partially because the underlying cartography, geometry, and spatial data on which they are built are themselves ill-suited to harnessing timing. Some work has begun to examine how VGEs can be converged with geosimulation to provide process geographies directly from data so that animated events in the VGE are sourced in underlying dynamic models [67–69]. A range of technologies from gaming and special effects, for example, have been used to facilitate this, drawing in particular on the synergies between agent-based artificial intelligence from geosimulation and finite state machines in gaming and special effects [70–73]. It has also been shown that these virtual geosimulation environments can be realized as VR, with direct interactive involvement by human users through inverted augmentation and experiential geosimulation, for example [68,74].

However, the dynamic processes that drive geosimulation rely heavily upon models [75], and few universal community models exist for pedestrian–vehicle interactions to drive them. This in turn is a byproduct of the difficulty of observing pedestrian–vehicle interactions in practice and the infeasibility of collecting empirical data because the interactions take place without warning and happen in fleeting moments of space and time. Building models is further confounded by the fact that pedestrian–vehicle interactions are the product of decisions and behaviors that happen at individual human level, but also that unfold in the shifting context of site dynamics, with non-linearities abounding up- and downscale between them [76–80].

We reason that some novel pathways for resolving these issues could come from the idea of nonrepresentational theory (NRT) [11,81,82]. Of note, NRT has begun to coalesce with long-standing traditions of cognitive mapping from behavioral geography [83,84] and

recent advances in deep learning, embodied artificial intelligence, and physical artificial intelligence [85,86] in ways that may be actionable in supporting new integrations between survey methods, XR-based modeling in VGEs, and human pedestrian and driver behavior understanding [23]. In brief, NRT posits that human behavior, particularly in geographic context, can be usefully considered through embodiment and enactment, conceptualized directly through the lens of lived experiences in everyday interactions. XR is a natural modality for facilitating NRT, and it shares significant analytical overlap with some of tenets of NRT, particularly the idea of enactment as being something that is centered on the body in motion [87,88], as choreography [89], as tourism [90]; and the potential for ethnography [11], photography [91], and videography [92,93] as modalities for studying NRT is expansive.

Although the promise of bringing these different theories, conceptualizations, and methodologies into alignment is clear, practicalities of building empirical support for NRT are thorny. Here, we point out that the cognitive and perceptual components of NRT as experience are relatively opaque to academic inquiry, particularly in ways that are empirical and generalizable to SBE&S [85]. For one-to-one mappings between individuals (crossers and drivers) and their dynamics, few reliable survey instruments are available [94–97], and what are available tend to be tedious to use for data collection [2,3,98]. In this paper, we propose that simple cellphone video, taken by people as they enact their day-to-day and momentary embodied experiences, could be used to provide data of lived geographies in ways that leverage the theoretical and explanatory advantages of NRT. By coupling video to computer vision and deep learning, we argue, empirical signatures and indices of those live (and lived) geographies can be built. When mapped to VGEs, synthetic one-to-one representations of context can be served as a counterpart setting, with the benefit that the virtuality opens up explorational exchange between reality, hypothesis-testing, modeling, and planning and prediction.

In what follows, we will show how this can be achieved (in prototypical format) in Road2Synth: as XR atop ethnographic VGEs, with synthetic stand-ins for pedestrian and driver perception, pitched toward road safety science scenarios that are difficult to access through traditional observational methodology and that are dangerous to experiment with in the real world [5,99–106]. We show that real-world on-the-ground video from urban streetscapes can be used to reconstruct VGEs directly, and that the VGE can serve as the management infrastructure for modeling safety phenomena. Further, we show that a pipeline that exchanges data between video, ethnography, VGEs, and XR can provide an explanatory medium for understanding the complex processes that could explain the behavioral geography of entities that are captured in video, as well as providing insight into the time geography of the events and phenomena that shape and are driven by them. We show this as applied science, demonstrating how Road2Synth can be used to identify and explain the phenomenon of PET: the fleeting window of space and time that unfolds in road-crossing interactions between pedestrians, drivers, crossings, and vehicles [49,107].

3. Understanding Pedestrian–Vehicle Interactions in Crossings

Not all pedestrian–vehicle interactions are a cause for concern, but those that are can be serious. Annually, in the United States, pedestrian injuries and fatalities due to crashes reached 71,635 and 7080, respectively, in 2024 [99]. The number of people that are hit by vehicles has been increasing (remarkably so) in the last 20 years, and most of this increase is in urban areas. We show results for the United States in Figure 1, but these trends are similar across the world [108–113]. Unfortunately, the trend has persisted despite the increasing sophistication of advanced driver awareness systems [114], Vision Zero campaigns [115–117], and efforts by policymakers to reverse course.

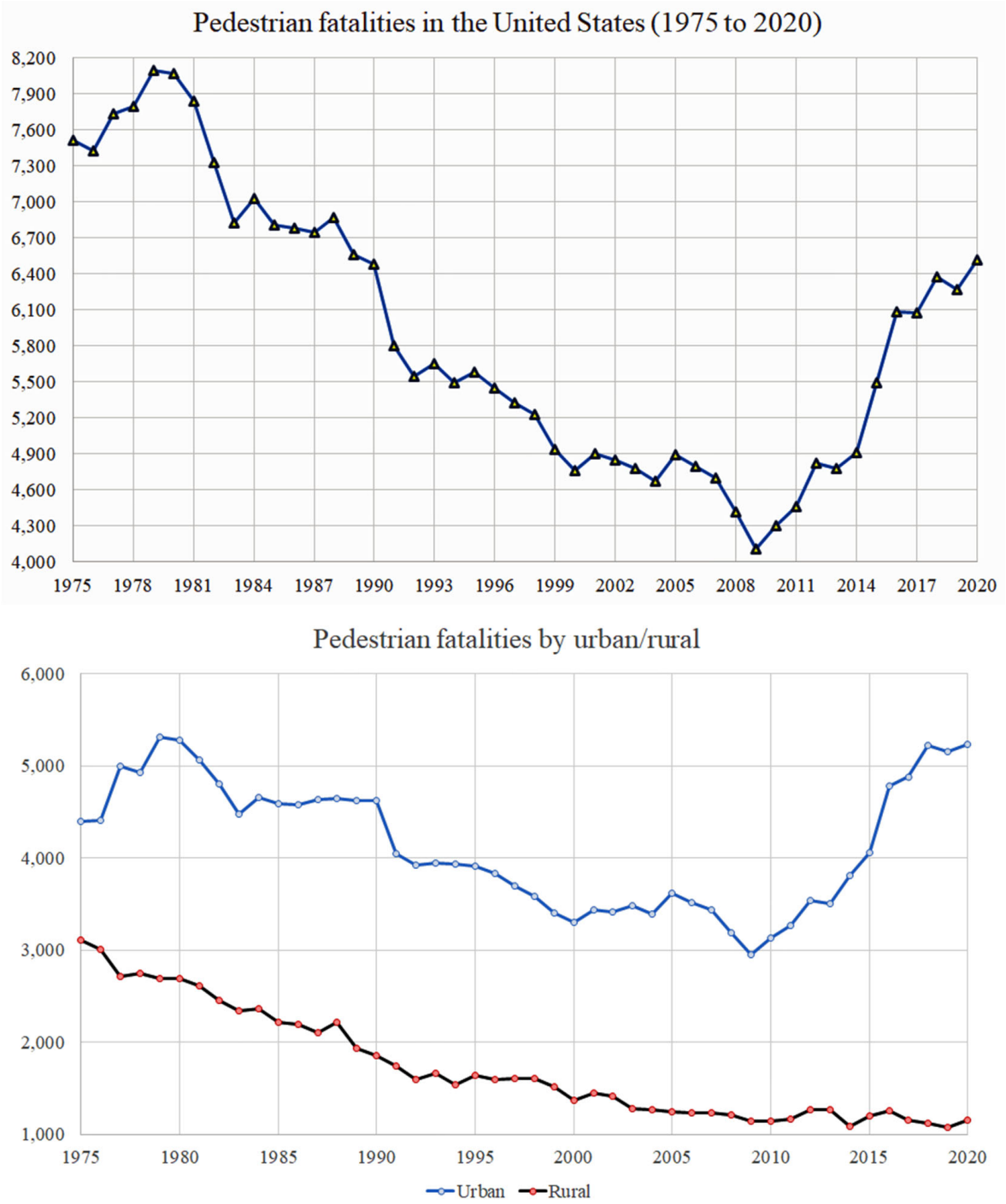


Figure 1. Pedestrian injuries and fatalities in the United States (underlying data are from the National Highway Traffic Safety Administration [99]).

There are many factors that might contribute to pedestrian–vehicle collisions, and surrogate measures are therefore useful to benchmark the circumstances involved. PET is one such measure that encapsulates some of the more vexing conditions surrounding collision potential. PET complements measures of time to collision (TTC) [118] between pedestrians and vehicles. TTC requires two road users to be on a projected collision course; it estimates the time for contact if they do not take measures to change their trajectory or speed. PET, however, measures the time that separates the movement (and therefore the interaction potential) of road users through a shared space. This shared space can be considered as a conflict zone, which is useful if we wish to target that space for a design intervention

(painting road markers, controlling access to the space, permitting or disallowing turns into the space, indicating timing with crossing signals to communicate when people should leave the space, etc.) [119]. PET is particularly relevant in considering the safety of pedestrian intersections in which vehicles begin to move near crossers, following the phase change of a traffic light from stop to go, as well as when vehicles turn against or into pedestrian streams while pedestrians have walk signals in the same space [120–123]. PET is straightforward to interpret: small PET values indicate higher propensity for conflict, because a pedestrian and a driver have less time (and by inference also space) to avoid a collision. Further, TTC and PET are usefully complementary. TTC looks forward in time geography. It is an indicator for how soon a pedestrian and vehicle might collide if neither changed course or speed. TTC therefore provides a surrogate for upcoming conflicts that one hopes later evasive action would avert. PET looks backward in time geography. It specifies how close in time two road users actually passed through the same space, and does so whether or not they were on a track to potential collision. An important point here is that PET is grounded to a space in the road. TTC, on the other hand, is not, and derives instead from road users' current positions and motion, wherever they are in the roadway, and only does so when they are on a collision course. By looking at thresholds on PET, one can examine near misses between vehicles and crossers [124,125], areas of the roadway in which there is routinely (perhaps structurally) little room for avoidance [118,123,126–128], infrastructure and design concerns at crossings [129], possible design interventions to target them [130,131], and a range of other preventative factors. In this paper, in particular, we are interested in the ability of Road2Synth to support explanations of PET that scale from exocentric (infrastructure) views to allocentric (relationship) and egocentric (behavioral) views (Figure 2).

There are three views that we consider XR can provide to enhance our understanding of PET. First, XR could help to "stage" the space-time positioning of entities in a road scene, providing site awareness and situational awareness for PET-relevant events (with GIS support). We will refer to this exocentric vista as the "scene view," and we will show that positions in exocentric perspective can be drawn directly from single-perspective monocular roadside video taken from a straightforward cellphone device. The same device, using a quick walk-around video, can additionally supply enough data to generate a three-dimensional reconstruction of the intersection (as a VGE). Second, a range of relativistic geographies unfold within the scene view, with affordance [132] between things and other things, which we delineate as an "allocentric view" [133,134]. This allocentricity, we argue, is critical in understanding PET, as it frames who has right of way [135], who has turn precedent in interactions [136], who should yield to whom in what events [137,138], who ought to move or stop first [125,139], and so forth. Indeed, PET itself is an allocentric measure, and so deriving it from exocentric perspectives is useful in and of itself. Third, XR could assist in modeling the driver perspective on the PET bubble, as well as the counterpart pedestrian perspective [140,141]. But these perspectives are not natively part of the scene view and require computing to reconstruct. We will term this as the "egocentric view" and we will demonstrate that this view can be built from VR cameras applied after reconstruction. Taken together, the scene view, allocentric view, and egocentric view allow for realistically vicarious insight into PET to be built, at the characteristic timing of potential conflict. Moreover, as we will show, these views can be derived from cellphone video captured opportunistically and with ease: roadside recordings and a quick walk-around capture. The trajectories and annotations that Road2Synth supports (and that helps to populate these views) can also then feed statistical models of the factors associated with PET in the real, captured scenes. Further, the detailed census, atlas, and choreographies that can be gathered from Road2Synth allow for statistical factors of PET to be examined, contextually, in detail.



Figure 2. (**Left**): The concept model for detecting egocentric (person-to-thing) and allocentric (thing-to-thing) interactions in small spaces and times from exocentric (third-person) vistas. (**Right**): Real photo images from roadside fixed camera video.

We regard these three views as important targets for Road2Synth for several reasons. First, producing them in XR from monocular video has a computational and methodological value proposition, particularly in unifying the views for data exchange and measurement. The second reason is practical. Vehicles and crossers must share the road at crossing sites, which means that rules must be in place to halt the movement of one while permitting the movement of the other. In busy crossings, this stop-and-go oscillation has a negative effect upon traffic flow, and policies are put in place to essentially allow for mixed traffic. This includes concurrent phasing [142], in which both vehicles and crossers have signals that direct them to proceed through a crossing; and permissive phasing [126], in which a driver is allowed to continue their trajectory through a crossing, but subject to yield orders. Critically, both concurrent and permissive phasing may take place during ninety-degree turns at intersections [122,143,144], where drivers have a significant perceptual

disadvantage compared to the vista afforded to them when moving straight ahead. In our study area of New York City, a third policy is in effect at many intersections: the leading pedestrian interval [145], in which pedestrian crossers get a head start from signaling that allows them to enter a crosswalk and proceed for several seconds before vehicles are signaled to also enter the same space. Both need to pay attention in these scenarios, often while watching multiple things in the road and the surrounding scene. Understanding what competes for their attention, when, and where during the time geography of potential conflict is important in determining whether those policies are safe. Here, we contend that XR could play a useful role in illuminating those conditions.

One way of considering different strategies of pedestrian–vehicle mixing during crossing zone conflicts is that they end up authorizing both drivers and pedestrians to enter into a time geography of potential conflict, and leave it to both to work out the right of way in the moment. However, as we detail below, there is not parity in awareness or in movement between the two [13,136,140,141,146,147]. During road-crossing episodes, perceptual mismatches form between drivers and pedestrians. Road2Synth, by bringing all views together, can help establish options for bridging gaps in shared experience. Briefly, below, we explain the driver and pedestrian fields of view. In Sections 4.6–4.9, we will show that these perspectives can be reconstructed through the synthetic "eyes" of both, using Road2Synth on real video data.

3.1. Driver Perspectives on Post Encroachment Time

Vehicle drivers are necessarily embodied to their vehicle. As a consequence, their body movements are entangled to that vehicle, with inherited restrictions. Consider that drivers must sit in a fixed position, elevated from road surfaces with limited ability to shift and move from the driver's seat. The rules of the road for vehicles are unique, controlling the physics of driving, the spaces that vehicles may enter, and the signaling and signage that must be obeyed during movement. As a result, drivers' perception is strongly typed to the movement of the vehicle. Low-bandwidth stimuli such as acceleration [148] and engine noise [149] provide much of their awareness of their surroundings, and do so by superimposing on a driver's innate sense of movement and motion [150]. Further, drivers' vision is obstructed by limited views of the road and the events unfolding within it. The structural frames of vehicles block full vision in most cases, while in modern vehicles an array of vehicle telemetry (and perhaps even entertainment) systems compete for their attention [151]. Moreover, drivers' ability to engage in dyadic relationships with other people in the road is reduced to either assumptions of inference or to furtive glimpses of non-verbal communication (NVC) [152], usually through unverified or weakly confirmed eye contact in glances and gazes at and from pedestrians [141]. Vehicles have few ways in which they can convey their next steps in the road: they are mostly limited to indicator lights when turning. Taken together, these limitations of enclosure, perceptual shifting, constrained visibility, and muting of naturalistic NVCs make it challenging for drivers to perform cognitive mapping of egocentric and allocentric affordances to road dynamics in ways that have parity in uptake for the pedestrians that they share space and time with. Moreover, this relative mismatch in behavioral geography unfolds in fleeting moments of space and time, often fractions of a second and centimeters of space [153], at a critical juncture in the negotiation of perception, perspective, and expectations for next actions in the negotiated geography between vehicles and road-crossers.

3.2. Pedestrian Perspectives on Post Encroachment Time

For pedestrians, road spaces present differently as lived experiences. Pedestrians are not as confined as drivers, they have fewer rules, and those that are applied to their behavior

may sometimes be broken without recourse. For example, pedestrians may ignore crossing signals and jaywalk [5], and the physics of their locomotion is constrained only by their abilities [17,154,155]: they can stop, sprint, and about-face maneuvers with relative ease and discretion. Pedestrian vistas scale dynamically while walking: from some locations and times, pedestrians may have clear views, but those can quickly become occluded, including total occlusion [156]. Their attention is often pulled toward many competing elements around them, including drawing their focus away from the road, from crossing signals, and from vehicles nearby. In many cases, diverting gaze while crossing is a natural part of locomotion when negotiating unnatural terrain (curbs, storm drains, manhole covers) [138]. Pedestrians are also influenced by social signals in different ways [157], including phenomena in which they react to other pedestrians' interactions with vehicles, in a sort of meta-reaction that can create macro-scale phenomena such as collective motion in unsafe conditions [158]. In some cases, these meta-behaviors are socially motivated [159], particularly for younger crossers, with some evidence of peer pressure at play during risky crossings [103,160,161].

4. Methodology

The phenomenon that Road2Synth essentially tries to model, then, is as follows. Pedestrians and vehicles course into the same small space and time in roadscapes, and they are asked to negotiate the time geography that they share in a way that produces collision free movement. This takes place with soft supervision, in the sense that both sets of road users are advised to follow indicators such as zebra crossings, road markers, crossing signals, and priority of movement, but may choose not to.

We address this large set of factors in Road2Synth in three ways. First, we aim to extract as many of the site factors as we can from video, using combinations of deep learning and annotation. Further, we map these to a shared spatial and temporal frame in a GIS (and then in a VGE) so that they can be analyzed statistically. Second, we distill dyadic behavior of pedestrians and vehicles in significant parts of the road to a threshold function (PET). This establishes an envelope criterion for examining the fleeting moments of space and time in which vehicles and crossers come into near contact, and then isolating the factors at play in influencing the time geography of that conflict (the site factor variables from deep learning and annotation). Third, we use XR reconstruction to build individualized, synthetic, perspective-based views that map to the lived experiences of pedestrians and vehicles within those same factors. In this way, Road2Synth flits across the three views that we outlined in Section 3: Scene, Allocentric, and Egocentric, providing data for each and linkages in between. Importantly, we propose to inform these questions with single-perspective monocular cellphone video taken from the roadside (at sidewalks), and we aim to generate data that can inform analysis at shared space-time parity to the scale of pedestrian–vehicle conflict (at around two meters).

4.1. Site Selection

We sampled a variety of road-crossing intersections across New York City and its outlying boroughs (which include suburban geography) for data collection (Table 1). 24 sites were selected in total to prototype the Road2Synth pipeline and to test generalizability across factors (Figure 3). Data collection for three of the sites encountered hardware problems, and so we use only 21 sites in the downstream analysis.

Table 1. Fieldwork site visit characteristics (21 sites; 1430 observations; 220 videos total).

Location	*Neighborhood*	*Season*	*Precipitation*	*Temperature (°F)*	*Time of Day*	*Day of Week*	*Signalized Crossing?*	*Pedestrian Density*	*Urban Geography*
Bronx									
East 149th Street and Grand Concourse	Mott Haven	Winter	Snow	36	Afternoon	Weekend	Y	Low	Commercial
Transit hub area; it contains dense retail uses outside transit stops; foot traffic and vehicles are mixtures of locals and through traffic									
Brooklyn									
Fort Greene Place and Lafayette Avenue	BAM Cultural District	Winter	Fair	28	Afternoon	Weekday	N	Low	Commercial
Cultural/tourist area, with museum, theaters, restaurants, close to a transit interchange. Foot traffic/vehicles mix locals, visitors, passersby									
Pacific Street and Court Street	Boerum Hill	Fall	Windy	51	Afternoon	Weekend	N	Low-Med	Residential
Townhouse-type residential area with adjacent shops and eating establishments; foot traffic and vehicles are mostly locals									
Flatbush Avenue Extension and Myrtle Avenue (North)	MetroTech	Fall	Sunny	50	Morning	Weekday	Y	Low	Office
Myrtle Avenue and Flatbush Avenue Extension (South)		Fall	Sunny	51	Morning	Weekday	Y	Low	Office
Core of downtown Brooklyn, with two university campuses and corporate headquarters in close proximity; foot traffic is largely commuter-based									
Tillary Street and Jay Street	Downtown Brooklyn	Fall	Sunny	52	Morning	Weekday	Y	Low	Office
Core of downtown Brooklyn, with two university campuses and corporate headquarters in close proximity; foot traffic is largely commuter-based									
Fulton Street and Lafayette Avenue	Fort Greene	Winter	Fair	33	Afternoon	Weekday	Y	Low	Commercial
High street/main street commercial artery in residential area with historic significance. Foot traffic/vehicles mostly locals and passersby									
Third Avenue and 37th Street	Sunset Park	Winter	Cloudy	33	Afternoon	Weekday	Y	Low	Industrial
Industrial and "big box" retail area with large anchor warehouse store and industry. Foot traffic is customers; vehicles include freight									
Manhattan									
West 42nd Street and Sixth Avenue	Bryant Park	Fall	Fair	48	Evening	Weekend	Y	Very high	Tourist
Inner-city park and site of the NY Public Library with corporate offices nearby. Foot traffic is mostly workers; vehicle traffic is mixed									
Bowery and Canal Street (West)	Chinatown	Winter	Cloudy	33	Afternoon	Weekday	Yes	Med	Mixed
Canal Street and Bowery (East)		Winter	Fair	35	Afternoon	Weekday	Yes	Med	Mixed
Dense mixture of residential and ground-level retail. Foot traffic is local shoppers and largely senior; mixed vehicle traffic									
Columbus Avenue and West 72nd Street (East)	Lincoln Square	Winter	Light Rain	36	Afternoon	Weekend	Yes	Med	Mixed
West 72nd Street and Columbus Avenue (West)		Winter	Light Rain	36	Afternoon	Weekend	Yes	Med	Mixed
Mixed-use residential and broadly commercial with apartments and retail. Foot traffic is locals and workers; vehicles are passersby									

Table 1. *Cont.*

Location	*Neighborhood*	*Season*	*Precipitation*	*Temperature (°F)*	*Time of Day*	*Day of Week*	*Signalized Crossing?*	*Pedestrian Density*	*Urban Geography*
Clinton Street and Delancey Street	Lower East Side	Winter	Fair	35	Afternoon	Weekday	Yes	High	Mixed
Mixed-use residential/commercial/retail area at entrance to a major city bridge. Foot traffic is mixed; vehicle traffic is cross-town passage									
West 43rd Street and Seventh Avenue	Times Square	Fall	Fair	48	Evening	Weekend	Y	Very high	Tourist
Seventh Avenue and West 44th Street		Fall	Fair	48	Evening	Weekend	Y	Very high	Tourist
International tourist site. Foot traffic is mostly visitors; vehicles are passersby and cabs; both are very busy day and night									
West 96th Street and Amsterdam Avenue	Upper West Side	Winter	Light Rain	36	Afternoon	Weekend	Yes	Low-Med	Residential
Residential site (apartments) with some local retail. Foot traffic is local; vehicles are passersby									
Sixth Avenue and West Third Street	West Village	Winter	Light Rain	35	Evening	Weekend	Yes	Low	Commercial
Retail area at the boundary of a residential district. Foot traffic is locals and school commuters. Vehicle traffic is mixed									
Queens									
23rd Street and Jackson Avenue (North)	Hunters Point	Winter	Rain	39	Evening	Weekday	Yes	Low-Med	Residential
Jackson Avenue and 23rd Street (South)		Winter	Rain	39	Evening	Weekday	Yes	Low-Med	Residential
Mixed-use area in inner suburb, with office buildings and residential apartments near transit interchange. Foot and vehicle traffic is mixed									
Northern Boulevard and 48th Street	Long Island City	Winter	Fair	41	Afternoon	Weekday	Yes	Low-Med	Residential
Vehicle-dependent area along a major roadway (car detailing businesses and a supermarket). Foot traffic is local; vehicle traffic is throughput									
(Below) Sites that were visited, but not considered in the statistical model:									
Location	*Neighborhood*	*Season*	*Precipitation*	*Temperature (°F)*	*Time of day*	*Day of week*	*Signalized crossing?*	*Pedestrian density*	*Urban geography*
Brooklyn									
86th Street and 18th Avenue	Bensonhurst	Winter	Fair	33	Evening	Weekday	Yes	Low	Mixed
Mostly residential housing with some local retail. Foot traffic and vehicle traffic are local									
Flatbush Avenue Extension and Tillary Street	Downtown Brooklyn	Fall	Sunny	52	Morning	Weekday	Yes	Low	Office
Downtown area with office towers and hotels. Foot traffic is locals and commuters; vehicles are passing through; many transit vehicles									
Jay Street and Tillary Street	Downtown Brooklyn	Fall	Sunny	52	Morning	Weekday	Yes	Low	Office
Downtown university and office area. Foot traffic is workers and students; vehicle traffic is throughput with a lot of transit									

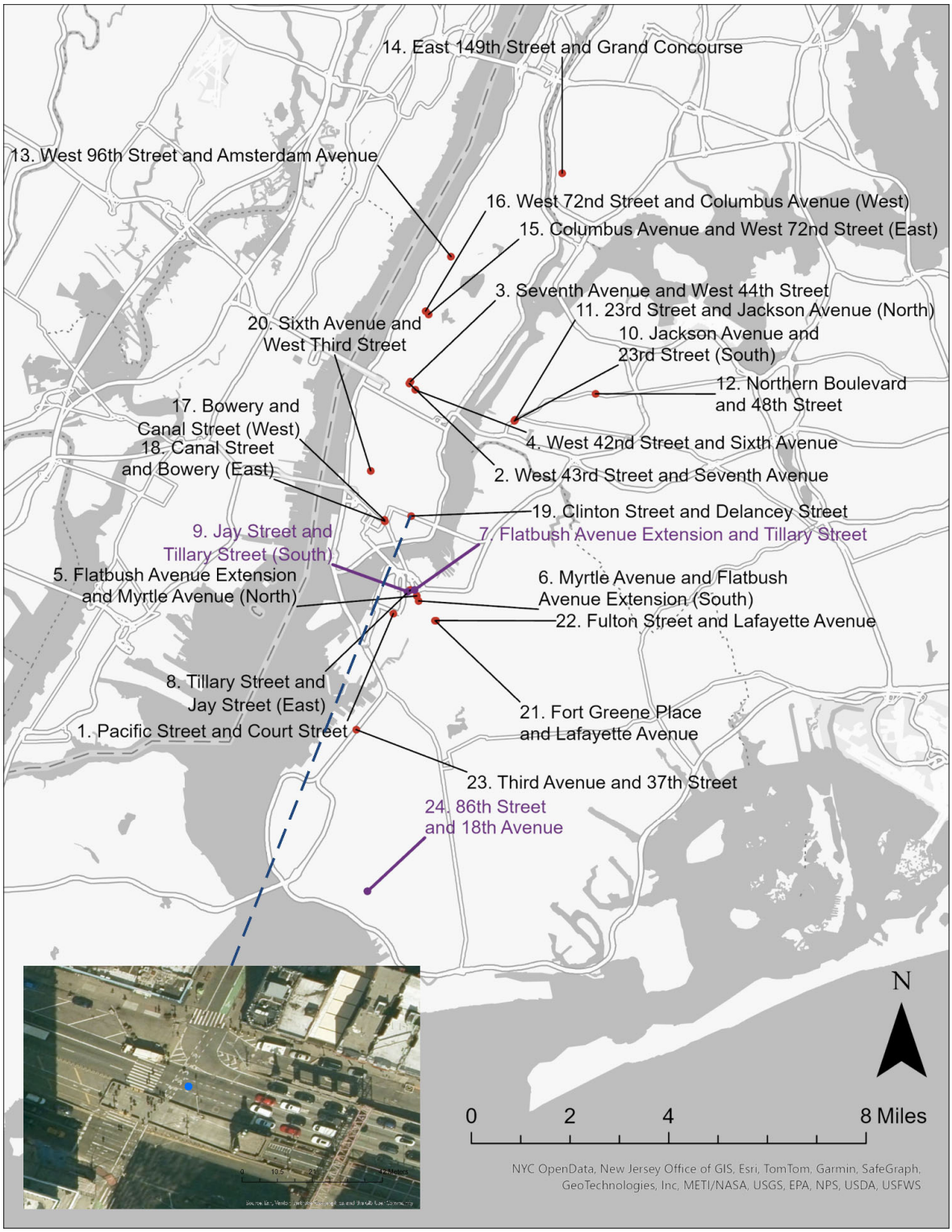


Figure 3. Fieldwork sites covered a wide swath of New York City and its outlying suburbs, encompassing a variety of sites and situations (citywide).

In selecting final sites for data-collection, we weighed four interlocking factors: *neighborhood characteristics* (trip types, demographics, age of urban infrastructure, land-use mix) [4,5]; *population density* (to account for low volumes of crossing density and for high volumes) [162]; *intersection design* (signalized and non-signalized crossings [163], pelican crossings [15,164], two-lane and multilane roads [128], road overpasses and bridges, bicycle

lanes [165], traffic turn rules [144]); and *historical accident and fatality data* (to include the city's most dangerous intersections as well as Vision Zero [115,116] interest sites).

4.2. Fieldwork

We visited the sites across varying time of day, time of week, season, and weather conditions (Table 1). At each site, we recorded simple cellphone video from tripods on the sidewalk looking into crossing sites at human eye level (Figure 4). (We also recorded walk-around video by traversing the perimeter of the intersection.) In New York City, it is permissible to record video in public with one-party consent (see Section 9 on privacy and ethics for the study). Further, the fieldwork protocol was approved by the authors' Institutional Review Board Human Subjects Protocols before fieldwork was initiated.



Figure 4. Some example video scenes from our fieldwork, showing variation in crowding, time of day, traffic conditions, weather, season, and infrastructure (all faces, logos, and license plates are distorted for anonymity).

Video was collected for both near and far vantages of crossing sites at corner crossings. Video was specifically recorded at human eye level to mimic pedestrian embodiment in the site dynamics. In total, we recorded 227 videos (215 ended up being usable), capturing data on complete signal phases (traffic state, state transitions, crossing signal phases, pedestrian- and crowd crossing epochs, and vehicle right-turn events). Different sites yielded different censuses of road processes and their event components. Most locations yielded 1–20 entities per crossing epoch, while locations such as Times Square (a massively busy tourist site) produced averages of 100 per signal phase. Our data collection captured crossing scenarios for individuals, groups, dyads, and crowds, composed of a full range of demographics. Additionally, we collected video for a broad range of vehicle types and driver behaviors (adherence to signals, non-adherence, rushed behavior, phase shifts in behavior, and sudden event onset). From the fieldwork, we identified anecdotally dominant components of PET for (1) vehicle factors, (2) pedestrian factors, (3) site factors, and (4) infrastructure factors (We describe these operationally in the experiments that we introduce in Section 5.)

New York City provides a useful test environment for scientific research for the same reasons that it is so popularly used for filmmaking: almost any streetscape scenario or configuration is available within the city. There are advantages and disadvantages to focusing on a single city (albeit one of the world's largest cities). The main drawback is that it narrows our prototype to a single set of rules of the road and biases it to the traffic patterns that the city produces, as well as the collective norms of pedestrian behavior. But New York City is massively varied and diverse, and we were able to engage fieldwork across a wide swath of the city.

For example, New York City, outside Manhattan, is actually quite suburban, and this allowed us to sample many different types of urban geographies and traffic infrastructures. Consider that the 21 analyzed sites span four boroughs (with ten sites in Manhattan, seven in Brooklyn, three in Queens, and one in the Bronx), each with varying neighborhood character and land use profiles and mixes (see details in Table 1). They include the entertainment core at Times Square, with its nearby Broadway theaters, tourist hotels, and major retail stores. More historical areas are represented by Bryant Park, beside the main branch of the New York Public Library (NYPL). Major office and commuter areas are covered by the MetroTech office complex in Brooklyn. University environments were represented around New York University and City Tech at different campus locations. Nightlife areas were covered by the performing-arts district that is anchored to the site of the Brooklyn Academy of Music (BAM). We also visited several residential neighborhoods, including single-family townhouse residential blocks in Boerum Hill and mixed residential–commercial sites with large numbers of apartment complexes in the Columbus Avenue area around Central Park. For varying demographics, we visited the commercial area of Chinatown, which hosts a large senior population. For urban geography with relatively new infrastructure, we visited new-build high-rises in Long Island City. For industrial spaces, we visited the automobile repair-oriented retail arterial on Northern Boulevard. For retail areas with large numbers of shoppers, we sampled the South Bronx retail and transit hub at the 149th Street–Grand Concourse (which also hosts an important subway interchange) and a big-box (car-oriented) retail area in Sunset and Industry City, which host a number of warehouse-style carpark-based stores. Considered for variation by land use, the fieldwork sites cover a wide range: commercial uses (4); industrial uses (1); mixed uses (5); office uses (3); residential uses (5); and tourist uses (3); with further variation in the pedestrian densities at each, ranging from low (eight sites) to high (four), and densities in between.

We additionally chose fieldwork sites to vary the road infrastructure samples, of which New York City has many configurations. Across all sites, we sampled different lane configurations, including one lane (and one-way streets), up to multiple roadbeds at a single site (the Grand Concourse with eight lanes in different configurations). We additionally visited intersections with major shifts in arterials (including entrances/exits to bridges and highways. Additionally, we sampled areas in which the civic infrastructure for pedestrians and for vehicles come into stark juxtaposition, e.g., sidewalks near expressway on-ramps, and local arterial roads that abut pedestrian plazas. Our fieldwork additionally included bicycle lanes between sidewalks and roads (ten protected lanes and four painted lanes in our sampling). We also considered variation in crosswalk signaling infrastructure. Of the 21 sites that we used for analysis, 19 had signals and 2 had none.

We also varied fieldwork to account for shifts in natural conditions. We deliberately staggered our site visits to cover two seasons, different days of the week, and different times of day, evening, and night. In the process, we gathered data for a range of precipitation conditions (including rain and snow) and temperatures (28 to 52 °F).

4.3. Deep Learning as a First Pass of Scene Understanding

We deployed machine learning to generate an automated contextual canvas (of pixels), census (of entities), atlas (of geography), and choreography (of maneuvers) for each videoed scene (Figure 5). For the pixel space, we generated a convolutional understanding of video frames [166]. Then, from the convolutional space, we ran object recognition to classify regions of the video that were occupied by pedestrians, objects that were being carried or pushed by pedestrians, and vehicles. We used deep learning on the convolutional pixel space to individually (anonymously) identify individual pedestrians and vehicles in a given scene frame, classify them (into vehicle types), and to box-bound them to the pixel geometry, with confidence intervals for object type and class membership. We relied on the "you only look once (YOLO)" scheme [167], *YOLO 26n* [168] to handle this in the pipeline. There are many such deep learning approaches to build feature recognition on frames, but YOLO is fast, reliable, and generalizable. *YOLO 26n* operates on a single convolutional neural network (CNN), which is applied (holistically) to each image frame in a video. For classification, we used *YOLO 26n*'s native connections to *Objects 365v1* [169] for pre-training, and *Common Objects in Context* (COCO) [170] for refinement [171].

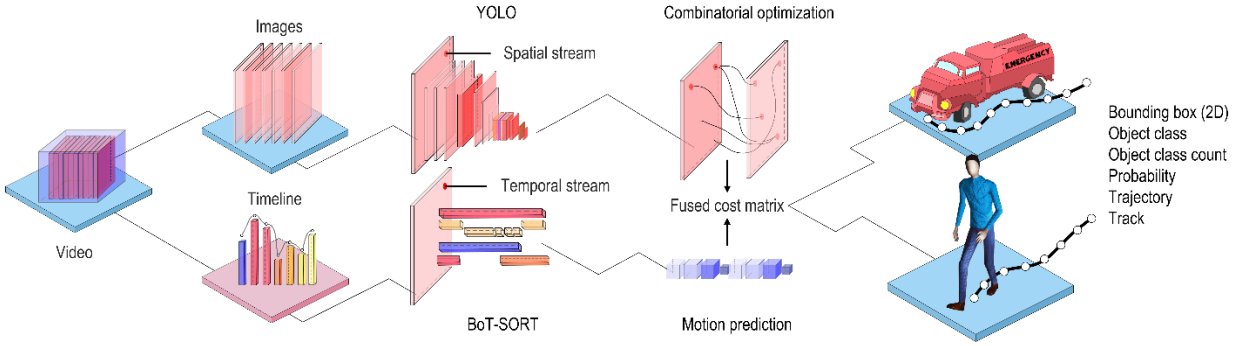


Figure 5. The deep learning pipeline for Road2Synth.

With *YOLO 26n* bounding boxes as persistent positional indices of individual pedestrians and vehicles in videos, we used multiple object tracking (MOT) to measure trajectories through the space-time geometry of a video pixel space. Specifically, we employed the detect-and-track scheme from the simple online real-time tracking (SORT) [172] MOT approach, using BoT-SORT [173] to provide between-frame tracking persistence where there were occlusions in the video. Within MOT, generally, footstep-adjacent space-time trajectories through successive frames are built by tracking the bottom of bounding boxes. Frame-by-frame persistence is assisted with a Kalman [174] filter and bipartite graph matching [175,176], using usual tracklet approaches [85,172]. Additionally, we made use of YOLO-based feature recognition to identify, classify, bound, and track the objects that pedestrians were transporting in a given scene. This was used to reason about the role of encumbrance on crossing behavior and events, and included classes of baggage, cellphones, mobility aids, and strollers.

4.4. Manual Space-Time Ethnography as a Second Pass of Scene Understanding

With a diverse set of video evidence for real-world road scenes, and automated deep learning for scene segmentation, classification, and tracking, we built a customized user annotation and labeling tool for post hoc markup and study of crossing data (Figure 6). The tool supports space-time ethnography with exploratory reach to individual pedestrians and vehicles, upscale to the crowd, and to the holistic road intersection scene. The annotation functionality is extensible, but initially we focused on supporting space-time

behavioral codes (which we cataloged manually through researcher observation across all 24 visited sites).

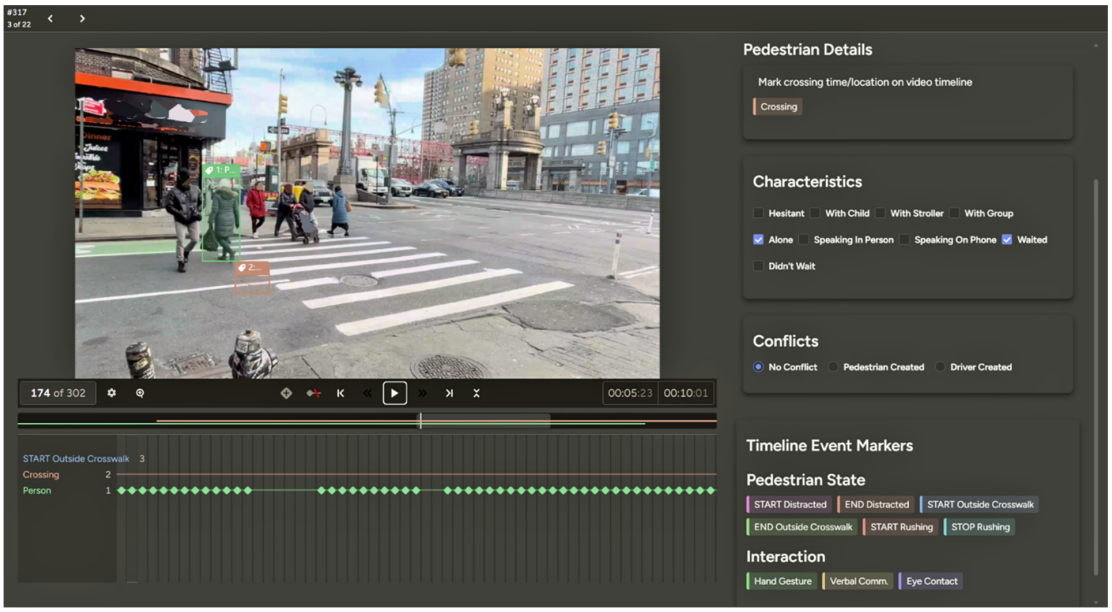


Figure 6. Semiautomated, deep learning assisted, annotation of site videos.

These codes included feature classification and prospective trajectory-ahead tracks through the video, both obtained by deep learning. We additionally added context-assignment tools for manual coding of the video. These included the ability to bound and label features of relevance (to manually bound pedestrian features and vehicle features if desired and to specify road features such as pedestrian-crossing areas (Figures 6 and 7)). Further, users were provided with the ability to scrub through video, pull up on-screen displays of deep learned features for any and all entities and objects, and to manually annotate the scene from pre-factorized attributes and characteristics. For pedestrians, users could note movement, relative positioning, interaction, distraction, relative time geography, and encumbrance. For vehicles, users could annotate movement, relative position, interactions, relative time geography, and vehicle type. Within each characteristic, we further supported break-out detail with additional granularity to permit labeling of sub-classes: non-verbal communication (NVC), sound, distance, topology (inside, outside, etc.), and events (Table 2).

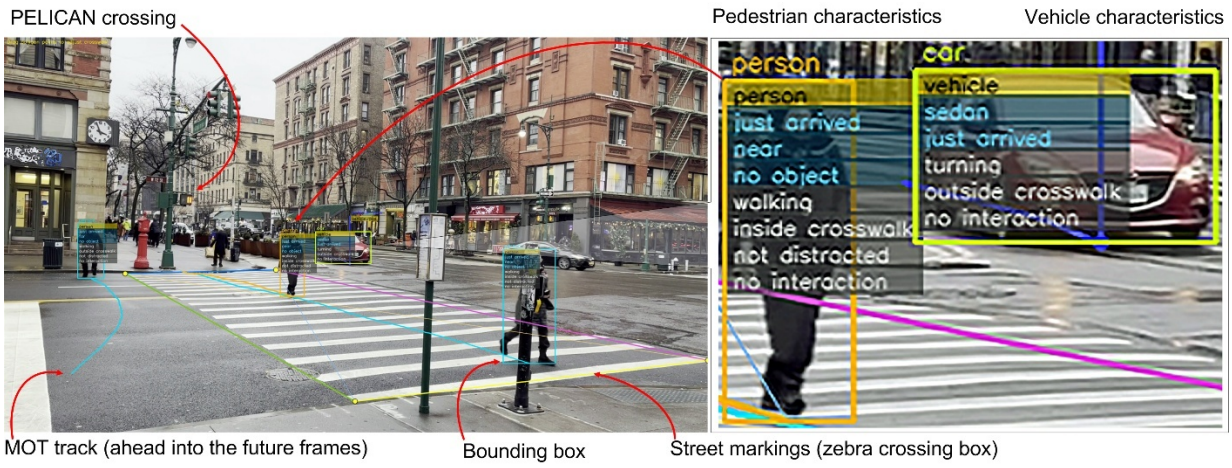


Figure 7. Deep learning and user-annotated pedestrian behavior and event labeling.

Table 2. User annotation classes in Road2Synth (variables factored from all 24 sites).

Entity	*Attribute Captured*	*Label*	*Rate*
Person	Locomotion	Walking, waiting	Per frame
	Crossing phase	Outside crosswalk, inside crosswalk	
	Distraction	Not distracted, distracted	
	NVC	No interaction, gesture, verbal, eye-contact	
	Timing	Just arrived, waited	Per track
	Starting point	Near, far	
	Encumbrance	No object, carrying object	
Vehicle	Steering	Turning, waiting	Per frame
	Road position	Outside crosswalk, inside crosswalk	
	Interaction	No interaction, honk, gap-find, gesture, verbal, eye-contact	
	Vehicle type	SUV, truck, bus, sedan, van, pickup truck	Per track
	Timing	Just arrived, waited	

4.5. Space-Time Homography

Because near collisions between pedestrian and vehicles and the dynamics of PET that unfold between them are a function of their choreography and proximity in crossing space, we sought to reconstruct trajectories of all entities in the video. Our fieldwork cellphone videos are monocular and recorded from a single, stable (tripod), eye-level perspective. Therefore, the video is almost always subject to occlusion and to related trajectory loss (or partial trajectory loss and reappearance) in crowded scenes. Irrespective of these innate problems (see further discussion in Section 6 on limitations), this is exactly the sort of video that is straightforward to obtain and to generalize for data collection at almost every site, and it is representative of the viewpoints that pedestrians face at the roadside. And so, "working with what we had" (occlusions and all) is a design choice.

Pedestrian trajectories are necessary to reconstruct the motion path for assets in the graphics scene graph. The trajectories that are generated by deep learning are in pixel space, but obviously pedestrians and vehicles interact in geographic space, and so a mapping between the two spaces is needed to resolve them. Moreover, we need geographic trajectories to register entities in XR. We used *FFmpeg* [177] to standardize video by decoding, filtering, and converting to build parity across frame rate and resolution. To map entities' positions in pixel space to their corresponding locations in geographic space, we performed a series of transformations. Our first step in doing so was to establish (in-image) positions for pedestrians and vehicles on the video frame road surface picture; next, we sought to map those points to a geographic space that could represent the real-world counterpart.

We took the output frames from the deep learning tasks, and we used them to establish the midpoint of the bottom edge of the pedestrian bounding box (where pedestrian footfall comes into reliable and repeated contact with road surfaces). These midpoints exist, then, in the image and also at a contact point with the road. Because our videos were captured by human observers standing at the roadside with an eye-level tripod position, the lower midpoint also sources each video as a vertical plane anchored to the horizontal of the crosswalk, and a single homography applies in each video. An equivalent point-position is not recoverable in the same way for vehicles, as the dimensions of vehicles are subject to distorted view in video as they turn and warp perspective relative to the camera. To address this, we used the midpoint of the front bumper of vehicles, with a vertical drop line to the road, which annotators marked at intervals along each vehicle's track (our annotation tool then interpolates between these marks to build the track).

We transformed positions from the pixel space to the ground plane (in two dimensions) by homography (using *OpenCV* [178]). Given a YOLO bounding box with corner positions (x_{min},y_{min},x_{max},y_{max}) in (Cartesian) pixel space, we used the foot contact pixel point (u,v) at the bottom center-point of the bounding area and specified it in homogeneous coordinates, $\mathbf{p} = [u,v,1]^T$. The product of a homography matrix H and $\mathbf{p}$, divided by scale factor λ, then gives us coordinates in the ground plane (scene space) as X and Y. In Equation (1), the entries h_{ij} of H are determined by the camera angle, position, and lens. These entries may not be known in advance, in which case they can be estimated by the direct linear transform (DLT) from $n \geq 4$ control points, i.e., positions that are known to the pipeline in image coordinates and in ground-plane coordinates (no three of either being collinear) by stacking the constraint matrices A_i into A and then calculating $\hat{\mathbf{h}}$ as the least-squares estimate of the vectorized homography $\mathbf{h}$ (see Equation (1)).

$$\begin{gathered}
\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{x_{min}+x_{max}}{2} \\ y_{max} \\ 1 \end{bmatrix}, \text{ where } p_3 = 1 \text{ expresses the 2D point in homogeneous coordinates.} \\
\lambda \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}}_{H} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \text{ where } X = \frac{h_{11}u+h_{12}v+h_{13}}{h_{31}u+h_{32}v+h_{33}},\ Y = \frac{h_{21}u+h_{22}v+h_{23}}{h_{31}u+h_{32}v+h_{33}}. \\
\lambda = h_{31}u + h_{32}v + h_{33} \text{ is the scale factor, and } \lambda > 0 \text{ is the validity condition.} \\
\boldsymbol{h} = [h_{11}, h_{12}, h_{13}, \ldots, h_{33}]^T \text{ and } \hat{\mathbf{h}} = \underset{\|\mathbf{h}\|=1}{\arg\min} \|A\mathbf{h}\|, \text{ where :} \\
A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} \text{ and } A_i = \begin{bmatrix} -u_i & -v_i & -1 & 0 & 0 & 0 & u_iX_i & v_iX_i & X_i \\ 0 & 0 & 0 & -u_i & -v_i & -1 & u_iY_i & v_iY_i & Y_i \end{bmatrix}
\end{gathered} \tag{1}$$

For roadway control points, we made use of crosswalk polygons (generally demarcated by sidewalk ramps on either side of the road or directly marked as zebra crossings or vehicle stop lines). User-annotators directly drew these polygons in the annotation phase (thereby framing them into the image space). For the polygon corners, the pipeline establishes ground-plane coordinates from side and diagonal lengths from Google Street View (we discuss this further in Section 5.1). To transform BoT-SORT trajectories into the ground plane (as $\mathcal{T}_j$), we used the homography described in Equation (1), which the pipeline applies successively to each tracked pedestrian ID in the scene. For each MOT-tracked pedestrian with track ID j, and for each frame $k \in \mathcal{K}_j$ of a video crossing signal cycle, Road2Synth generates the foot contact point (u,v) from j's bounding box and writes it in homogeneous form, as $\mathbf{p}_j(k)$ (Equation (2)). Vehicle (j') trajectories are formed in the same manner, but using the ground point below the middle of the front bumper, as (u,v). Multiplying by the homography H then yields the ground-plane position in homogeneous form, as $H\,\mathbf{p}_j(k)$. Below, the perspective division $\pi(\cdot)$ maps homogeneous coordinates back to Cartesian coordinates by dividing by the scale factor λ to produce ground-plane coordinates (X,Y). The trajectory $\mathcal{T}_j$ is then generated as a time-ordered sequence of ground-plane positions, $\mathbf{P}_j(k)$, mapped in order of the time of frame k, i.e., t_k, where $k = 0$ is the first frame, $t_k = \frac{k}{\text{fps}}$, and $\mathcal{K}_j$ is the set of frames in which pedestrian j is detected.

$$\begin{gathered}
\mathbf{p}_j(k) = \left[u_j(k), v_j(k), 1\right]^T \\
H\,\mathbf{p}_j(k) = \lambda[X,\ Y,\ 1]^T = [\lambda X,\ \lambda Y,\ \lambda]^T \\
\pi\left([\lambda X,\ \lambda Y,\ \lambda]^T\right) = \left[\tfrac{\lambda X}{\lambda},\ \tfrac{\lambda Y}{\lambda}\right]^T = [X,\ Y]^T \\
\mathcal{T}_j = \left\{\left(t_k,\ \mathbf{P}_j(k)\right) : k \in \mathcal{K}_j\right\}
\end{gathered} \tag{2}$$

The projected coordinate system was standardized by rotating and flipping each instance so that right-turning vehicles appeared in the bottom-left quadrant and pedestrians crossed along the vertical y-axis. We used this as a standardization to make trajectories comparable across intersections and to assist in retaining relative placement of surrounding environmental features, chiefly buildings, roads, sidewalks, and street objects such as poles and street furniture (We kept the original coordinates for GIS and georeferenced reconstructions.)

A range of confounding factors may contribute jitter to the trajectories, including bounding-box warping as pedestrians shift in relative focus in the camera view, occlusions as people and things shift in the frame, and some frame-level tracking noise that persists after Kalman filtering [174]. To control for this, we used Savitzky–Golay smoothing [179] (via *SciPy* [180]) on trajectory jitter, which reduces frame-by-frame jitter while preserving the shape of trajectories (although it does not correct large tracking jumps and there is some straightening at fine resolutions below 4 m, as we discuss in Section 6). This jitter-smoothing is particularly important for translation to our XR components of the Road2Synth pipeline, because trajectories are a base for animating avatar-representations of pedestrians and vehicles in synthetic form.

In our fieldwork, we visited 24 sites, and 21 of those sites returned video that we used in analysis. In total, Road2Synth built 4403 trajectories (3668 from pedestrians and 735 from vehicles) from 215 signal-cycle epochs in our collection of roadside video, with 1592 pedestrian–vehicle interaction pairs. We remind the reader that these are additionally annotated (see Section 4.4) so that they contain information about when pedestrians began to cross the road, the qualitative locations of the streetscape in which crossing took place, and what the context of the ambient crossing surrounding environment was at the time of crossing. In addition, information of the video site is available per trajectory $\mathcal{J}_j$ and per frame t_k as in Table 2, and the automated labels from deep learning are available. In this way, Road2Synth adds to the value of the original video by producing a wide swath of compound data, all of which can be translated into GIS and VGE (and XR).

4.6. Reconstruction

The XR endpoint for Road2Synth is a VGE, which we created in *Unity* as a prototype. Our goal in doing so was to reconstruct a synthetic instance of the surveyed roadside scene in an immersive, explorable virtual environment that could be used to examine, study, and hypothesize about safety issues in ways that are not otherwise accessible in the real world. Our key innovation in this step was to establish a scientific laboratory and sandbox for safety science, with direct connection to fieldwork data and annotations registered to the GIS. In particular, we aimed to use the virtual camera functions of *Unity* to generate synthetic perspectives of road users by essentially projecting individualized vistas through their "synthetic eyes." In this way, we can directly recover/reconstruct the egocentric view that we described in Section 3 (both for the driver perspectives of Section 3.1 and the pedestrian perspectives of Section 3.2).

To move through the pipeline from the trajectory-homography space to the XR endpoint, for a test-case intersection scene, we used *Luma* [181] to produce quick-return photogrammetric models from a walk-around video. Our use of *Luma* is designed to test solutions that can be launched on a cellphone application (reconstruction happens off-device, in the cloud). In this way, it provides a good test of ease of integration with Road2Synth's cellphone video entry point to the pipeline, although *Luma* could be replaced with other toolkits.

The exact techniques used in *Luma* are proprietary, but we understand from documentation [182,183] that it relies initially on structure from motion (SfM) [184] and three-dimensional Gaussian splatting [185] to quickly transfer the static elements of video

scenes to a pseudo-volumetric geographic space (Figures 8 and 9). For the site with the most unsafe PET in our study, we recorded a walk-around video (from what we refer to as the "capture camera" to distinguish its video from roadside fixed footage) by traversing the four-sided perimeter of a given intersection during moments of traffic and pedestrian calm. The end product, shown in Figure 9, is a mesh that captures the fixed elements of the test site, as a streetscape (roads, sidewalks, curbs, ramps) and the objects that occupy it (vendor stands, traffic lights, lampposts, fire hydrants, crossing signals, benches, mailboxes, bollards, barriers, crossing islands), as well as its visual appearance (including road signs, lane markings, zebra crossings). We can then render dynamic elements from Road2Synth into this mesh as coming directly from the trajectories discussed in Section 4.5. In this way, both the fixed and dynamic elements of the real-world scene from video are transposed into XR, and in Figures 9 and 10 we show them rendered directly in *Unity* (see Section 4.7).

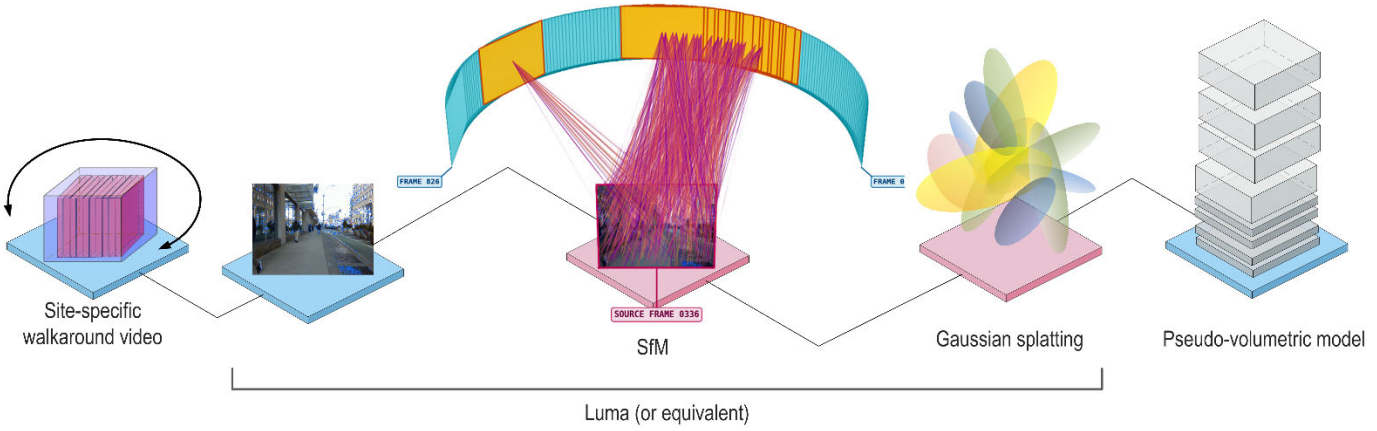


Figure 8. The reconstruction pipeline for building fixed-scene models from walk-around input video.

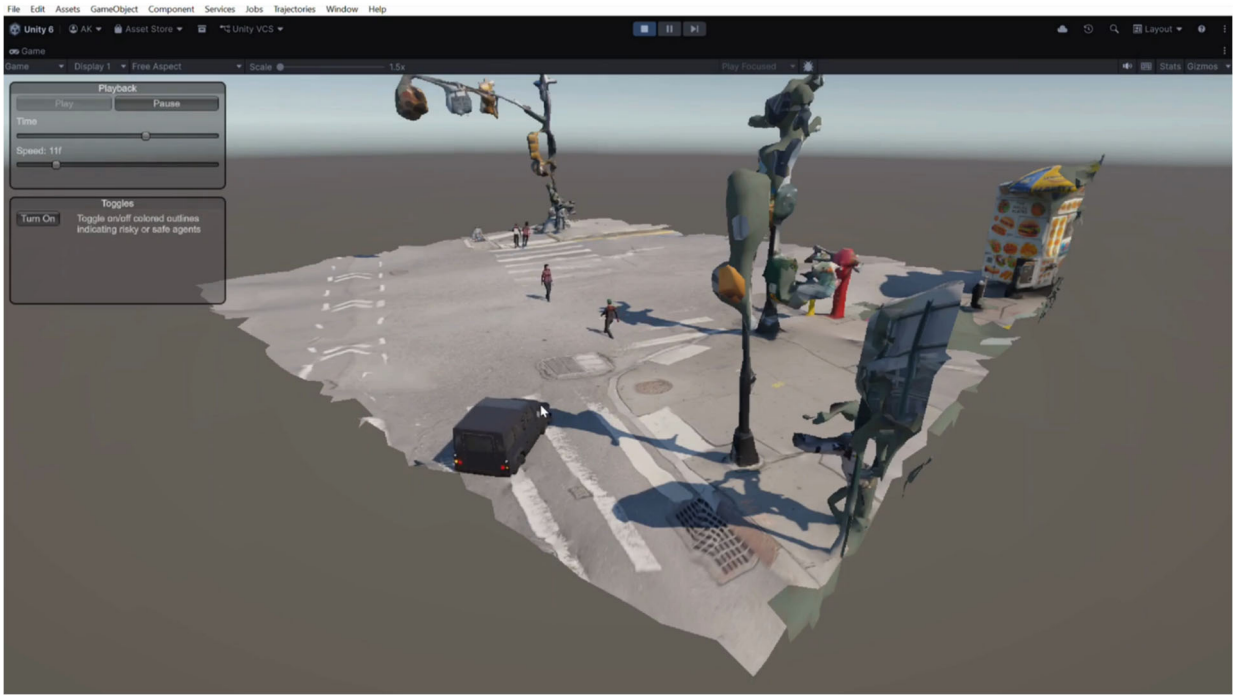


Figure 9. The resulting 3D fixed infrastructure mesh (roads, sidewalks, street furniture, lampposts, traffic and crossing lights, road markings) produced by 3D Gaussian splatting (*Luma*), rendered in *Unity* from a (non-visited) exocentric viewpoint. Dynamic elements (pedestrians and vehicles) are introduced dynamically from their trajectories and rendered into the mesh.

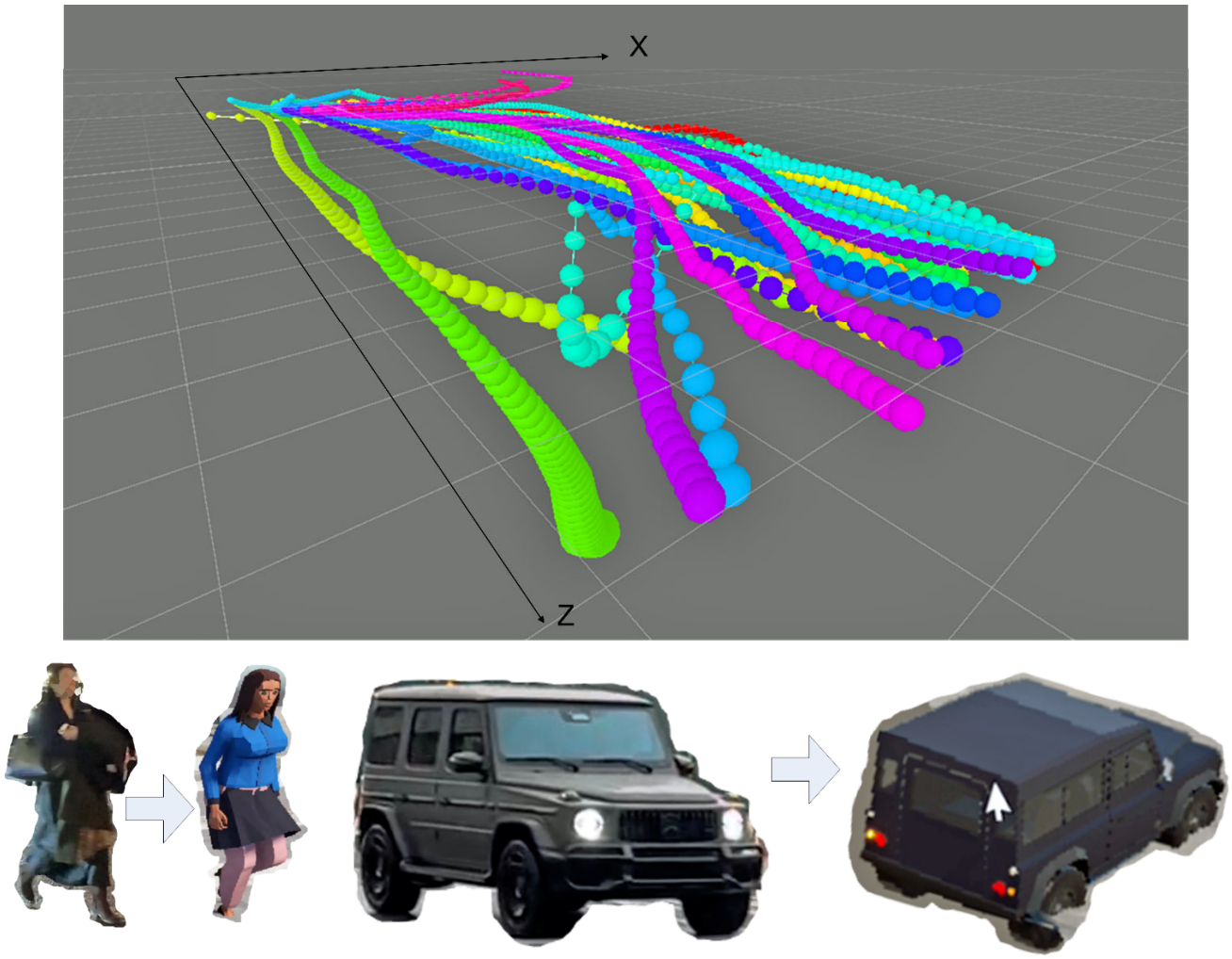


Figure 10. Trajectories from 2D monocular video are transformed by homography and read-in directly to Unity space. Avatars (and mocap samples) are mapped as stand-ins for real-world pedestrians and vehicles.

In what follows, we briefly describe the general procedure for accomplishing this using the elements of Road2Synth to show the reader that the pipeline is extensible to reconstruction. Other components could be used in place of *Luma*, but our point is that such packages can be used for convenience, yielding in this case a lightweight capture task (a single walk-around video) that can run from a user's cellphone (in this case a phone app interface to cloud reconstruction). In our prototype, *Luma* is used to reconstruct the statics of the site, while Road2Synth works with the data it generates to "enliven" it with dynamics captured from real events in next XR steps downstream in the pipeline. Note also that Road2Synth does so while preserving the space-time GIS metric properties of captured events and positions (see Section 4.8) such that the reconstruction by XR helps to convert the data into VGE.

4.6.1. Structure from Motion

We understand *Luma* to rely initially upon structure from motion (SfM) [184] to produce camera poses (i.e., where the camera is and where it is pointing) and three-dimensional points from input video. Below, we describe SfM in generic terms so that its ability to work with the data generated in earlier stages of the Road2Synth pipeline are clear. However, note that different schemes to take up these data and pass them to reconstruction techniques could be used instead. Road2Synth, as a pipeline, is agnostic in this regard.

Generally stated, SfM is used to recover the capture camera intrinsic matrix K (which we note in our fieldwork as being the same per frame, because a single camera with the same settings recorded each video), the camera rotation (R_q), the camera position ($\mathbf{C}_q$) in each capture frame q, and the three-dimensional positions ($\mathbf{M}_m$) of the scene points. We used the same cellphone camera (and same settings) for fixed tripod recordings and then again as a handheld for walk-around video. (For ease of notation, we will denote the intrinsic matrix of the fixed-perspective roadside camera and the walk-around camera as $K_{fixed} = K$ hereafter.) For each pair of considered capture frames q and q', then SfM estimates the fundamental matrix $F_{qq'}$ from matched keypoints; this and the intrinsic matrix yield the essential matrix $E_{qq'}$ (by Equation (3)):

$$E_{qq'} = K^T F_{qq'} K \tag{3}$$

SfM provides a convenient way to recover relative rotation (and direction) of the translation between frames q and q', from $E_{qq'}$. Matched points can then be triangulated: by adding frames successively and then estimating each subsequent frame's pose (from matches to points that were already reconstructed) and then triangulating for newly appearing points. By integrating with *Luma*, then, Road2Synth can read data from observed (Cartesian in the image) pixel-space coordinates $u_m(q)$ and $v_m(q)$ of scene point m in capture frame q, and the set $\mathcal{Q}_m$ of capture frames in which scene point m is observed. (Bundle adjustment in SfM can provide a solution for K,R_q,$\mathbf{C}_q$,$\mathbf{M}_m$ by minimizing the reprojection error, as in Equation (4) (where $\pi(\cdot)$ is the perspective division already discussed.)

$$\min_{K,\, R_q,\, \mathbf{C}_q,\, \mathbf{M}_m} \sum_m \sum_{q \in \mathcal{Q}_m} \left\| \begin{bmatrix} u_m(q) \\ v_m(q) \end{bmatrix} - \pi\left(K R_q \left(\mathbf{M}_m - \mathbf{C}_q \right) \right) \right\|^2 \tag{4}$$

The fixed-vantage camera and walk-around camera are the same, so the intrinsic matrix in the homography of Equation (1) carries into Equations (3) and (4). The end result is that Road2Synth can read a video, then use *Luma* to return a camera pose (R_q,$\mathbf{C}_q$) for every frame q, as well as a sparse point cloud of scene points $\mathbf{M}_m$. These can be used to build XR reconstructions from the original data, and we show how this can be done with Gaussian splatting.

4.6.2. Three-Dimensional Gaussian Splatting

The points $\mathbf{M}_m$ are sparse (and not suitable for XR), but come from real data. Different computational techniques can be used to generate reconstructions from them. For our prototype, we used *Luma*'s three-dimensional Gaussian splatting [181] to run dense reconstruction, which we ported into *Unity* to be realized as XR. We discuss the general technique for Gaussian splatting as follows, again as an indicator of how Road2Synth can integrate with XR-adjacent methods. (Other techniques could be used instead, e.g., neural radiance fields, NeRFs [186]).

Taking the sparse points $\mathbf{M}_m$ as a starting point in the pipeline, a reconstructed scene can be represented with a (massive) set of three-dimensional Gaussians, becoming much more dense that the initial set. In brief, each Gaussian g is interpreted as having a mean position of $\boldsymbol{\mu}_g$, an ellipsoidal shape (with covariance Σ_g coming from scaling and rotation), an opacity o_g, and view-dependent coloring $\mathbf{c}_g$ that is derived from spherical harmonics. Using K and the camera poses (R_q,$\mathbf{C}_q$) recovered via SfM, the splatting then renders the Gaussians into each frame q. Importantly, the center of g projects to the Cartesian pixel-space coordinates and holds onto the same camera characteristics and to the perspective division from Equations (2)–(4), with each Gaussian then informing the rendering of the pixels that it covers. For an arbitrary pixel position in a capture frame (which we denote as $\boldsymbol{\xi}$), g contributes its opacity, weighted by its projected 2D Gaussian, as $\alpha_g(\boldsymbol{\xi})$, i.e., how

much Gaussian g covers a given pixel $\boldsymbol{\xi}$ (Equation 5). The Gaussians that cover pixel $\boldsymbol{\xi}$, numbered $g = 1, \ldots, N_G$ from front to back in distance from the camera (with $g = N_G$ being the one farthest from the camera) are then blended in Gaussian splatting to take on the rendered pixel color at frame q, i.e., $\hat{\mathbf{I}}_q(\boldsymbol{\xi})$. The Gaussians are then optimized by nudging them until each $\hat{\mathbf{I}}_q(\boldsymbol{\xi})$ matches $\mathbf{I}_q(\boldsymbol{\xi})$ as closely as possible (Equation (5)). We note a couple of salient points regarding this reconstruction stage of Road2Synth. First, the cellphone videos that are "fed" into the pipeline contain not only substantive data of the road users in pedestrian–vehicle interactions, but additionally contain image data (from the walk-around) that can be used (via SfM) to generate computational models of the *actual site* from which the video is recorded. In much the same way that we use frame-by-frame associations to build trajectories, we can iterate across video frames and use the resulting sparse points to construct dense representations of the fixed scene using techniques such as Gaussian splatting. Indeed, Gaussian splatting can also draw from the coloring of the scene to render reconstructions that match the moments of site capture, including illumination and texture. We note in particular that these reconstructions can be re-rendered for viewpoints that have never actually been visited by the camera or the camera-holder, and this is useful when we seek to port the reconstruction to *Unity* as an explorative VGE. Further, rendering by Gaussian splatting can be achieved in real time on modest graphics hardware because the sorting and blending of projected Gaussians maps well to GPUs [185]. In particular, the reconstruction can be rendered from a chosen viewpoint, retained as Gaussians for real-time display, or converted into surface meshes.

$$\alpha_g(\boldsymbol{\xi}) = o_g \exp\left(-\frac{1}{2}\left(\boldsymbol{\xi} - \begin{bmatrix} u_g \\ v_g \end{bmatrix}\right)^T \left(\Sigma_g^{2D}\right)^{-1}\left(\boldsymbol{\xi} - \begin{bmatrix} u_g \\ v_g \end{bmatrix}\right)\right) \tag{5}$$

where Σ_g projects to the two-by-two image covariance Σ_g^{2D}, $\begin{bmatrix} u_g \\ v_g \end{bmatrix} = \pi\left(KR_q\left(\boldsymbol{\mu}_g - \mathbf{C}_q\right)\right)$, and $\Sigma_g^{2D} = J_g R_q \Sigma_g R_q^T J_g^T$, with J_g being the 2 × 3 Jacobian of the projection (including K), evaluated at the center of g.

To test the concept, we did this for one field site. Fixed-infrastructure scenes were generated to match the counterpart civic infrastructure and the geography of the roadside scene, then georeferenced to the GIS at control points (polled from assisted GPS to reduce urban canyon artefacts) to produce 1:1 scale to reality. Gaussian splatting makes the reconstruction fast, but with the trade-off that it produces some visual artifacts (Figure 9). Our cursory approach could be replaced with more finely tuned reconstruction, but we wished to build easy-to-use geomatic modeling capabilities. Elsewhere, we show how VGEs can be built from video with high degrees of geomatic and geometric match to the real world [70], including for augmented reality platforms [74], but the process is not quick and not easily embedded in user-facing dashboard platforms, as used in the pipeline that we are demonstrating here. Embedding fast fixed-detail transfer is a topic for future research and development, and new techniques such as *BlendFusion* could be useful in this regard [187]. Similarly, other XR techniques could be used, particularly AR [74], which is a promising area of future inquiry for Road2Synth, particularly in its applied implications for situational awareness.

Note that all of the data from the earlier portions of the Road2Synth pipeline are retained in the reconstruction step. Trajectories keep their space-time positions from deep learning, and annotations from ethnography carry forward. *Luma* adds value from the original imagery, as a 3D reconstruction and via raw image color and texture from the naturalistic observation of the scene. Importantly, these aspects of the fieldwork can now be viewed from new vantages because of the reconstruction, becoming accessible as egocentric views. And, GIS metrics can be built and retained. Crosswalk corners (user-verified)

appear in both of the fixed and walk-around vantages. Using these (and the distances from annotation), we can also cross-register data from each stage of the pipeline to GIS (Section 4.8).

4.7. Trajectory-Avatar Animation

For XR use, each trajectory $\mathcal{J}_j$ can be ported directly into a *Unity* framework $\left(\mathbf{P}_j^{\text{Unity}}(k)\right)$, subbing (X,Y) into *Unity*'s world coordinates (x and z) (Equation 6). The Gaussian-splatted model from Section 4.6.2 can be converted into a mesh, and the trajectories can then be aligned to that mesh at the points of foot contact and vehicle contact with its surfaces (we set them to the Unity plane height for the road in Equation (6), which pertains to both pedestrians j and vehicles j' and so we show a base condition of 0, awaiting the insertion of avatars at given frames where they appear). Further, because we have a timestamp for each position in the GIS, we can also replay movement of all entities directly in *Unity* (using spatial data structures to load and unload them from the scene graph efficiently).

We used stock avatars for pedestrians and vehicles to animate road users in the resulting *Unity* VGE (Figure 10). Because we had general vehicle types from the annotation round, we matched vehicle avatars to YOLO car, bus, truck, and user-labeled categories (for pedestrians, we used stand-ins). In future work, we hope to explore direct 1:1 mappings to vehicle make and model databases using *VMMRdb* [188] to generate matching vehicle avatars, and to explore using pedestrian attribute recognition to generate commensurate pedestrian avatars [189]. For the prototype that we present here, avatars were animated along scene trajectories using traditional keyframe animation using *Unity*'s built-in motion capture libraries and motion retargeting [190]. Elsewhere, we have shown that the same architecture can support bespoke motion capture of human-participant users interacting with road scenes in VR [85,191].

$$\mathbf{P}_j^{\text{Unity}}(k) = (sX,\ 0,\ sY) \tag{6}$$

4.8. Shared Manifold as GIS

To support data exchange and data interpretation across the various lenses of examination that we have as road phenomena (video space, pixel space, bounding box space, trajectory space, scene space, virtual and geometric camera space, GIS, *Unity* space), we engaged a shared manifold via slipstreaming. Notably, this allows for handling of both geographic data as well as scene graph data [69]. Video-derived positions were projected into this manifold using homography and stored in a GIS. Homography supports mathematical transformations as mapping between planes while preserving straight lines. So, in performing these mappings, homography allows for geometric operations to be performed across diverse planes, as if they were in a shared geometric space. It is this shared space that we consider as a manifold, and homography allows us to project into that manifold (extensibly) from the roadside camera's vista, while SfM and matched control-point alignment bring in the reconstructed scene. Specifically, we can then project "out of" this manifold to VR and VGEs (as in Figure 10).

To converge space and time while preserving the manifolds, we treated trajectories as locally discretized space-time paths [192,193], as portions of continuous movements and locomotion in the real-world. Collectively, then, we can regard the combination of GIS, space-time paths, slipstreaming, and VR as a dynamic VGE (and we remind the reader that building dynamics into VGEs is an open topic of investigation in geographic information science). This VGE becomes interoperable across mathematics, spatial analysis, event queries, visual analytics, and computational geometry [35,48].

Road2Synth uses two sets of control points. To assist homography, we ask annotators to draw the crosswalk polygon in a given roadside video (we will detail this further in

Section 5.1). This supplies image coordinates (u_i,v_i) of the control points, while their ground coordinates (X_i,Y_i) come from the crosswalk side and diagonal length (measured in the annotation tool by pulling from Google Street View). Homography then maps each foot point (or vehicle contact point) $\mathbf{p}_j(k)$ (or $\mathbf{p}_{j'}(k)$) into a metric ground-plane frame. For the VGE reconstruction, we paired the walk-around video to assisted GPS (with the cellphone camera held at each control point). This, then, allowed us to georeference the Gaussian-splatted scene in the GIS projected coordinate system. The two frames are then linked by matching the crosswalk corners, which appear in both, so that trajectories can be replayed in the reconstructed scene. For identity persistence, we recorded track identities in a timestamped (t_k) identity table. Each trajectory $\mathcal{T}_j$ was stored in the GIS as polyline features (with t_k attributes), with each pedestrian and vehicle trajectory track within each video carrying joined attributes of the BoT-SORT track ID and data from the fieldwork, deep learning, and annotation phases, with the per-frame YOLO detections held in the timestamped identity table.

In Table 3, we show a range of calculated distances that are possible from the GIS (We can also calculate timing, as in Table 4.) This is a significant benefit for Road2Synth, in that it directly allows for positioning from the "scene view" that we discussed in the paper opening. We will show later, for example, that TTC measures can be built (directly) from this database.

Table 3. Distances and counts for pedestrians, vehicles, pairings while crossing (by site; distances in meters).

	Pedestrian to Pedestrian			***Pedestrian to Vehicle***	***Vehicle to Vehicle***	
Site	*Total crosser frames with >1 crosser (each counted per frame at 30 fps)*	*In frames at left, median straight-line distance to closest crosser [IQR over the frames]*	*Pedestrian pairs (in crosswalk in at least one shared frame; summed over recordings)*	*Pedestrian–vehicle pairs (pedestrian in crosswalk, vehicle anywhere in at least one shared frame; summed over recordings)*	*Pairs of vehicles in same frame; summed over recordings [those closer than 1 m]*	*For pairs at left, median smallest gap between vehicle proxy shapes*
23rd Street and Jackson Avenue (North) **	33,415	2.28 [1.27–4.36]	448	8	0	N/A
Bowery and Canal Street (West) **	17,652	1.14 [0.77–2.62]	102	49	8 [0]	5.35
Canal Street and Bowery (East)	89,174	1.43 [0.97–2.35]	1810	275	15 [0]	4.47
Clinton Street and Delancey Street	16,190	0.92 [0.58–1.76]	362	66	13 [4]	2.56
Columbus Avenue and W 72nd Street (East)	15,699	1.05 [0.79–1.61]	319	30	3 [0]	5.05
East 149th Street and Grand Concourse	4763	0.74 [0.57–1.45]	53	20	2 [0]	2.86
Flatbush Avenue Extension and Myrtle Avenue (North)	9548	2.19 [1.20–3.64]	122	111	54 [16]	3.21
Fort Greene Pl and Lafayette Avenue	2225	1.39 [0.75–2.81]	15	6	1 [0]	5.25
Fulton Street and Lafayette Avenue	11,624	1.33 [0.79–2.72]	85	46	7 [2]	1.31
Jackson Avenue and 23rd Street (South) **	6515	2.86 [1.77–6.84]	44	33	9 [0]	10.39
Myrtle Avenue and Flatbush Avenue Extension (South)	672	3.48 [2.76–4.14]	4	25	21 [0]	10.38
Northern Blvd and 48th Street	35,046	1.54 [0.82–3.73]	280	124	32 [3]	5.60
Pacific Street and Court Street	2179	1.85 [1.06–2.85]	32	3	0	N/A

Table 3. *Cont.*

		Pedestrian to Pedestrian		*Pedestrian to Vehicle*	*Vehicle to Vehicle*	
Seventh Avenue and West 44th Street	127,227	0.86 [0.57–1.20]	5953	303	9 [2]	2.59
Sixth Avenue and West Third Street	16,979	1.09 [0.76–1.83]	129	38	15 [0]	5.60
Third Avenue and 37th Street	11,486	0.75 [0.13–1.13]	59	66	36 [8]	3.61
Tillary Street and Jay Street (East)	2320	2.16 [1.08–4.69]	20	10	2 [0]	8.34
West 42nd Street and Sixth Avenue *	131,220	0.66 [0.42–0.98]	9652	211	2 [0]	1.82
West 43rd Street and Seventh Avenue *	210,744	0.72 [0.46–1.05]	18,630	833	8 [4]	0.97
West 72nd Street and Columbus Avenue (West)	104,733	1.16 [0.75–2.12]	1713	126	6 [0]	6.17
West 96th Street and Amsterdam Avenue	92,469	1.35 [0.80–2.39]	1152	369	53 [9]	3.91
All 21 sites *	**941,880**	**0.95 [0.61–1.61]**	**40,984**	**2752**	**296 [48]**	**4.06**
Excluding three recordings with measurement problems (*) and 37 recordings with questioned distance calibration (**) (175 left after removal of 40 of 215 total recordings with trajectories)	**884,298**	**0.92 [0.59–1.51]**	**40,390**	**2662**	**279 [48]**	**3.96**

Table 4. Speeds and counts for pedestrians and vehicles while crossing (by site; speeds in meters per second).

	Crossing Pedestrians Inside the Crosswalk (When Moving; at 30 fps)		*Vehicles in the Crosswalk (When Moving; at 30 fps)*	
Site	*Pedestrians*	*Median per-pedestrian walking speed [IQR across all pedestrians in recordings at site]*	*Vehicles*	*Median per-vehicle travel speed [IQR across all vehicles in recordings at site]*
23rd Street and Jackson Avenue (North) **	144	1.95 [1.66–2.30]	3	3.85 [3.16–5.05]
Bowery and Canal Street (West) **	51	1.40 [1.30–1.55]	36	3.60 [3.09–4.72]
Canal Street and Bowery (East)	238	1.36 [1.16–1.62]	38	4.26 [3.22–5.63]
Clinton Street and Delancey Street	113	1.26 [1.04–1.45]	37	2.91 [2.49–3.64]
Columbus Avenue and W 72nd Street (East)	70	1.41 [1.27–1.60]	9	3.68 [2.54–3.87]
East 149th Street and Grand Concourse	51	1.36 [1.26–1.48]	21	3.49 [3.26–3.89]
Flatbush Avenue Extension and Myrtle Avenue (North)	79	1.58 [1.43–1.69]	54	3.46 [2.25–4.36]
Fort Greene Pl and Lafayette Avenue	46	1.46 [1.33–1.63]	9	3.78 [3.20–4.35]
Fulton Street and Lafayette Avenue	43	1.49 [1.28–1.79]	26	3.93 [2.31–4.69]
Jackson Avenue and 23rd Street (South) **	38	1.90 [1.44–2.35]	40	4.57 [3.20–6.56]
Myrtle Avenue and Flatbush Avenue Extension (South)	25	1.57 [1.40–1.80]	92	5.27 [4.58–6.04]
Northern Blvd and 48th Street	89	1.64 [1.41–1.93]	60	4.44 [3.43–5.42]
Pacific Street and Court Street	21	1.20 [1.06–1.40]	1	2.17
Seventh Avenue and West 44th Street	416	1.19 [1.05–1.38]	20	2.47 [1.47–3.49]
Sixth Avenue and West Third Street	61	1.40 [1.24–1.68]	47	4.37 [3.75–4.91]
Third Avenue and 37th Street	41	1.44 [1.26–1.55]	70	3.52 [2.98–3.93]
Tillary Street and Jay Street (East)	17	1.39 [1.27–1.53]	14	3.50 [3.24–4.17]
West 42nd Street and Sixth Avenue *	394	1.02 [0.86–1.23]	5	2.79 [1.79–2.90]

Table 4. *Cont.*

	Crossing Pedestrians Inside the Crosswalk (When Moving; at 30 fps)		*Vehicles in the Crosswalk (When Moving; at 30 fps)*	
West 43rd Street and Seventh Avenue *	762	1.00 [0.86–1.18]	16	2.37 [1.19–2.95]
West 72nd Street and Columbus Avenue (West)	300	1.37 [1.20–1.54]	27	3.80 [3.50–4.32]
West 96th Street and Amsterdam Avenue	256	1.36 [1.16–1.54]	95	3.15 [2.59–3.74]
All 21 sites *	**3255**	**1.25 [1.02–1.50]**	**720**	**3.76 [2.87–4.71]**
Excluding three recordings with measurement problems (*) and 37 recordings with questioned distance calibration (**) (175 left after removal of 40 of 215 total recordings with trajectories). Note that pedestrian and vehicle instances are different than their *trajectory/track* counts.	**3022**	**1.23 [1.00–1.45]**	**641**	**3.73 [2.83–4.66]**

Mapping our Road2Synth data into GIS has additional benefits. In Figure 11, we show, for example, where unsafe PETs in our analysis were clustered. From the GIS, we were able to compare these locations to historical crash pedestrian injury data for New York City. This shows that GIS could potentially be used for site selection of accident anomalies, and Road2Synth could be used to examine individual site factors, pedestrian behavior, driver behavior, and to model potential interventions.

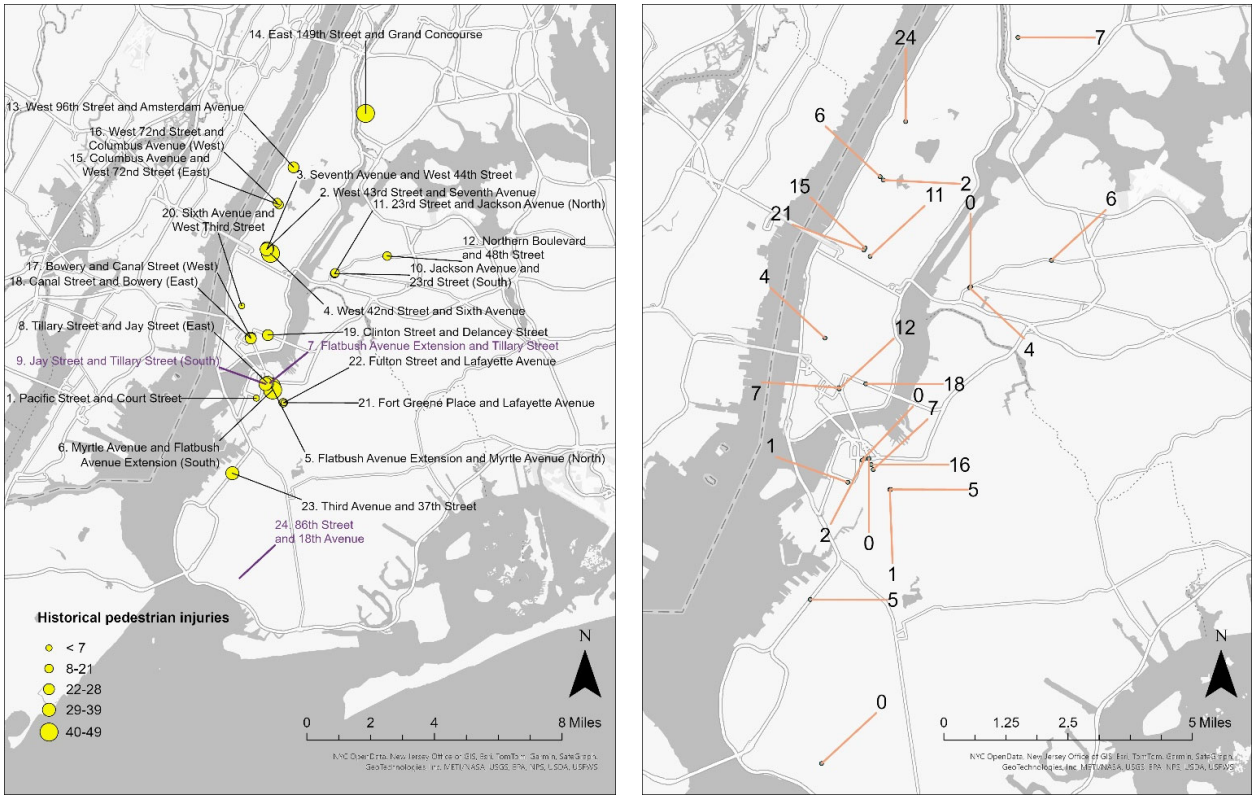


Figure 11. (**Left**) Historical pedestrian-injury data within 50 m of the intersection center point (reported with coordinates in police data from New York City Open Data Motor Vehicle Collision database) covering July 2012 to 23 October 2026 (441 pedestrian injuries in 435 crashes with pedestrian injuries or fatalities; 9 pedestrian fatalities not shown). (**Right**) Unsafe PETs from Road2Synth (values are detected unsafe PETs in pedestrian–vehicle pairs).

4.9. Pedestrian and Driver Perspective Reconstruction

Recall that the initial fieldwork video is collected from a fixed vantage at the roadside, and the dynamics of pedestrian–vehicle interactions unfold within its view

(Figures 1 and 12). At its essence, this is an exocentric (third-person) perspective. However, our manifold approach allows us to reconsider perspectives within the spaces captured by the walk-around video and reconstructed with Gaussian splatting, populated by the trajectories observed from the initial exocentric vantage. We mention again that the Gaussian splatting approach allows us to render the reconstructed scene from viewpoints that the capture camera did not visit. In other words, we have considerable flexibility to vicariously reconsider the "internal micro-geography" that sits within the exocentric viewshed.



Figure 12. Road2Synth transfers from real pedestrian and driver fields of view to synthetic Unity views from exocentric to egocentric perspectives not originally present in the input video.

Taking advantage of the manifold, we used virtual cameras to reconstruct individualized pedestrian and driver perspectives for each entity in the VR-VGE (Figure 12). We regard this as an extensible scientific contribution for the pipeline, as it establishes new modalities for reconstructing and studying human perspectives on pedestrian–vehicle interactions vicariously, as they may have unfolded in reality. Although not part of this proof of concept, at the same time a range of hypothetical parameters and factors could be brought into the analysis, such as site factors of signal visibility, dynamic factors of occlusion and distraction, and even weather and lighting differences if we apply them in *Unity* as shaders or as downstream simulations, or render them in real time from Gaussians.

We introduced virtual cameras into the manifold space, with each virtual camera maintaining an intrinsic matrix (which we term "K_{view}" here) with rotation $R_j(k)$ and position $\mathbf{C}_j(k)$ for track j at frame k (Equation 7). Recall that we have ground-plane positions in the manifold as $\mathbf{P}_j(k) = \left[X_j(k),\ Y_j(k)\right]^T$, with track j at frame k, and Z height for each avatar above the ground plane (which can also be a mesh for local height). From this foundation, we placed virtual cameras at eye height above the foot contact point for pedestrian avatars and at an offset from the vehicle's contact point (the midpoint of the length of the front bumper of the vehicle) for virtual drivers, both with gaze rays in the direction of travel from their individual trajectories $\widetilde{J_j}$.

Using the virtual cameras, we can render either the reconstruction from Gaussian splatting fused with *Unity* avatars or we can render the scene from a pedestrian- or driver-perspective *Unity* scene graph at any chosen point from the VGE–GIS mesh (or timeline). Our intent, then, is to have all of the different views of pedestrian–vehicle interactions

(scene, allocentric, egocentric) at parity in the same reconstruction. From the perspective of the pedestrian avatar or driver avatar, a scene point **M** (for points in front of the viewer) then appears in the view of entity j at the Cartesian pixel-space coordinates $\boldsymbol{\xi}_j(k)$, as in Equation (7) (which now shows the projection from Equation (4), but also includes the pose for each pedestrian and vehicle, and K_{view}).

$$\boldsymbol{\xi}_j(k) = \pi\Big(K_{view}\, R_j(k)\big(\mathbf{M} - \mathbf{C}_j(k)\big)\Big) \tag{7}$$

5. Experiments

To demonstrate the scientific use case for the pipeline, we built PET metrics for each of the 1430 pedestrian–vehicle interaction pairs from our study across 21 different sites (17 intersections). We imply in what follows that these empirical results could be further supported downstream with egocentric view analysis in the VR-VGE (as shown in Section 4). For parsimony, here we describe scientific analysis of PET to demonstrate that empirical work from the real world can be docked to VR-VGE, which could carry that analysis forward with what-if scenarios in XR reconstructions (moving from one part of the pipeline to the next).

5.1. Consistency Checks

Road2Synth is designed to be lightweight, as a supplement to survey-grade research. To check the pipeline's consistency, internally, we ran three checks. First, because the final phases of the pipeline rely on trajectories, which we smoothed to reduce frame jitter, we evaluated how smoothing moved positions. Across all pedestrian tracks ($n = 3448$), smoothing had a median of 3.6 cm shift in position, and for vehicles ($n = 742$) it was 14 cm. There was some variation based on season (1.9 cm for pedestrians in the fall, 5.7 cm in the winter; 9.2 cm for vehicles in the fall, 17.4 cm in the winter); precipitation (2.5 cm for pedestrians in dry conditions, 5.5 cm in rain or snow; 13.5 cm for vehicles in dry conditions, 15.5 cm in rain or snow); and time of day (5.4 cm for pedestrians in the morning, 5.1 cm in afternoon, 2 cm in evening and night; 9.8 cm for vehicles in the morning, 17.5 cm in the afternoon, 12.3 cm in the evening).

Second, we evaluated the consistency of measurements in our on-screen annotation pipeline. We note that pipeline distances will "inherit" the calibration from each recording, which rests on crosswalk lengths measured in Google Street View. Given this calibration, we evaluated as follows. For each crosswalk polygon (drawn by a user-annotator), we measured six lengths of each crosswalk (sides and diagonals). Our annotation tool uses the sides and the first diagonal to place the polygon corners, so the second diagonal can be used as an independent check. Across 17 of the 21 sites, the second diagonal differed from the value implied by the other five polygon lengths by a median of 22 cm (range was 0.1 to 46 cm, with standard deviation of 18 cm). This check is not an error measurement, but it shows that the measurements in the pipeline are internally consistent to about 1%. We report 17 out of 21 because we excluded four sites that had entry or measurement errors: a mistyped 222 m diagonal, six lengths that cannot form a single quadrilateral, a stored polygon that does not match its own measurements, and a second diagonal that was 85 cm shorter than the other five lengths imply. We note that the crosswalk-measurement errors at four sites do not change our results downstream in the pipeline analysis. For two sites, the error is in the second diagonal, which the annotation placement does not use. For the other two sites, we checked the walking speeds for plausibility and they agreed to within 0.06 m/s with those in the camera recording. We additionally note that these errors would stretch distances and speeds by the same factor and not impact TTC or PET calculations. We mention the details of these errors to also highlight how Road2Synth can generate viable

downstream results, even when there are user annotation mistakes. This is important if we consider Road2Synth as a tool that casual users might access.

Third, we evaluated the consistency of sinuosity of the pedestrian trajectories generated by the pipeline, and we compared them to a GPS archive of sinuosity metrics for road-crossing pedestrians that is reported in [194]. Importantly, this archive includes road-crossing data for New York city and its surrounding boroughs (20 total trajectories, including two suburban crossings, two mid-block crossings, and two one-way crossings). We used the same sinuosity metrics as that archive (the VFractal mean fractal dimension developed by [195]). Because the archive data were collected at a different scale, we averaged fractal dimension over scales of 0.5 m to 4 m. Across our sites, the median mean fractal dimension was 1.0066 for our 2336 (raw, not smoothed) pedestrian tracks compared to 1.0071 for the archive tracks. The equivalence check implies that the pipeline's raw pedestrian trajectories have the sinuosity and geometry shape of real road-crossing paths (the difference is −0.0004; with a 90% CI). We did not perform a comparison of vehicle trajectory sinuosity because we did not have a published comparison set.

5.2. Pedestrian–Vehicle Interaction Thresholding as Time to Collision and as Post Encroachment Time

We calculated TTC and PET for all smoothed trajectories in the dataset between each pedestrian and each vehicle that appeared in the same scene. Again, TTC is prospective, as a measure of how soon both would meet if they did not alter course from the measurement location and time, whereas PET is retrospective, as a measure of how far apart in time a pedestrian and a vehicle passed through a defined area (which we interpret as a conflict zone).

For TTC, we evaluated smoothed trajectories ten times per second, while a given pedestrian was positioned within one meter of a crosswalk. The pedestrian and vehicle velocities were then estimated from positions at one third of a second before and after each of these moments in space and time. We represented each pedestrian as a disc (0.3 m in radius) and we established proxy areas for vehicles based on their YOLO and annotator classification type (for bus, we used 31.72 m^2; 11.4 m^2 for pickup truck; 8.88 m^2 for sedan; 9.56 m^2 for SUV; 18.75 m^2 for truck; and 10.4 m^2 for van), with a rectangle extended backward from their front bumper along the vehicle trajectory heading. At each moment, then, TTC is the time it would take for the pedestrian disc to make first contact with the vehicle rectangle if nothing changes in their trajectories or speed (with a ten-second look-ahead time). For pedestrians and vehicles, we considered a straight-line trajectory. Then, for vehicles we additionally considered a proxy turning-sensitive arc (as our applications are on right-turning vehicles). Across all pedestrian–vehicle dyads, across all 21 sites, 19% were calculated to be on a collision course (median minimum TTC of 4.7 s) at some moment in the assessment under straight-line projection, and 26% were found under the turning arc projection (median minimum TTC of 4.8 s).

For PET, we evaluated smoothed trajectories at their full 30 frames per second. For each pedestrian that initiated a crossing near a right-turn lane, we paired their trajectory with the trajectory of each vehicle that was detected in the video at the same time. This produced 1632 pairs (discounting two mis-scaled recordings). We considered a conflict point as the midpoint of the pedestrian and vehicle's closest approach (calculated for every pedestrian–vehicle position in the GIS, regardless of timing). From this point, we then built a conflict area as a disc of two meters in radius. This in essence serves to safety-buffer the space between pedestrians and vehicles as a proxemic area [196–198].

We consider two meters as being an area that could lead to a problematic interaction. PET is then the time difference between a given pedestrian's last frame timestamp inside the

conflict disc and the vehicle's first frame timestamp inside the same disc. In other words, PET is the time that separates the pedestrian's exit from the conflict area from the vehicle's entry to the conflict area. Because of this, we obtain a PET only when both a pedestrian and a vehicle enter the conflict area, which happens when their actual traversed paths (not their future paths as would be the TTC case) come within four meters of one another. Moreover, we record PET only when the pedestrian leaves before the vehicle arrives in the conflict area. Short PETs leave little room for error in pedestrian–driver interaction. This serves as a surrogate for a compendium of factors, including the potential steering time [199], collision avoidance time [164,200–202], and the potential notification time available to both parties to correct. Of the pairs in the analysis, 1430 satisfied these conditions and received a PET. We set a PET threshold of 3 s, which caught 12.2% of cases as falling below it. This threshold is specific to the two-meter conflict area: a larger area would shorten PETs and find more interactions to be unsafe (For example, 54% of the 1430 pairs had a PET above 10 s.) To examine sensitivity, we also rescaled distance by $+/-$ 20%, which we found changed PET by at most a median of 0.5 s.

PET is stated procedurally in Algorithm 1 for a single pedestrian–vehicle dyad (j,j'), where j is the pedestrian and j' is the vehicle, as before. The vehicle's trajectory $\mathcal{T}_{j'}$ is formed as in Equation (2), with the ground point below the middle of its front bumper as (u,v). In Algorithm 1, $k \in \mathcal{K}_j$ denotes a frame of the pedestrian's track in the test (while $k' \in \mathcal{K}_{j'}$ denotes this for the pedestrian's) and k^* is the optimal value of k as the minimized straight-line distance between two positions. $\overline{P}_{jj'}$ is the straight-line midpoint between the two closest positions of j and j'. $\mathcal{W}_j$ is a subset of the positions in $\mathcal{K}_j$ that satisfy a test of being within distance r of $\overline{P}_{jj'}$. Line 1 in Algorithm 1 addresses the task of finding the closest approach of the two paths. Line 2 discards dyads whose paths do not come within two radiuses (four meters). Lines 3–4 construct the conflict area and then find the frames in which the pedestrian and vehicle are inside. Lines 5–6 calculate PET and retain it only in cases where the pedestrian leaves first. Algorithm 1 also resolves two choices that the PET description leaves open: that the conflict point is found from the two whole paths, regardless of timing, and that a position that sits exactly on the conflict area boundary is counted as being "inside" the conflict zone for PET.

Algorithm 1. PET calculation for pedestrian–vehicle dyad (j, j')

Inputs: $\mathcal{T}_j$; $\mathcal{T}_{j'}$ (Equation (2)); fps; radius r of conflict zone
Output: $\mathrm{PET}_{jj'}$ in seconds, **OR** undefined

1 $(k^*, k'^*) \leftarrow \arg \min_{\substack{k \in \mathcal{K}_j \\ k' \in \mathcal{K}_{j'}}} \left\| P_j(k) - P_{j'}(k') \right\|$
2 **IF** $\left\| P_j(k^*) - P_{j'}(k'^*) \right\| > 2r$ **THEN RETURN** undefined
▷ *the paths never shared the conflict area*
3 $\overline{P}_{jj'} \leftarrow \frac{1}{2}\left(P_j(k^*) + P_{j'}(k'^*)\right)$
4 $\mathcal{W}_j \leftarrow \left\{k \in \mathcal{K}_j : \left\| P_j(k) - \overline{P}_{jj'} \right\| \le r\right\}$; $\mathcal{W}_{j'} \leftarrow \left\{k' \in \mathcal{K}_{j'} : \left\| P_{j'}(k') - \overline{P}_{jj'} \right\| \le r\right\}$
5 $PET_{jj'} \leftarrow \left(\frac{\min \mathcal{W}_{j'} - \max \mathcal{W}_j}{\mathrm{fps}}\right)$
6 **IF** $PET_{jj'} > 0$ **THEN RETURN** $PET_{jj'}$ **ELSE RETURN** undefined
▷ *the vehicle entered the conflict area before the pedestrian left it; the vehicle went first or both vehicle and pedestrian were inside the conflict area at the same time*

5.3. Statistical Modeling

We built a statistical model (Equation (8)) from the study data, as a logistic regression, to estimate the odds that an interaction between a pedestrian and a vehicle was unsafe, defined as a PET below three seconds (as in Algorithm 1). Specifically, the model estimates how factors of the scene change the odds that a vehicle and a pedestrian will come into

potential conflict within a three-second bubble of time in the crosswalk area of the sites that we visited. These conflict time geographies act, then, as a surrogate for risk.

Results are shown in Table 5, where we state the estimated odds ratios per coefficient as follows. For signal-phase variables, they are expressed relative to a solid "Walk" crossing signal category. For vehicle-type variables, they are expressed relative to the "Sedan" category. To offset the possibility of misinterpretation of confidence intervals due to a small number of sample sites (21 sites, although we have many data points within each), we used two *R* packages: *clubSandwich* [203] for CR2 confidence intervals and for *p*-values (clustered based on site), and then *clusterSEs* [204] to check robustness. The robustness check used a pairs cluster bootstrap (by site) of 4999 draws (We also mark coefficients with "@" in Table 5 where there was sparse entries for a given category; we did not interpret these). Overall, the model fitness calculation yielded McFadden R^2 of 0.21 and AUC of 0.82 (for $n = 1430$ pairs (174 of them were unsafe by our PET threshold of 3 s) across 21 sites). To examine sensitivity to clustering, we analyzed model errors by intersection instead of site (17 intersections instead of 21 sites), which yielded the same findings for coefficients and almost the same *p*-values on both "Don't Walk" phases and crowd density.

Table 5. Statistical model results and robustness tests.

Coefficient in Equation (10)	*Variable (see Table 2)*	*β-Value*	*Odds Ratio*	*95% CI (from CR2)*	*p-Value from CR2 [Bootstrap]*
		Signal phase categories			
β_1	Solid "Don't Walk"	1.894	6.647	0.731–60.413	0.075 [0.093]
β_2	No Signal[@]	1.725	5.614	N/A	0.227 [0.243]
β_3	Flashing "Don't Walk"	0.932	2.538	1.529–4.215	0.002 [0.006]
		Vehicle type categories			
β_4	Truck[@]	0.665	1.945	0.303–12.464	0.316 [0.264]
β_5	Bus[@]	0.252	1.286	N/A	0.754 [0.723]
β_6	Pickup truck	0.137	1.147	0.515–2.553	0.704 [0.679]
β_7	SUV	−0.048	0.953	0.568–1.600	0.842 [0.825]
β_8	Van	−0.127	0.880	0.299–2.596	0.790 [0.789]
		Pedestrian characteristics			
β_9	Carrying an object	−0.005	0.995	0.379–2.612	0.990 [0.987]
β_{10}	Pedestrian Waited	−0.413	0.662	0.311–1.408	0.246 [0.246]
β_{11}	Pedestrian Distracted (Fraction)	0.992	2.697	0.251–28.988	0.294 [0.278]
β_{12}	Pedestrian Mean Speed	0.014	1.014	0.639–1.610	0.945 [0.939]
β_{13}	Pedestrian is in Crosswalk (Fraction)	−0.232	0.793	0.272–2.310	0.642 [0.641]
β_{14}	Average # Pedestrians Around Crossing	−0.466	0.628	0.532–0.741	<0.001 [<0.001]
		Vehicle actions			
β_{15}	Vehicle Waited	−0.129	0.879	0.452–1.711	0.631 [0.605]
β_{16}	Vehicle Waiting	−0.249	0.780	0.414–1.469	0.362 [0.291]
β_{17}	Vehicle Mean Speed	−0.185	0.831	0.677–1.020	0.072 [0.059]

Above, odds ratios are stated relative to the steady "Walk" phase for signal phase and to sedans for vehicle type. @ denotes sparse categories ("No Signal", 3 pairs; "Bus", 14 pairs; "Truck", 26 pairs). For categories with answers of "Yes" or "No" only (carrying an object, pedestrian waited, vehicle waited, vehicle waiting), the ratios compare yes with no. For variables with fractions (pedestrian distracted, pedestrian in crosswalk), the ratios compare total time with zero time (fractions are 0 to 1). For continuous variables, ratios are per unit increase (for mean speed, units are in meters per second; for pedestrian counts they are the addition of an extra pedestrian to the crossing area).

Substantively, our model findings are as follows. Only two variables turned out to have associations with statistical significance to unsafe PET: crossing signals that communicate a solid "Don't Walk," and crowding of pedestrians around the crosswalk. This matches the

theoretical literature and other case study reports of errant behavior in the "crossing stage" and peer group assembly in the "approach stage" [12,14,15,128,129,205,206]. Two variables were close to statistical significance: again, in the crossing signal phase, but also for vehicle speed. We discuss these, as they are also relevant in the supporting literature.

The association between "Don't Walk" crossing signals and unsafe PET is relatively pronounced among all categories. During solid "Don't Walk" signals, pedestrians should not begin a crossing at all (i.e., if they move into the road, it is by jaywalking). When the crossing signal enters its flashing "Don't Walk" phase, pedestrians are advised to complete a crossing if in the crosswalk already, and to otherwise not proceed into the road at all and wait for the next crossing cycle. Movement of pedestrians into the crosswalk in either case of "Don't Walk" (solid or flashing for USA signals) is found in our analysis to be associated with increased odds of unsafe PET, particularly (and significant statistically) for flashing signals. Our model compares these situations to a solid "Walk" signal as follows: odds of an unsafe PET were found to be 2.5 times higher during flashing "Don't Walk" (versus solid "Walk") and 6.6 times higher during solid "Don't Walk" (versus solid "Walk"). Although the flashing case was significant, the solid "Don't Walk" crossing signal was not found to be statistically significant (and we recorded only 48 total "Don't Walk" pedestrian–vehicle dyads, at ten sites). (The "No signal" category had only three pedestrian–vehicle pairs in our study, so we do not discuss it further.)

The model also shows that there is some collective safety in crowds during road-crossing and this again matches other findings on peer effects at crossings [207]. Holding other categories and variables constant, the model reveals that each additional pedestrian that appears around a crossing is associated with 0.63 times odds of an unsafe PET. In other words, each addition of a pedestrian to the crossing assembly stage is associated with a reduction in unsafe PET odds by 37%. If we run through a diagnostic exercise with these odds, this infers that for our Times Square crossings (our most packed number of crossers around the intersection), if we take the most crowded tenth of pedestrian–vehicle interactions, the model estimates that the chances of unsafe PET drop to below 1% (Indeed, not one of the 55 pairs from the top tenth most crowded observations was unsafe by PET).

Vehicle speed showed a weak association with unsafe PET. Again, holding all other variables constant, each addition of a unit of vehicle speed (1 m per second) is associated with 0.83 times odds of an unsafe PET. Put another way, each addition of unit speed by a vehicle is associated with a reduction in unsafe PET odds by 17%. But this association was not statistically significant. Nonetheless, this tendency warrants some discussion. The finding for statistical association between speed and unsafe PET seems counterintuitive at face value. One possible explanation is that pedestrians are less likely to engage in conflict with vehicles that have picked up pace, while they may chance it if vehicles are inching and crawling into the conflict zone, as is common in right turns after a vehicle has slowed or stopped and the driver is trying to nudge their way through a pedestrian stream (We also point out that this reflects when close encounters occur, not that higher speeds are safer; holistically, the severity of any collision rises with vehicle speed.) Another possible explanation, consistent with a known problem at urban crossings [208], is that slow-moving vehicles edge through gaps between crossing pedestrians. This is exactly the type of phenomenon that is open to exploration in our XR pipeline. For example, the reconstructed dynamics that we show in Figure 12 illustrate ways in which a driver sneaking strategy may lead to problems: when vehicles edge close to pedestrians, the pedestrians' bodies can become obscured from the driver's view (e.g., due to the A-pillar blind spot in Figure 12). Similarly, it can be seen by examining cameras vicariously through the eyes of pedestrians that their direction of travel often points away from approaching vehicles,

which suggests that opportunities for eye contact and other NVCs may be reduced due to gaze disconnect.

We note that some of the data in Table 2 are not modeled statistically (chiefly, the codes for pedestrian–driver interactions), but we show these data to demonstrate the ethnographic markup capabilities of Road2Synth. Pedestrian–vehicle interaction codes are captured only for events that take place during an actual dyadic encounter, and therefore miss the events that led up to an encounter: as such, we do not include them in modeling PET. We also found that NVC type pedestrian–driver interactions were sparse in our data (likely because they are relatively rare and/or because they were difficult for annotators to notice in video). How these interactions interplay in conflicts, though, is possible to examine through the *Unity* connection to XR that Road2Synth provides, and this is a topic for further investigation.

$$\begin{aligned}\text{logit}[\Pr(\text{PET} < 3\,\text{s})] = \; & \beta_0 + \beta_1\,\text{DontWalk} + \beta_2\,\text{NoSignal} + \beta_3\,\text{FlashingDontWalk} \\ & + \beta_4\,\text{Truck} + \beta_5\,\text{Bus} + \beta_6\,\text{Pickup} + \beta_7\,\text{SUV} + \beta_8\,\text{Van} \\ & + \beta_9\,\text{Carrying} + \beta_{10}\,\text{PedWaited} + \beta_{11}\,\text{PedDistracted} \\ & + \beta_{12}\,\text{PedSpeed} + \beta_{13}\,\text{PedInside} + \beta_{14}\,\text{PedsAround} \\ & + \beta_{15}\,\text{VehWaited} + \beta_{16}\,\text{VehWaiting} + \beta_{17}\,\text{VehSpeed}\end{aligned} \tag{8}$$

where: $\text{logit}[\Pr(\text{PET} < 3\,\text{s})] = \ln \frac{\Pr(\text{PET} < 3\,\text{s})}{1-\Pr(\text{PET} < 3\,\text{s})}$ and $\text{OR}_m = e^{\beta_m}$ $(m = 1, \ldots, 17)$ is the odds ratio for term m, i.e., the factor by which odds of an unsafe PET change per unit increase in m (relative to its reference category) with the other terms held constant.

With Road2Synth data stored in a GIS, we are able to compare unsafe PETs with historical crash-related pedestrian-injury and fatality data. New York City maintains an open data repository of police-reported data, some of which are geocoded in ways that allow those reports to be tallied to our fieldwork sites (Recall that we chose accident hotspots as one of our site selection factors.) Analyzing data from 2012 to 2026 and assigning totals to our intersections (within a fifty-meter polling radius), we calculated a descriptive Spearman rank correlation between unsafe PET and historic data where we had mutual data (17 intersections, with multiple intersection sides recorded as distinct at some sites; see Figure 3). We found that our fieldwork intersections with more unsafe PETs also reported higher historical totals of pedestrian injury and fatality, and this association remained when we controlled for the number of pedestrian–vehicle pairs in our fieldwork. The correlation was +0.53 (with $p = 0.03$ by permutation) (+0.52 and $p = 0.04$ if we consider pedestrian–vehicle pairs). There are problems with relying on this evaluation empirically (small sample size, assignment by filter radius, historical data, possible issues in reporting, crashes with multiple injuries or fatalities, our fieldwork sites already selected partially for accident prevalence), but nonetheless it points to the sort of site-factor connections that can be made from Road2Synth.

6. Limitations

Road2Synth is a prototype, and it is not without limitations in its current form. First, issues regarding tracking errors could be addressed. We note that we did not have access to hand-labeled ground truth for the testing that we show in this paper, and as a result we did not directly measure the accuracy of Road2Synth's detection and tracking. However, we can report on (internal) consistency checks. As is common in computer vision of real-world dynamic scenes, we recorded tracking errors depending on recording conditions. For example, for Fall recordings we found jumps in 11% of pedestrian tracks and in 57% for rain or snow. However, jumps were only 5% across all vehicle tracks. But these jumps by conditions cannot be easily untangled from other site conditions that differed across recording location. To evaluate the resiliency of Road2Synth, we engaged an exercise in which we held back the set of 428 pedestrian–vehicle pairs for which there was a jump in

either track, finding that our statistical model results still hold even with these removals. The odds ratios remained as strong associations if they were removed, and the ratio value changed in a relatively small amount. For example, the statistical association of flashing "Don't Walk" signals with unsafe PET had an odds ratio of 3.0 (higher than the non-removal 2.5) and the odds ratio for crowd density was 0.61 (lower than the non-removal 0.63). In both cases, statistical significance also held. We note that the association between precipitation and jumps lost statistical significance once we controlled for the influence of season. Still, we acknowledge that seasonal and precipitation conditions do likely impact pedestrian–vehicle interactions in meaningful ways, and finding ways to isolate these effects and then to remedy their influence on jumps is a topic for future work.

Frame-by-frame tracking and other effects can also lead to jitter in computed trajectories. We addressed this by jitter-sensitive smoothing, but this is not enough to correct jumps and we found that smoothing straightened trajectory paths at scales below four meters. This is a limitation at the scale of our chosen PET two-meter conflict area and warrants further work to control for. Nonetheless, we point out that the *raw* trajectories calculated by Road2Synth do actually match-up with the sinuosity of real-world walking paths, as our calculations of Road2Synth pedestrian trajectories VFractal metrics showed when compared to published GPS archives of crossing pedestrians in New York City.

Occlusions are an ongoing headache for computer vision in busy and dynamic urban scenes, with potential negative impact on the efficacy of tracking results. In several of our fieldwork sites (Times Square, for example), the potential for lots of occlusions presents and is simply unavoidable when looking at pedestrian–vehicle interactions from one line of sight at the roadside. Our principal remediation for this is bound up in BoT-SORT, which uses Kalman filtering to recover loss due to occlusion. This provides some healing of trajectories, in the sense that occlusions are usually brief and BoT-SORT returns whole polylines. But this is a solution on tracking, rather than addressing the occlusion itself. Jumps were only 5% of vehicle tracks and we therefore regard them as less affected by occlusion than pedestrians (vehicle bumpers were hand-annotated by human users). However, we can report some promising results for occlusion due to crowding in the crosswalk. If one considers separations of below 0.4 m as being the rough foot-point spacing of two adults that are standing shoulder-to-shoulder and if one reasons that this distance is a lower bound of reliability (occluding their foot tracking), it sets a benchmark that is outside the range of our claims regarding PET (which are at an order of magnitude above this, at four meters). Across our discussed 21 fieldwork sites, we examined the 18 that were free of calibration concerns, and we found that the median distance between closest crossing pedestrians was 0.92 m (as in Table 3, these came from GIS). This is safely above the 0.4 m threshold (a relatively small fraction, 12%, of the frames for crossing situations showed pedestrians that were closer).

Road2Synth includes a semiautomated annotation tool, but aspects of it do rely on user input. We did not test its performance with an annotator outside our team, and as such we do not report measures of inter-annotator agreement. For our statistical results, this is not a critical shortcoming, as the annotations that are contributed by users are for crossing signal phase and vehicle bumpers. Vehicle bumpers are not a determinant in the models, and signal phases are not a subjective judgment call of the annotator (they flash their states in the video, as they do in the real world). Other variables, such as the position of pedestrians relative to each other and crowd density around the crosswalk, come directly from the automated components of the pipeline, and there is therefore no annotator testing required. Similarly, crosswalks are drawn by annotators, but they are calibrated on Google Street View data (and our testing showed agreement to 1.2%). We do rely on annotator judgment for behavioral codes (distraction, encumbrance, waiting, NVCs), but these enter

our models as category covariates only. And, none turned out as statistically significant in our reported findings. Nonetheless, rotating annotation among different users (including non-trained users) would be a valuable test for Road2Synth, particularly if we consider its potential use as a scientific tool with broader user uptake. And, further investigation into behavioral codes and their association with PET will require careful consideration of how annotators reach their decisions in the pipeline.

It would be useful to look more closely at the pipeline's effectiveness in measurement accuracy. Distances in the pipeline, as we discussed, inherit each recording's calibration, and this depends in our tests on Google Street View rather than formal surveying for control points. Our consistency checks catch recordings with user calibration concerns, and we excluded them from the analysis and findings that we report here or reported them in separate terms. Further, we showed that even rescaling distances by +/- 20% has only a median 0.5 s change to PET. Moreover, we mention that because we used tripods for roadside cellphone video recording, a single homography holds in each of the recordings in the work we show here. Nonetheless, improving measurement will always be a worthy target for future work.

The fully worked XR demonstration (Gaussian-splatting) is just for a single site, to prove the concept. Further, it relies on a commercial tool. The latter was intentional, to demonstrate that Road2Synth can work with cellphone apps. We have shown (based on the specifications for the software) that we can successfully integrate a range of reconstruction methods with Road2Synth and these include SfM and Gaussian splatting, which we demonstrate can integrate with the homography and GIS from Road2Synth and which can be interwoven into *Unity* to remap trajectories from our scheme into commercial XR systems. We regard this as a proof of utility and we point out that many different software approaches (open source or commercial) could be extensibly placed in that part of the pipeline, including Sim2Real [209,210] schemes for autonomous driving, Wi-Edge approaches in smart city applications [26], monitoring and alert systems for Internet of Things applications [211,212], geosimulation with agent-based models [60,67,68], and other approaches.

An important point that we wish to stress is as follows. In all limitations, we remind the reader that in this pipeline we are trading the ease of using cellphone video in the toolkit against the perfection of a formal survey-grade exercise. Road2Synth is intended to provide easy access to XR-assisted safety analysis, rather than a definitive solution. We see the pipeline as operating alongside formal survey-grade approaches using LiDAR for measurement, multiple cameras, overhead video, and human calibration and testing.

7. Findings

Our findings from this proof-of-concept exercise, then, are as follows. First, we were successful in building a pipeline that can take roadside cellphone video and run it through incremental methodological phases that support ethnography, computer vision and deep learning, homography, GIS and spatial analysis, statistical modeling, 3D scene reconstruction, and XR animation, with retention of original data and the addition of value-added information along the way. Further, we were able to run Road2Synth across a very large volume of video data. We show worked examples for 21 field sites, with 215 cellphone recording captures that returned 4403 trajectories. This is shown across New York City, which indicates it can be deployed citywide while handling shifts in scene properties, demographics, traffic rates, weather and time of day changes, and variability in road infrastructure. Indeed, our claim that Road2Synth could serve as a semiautomated generator of roadside atlases, censuses, and choreographies of pedestrian–vehicle safety while crossing has some genuine support in our tests (Applications outside New York City may extend this support

further.) We contend that this support holds up in generalizability at city scale and in depth of detail across many thousands of observations.

Second, Road2Synth, although prototypical, holds up against several consistency checks. Further practical support for this comes from analysis of historical pedestrian–vehicle crash data, with Road2Synth's unsafe PET metrics returning correlations to accident patterns (broadly) from historical data that were not part of the observations. Internal consistency checks hold up on the pipeline's crosswalk measurements, on jitter, and on sinuosity of trajectories, showing reasonable goodness of fit to expected behavior for the pipeline in applications to real scenarios.

Third, Road2Synth shows some promise as a hypothesis-testing instrument. Our statistical models on the data that it collects show associations in factors for crossing safety, including the effects of signal phase and crowding on increasing or lowering odds of unsafe PET. Further, at a holistic level, Road2Synth shows that it could be used for geographical site selection for safety risk in PET across a city, as well as serving as a tool to diagnose particular site and individual factors that contribute to that risk at the resolution of the crosswalk, the signal, and other hyper-local factors. Considered across both scales, Road2Synth shows promise as a flexible and scaling situational awareness tool for roadside safety science.

There are several advantages for XR. In Section 3, we discussed three XR-mediated views of pedestrian–vehicle interactions: scene view (exocentric), allocentric view (thing to thing), and egocentric view (from pedestrian and driver perspectives). Our findings at each of these views show that XR can provide more than visualization at the end of the pipeline. The XR scene view, for example, can be populated with detailed positional data that facilitate the generation of VGEs from Road2Synth, with reach to individualized presence in collective space-time paths through a scene. Our tests of sinuosity, for example, show that Road2Synth produces realistic shapes of trajectories when compared to real-world data from New York City road-crossing. Walking speeds also match to realistic values. The XR allocentric view can be built directly on PET, with the result that it can catalog (empirically) the time-space interactions between entities in the roadscape. In statistical modeling, for example, we showed the effect of crowding in reducing unsafe PETs. And we showed the allocentric relationships between crossing and crossing signals, with empirical association. The XR egocentric view, then, can help to explain these behaviors, events, and associations, and we reason that couplings to geosimulation-based VGEs would allow theories and hypotheses to be developed in sandbox fashion, with potential reach back to real-world outcomes. Indeed, using VR, real human users could drop into some of the scenes from Road2Synth, and their likely behavior could be gauged ahead of real-world roll-out of design interventions.

8. Conclusions

We examined the potential for easy-to-capture video to be folded into scientific pipelines for the construction of XR environments. Using a proof of concept from traffic safety science, we showed that two-dimensional exocentric roadside video (paired with a brief walk-around video capture) can be used to reconstruct (and to dynamically populate and synthetically enliven) immersive three-dimensional scenes in *Unity*. At its core, this shows that video can be used to capture, frame, and reconstruct scenes with the moving dynamics of the pedestrians and vehicles that it images, and that traditional ethnographic research can be carried along in that process, making use of computer vision and deep learning to add explanatory value along the way. One key methodological element of achieving this in pipeline fashion is to resolve difference in spaces that collect among varying viewpoints, relations between objects and entities, and mathematical and computing modalities. We showed that a shared manifold tied to GIS can assist in establishing convergence between

these spaces. Further, we demonstrated that once observations and records are registered to this manifold, they can be rendered from multiple perspectives and in multiple media (including statistical models). We argue that in doing so, XR can take on attributes of VGEs, with considerable advantages for fieldwork-based science, which invariably benefits from strong linkages between the virtuality of XR and tangible ground truth in the real world.

This paper catalogs the development of a working prototype, which we term as "Road2Synth," that demonstrates the utility of video-anchored XR for applied safety science. In particular, we showed that the fleeting moments of space and time that account for conflict risks between pedestrians and vehicles in urban crossings can be captured, with significant explanatory detail, using video and ethnographic annotation. Deep learning can assist, providing automation and indexing that ease the exchange of the information contained in video into what-if reconstructed XR environments. Within VR, for example, this information can take on new meaning and explorative value in teasing out regularities and associations that are captured by statistical models.

In the motivation for this work, we proposed that XR could deepen understanding of PET by enabling three complementary views of traffic intersections and pedestrian crossings. An exocentric scene view can frame positions of identified and tracked road users in a crossing scene, and we have shown that Road2Synth can assist in deriving those positions from video, mapping them to a GIS, and tracking them to a unified geography. Importantly, this insight can be drawn directly from a single-perspective roadside video on a straightforward cellphone device. Moreover, a walk-around video can be used to build three-dimensional VGEs for the same site. Using Road2Synth, a counterpart allocentric view can be generated to capture the relationships between road users and to frame their potential for conflict: who has a right of way, who yields to whom when conflicts arise, who moves first, and how all of these events interoperate with crossing signals. We showed that PET, which records timing of who passes through intersection conflict areas and by how much (including under shifting signal phases), can be calculated and analyzed directly in this allocentric view. Further, we show that Road2Synth can build an egocentric view, which we use to place virtual cameras into the synthetic eyes of individual pedestrians and individual drivers, connecting PET to individual moments of lived experience (albeit vicariously) of road users. This brings fieldwork into direct bearing with NRT and other conceptual and theoretical modalities for examining people's behavioral geography, with direct grounded (and empirical) support. Consider, for example, that in this way we are able to recover synthetic vistas and vantage points from the roadside in ways the original video never recorded.

Our demonstration is a first step and we acknowledge that significant future work could be adopted to advance the idea further. This could include applications to other human and behavioral geographies that unfold over high resolutions of space and time, for example in crowd dynamics [213]; social and socio-psychological interactions between people, places, and things [214]; or retail touchpoints in customer journeys [215]. Other endpoints for the pipeline could easily be envisaged, many of which would prove compatible with XR. These include human agent-based simulations [216], exocentric and egocentric computer gaming [217], and driving simulators [218]. It is reasonable to consider that systems like Road2Synth could be used to harness explainable AI [219], particularly given the broad implications for simulation and inference in XR atop video scenes from the real world. It is also feasible that pipelines such as Road2Synth could be placed in installations close to the phenomena that they can explain and explore, for example, as roadside edge computing [212,220], where they could be useful in advancing existing capabilities for smart and sentient streetscapes.

In the applied science that we have shown in this paper, the use of Road2Synth opens up some intriguing questions with society-facing value propositions. The first centers on understanding human risk-taking in roadside crossing contexts [103,144,205,221,222]. For example, there are long-standing open questions of how we can build policies for vulnerable crossers, particularly the young and the elderly [103,223,224]. The systems that route vehicles and crossers through the road only function as well as adherence to the prompts and rules that they effect. However, there are known lapses in adherence on both counts in both young and old populations. In particular, younger crossers are more predisposed to overestimating their ability to fulfill traffic gaps with small space-time windows of opportunity than the general population might be [221,225–227]. Elderly crossers, especially those at the upper end of the age demographic, are more likely to misjudge the open traffic gaps available to them when stepping into the road from a sidewalk, and they are more prone to miscalculating time to collision and distance to collision than non-elderly crossers might be [105,106,228]. These factors are due to a mixture of declining faculties, locomotion flexibility, and balance, which make curbsides and mid-roadway islands a risk factor for some populations [223,229,230]. Across all pedestrians, distractions from handheld electronic devices are also causing challenges for perceptual awareness [231].

Our Road2Synth pipeline shows how one might be able to study entire holistic road scenes by starting, first and foremost, with the pedestrian perspective, then working one's way back to draw-in explanatory context. This is a relatively unexplored perspective in safety science, because data from individual viewpoints are hard to come by. Road2Synth could be useful in providing such data. As we have shown, analyses built on the platform can add explanatory value to existing traffic-oriented indicators and risk factors, such as TTC and PET, and these metrics could interplay with traditional statistical models in traffic engineering. Design factors that could remediate or remedy hazardous crosser–vehicle interactions constitute a second policy vantage that Road2Synth enables. The XR-based endpoint for Road2Synth could support a variety of design interventions, which could be explored as what-if possibilities ahead of construction in the real world. This includes painting guidance onto roads, as turn arrows and stopping points for drivers, as well as zebra crossing zones and directional advice (e.g., "Look right") for pedestrians. A number of traffic-calming measures have been considered in urban design [18,119,232–234]. Similar designs could be examined in Road2Synth, including bulb-outs that extend the sidewalk into the road at crossing sites, pedestrian refuge islands that give pedestrians a waiting space in the middle of road lanes, chicanes that force drivers to slow down, raised intersections that serve as a speed bump and level pedestrian–driver inter-visibility, and reconfigured turn geometry that forces a tighter turning radius (and therefore slower vehicle speeds at turns). Curbside elements could also be considered to shoehorn pedestrians into stopping and looking up before entering crossings, including using bollards. In many cases, these design elements are integrated with signaling infrastructure, e.g., in pedestrian light-controlled (pelican) crossings, in which crossers can directly call for a halt in traffic using buttons at the roadside. For right-turn conflicts, which cause particular problems for pedestrian safety, the concentration of short PETs during flashing "Don't Walk" signals that we identified are perhaps indicative of signal-timing interventions that could be made to improve safety, e.g., through splitting turn phases and adjusting clearance timing. Road2Synth provides one mechanism by which those timings could be examined synthetically before application in reality. Indeed, as our *Unity* implementation demonstrates, Road2Synth allows for recorded paths to be replayed directly through the "eyes" of synthetic counterpart road users, which could be coupled with real human-participant interactive scenario testing to see how people would respond to design interventions. This is a topic that we take up in a complementary system, Sidewalk2Synth [194]. Counterpart strategies for drivers could be examined to

see how convergence between pedestrian, driver, and site factors might converge around a particular design solution or intervention. Indeed, this might focus dyadic attention on the crossing phase so that all parties have the information, insight, context, and time that they need to act safely in its small time geography.

9. Ethics and Privacy Concerns

Recording video in public places is legal in New York City. In many of the locations that we visited (particularly tourist areas), it is routine throughout the day and night. Nonetheless, we conducted all research in this project and covered by this paper using an approved human subjects protocol under the supervision of our institutional review board (which is independent body of our institution). All research participants underwent humans subjects research training and obtained CITI certification before engaging in the research.

To safeguard the anonymity of the data in the project, all videos are stored on a password-protected folder, with gated user access and SSO credentialling, in our institutional storage system that is reserved for high-risk data. Our data do not contain undue risk (similar street scene data are, for example, routinely published to Web video sites and feature in movies and news programs). Nonetheless, out of an abundance of caution, we stored all data on our Level 3 system (for the highest risk data). To safeguard the privacy of any individuals, all images in this paper have been filtered to scramble the appearance of faces, license plates, and logos using a crystallize filter. For this reason, we are not releasing any of the data to public repositories and we only show aggregated statistics in this paper or anonymized tracks.

We note that our ability to re-render scenes as VGE allows us to naturally anonymize people and things that are recorded in fieldwork: we can, for example, literally replace a pedestrian with a synthetic stand-in, as we have shown in this paper. Finally, we raise the consideration that the merits of studying crossing safety and attempting to better equip communities and municipalities with ease-of-use tools to safeguard pedestrians when moving through traffic counterbalances concerns of recording videos of everyday public life in ways that are legal and often normal in those locations.

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

Funding: This research received no external funding.

Data Availability Statement: Source data is withheld due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Middleton, J. 'Stepping in Time': Walking, Time, and Space in the City. *Environ. Plan. A Econ. Space* **2009**, *41*, 1943–1961. [CrossRef]
2. Whyte, W.H. *The Social Life of Small Urban Spaces*; The Conservation Foundation: Washington, DC, USA, 1980.
3. Goffmann, E. *Behavior in Public Places*; Free Press: New York, NY, USA, 1963.
4. Moudon, A.V.; Hess, P.M.; Snyder, M.C.; Stanilov, K. Effects of Site Design on Pedestrian Travel in Mixed-Use, Medium-Density Environments. *Transp. Res. Rec.* **1997**, *1578*, 48–55. [CrossRef]
5. Hess, P.; Moudon, A.; Snyder, M.; Stanilov, K. Site Design and Pedestrian Travel. *Transp. Res. Rec. J. Transp. Res. Board* **1999**, *1674*, 9–19. [CrossRef]

6. Talen, E. Neighborhoods as Service Providers: A Methodology for Evaluating Pedestrian Access. *Environ. Plan. B Plan. Des.* **2003**, *30*, 181–200. [CrossRef]
7. Stutts, A.; Kretschmer, M.A.; Schultz, G.G. Improving Safety at Signalized Intersections: Results from a State Dot Survey. In *Proceedings of the International Conference on Transportation and Development (Ictd), Detroit, Mi, USA, 28 June–1 July 2026*; Wieferich, B., FGeara, T., Ryan, A., Eds.; American Society of Civil Engineers: Reston, VA, USA, 2026; pp. 615–626.
8. Marsden, J.P.; Montgomery, S.R. A General Survey of the Walking Habits of Individuals. *Ergonomics* **1972**, *15*, 439–451. [CrossRef] [PubMed]
9. FHWA. *Nationwide Personal Travel Survey*; Department of Transportation Federal Highway Administration: Washington, DC, USA, 1995.
10. Sisiopiku, V.P.; Akin, D. Pedestrian Behaviors at and Perceptions Towards Various Pedestrian Facilities: An Examination Based on Observation and Survey Data. *Transp. Res. Part F Traffic Psychol. Behav.* **2003**, *6*, 249–274. [CrossRef]
11. Vannini, P. Non-Representational Ethnography: New Ways of Animating Lifeworlds. *Cult. Geogr.* **2015**, *22*, 317–327. [CrossRef]
12. Aghabayk, K.; Esmailpour, J.; Jafari, A.; Shiwakoti, N. Observational-Based Study to Explore Pedestrian Crossing Behaviors at Signalized and Unsignalized Crosswalks. *Accid. Anal. Prev.* **2021**, *151*, 105990. [CrossRef] [PubMed]
13. Lanzer, M.; Gieselmann, M.; Mühl, K.; Baumann, M. Assessing Crossing and Communication Behavior of Pedestrians at Urban Streets. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *80*, 341–358. [CrossRef]
14. Mohammed, M.H. Analysis of Pedestrians' Behavior at Pedestrian Crossings. *Saf. Sci.* **2001**, *38*, 63–82. [CrossRef]
15. Gang, R.; Zhou, Z.; Zhang, W.; Wang, Y. Crossing Behaviors of Pedestrians at Signalized Intersections: Observational Study and Survey in China. *Transp. Res. Rec.* **2011**, *2264*, 65–73. [CrossRef]
16. Cohen, A.; Bar-Gera, H.; Parmet, Y.; Ronen, A. Guardrail Influence on Pedestrian Crossing Behavior at Roundabouts. *Accid. Anal. Prev.* **2013**, *59*, 452–458. [CrossRef] [PubMed]
17. Oudejans, R.R.; Michaels, C.F.; van Dort, B.; Frissen, E.J. To Cross or Not to Cross: The Effect of Locomotion on Street-Crossing Behavior. *Ecol. Psychol.* **1996**, *8*, 259–267. [CrossRef] [PubMed]
18. Collett, P.; Marsh, P. Patterns of Public Behavior: Collision Avoidance on a Pedestrian Crossing. In *Nonverbal Communication, Interaction and Gesture Selections from Semiotica (Approaches to Semiotics; 41)*; Kendon, A., Ed.; Mouton: The Hague, The Netherlands, 1981; pp. 199–217.
19. Hudson-Smith, A.; Hügel, S.; Roumpani, F. Self-Monitoring, Analysis and Reporting Technologies: Smart Cities and Real-Time Data. In *The Routledge Companion to Smart Cities*; Willis, K.S., Aurigi, A., Eds.; Routledge: London, UK, 2020; pp. 383–394.
20. Shull, P.B.; Jirattigalachote, W.; Hunt, M.A.; Cutkosky, M.R.; Delp, S.L. Quantified Self and Human Movement: A Review on the Clinical Impact of Wearable Sensing and Feedback for Gait Analysis and Intervention. *Gait Posture* **2014**, *40*, 11–19. [CrossRef] [PubMed]
21. Kiefer, P.; Giannopoulos, I.; Raubal, M. Where Am I? Investigating Map Matching during Self-Localization with Mobile Eye Tracking in an Urban Environment. *Trans. GIS* **2014**, *18*, 660–686. [CrossRef]
22. Swan, M. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *J. Sens. Actuator Netw.* **2012**, *1*, 217–253. [CrossRef]
23. Torrens, P.M. Ten Traps for Non-Representational Theory in Human Geography. *Geographies* **2024**, *4*, 253–286. [CrossRef]
24. Gawande, U.; Hajari, K.; Golhar, Y. Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Activity Recognition. *Procedia Comput. Sci.* **2023**, *218*, 2438–2447. [CrossRef]
25. Gavrila, D.M. Sensor-Based Pedestrian Protection. *IEEE Intell. Syst.* **2001**, *16*, 77–81. [CrossRef]
26. Torrens, P.M. Smart and Sentient Retail High Streets. *Smart Cities* **2022**, *5*, 1670–1720. [CrossRef]
27. Thrift, N. The 'Sentient' City and What It May Portend. *Big Data Soc.* **2014**, *1*, 2053951714532241. [CrossRef]
28. Crang, M.; Graham, S.N. Sentient Cities Ambient Intelligence and the Politics of Urban Space. *Inf. Commun. Soc.* **2007**, *10*, 789–817. [CrossRef]
29. Batty, M. Big Data, Smart Cities and City Planning. *Dialogues Hum. Geogr.* **2013**, *3*, 274–279. [CrossRef]
30. Taleb, T.; Dutta, S.; Ksentini, A.; Iqbal, M.; Flinck, H. Mobile Edge Computing Potential in Making Cities Smarter. *IEEE Commun. Mag.* **2017**, *55*, 38–43. [CrossRef]
31. Batty, M. *Artificial Intelligence and Smart Cities*; SAGE Publications Sage UK: London, UK, 2018.
32. Batty, M. Smart Cities, Big Data. *Environ. Plan. B Plan. Des.* **2012**, *39*, 191–193. [CrossRef]
33. Graham, M.; Zook, M.; Boulton, A. Augmented Reality in Urban Places: Contested Content and the Duplicity of Code. *Trans. Inst. Br. Geogr.* **2013**, *38*, 464–479. [CrossRef]
34. Feiner, S.; MacIntyre, B.; Höllerer, T.; Webster, A. A Touring Machine: Prototyping 3d Mobile Augmented Reality Systems for Exploring the Urban Environment. *Pers. Technol.* **1997**, *1*, 208–217. [CrossRef]
35. Chen, M.; Claramunt, C.; Çöltekin, A.; Liu, X.; Peng, P.; Robinson, A.C.; Wang, D.; Strobl, J.; Wilson, J.P.; Batty, M.; et al. Artificial Intelligence and Visual Analytics in Geographical Space and Cyberspace: Research Opportunities and Challenges. *Earth-Sci. Rev.* **2023**, *241*, 104438. [CrossRef]

36. Çöltekin, A.; Lochhead, I.; Madden, M.; Christophe, S.; Devaux, A.; Pettit, C.; Lock, O.; Shukla, S.; Herman, L.; Stachoň, Z. Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 439. [CrossRef]
37. Liu, B.; Ding, L.; Meng, L. Spatial Knowledge Acquisition with Virtual Semantic Landmarks in Mixed Reality-Based Indoor Navigation. *Cartogr. Geogr. Inf. Sci.* **2021**, *48*, 305–319. [CrossRef]
38. Baudrillard, J. *Simulacra and Simulation*; University of Michigan Press: Ann Arobor, MI, USA, 1994.
39. Torrens, P.M.; Kim, R.; Torrens, F.S.; Shinozaki-Conefrey, K. Geosimulation by Illusion: An Anchorpoint Hypothesis for Redirected Walking Through Expansive Virtual Geographies. In *Proceedings of the GeoSim '26: Proceedings of the 9th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation, Riverside, CA, USA, 3–6 November 2026*; Kim, J.-S.K., Züfle, A., Eds.; Association for Computing Machinery: New York, NY, USA, 2026; pp. 39–49.
40. McComas, J.; MacKay, M.; Pivik, J. Effectiveness of Virtual Reality for Teaching Pedestrian Safety. *Cyberpsychol. Behav.* **2002**, *5*, 185–190. [CrossRef] [PubMed]
41. Kwon, J.-H.; Kim, J.; Kim, S.; Cho, G.-H. Pedestrians Safety Perception and Crossing Behaviors in Narrow Urban Streets: An Experimental Study Using Immersive Virtual Reality Technology. *Accid. Anal. Prev.* **2022**, *174*, 106757. [CrossRef] [PubMed]
42. Townsend, A.M. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*; W. W. Norton: New York, NY, USA, 2013.
43. Torrens, P.M. Computational Streetscapes. *Computation* **2016**, *4*, 37. [CrossRef]
44. Ricci, M.; Mosca, N.; Rafik, M.; Di Summa, M. How Extended Reality Is Shaping Smart Cities: A Systematic Literature Review. *Appl. Sci.* **2026**, *16*, 679. [CrossRef]
45. Neves, J.N.; Câmara, A. Virtual Environments and Gis. *Geogr. Inf. Syst.* **1999**, *1*, 557–565.
46. Chen, M.; Sheng, Y.H.; Wen, Y.N.; Sheng, J.W.; Su, H.J. Virtual Geographic Environments Oriented 3d Visualization System. *J. Syst. Simul.* **2008**, *20*, 5105–5108.
47. Lin, H.; Batty, M. *Virtual Geographic Environments: A Primer*; ESRI Press: Redlands, CA, USA, 2011.
48. Lin, H.; Batty, M.; Jørgensen, S.E.; Fu, B.; Konecny, M.; Voinov, A.; Torrens, P.M.; Lu, G.; Zhu, A.-X.; Wilson, J.P.; et al. Virtual Environments Begin to Embrace Process-Based Geographic Analysis. *Trans. GIS* **2015**, *19*, 493–498. [CrossRef]
49. Peesapati, L.N.; Hunter, M.P.; Rodgers, M.O. Can Post Encroachment Time Substitute Intersection Characteristics in Crash Prediction Models? *J. Saf. Res.* **2018**, *66*, 205–211. [CrossRef] [PubMed]
50. Oden, J.T.; Belytschko, T.; Fish, J.; Hughes, T.J.R.; Johnson, C.; Keyes, D.; Laub, A.; Petzold, L.; Srolovitz, D.; Yip, S.; et al. *Simulation-Based Engineering Science: Revolutionizing Engineering Science Through Simulation*; National Science Foundation: Alexandria, VA, USA, 2006.
51. Casti, J.L. *Would-Be Worlds: How Simulation Is Changing the Frontiers of Science*; John Wiley & Sons: New York, NY, USA, 1997.
52. Qi, S.; Menozzi, M. Investigating Pedestrian Crossing Decision with Autonomous Cars in Virtual Reality. *Z. Für Arbeitswissenschaft* **2023**, *77*, 218–229. [CrossRef]
53. Figueroa-Medina, A.M.; Valdés-Díaz, D.; Colucci-Ríos, B.; Cardona-Rodríguez, N.; Chamorro-Parejo, A. Analysis of Walking Speeds and Success Rates on Mid-Block Crossings Using Virtual Reality Simulation. *Accid. Anal. Prev.* **2023**, *183*, 106987. [CrossRef] [PubMed]
54. Luu, D.T.; Eom, H.; Cho, G.-H.; Kim, S.-N.; Oh, J.; Kim, J. Cautious Behaviors of Pedestrians While Crossing Narrow Streets: Exploration of Behaviors Using Virtual Reality Experiments. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *91*, 164–178. [CrossRef]
55. Guo, X.; Angulo, A.; Robartes, E.; Chen, T.D.; Heydarian, A. Orclsim: A System Architecture for Studying Bicyclist and Pedestrian Physiological Behavior through Immersive Virtual Environments. *J. Adv. Transp.* **2022**, *2022*, 2750369. [CrossRef]
56. Angulo, A.; Robartes, E.; Guo, X.; Chen, T.D.; Heydarian, A.; Smith, B. Validation of a Virtual Reality Simulator with Real-World Observations for Pedestrian Safety at Midblock Crossings. *SSRN Electron. J.* **2022**. Available online: https://www.researchgate.net/publication/359191890_Validation_of_a_Virtual_Reality_Simulator_with_Real-World_Observations_for_Pedestrian_Safety_at_Midblock_Crossings (accessed on 2 November 2026). [CrossRef]
57. Sebastian, W.; Julia, B.; Fabian, J.; Angelika, T.-O.; Christian, H.; Bernhard, P.; Patrick, S. Ivroad: Immersive Virtual Road Crossing as an Assessment Tool for Unilateral Spatial Neglect. *Comput. Graph.* **2021**, *99*, 70–82. [CrossRef]
58. Rajaram, B.; Williams, B.; Owens, J.; Gibbons, R. The Reality of Virtual Reality: A Comparison of Pedestrian Behavior in Real and Virtual Environments. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2018**, *62*, 2056–2060. [CrossRef]
59. Deb, S.; Carruth, D.W.; Sween, R.; Strawderman, L.; Garrison, T.M. Efficacy of Virtual Reality in Pedestrian Safety Research. *Appl. Ergon.* **2017**, *65*, 449–460. [CrossRef] [PubMed]
60. Torrens, P.M.; Kim, R. Evoking Embodiment in Immersive Geosimulation Environments. *Ann. GIS* **2024**, *30*, 35–66. [CrossRef]
61. Benenson, I.; Torrens, P.M. A Minimal Prototype for Integrating GIS and Geographic Simulation through Geographic Automata Systems. In *Geodynamics*; Atkinson, P., Foody, G., Darby, S., Wu, F., Eds.; CRC Press: Boca Raton, FL, USA, 2005; pp. 347–369.
62. Goodchild, M.F. Virtual Geographic Environments as Collective Constructions. *Virtual Geogr. Environ.* **2009**, 15–24.
63. Batty, M. Virtual Geography. *Futures* **1997**, *29*, 337–352. [CrossRef]

64. Batty, M.; Chapman, D.; Evans, S.; Haklay, M.; Kueppers, S.; Shiode, N.; Smith, A.; Torrens, P.M. Visualizing the City: Communicating Urban Design to Planners and Decision-Makers. In *Planning Support Systems in Practice: Integrating Geographic Information Systems, Models, and Visualization Tools*; Brail, R.K., Klosterman, R.E., Eds.; ESRI Press and Center for Urban Policy Research Press: Redlands, CA, USA; New Brunswick, NJ, USA, 2001; pp. 405–443.
65. Lin, H.; Chen, M. Managing and Sharing Geographic Knowledge in Virtual Geographic Environments (Vges). *Ann. GIS* **2015**, *21*, 261–263. [CrossRef]
66. Torrens, P.M. Process Models and Next-Generation Geographic Information Technology. In *Gis Best Practices: Essays on Geography and Gis*; ESRI Press: Redlands, CA, USA, 2009; Volume 2, pp. 63–75.
67. Torrens, P.M.; Kim, R.; Shinozaki-Conefrey, K. Experiential Geosimulation. In Proceedings of the 8th ACM SIGSPATIAL International Workshop on Geospatial Simulation, Minneapolis, MN, USA, 3–6 November 2025; pp. 12–20.
68. Torrens, P.M.; Gu, S. Real-Time Experiential Geosimulation in Virtual Reality with Immersion-Emission. In *Proceedings of the 4th Acm Sigspatial International Workshop on Geospatial Simulation*; Association for Computing Machinery: Beijing, China, 2021; pp. 19–28.
69. Torrens, P.M. Slipstreaming Human Geosimulation in Virtual Geographic Environments. *Ann. GIS* **2015**, *21*, 325–344. [CrossRef]
70. Torrens, P.M. Intertwining Agents and Environments. *Environ. Earth Sci.* **2015**, *74*, 7117–7131. [CrossRef]
71. Torrens, P.M. Behavioral Intelligence for Geospatial Agents in Urban Environments. In *IEEE Intelligent Agent Technology (Iat 2007)*; Lin, T.Y., Bradshaw, J.M., Klusch, M., Zhang, C., Eds.; IEEE: Los Alamitos, CA, USA, 2007; pp. 63–66.
72. Hudson-Smith, A. Incoming Metaverses: Digital Mirrors for Urban Planning. *Urban Plan.* **2022**, *7*, 343–354. [CrossRef]
73. Crooks, A.; Hudson-Smith, A.; Dearden, J. Agent Street: An Environment for Exploring Agent-Based Models in Second Life. *J. Artif. Soc. Soc. Simul.* **2009**, *12*, 10.
74. Torrens, P.M.; Gu, S. Inverse Augmentation: Transposing Real People into Pedestrian Models. *Comput. Environ. Urban Syst.* **2023**, *100*, 101923. [CrossRef]
75. Benenson, I.; Torrens, P.M. *Geosimulation: Automata-Based Modeling of Urban Phenomena*; John Wiley & Sons: London, UK, 2004.
76. Batty, M.; Torrens, P.M. Modeling Complexity: The Limits to Prediction. *CyberGeo* **2001**, *201*, 1035.
77. Allen, P.M.; Torrens, P.M. Knowledge and Complexity. *Futures* **2005**, *37*, 581–584. [CrossRef]
78. Helbing, D.; Molnár, P. Self-Organization Phenomena in Pedestrian Crowds. In *Self-Organization of Complex Structures: From Individual to Collective Dynamics*; Schweitzer, F., Ed.; Gordon and Breach: London, UK, 1997; pp. 569–577.
79. Helbing, D.; Molnár, P.; Farkas, I.J.; Bolay, K. Self-Organizing Pedestrian Movement. *Environ. Plan. B Plan. Des.* **2001**, *28*, 361–383. [CrossRef]
80. Helbing, D.; Buzna, L.; Johansson, A.; Werner, T. Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions. *Transp. Sci.* **2005**, *39*, 1–24. [CrossRef]
81. Thrift, N. *Non-Representational Theory: Space, Politics, Affect*; Routledge: New York, NY, USA, 2008.
82. Smith, R.G. Baudrillard's Nonrepresentational Theory: Burn the Signs and Journey without Maps. *Environ. Plan. D Soc. Space* **2003**, *21*, 67–84. [CrossRef]
83. Golledge, R.; Stimson, R.J. *Spatial Behavior: A Geographic Perspective*; The Guilford Press: New York, NY, USA, 1997.
84. Golledge, R.G. Environmental Cognition. *Handb. Environ. Psychol.* **1987**, *1*, 131–174.
85. Torrens, P.M. Deep Geographic Embodiment. *Inf. Geogr.* **2025**, *2*, 100034. [CrossRef]
86. Montello, D.R. (Ed.) Artificial Intelligence and Behavioral Geography. In *Handbook of Behavioral and Cognitive Geography*; Edward Elgar Publishing: Cheltnam, UK, 2018; pp. 357–372.
87. Sheller, M. From Spatial Turn to Mobilities Turn. *Curr. Sociol.* **2017**, *65*, 623–639. [CrossRef]
88. Büscher, M.; Urry, J. Mobile Methods and the Empirical. *Eur. J. Soc. Theory* **2009**, *12*, 99–116. [CrossRef]
89. Pred, A. The Choreography of Existence: Comments on Hägerstrand's Time-Geography and Its Usefulness. *Econ. Geogr.* **1977**, *53*, 207–221. [CrossRef]
90. Endo, H. The Touristic Turn in Social and Human Sciences: Searching for the Generative Discipline. *Tour. Stud. Rev.* **2013**, *1*, 129–144.
91. Arnold, E. Photography, Composition, and the Ephemeral City. *Area* **2021**, *53*, 659–670. [CrossRef]
92. Kaklauskas, A.; Zavadskas, E.K.; Bardauskiene, D.; Cerkauskas, J.; Ubarte, I.; Seniut, M.; Dzemyda, G.; Kaklauskaite, M.; Vinogradova, I.; Velykorusova, A. An Affect-Based Built Environment Video Analytics. *Autom. Constr.* **2019**, *106*, 102888. [CrossRef]
93. Simpson, P. 'So, as You Can See...': Some Reflections on the Utility of Video Methodologies in the Study of Embodied Practices. *Area* **2011**, *43*, 343–352. [CrossRef]
94. Cutting, J.E. Perceptual Artifacts and Phenomena: Gibson's Role in the 20th Century. In *Advances in Psychology*; Masin, S.C., Ed.; Elsevier Science Publishers: New York, NY, USA, 1993; pp. 231–260.
95. Gibson, J.J. *The Ecological Approach to Visual Perception*; Houghton, Mifflin and Company: Boston, MA, USA, 1979.
96. Gibson, J.J. *The Senses Considered as Perceptual Systems*; Houghton-Mifflin: Boston, MA, USA, 1966.

97. Gibson, J.J. *The Perception of the Visual World*; Houghton-Mifflin: Boston, MA, USA, 1950.
98. Goffmann, E. *Relations in Public: Microstudies in the Public Order*; Basic Books: New York, NY, USA, 1971.
99. National Highway Traffic Safety Administration. *Pedestrians: 2024 Data*; National Highway Traffic Safety Administration: Washington, DC, USA, 2026.
100. IIHS. Fatality Facts 2020: Pedestrians. Insurance Institute for Highway Safety, Highway Loss Data Institute. Available online: https://www.iihs.org/topics/fatality-statistics/detail/pedestrians (accessed on 23 February 2023).
101. CDC. Pedestrian Safety. U.S. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Available online: https://www.cdc.gov/transportationsafety/pedestrian_safety/index.html (accessed on 23 February 2023).
102. Nesoff, E.D.; Milam, A.J.; Pollack, K.M.; Curriero, F.C.; Bowie, J.V.; Gielen, A.C.; Furr-Holden, D.M. Novel Methods for Environmental Assessment of Pedestrian Injury: Creation and Validation of the Inventory for Pedestrian Safety Infrastructure. *J. Urban Health* **2018**, *95*, 208–221. [CrossRef] [PubMed]
103. Liu, Y.-C.; Tung, Y.-C. Risk Analysis of Pedestrians' Road-Crossing Decisions: Effects of Age, Time Gap, Time of Day, and Vehicle Speed. *Saf. Sci.* **2014**, *63*, 77–82. [CrossRef]
104. Koh, P.P.; Wong, Y.D.; Chandrasekar, P. Safety Evaluation of Pedestrian Behaviour and Violations at Signalised Pedestrian Crossings. *Saf. Sci.* **2014**, *70*, 143–152. [CrossRef]
105. Dommes, A.; Cavallo, V. The Role of Perceptual, Cognitive, and Motor Abilities in Street-Crossing Decisions of Young and Older Pedestrians. *Ophthalmic Physiol. Opt.* **2011**, *31*, 292–301. [CrossRef] [PubMed]
106. Harrell, W.A. Factors Influencing Pedestrian Cautiousness in Crossing Streets. *J. Soc. Psychol.* **1991**, *131*, 367–372. [CrossRef]
107. Zheng, L.; Ismail, K.; Meng, X. Traffic Conflict Techniques for Road Safety Analysis: Open Questions and Some Insights. *Can. J. Civ. Eng.* **2014**, *41*, 633–641. [CrossRef]
108. Senserrick, T.; Boufous, S.; De Rome, L.; Ivers, R.; Stevenson, M. Detailed Analysis of Pedestrian Casualty Collisions in Victoria, Australia. *Traffic Inj. Prev.* **2014**, *15*, S197–S205. [CrossRef] [PubMed]
109. Solagberu, B.A.; Balogun, R.A.; Mustafa, I.A.; Ibrahim, N.A.; Oludara, M.A.; Ajani, A.O.; Idowu, O.E.; Osuoji, R.I. Pedestrian Injuries in the Most Densely Populated City in Nigeria—An Epidemic Calling for Control. *Traffic Inj. Prev.* **2015**, *16*, 184–189. [CrossRef] [PubMed]
110. Ogendi, J.; Odero, W.; Mitullah, W.; Khayesi, M. Pattern of Pedestrian Injuries in the City of Nairobi: Implications for Urban Safety Planning. *J. Urban Health* **2013**, *90*, 849–856. [CrossRef] [PubMed]
111. Zegeer, C.V.; Bushell, M. Pedestrian Crash Trends and Potential Countermeasures from around the World. *Accid. Anal. Prev.* **2012**, *44*, 3–11. [CrossRef] [PubMed]
112. Lam, W.W.; Loo, B.P.; Yao, S. Towards Exposure-Based Time-Space Pedestrian Crash Analysis in Facing the Challenges of Ageing Societies in Asia. *Asian Geogr.* **2013**, *30*, 105–125. [CrossRef]
113. Arregui-Dalmases, C.; Lopez-Valdes, F.J.; Segui-Gomez, M. Pedestrian Injuries in Eight European Countries: An Analysis of Hospital Discharge Data. *Accid. Anal. Prev.* **2010**, *42*, 1164–1171. [CrossRef] [PubMed]
114. Geronimo, D.; Lopez, A.M.; Sappa, A.D.; Graf, T. Survey of Pedestrian Detection for Advanced Driver Assistance Systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1239–1258. [CrossRef] [PubMed]
115. City of New York. Vision Zero. Available online: https://www1.nyc.gov/site/visionzero/index.page (accessed on 18 June 2019).
116. Vision Zero Action Plan. City of New York: New York, NY, USA, 2014. Available online: https://www.nyc.gov/html/visionzero/pdf/nyc-vision-zero-action-plan.pdf (accessed on 4 January 2026).
117. Johansson, R. Vision Zero–Implementing a Policy for Traffic Safety. *Saf. Sci.* **2009**, *47*, 826–831. [CrossRef]
118. Golakiya, H.; Chauhan, R.; Santosh Bari, C.; Dhamaniya, A. Pedestrian Safety Analysis at Urban Midblock Section under Mixed Traffic Conditions Using Time to Collision as Surrogate Safety Measure. *Curr. Sci.* **2022**, *123*, 1117–1128. [CrossRef]
119. New York City Department of City Planning. *New York City Pedestrian Level of Service Study: Phase I*; New York City Department of City Planning: New York, NY, USA, 2006.
120. Autey, J.; Sayed, T.; Zaki, M.H. Safety Evaluation of Right-Turn Smart Channels Using Automated Traffic Conflict Analysis. *Accid. Anal. Prev.* **2012**, *45*, 120–130. [CrossRef] [PubMed]
121. Islam, S.M.; Washington, S.; Kim, J.; Haque, M. A Comprehensive Analysis on the Effects of Signal Strategies, Intersection Geometry, and Traffic Operation Factors on Right-Turn Crashes at Signalised Intersections: An Application of Hierarchical Crash Frequency Model. *Accid. Anal. Prev.* **2022**, *171*, 106663. [CrossRef] [PubMed]
122. Harwood, D.W.; Bauer, K.M.; Potts, I.B.; Torbic, D.J.; Richard, K.R.; Rabbani, E.R.; Hauer, E.; Elefteriadou, L.; Griffith, M.S. Safety Effectiveness of Intersection Left-and Right-Turn Lanes. *Transp. Res. Rec.* **2003**, *1840*, 131–139. [CrossRef]
123. Jiang, C.; Qiu, R.; Fu, T.; Fu, L.; Xiong, B.; Lu, Z. Impact of Right-Turn Channelization on Pedestrian Safety at Signalized Intersections. *Accid. Anal. Prev.* **2020**, *136*, 105399. [CrossRef] [PubMed]
124. Kitazawa, K.; Fujiyama, T. Pedestrian Vision and Collision Avoidance Behavior: Investigation of the Information Process Space of Pedestrians Using an Eye Tracker. In *Pedestrian and Evacuation Dynamics 2008*; Klingsch, W.W.F., Rogsch, C., Schadschneider, A., Schreckenberg, M., Eds.; Springer: Berlin, Germany, 2010; pp. 95–108.

125. Cutting, J.E.; Vishton, P.M.; Braren, P.A. How We Avoid Collisions with Stationary and Moving Obstacles. *Psychol. Rev.* **1995**, *102*, 627–651. [CrossRef]
126. Qi, Y.; Yuan, P. Pedestrian Safety at Intersections under Control of Permissive Left-Turn Signal. *Transp. Res. Rec.* **2012**, *2299*, 91–99. [CrossRef]
127. Zhuang, X.; Wu, C. The Safety Margin and Perceived Safety of Pedestrians at Unmarked Roadway. *Transp. Res. Part F Traffic Psychol. Behav.* **2012**, *15*, 119–131. [CrossRef]
128. Li, Y.; Fernie, G. Pedestrian Behavior and Safety on a Two-Stage Crossing with a Center Refuge Island and the Effect of Winter Weather on Pedestrian Compliance Rate. *Accid. Anal. Prev.* **2010**, *42*, 1156–1163. [CrossRef] [PubMed]
129. de Lavalette, B.C.; Tijus, C.; Poitrenaud, S.; Leproux, C.; Bergeron, J.; Thouez, J.-P. Pedestrian Crossing Decision-Making: A Situational and Behavioral Approach. *Saf. Sci.* **2009**, *47*, 1248–1253. [CrossRef]
130. Sarker, M.S.; Carsten, O.; Huang, Y.; Hajiseyedjavadi, F. Promoting Pedestrian Safety in Bangladesh: Identifying Factors for Drivers' Yielding Behavior at Designated Crossings Using Behavior Change Theories. *Traffic Inj. Prev.* **2024**, *25*, 976–985. [CrossRef] [PubMed]
131. Kang, B. Identifying Street Design Elements Associated with Vehicle-to-Pedestrian Collision Reduction at Intersections in New York City. *Accid. Anal. Prev.* **2019**, *122*, 308–317. [CrossRef] [PubMed]
132. Jordan, T.; Raubal, M.; Gartrell, B.; Egenhofer, M. An Affordance-Based Model of Place in Gis. In *Proceedings of the Eighth International Symposium on Spatial Data Handling*; Poiker, T.K., Chrisman, N.R., Eds.; International Geographical Union: Vancouver, Kanada, 1998; pp. 98–109.
133. Meilinger, T.; Vosgerau, G. Putting Egocentric and Allocentric into Perspective. In *Spatial Cognition Vii (Lecture Notes in Computer Science*; Hölscher, C., Shipley, T.F., Olivetti Belardinelli, M., Bateman, J.A., Newcombe, N.S., Eds.; Springer: Berlin, Germany, 2010; Volume 6222, pp. 207–221.
134. Plank, M.; Müller, H.J.; Onton, J.; Makeig, S.; Gramann, K. Human Eeg Correlates of Spatial Navigation within Egocentric and Allocentric Reference Frames. In *Spatial Cognition Vii. Spatial Cognition 2010. Lecture Notes in Computer Science 6222*; Hölscher, C., Shipley, T.F., Belardinelli, M.O., Bateman, J.A., Newcombe, N.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 191–206.
135. Hatfield, J.; Fernandes, R.; Soames Job, R.F.; Smith, K. Misunderstanding of Right-of-Way Rules at Various Pedestrian Crossing Types: Observational Study and Survey. *Accid. Anal. Prev.* **2007**, *39*, 833–842. [CrossRef] [PubMed]
136. Dey, D.; Terken, J. Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 24–27 September 2017*; Politis, I., Large, D., Eds.; Association for Computing Machinery: New York, NY, USA, 2017; pp. 109–113.
137. Tian, K.; Wei, C.; Lyu, W.; Wang, Y.; Lee, Y.M.; Merat, N.; Romano, R.; Markkula, G. Interacting with Yielding Vehicles: A Perceptually Plausible Model for Pedestrian Road Crossing Decisions. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 9654–9668. [CrossRef]
138. Wang, R.F.; Cutting, J.E. Where We Go with a Little Good Information. *Psychol. Sci.* **1999**, *10*, 71–75. [CrossRef]
139. Kumar, A.; Ghosh, I. Analysis of Spontaneous Order of Pedestrian–Vehicle Conflicts at Signalized Intersections. *Transp. Res. Rec.* **2020**, *2674*, 440–454. [CrossRef]
140. Chen, W.; Zhuang, X.; Cui, Z.; Ma, G. Drivers' Recognition of Pedestrian Road-Crossing Intentions: Performance and Process. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *64*, 552–564. [CrossRef]
141. Onkhar, V.; Bazilinskyy, P.; Dodou, D.; De Winter, J.C.F. The Effect of Drivers' Eye Contact on Pedestrians' Perceived Safety. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *84*, 194–210. [CrossRef]
142. Furth, P.G.; Koonce, P.J.; Miao, Y.; Peng, F.; Littman, M. Mitigating Right-Turn Conflict with Protected yet Concurrent Phasing for Cycle Track and Pedestrian Crossings. *Transp. Res. Rec.* **2014**, *2438*, 81–88. [CrossRef]
143. Subedi, A.; Vahedi Saheli, M.; A Singleton, P.; Gaither, A.; Mekker, M. Examining the Interplay of Pedestrian and Right-Turning Driver Behaviors in Traffic Conflict Severity. *Transp. Res. Rec.* **2026**, *2680*, 218–236. [CrossRef]
144. Xin, F.; Wang, X.; Sun, C. Risk Evaluation for Conflicts between Crossing Pedestrians and Right-Turning Vehicles at Intersections. *Transp. Res. Rec.* **2021**, *2675*, 1005–1014. [CrossRef]
145. Van Houten, R.; Retting, R.A.; Farmer, C.M.; Van Houten, J. Field Evaluation of a Leading Pedestrian Interval Signal Phase at Three Urban Intersections. *Transp. Res. Rec.* **2000**, *1734*, 86–92. [CrossRef]
146. Crowley-Koch, B.J.; Van Houten, R.; Lim, E. Effects of Pedestrian Prompts on Motorist Yielding at Crosswalks. *J. Appl. Behav. Anal.* **2011**, *44*, 121–126. [CrossRef] [PubMed]
147. Schmidt, S.; Färber, B. Pedestrians at the Kerb–Recognising the Action Intentions of Humans. *Transp. Res. Part F Traffic Psychol. Behav.* **2009**, *12*, 300–310. [CrossRef]
148. Recarte, M.A.; Nunes, L.M. Perception of Speed in an Automobile: Estimation and Production. *J. Exp. Psychol. Appl.* **1996**, *2*, 291. [CrossRef]
149. Doleschal, F.; Badel, G.-T.; Verhey, J.L. Rumbling, Humming, Booming–Perception of Vehicle Interior Noise at Low Engine Speeds. *Appl. Acoust.* **2022**, *197*, 108915. [CrossRef]

150. Lacherez, P.; Turner, L.; Lester, R.; Burns, Z.; Wood, J.M. Age-Related Changes in Perception of Movement in Driving Scenes. *Ophthalmic Physiol. Opt.* **2014**, *34*, 445–451. [CrossRef] [PubMed]
151. Ezzati Amini, R.; Al Haddad, C.; Batabyal, D.; Gkena, I.; De Vos, B.; Cuenen, A.; Brijs, T.; Antoniou, C. Driver Distraction and in-Vehicle Interventions: A Driving Simulator Study on Visual Attention and Driving Performance. *Accid. Anal. Prev.* **2023**, *191*, 107195. [CrossRef] [PubMed]
152. Bourquin, E.A.; Wall Emerson, R.; Sauerburger, D.; Barlow, J. Conditions That Influence Drivers' Yielding Behaviour: Effects of Pedestrian Gaze and Head Movements. *Int. J. Orientat. Mobil.* **2016**, *8*, 13–26. [CrossRef]
153. Torrens, P.M.; Kim, R. Using Immersive Virtual Reality to Study Road-Crossing Sustainability in Fleeting Moments of Space and Time. *Sustainability* **2024**, *16*, 1327. [CrossRef]
154. Warren, J. Visually Controlled Locomotion: 40 Years Later. *Ecol. Psychol.* **1998**, *10*, 177–219.
155. Patla, A.E.; Vickers, J.N. How Far Ahead Do We Look When Required to Step on Specific Locations in the Travel Path during Locomotion? *Exp. Brain Res.* **2003**, *148*, 133–138. [CrossRef] [PubMed]
156. Ratti, C. Space Syntax: Some Inconsistencies. *Environ. Plan. B Plan. Des.* **2004**, *31*, 487–499. [CrossRef]
157. Pfeffer, K.; Hunter, E. The Effects of Peer Influence on Adolescent Pedestrian Road-Crossing Decisions. *Traffic Inj. Prev.* **2013**, *14*, 434–440. [CrossRef] [PubMed]
158. Torrens, P.M. Agent Models of Customer Journeys on Retail High Streets. *J. Econ. Interact. Coord.* **2023**, *18*, 87–128. [CrossRef] [PubMed]
159. Torrens, P.M.; Griffin, W.A. Exploring the Micro-Social Geography of Children's Interactions in Preschool: A Long-Term Observational Study and Analysis Using Geographic Information Technologies. *Environ. Behav.* **2013**, *45*, 584–614.
160. Dommes, A.; Cavallo, V.; Dubuisson, J.-B.; Tournier, I.; Vienne, F. Crossing a Two-Way Street: Comparison of Young and Old Pedestrians. *J. Saf. Res.* **2014**, *50*, 27–34. [CrossRef] [PubMed]
161. Zeedyk, M.S.; Wallace, L.; Spry, L. Stop, Look, Listen, and Think?: What Young Children Really Do When Crossing the Road. *Accid. Anal. Prev.* **2002**, *34*, 43–50. [PubMed]
162. Krizek, K.J. Lifestyles, Residential Location Decisions, and Pedestrian and Transit Activity. *Transp. Res. Rec.* **1981**, *2006*, 171–178.
163. Lipovac, K.; Vujanic, M.; Maric, B.; Nesic, M. The Influence of a Pedestrian Countdown Display on Pedestrian Behavior at Signalized Pedestrian Crossings. *Transp. Res. Part F Traffic Psychol. Behav.* **2013**, *20*, 121–134. [CrossRef]
164. Onelcin, P.; Alver, Y. Illegal Crossing Behavior of Pedestrians at Signalized Intersections: Factors Affecting the Gap Acceptance. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *31*, 124–132. [CrossRef]
165. Krizek, K.J.; Roland, R.W. What Is at the End of the Road? Understanding Discontinuities of on-Street Bicycle Lanes in Urban Settings. *Transp. Res. Part D Transp. Environ.* **2005**, *10*, 55–68. [CrossRef]
166. LeCun, Y.; Bengio, Y. Convolutional Networks for Images, Speech, and Time Series. In *The Handbook of Brain Theory and Neural Networks*; Arbib, M.A., Ed.; MIT Press: Cambridge, MA, USA, 1995; pp. 276–285.
167. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), Las Vegas, NV, USA, 26 June–1 July 2016*; Li, F.-F., Bajcsy, R., Tuytelaars, T., Eds.; IEEE: Los Alamitos, CA, USA, 2016; pp. 779–788.
168. Jocher, G.; Qiu, J.; Liu, M.; Lyu, S.; Akyon, F.C.; Kalfaoglu, M.E. Ultralytics Yolo26: Unified Real-Time End-to-End Vision Models. *arXiv* **2026**, arXiv:2606.03748.
169. Shao, S.; Li, Z.; Zhang, T.; Peng, C.; Yu, G.; Zhang, X.; Li, J.; Sun, J. Objects365: A Large-Scale, High-Quality Dataset for Object Detection. In *Proceedings of the 2019 IEEE/Cvf International Conference on Computer Vision (Iccv), Seoul, South Korea, 27 October–2 November 2019*; Lee, K.M., Forsyth, D., Pollefeys, M., Tang, X., Eds.; IEEE: Los Alamitos, CA, USA, 2019; pp. 8429–8438.
170. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft Coco: Common Objects in Context. In *13th European Conference, Zurich, Switzerland, 6–12 September 2014, Proceedings, Part V*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755.
171. Ultralytics. Ultralytics Yolo 26n (21 August 2026). 2026. Available online: https://docs.ultralytics.com/models/yolo26 (accessed on 2 January 2026).
172. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple Online and Realtime Tracking. In *Proceedings of the 2016 IEEE International Conference on Image Processing (Icip), Phoenix, AZ, USA, 25–28 September 2016*; Pereira, F., Sharma, G., Eds.; IEEE: Los Alamitos, CA, USA, 2016; pp. 3464–3468.
173. Aharon, N.; Orfaig, R.; Bobrovsky, B.-Z. Bot-Sort: Robust Associations Multi-Pedestrian Tracking. *arXiv* **2022**, arXiv:2206.14651.
174. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]
175. Goldberger, J.; Tassa, T. A Hierarchical Clustering Algorithm Based on the Hungarian Method. *Pattern Recognit. Lett.* **2008**, *29*, 1632–1638. [CrossRef]
176. Kuhn, H.W. The Hungarian Method for the Assignment Problem. *Nav. Res. Logist. Q.* **1955**, *2*, 83–97. [CrossRef]
177. Tomar, S. Converting Video Formats with Ffmpeg. *Linux J.* **2006**, 10.
178. Bradski, G.; Kaehler, A. *Learning Opencv: Computer Vision with the Opencv Library*; O'Reilly: Sebastapool, CA, USA, 2008.

179. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
180. Virtanen, P.; Gommers, R.; E Oliphant, T.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. Scipy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
181. Luma AI, Inc. *Luma 3d Capture Mobile App*; Luma AI, Inc.: Palo Alto, CA, USA, 2026.
182. Luma AI Inc. Lumaapi-Python (1 October 2024). 2024. Available online: https://github.com/lumalabs/lumaapi-python (accessed on 1 October 2024).
183. Luma Web Examples. Available online: https://github.com/lumalabs/luma-web-examples (accessed on 1 October 2025).
184. Goesele, M.; Snavely, N.; Curless, B.; Hoppe, H.; Seitz, S.M. Multi-View Stereo for Community Photo Collections. In *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision (Iccv 2007)*; Davis, L., Bouthemy, P., Ikeuchi, K., Eds.; IEEE: Rio de Janeiro, Brazil, 2007; pp. 1–8.
185. Kerbl, B.; Kopanas, G.; Leimkuehler, T.; Drettakis, G. 3d Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* **2023**, *42*, 139. [CrossRef]
186. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision–ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 405–421.
187. Li, Q.; Torrens, F.; Chen, K.; Sun, Q. Blendfusion: Procedural 3d Texturing Assistant with View-Consistent Generative Models. In *I3d Companion '25: Companion Proceedings of the Acm Siggraph Symposium on Interactive 3d Graphics and Games*; Billeter, M.M., Watson, B., Marroquim, R., Reznikov, L., Sin, Z., Vives, M., Eds.; Association for Computing Machinery: New York, NY, USA, 2025.
188. Tafazzoli, F.; Frigui, H.; Nishiyama, K. A Large and Diverse Dataset for Improved Vehicle Make and Model Recognition. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (Cvprw), Honolulu, HI, USA, 21–26 July 2017*; Chellappa, R., Hoogs, A., Zhang, Z., Eds.; IEEE: Los Alamitos, CA, USA, 2017; pp. 874–881.
189. Wang, X.; Zheng, S.; Yang, R.; Zheng, A.; Chen, Z.; Tang, J.; Luo, B. Pedestrian Attribute Recognition: A Survey. *Pattern Recognit.* **2022**, *121*, 108220. [CrossRef]
190. Gleicher, M. Retargetting Motion to New Characters. In *Proceedings of Siggraph 98: The 25th Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, 19–24 July 1998*; Yang, C.C., Murali, T.M., Eds.; Association for Computing Machinery: New York, NY, USA, 1998; pp. 33–42.
191. Kim, R.; Torrens, P.M. Building Verisimilitude in Vr with High-Fidelity Local Action Models: A Demonstration Supporting Road-Crossing Experiments. In *Proceedings of the 38th Acm Sigsim Conference on Principles of Advanced Discrete Simulation, Atlanta, GA, USA, 24–26 June 2024*; Loper, M., Pellegrini, A., Eds.; Association for Computing Machinery: New York, NY, USA, 2024; pp. 119–130.
192. Torrens, P.M. Moving Agent Pedestrians through Space and Time. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 35–66. [CrossRef]
193. Hägerstrand, T. Diorama, Path and Project. *Tijdschr. Voor Econ. En. Soc. Geogr.* **1982**, *73*, 323–339. [CrossRef]
194. Torrens, P.M.; Kim, R. Sidewalk2synth: Generating Synthetic Embodied Locomotion from Real-World Streetscapes. *Urban Inform.* **2025**, *4*, 16. [CrossRef]
195. Nams, V.O. The Vfractal: A New Estimator for Fractal Dimension of Animal Movement Paths. *Landsc. Ecol.* **1996**, *11*, 289–297. [CrossRef]
196. Hall, E.T. A System for the Notation of Proxemic Behavior. *Am. Anthropol.* **1963**, *65*, 1003–1026. [CrossRef]
197. Harrigan, J.A. Proxemics, Kinesics, and Gaze. In *The New Handbook of Methods in Nonverbal Behavior Research*; Harrigan, J.A., Rosenthal, R., Scherer, K., Eds.; Oxford University Press: New York, NY, USA, 2005; pp. 137–198.
198. Cook, M. Experiments on Orientation and Proxemics. *Hum. Relat.* **1970**, *23*, 61–76. [CrossRef]
199. Summala, H. Driver/Vehicle Steering Response Latencies. *Hum. Factors* **1981**, *23*, 683–692. [CrossRef]
200. Caird, J.K.; Hancock, P.A. The Perception of Arrival Time for Different Oncoming Vehicles at an Intersection. *Ecol. Psychol.* **1994**, *6*, 83–109. [CrossRef] [PubMed]
201. Ishaque, M.M.; Noland, R.B. Behavioural Issues in Pedestrian Speed Choice and Street Crossing Behaviour: A Review. *Transp. Rev.* **2008**, *28*, 61–85. [CrossRef]
202. Kadali, B.R.; Vedagiri, P. Effect of Vehicular Lanes on Pedestrian Gap Acceptance Behaviour. *Procedia-Soc. Behav. Sci.* **2013**, *104*, 678–687. [CrossRef]
203. Clubsandwich 0.5.1: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. Comprehensive R Archive Network (CRAN). Available online: https://jepusto.github.io/clubSandwich/ (accessed on 11 January 2026).
204. Esarey, J.; Menger, A. Practical and Effective Approaches to Dealing with Clustered Data. *Political Sci. Res. Methods* **2019**, *7*, 541–559. [CrossRef]

205. Sueur, C.; Class, B.; Hamm, C.; Meyer, X.; Pelé, M. Different Risk Thresholds in Pedestrian Road Crossing Behaviour: A Comparison of French and Japanese Approaches. *Accid. Anal. Prev.* **2013**, *58*, 59–63. [CrossRef] [PubMed]
206. Himanen, V.; Kulmala, R. An Application of Logit Models in Analysing the Behaviour of Pedestrians and Car Drivers on Pedestrian Crossings. *Accid. Anal. Prev.* **1988**, *20*, 187–197. [CrossRef] [PubMed]
207. Faria, J.J.; Krause, S.; Krause, J. Collective Behavior in Road Crossing Pedestrians: The Role of Social Information. *Behav. Ecol.* **2010**, *21*, 1236–1242. [CrossRef]
208. Ni, Y.; Wang, M.; Sun, J.; Li, K. Evaluation of Pedestrian Safety at Intersections: A Theoretical Framework Based on Pedestrian-Vehicle Interaction Patterns. *Accid. Anal. Prev.* **2016**, *96*, 118–129. [CrossRef] [PubMed]
209. Farley, A.; Zand, M.; Greenspan, M. Diffusion Dataset Generation: Towards Closing the Sim2real Gap for Pedestrian Detection. In *Proceedings of the 20th Conference on Computer and Robot Vision (Crv), Montreal, QC, Canada, 6–8 June 2023*; Li, K., Lin, H.-C., Eds.; IEEE: Los Alamitos, CA, USA, 2023; pp. 169–176.
210. Pasios, S.; Nikolaidis, N. Carla2real: A Tool for Reducing the Sim2real Gap in Carla Simulator. *arXiv* **2024**, arXiv:2410.18238.
211. Vyas, H.; Torrens, P.M. Tackling "Leaky" Facial and Object Recognition Operations over Federated Learning. In *Proceedings of the 2024 4th Interdisciplinary Conference on Electrics and Computer (Intcec)*; Bizon, N., Lopez-Guede, J.M., Selami, B., Eds.; IEEE: Los Alamitos, CA, USA, 2024; pp. 1–5.
212. Potdar, K.; Torrens, P.M. Modelling Spatio-Temporal Patterns in Pedestrian Behavior at the Edge with Jetson Soms. In Proceedings of the NVIDIA GPU Technology Conference (GTC), Washington, DC, USA, 4–6 November 2019; DCP91291.
213. Johansson, A.; Helbing, D.; Al-Abideen, H.; Al-Bosta, S. From Crowd Dynamics to Crowd Safety: A Video-Based Analysis. *Adv. Complex Syst.* **2008**, *11*, 497–527. [CrossRef]
214. Meier, B.P.; Schnall, S.; Schwarz, N.; A Bargh, J. Embodiment in Social Psychology. *Top. Cogn. Sci.* **2012**, *4*, 705–716. [CrossRef] [PubMed]
215. Torrens, P.M. Data Science for Pedestrian and High Street Retailing as a Framework for Advancing Urban Informatics to Individual Scales. *Urban Inform.* **2022**, *1*, 9. [CrossRef] [PubMed]
216. Torrens, P.M. Agent-Based Modeling and the Spatial Sciences. *Geogr. Compass* **2010**, *4*, 428–448. [CrossRef]
217. Amato, G.; Ciampi, L.; Falchi, F.; Gennaro, C.; Messina, N. Learning Pedestrian Detection from Virtual Worlds. In *Image Analysis and Processing–Iciap 2019. 20th International Conference, Trento, Italy, 9–13 September 2019, Proceedings, Part I*; Ricci, E., Bulò, S.R., Snoek, C., Lanz, O., Messelodi, S., Sebe, N., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 302–312.
218. Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V. Carla: An Open Urban Driving Simulator. *arXiv* **2017**, arXiv:1711.03938.
219. Maathuis, C.; Anca Cidota, M.; Datcu, D.; Marin, L. Integrating Explainable Artificial Intelligence in Extended Reality Environments: A Systematic Survey. *Mathematics* **2025**, *13*, 290. [CrossRef]
220. Greengard, S. Ai on Edge. *Commun. ACM* **2020**, *63*, 18–20. [CrossRef]
221. Underwood, J.; Dillon, G.; Farnsworth, B.; Twiner, A. Reading the Road: The Influence of Age and Sex on Child Pedestrians' Perceptions of Road Risk. *Br. J. Psychol.* **2007**, *98*, 93–110. [CrossRef] [PubMed]
222. Oxley, J.; Charlton, J.; Fildes, B. *The Effect of Cognitive Impairment on Older Pedestrian Behaviour and Crash Risk*; Monash University Accident Research Centre: Melbourne, Australia, 2005.
223. Zivotofsky, A.Z.; Eldror, E.; Mandel, R.; Rosenbloom, T. Misjudging Their Own Steps: Why Elderly People Have Trouble Crossing the Road. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2012**, *54*, 600–607.
224. Heasley, K.; Buckley, J.G.; Scally, A.; Twigg, P.; Elliott, D.B. Stepping up to a New Level: Effects of Blurring Vision in the Elderly. *Investig. Ophthalmol. Vis. Sci.* **2004**, *45*, 2122–2128. [CrossRef] [PubMed]
225. Holland, C.; Hill, R. The Effect of Age, Gender and Driver Status on Pedestrians' Intentions to Cross the Road in Risky Situations. *Accid. Anal. Prev.* **2007**, *39*, 224–237. [CrossRef] [PubMed]
226. Meir, A.; Oron-Gilad, T.; Parmet, Y. Can Child-Pedestrians' Hazard Perception Skills Be Enhanced? *Accid. Anal. Prev.* **2015**, *83*, 101–110. [CrossRef] [PubMed]
227. Zeedyk, M.S.; Kelly, L. Behavioural Observations of Adult–Child Pairs at Pedestrian Crossings. *Accid. Anal. Prev.* **2003**, *35*, 771–776. [CrossRef] [PubMed]
228. Wolff, M. Notes on the Behavior of Pedestrians. In *People in Places: The Sociology of the Familiar*; Birenbaum, A., Sagarin, E., Eds.; Praeger: New York, NY, USA, 1973.
229. Langlois, J.A.; Keyl, P.M.; Guralnik, J.M.; Foley, D.J.; Marottoli, R.A.; Wallace, R.B. Characteristics of Older Pedestrians Who Have Difficulty Crossing the Street. *Am. J. Public Health* **1997**, *87*, 393–397. [CrossRef] [PubMed]
230. Winter, D.A.; Patla, A.E.; Frank, J.S.; Walt, S.E. Biomechanical Walking Pattern Changes in the Fit and Healthy Elderly. *Phys. Ther.* **1990**, *70*, 340–347. [CrossRef] [PubMed]
231. Tapiro, H.; Oron-Gilad, T.; Parmet, Y. Cell Phone Conversations and Child Pedestrian's Crossing Behavior; a Simulator Study. *Saf. Sci.* **2016**, *89*, 36–44. [CrossRef]
232. Zacharias, J. Pedestrian Behavior and Perception in Urban Walking Environments. *J. Plan. Lit.* **2001**, *16*, 3–18. [CrossRef]

233. Hill, M.R. *Walking, Crossing Streets, and Choosing Pedestrian Routes*; University of Nebraska Press: Lincoln, Nebraska, 1984.
234. Fruin, J.J. *Pedestrian Planning and Design*; Metropolitan Association of Urban Designers and Environmental Planners, Inc.: New York, NY, USA, 1971.

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