Torrens and Kim *Urban Informatics* (2025) 4:16
https://doi.org/10.1007/s44212-025-00081-z

Urban Informatics
The International Society for Urban Informatics

RESEARCH **Open Access**

Sidewalk2Synth: generating synthetic embodied locomotion from real-world streetscapes

Paul M. Torrens[1*] and Ryan Kim[1]

Abstract
We demonstrate methods for generating synthetic scenarios of embodied locomotion on sidewalks, drawn from detailed observations of situation and of context on real-world streetscapes. We show that, through observation and sensing, quite rich and high-resolution data can be gleaned from small and fleeting windows on pedestrian locomotion as it unfolds in lived spaces. These insights can provide valuable explanations of how pedestrians experience and embody encounters in physical and social context at localized and individualized scales of space and time. Using agent AI, we demonstrate that this knowledge can be transferred into high fidelity models, capturing the essence of embodied locomotion and providing a basis for experimentation with what-if scenario as simulation. By implementing simulations as virtual reality media, we showcase an end-to-end experimental pipeline that allows real human participants to embody themselves in synthetic sidewalks, directly using their innate and tangible perception and locomotion. Our approach establishes a new pliability between real and synthetic embodied locomotion, which we argue can provide experimental maneuverability relative to ordinary questions, as well as to extraordinary scenarios that are challenging to examine on the ground. Sidewalk2Synth could also help to circumnavigate existing challenges in machine learning around training-based approaches that lack robust empirical evidence of priors and that are otherwise resistant to generalization outside specific places and times.

Keywords Embodiment, Artificial intelligence, Agents, Autonomous machine intelligence, Locomotion, Virtual reality, Streetscape

1 Introduction

"That's not really me in there/I would never do that" (Nine Inch Nails, 2016).

Streetscapes are an important sub-space of the built environment, formed among the interstitial geographies that take hold between building façades, sidewalks, curbs, and roadways. They are often teeming with intermingled activity that people routinely spill into, and course over. Pedestrians come to streetscapes from their homes or from establishments that they have visited, while others commingle with them to load and unload passengers and goods onto transport systems and in and out of businesses (Mishra et al., 2015; Sarker et al., 2015). A variety of workers rely on streetscapes as a production environment, including public works crews (Loukaitou-Sideris & Ehrenfeucht, 2011), crossing guards (Gutierrez et al., 2014), and salespeople (Bhargava & Donthu, 1999), who must take care to read the shifting dynamics of streetscapes as a servicescape. Many streetscapes carry specific meaning—as places (Tuan, 1975, 1979)—with cultural associations (Batty et al., 2003), with historical significance (de Valera, 1986), and with persistent

*Correspondence:
Paul M. Torrens
torrens@nyu.edu
[1] Department of Computer Science and Engineering and Center for Urban Science + Progress, Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA

character that can shape them into attraction sites for tourism (Chen et al., 2024; Urry, 2002).

Explaining the interplay among all these factors, collated across the perspectives of many differently motivated and acting individuals, each played out against the broad milieu of streetscape settings and scenarios presents significant overlapping complexity in interpretation. It is therefore understandable to investigate streetscape dynamics for commonalities, which could form a starting point for understanding. Streetscapes that are conventionally considered to be successful are often regarded as both accessible to locomotion and generous in the overlapping opportunities for encounter that they afford their visitors (Foltête & Piombini, 2007). Both considerations of how pedestrians build entanglement to their surroundings—with movement and engagement—can usefully be converged in the concept of "embodiment". At its core, embodiment is concerned with how one's behavior is physically enacted relative to encountered input (Kiverstein, 2012; Ziemke, 2013). Our consideration of embodiment is focused on two special cases. First, we investigate streetscape embodiment, wherein behavior is brought to life in the physical and human environments that streetscapes host. Second, we examine embodied locomotion, as enacted traversal through streetscapes and the encounters that happen along the way.

The high-level purpose of the paper is to show how one might simulate synthetic embodiment, rather directly, from locomotion context as it plays out on real streetscapes. We chase this goal with three aims. First, we endeavor to show the intimate connection between pedestrians' locomotion along streetscapes and their embodiment to the streetscape. Studying this connection, we reason, can assist in building knowledge of behavioral geography (Golledge & Stimson, 1997). Second, we present the case that the current generations of data products that could inform understanding of embodied locomotion could be very usefully supplemented with fresh insight that can come from computing (Torrens, 2016a, 2018a, 2022a). Here, we examine two main pathways for computer science to contribute novel actionable knowledge: automated observation of real embodied pedestrian locomotion, and simulation-assisted exploration of what-if factors that can be advanced in virtual reality. Third, the details of getting this to work suggest several promising lines of academic inquiry for the community, including empirical examination of whether behavior might be robustly implied from observation of encounters (Johansson et al., 2008), how immersive sensing technologies could be adapted to capture pedestrian encounters in very small and fleeting windows of space and time (Camara et al., 2020), and whether extended reality media can act as reliable experimental environments (Chen et al., 2023; Çöltekin et al., 2020). We endeavor to show that these three aims are indeed actionable through a unified pipeline that moves from observation, through computational analysis, via models, into simulation, and codified as empirical outputs. We refer this pipeline as "Sidewalk2Synth", as a portmanteau of the two boundaries of the problem space, beginning in the real world and ending in a synthetic, research-tinged, synthetic approximation of that reality, with significant flexibility to bounce back and forth with questioning between both. Bridging gaps between what is real and what might be usefully synthetic raises questions of authenticity, and so we also discuss how one might validate embodied locomotion in our pipeline.

Briefly, the paper is organized as follows. First, we provide background material to motivate the work and to set it in the context of existing scholarship. Second, we describe an observational instrument to obtain ground truth data from real world locomotion embodiment in outdoor settings, chiefly from pedestrian encounters in busy urban streetscape scenes. This involves both first-person and third-person observation with qualitative behavioral coding as well as sensor-based measurements. Third, we introduce a live experimental instrument for collecting motion capture data from real people in a studio setting, mocked-up to mimic the observed real-world scenes. Fourth, we detail a computational model designed to facilitate representation of the observations, qualitative data, and empirical measurements in a synthetic and simulation-based representation of the observed scenes, for which we have many (highly-finesse) dimensions of scale and experimental control (Fig. 1). This simulation is based on relatively strong agents (Franklin & Graesser, 1997), i.e., decision-making and situationally driven finite state machines that are developed for a match to real-world and theoretical behaviors (Torrens, 2010b). Fifth, we illustrate a live experimental scheme to immerse real human users directly in the simulation via virtual reality immersion with paired real-time telematics of user states. We show that this virtual instrument can be used to generate locomotion and embodiment scenarios for which we would not otherwise have experimental control in real-world settings (particularly those involving close encounters with fast-moving vehicles). Sixth, we describe a series of tests, imposed on the model environment, on user behavior, and on pairings between the two. The aim of these tests is to assess whether Sidewalk2Synth can faithfully reveal realistic user behaviors. Our tests show that this is the case, and on this basis, we examine some preliminary use-case scenarios, based around examining social embodiment (to peer and group effects) in road-crossing behavior. We conclude with some commentary

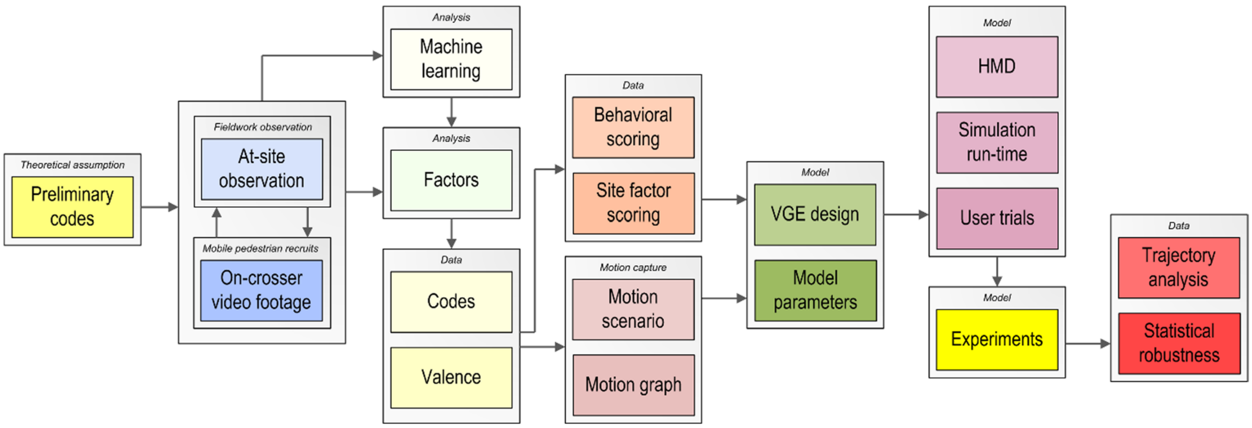


Fig. 1 The main steps in developing the Sidewalk2Synth pipeline flit between theoretical concept-development, fieldwork to test those concepts in natural settings, data science to add value to observational data, experiments to add new data by motion capture, development of virtual and agent models, run-time development for virtual reality hardware, simulation scenarios with recruited human subjects, data output, and trajectory analysis

on future work and the implied implementations of our experimental instrument for locomotion science. The general flow of activity in developing Sidewalk2Synth is illustrated in, and details are provided throughout the remaining text of the paper.

2 Background

Typically, embodiment evokes physical considerations of how people connect, through touch and sensation, to things around them. By extension, one can straightforwardly consider how people might project that physicality and alter their physical behavior to avail of tangible connections or to avoid a contact. Similarly, one might embody their interactions to their surroundings to signal a physical action, through choreography or non-verbal communication, for example (Andersen, 2008). There are a variety of disciplinary considerations of embodiment that are pursued in academic discourse (Simonsen, 2013). In embodiment, sensation and sensing are generally framed as being ego-centric, in that embodiment can come from an individual's interpretation of different dimensions of streetscape character (Crouch, 2000). Ideas about *environmental embodiment* are well established, and generally center around the idea of affordance: what one's physical embodiment enables you to do and avail of (Gibson, 1950, 1966, 1979). In some ways, we might usefully consider this as opportunistic embodiment, raising questions of what circumstances establish opportunities for embodiment, and when they present how those opportunities might connect with a pedestrian's behavior. This is a productive strain of reasoning because locomotion brings pedestrians into ever-shifting opportunities as encounters. Increasingly, embodiment is also considered as a property of psychology, especially as a factor in cognition (*embodied cognition*) (Adams, 2010; Anderson, 2003; Clark, 2008). Sub-themes include the role of embodiment in shaping emotional behavior (Michalak et al., 2009) and attachment, for example (Davidson & Milligan, 2004). Embodiment can also be interpreted through the lens of social psychology (*social embodiment*) (Goldman & de Vignemont, 2009; Lindblom, 2015; Meier et al., 2012) and sociality more generally (Niedenthal et al., 2005). The material conditions that convey or afford embodiment may be flexibly considered, as connections in a cyberspace (Dodge & Kitchin, 2005), transactions with media (Krishna & Schwarz, 2014), or even encounters that are curated formally through a defined user experience (Dourish, 2001) or as components of computing (Schick & Malmborg, 2010). Generally, these come under the theme of *embodied interaction*. Within each of these embodiment disciplines, one can then consider particularly typologies of embodiment. This is very well-covered by Ziemke (2013), who outlines six approaches to typifying embodiment. For the purposes of this paper, we deal with three of them. We entertain the notions of "physical embodiment" as a tangible form of lived interaction; of "historical embodiment" as the accumulation of knowledge, skill, mannerisms, habits, and norms developed from prior interactions; and of "social embodiment" as embodiment with social sensation.

Here, we introduce our own approach to embodiment, which retains connections to each of the Ziemke (2013) concepts. We reference this as *embodied locomotion,* and we consider it with application to the domain of streetscape science. Locomotion and embodiment are closely coupled in streetscape encounters (Middleton, 2010; Shaw, 2015). We consider that when people move in busy physical settings or in crowded social conditions, their embodiment shifts rapidly, flitting from one encounter to the next through dynamics of action, reaction, interaction, transaction, and even proaction. Examining the "when?", "where?", and "with-whom?" of these processes these shifts can potentially be incredibly useful in explaining potential reasons "why?". Revealing empirical properties of embodied locomotion to settle such questions is a challenge with many dimensions of consideration. First, embodiment is a highly personal phenomenon (Cresswell, 1999). Given the considerable number of people and events that pulse through a streetscape, individuality of embodiment invokes the law of requisite variety (Ashby, 1958) at burdensome scale. Second, embodiment, especially while moving, can unfold over very fragile, delicate, and fast-moving pockets of space and time, as fleeting encounters that are transitory even if they are significant (Crouch, 2000). Third, the context of embodiment can be difficult to generalize from one setting to another (Middleton, 2010). Consider, for example, how you would embody yourself on a walk that is routine, as compared to a new streetscape that you might visit as a tourist. Fourth, embodied phenomena easily assume properties of complex adaptive systems (Torrens, 2010a, 2015b), with all of the thorny issues of non-linearity that usually entails, and that can easily eschew tractable codification in scientific inquiry.

Given these challenges, models of embodied locomotion are a logical choice for experimenting with questions of "when?", "where?", "with-whom?", and "why?" But, existing model approaches generally take on coarse representations that are ill-suited in extension to microscales beyond their original design (Torrens, 2014a, 2014b). This relates to longstanding discord between top-down models and bottom-up models (Torrens & Nara, 2013; Torrens et al., 2013) on points of ecological fallacy (Openshaw, 1984; Wrigley et al., 1996), modifiable areal unit problems (Openshaw, 1983), and laws of requisite variety (Ashby, 1958). New theoretical ideas—chiefly Non-Representational Theory (NRT) (Thrift, 2008), which advocates for examination of how geographies and other disciplinary realities are generated through encounter, rather than reacting to representational formalisms—have emerged to tackle some of the conceptual challenges of framing embodiment, but they are overwhelmingly conceptual in exposition (Torrens, 2024). Nonetheless, many of these concepts are actionable as models, particularly those from early success in advancing NRT ideas of mobile embodiment (the so termed "mobilities turn" in geography and sociology) (Edensor, 2012; Sheller, 2017; Sheller & Urry, 2006). Matching model support could be very helpful in buttressing this conceptual work, and this is partially what we aim to present in this paper.

Inevitably, questions of data arise when discussing modeling embodiment. Unlike trip-type mobilities research (He et al., 2015; Hong et al., 2017; Krumm & Horvitz, 2007), which can poll from large archives of discrete location-based check-in data and activity tags (Liao et al., 2024), embodied locomotion requires, in essence, that the continuum of embodied behavior be sampled and that those samples cover encounters at parity with lived experiences. By extension, embodied locomotion has more to do with affective computing than data science (Clough, 2008; Griffin et al., 2007; Picard, 2000; Torrens & Griffin, 2013; Torrens et al., 2011, 2012). Looking to the future, it is also feasible to consider that humans are being joined on streetscapes by moving machines. These machines are also embodied (Chrisley, 2003; Dourish, 2001; Duan et al., 2022), with the result that human–computer embodiment in locomotion is also a near-term consideration. Consider, for example, how semi-autonomous vehicles and robots that are tasked with moving through streetscapes using synthetic perception and sensorimotor control (Bojarski et al., 2016; LeCun et al., 2005), which must sense and make sense of streetscape dynamics to effect motion that is efficient and compliant with situational norms of those streetscapes. At face value, this begets consideration of a form of streetscape AI, one which could share the same viewsheds as pedestrians and drivers and possibly reason about how to move. Any applications of a streetscape AI, then, would face the same challenges that NRT addresses in empirical application: heavy individualization and contextualization that would render both training and application of AI difficult to resolve (Guo & Liu, 2024). There is conceivably an opening in encounter-based modeling for an adjacent consideration of hyper-local situational awareness for streetscape AI. Ideally, this situational awareness would maintain existing pathways for synthetic perception (particularly the impressive benefits that can be garnered from deep learning on visual data drawn from streetscape viewsheds), while also facilitating reasoning on the very individualized, dynamic, and non-generalizable context that enliven and animate streetscapes in everyday life. With this reasoning, one quickly jumps toward considerations of modeling embodied locomotion as a form of autonomous machine intelligence (AMI) (LeCun, 2022). Indeed, it would be straightforward to argue that the configurator, perception, cost, short-term memory, and actor modules of LeCun's (2022) scheme for AMI could connect to his world model as embodied locomotion.

Our aims for the Sidewalk2Synth pipeline share some of the themes being developed for Sim2Real as a tool for training autonomous driving training (Doersch & Zisserman, 2019; Farley et al., 2023; Pasios & Nikolaidis, 2024).

Generally, Sim2Real is delivered as three-dimensional visual simulations that are used to produce training data (usually two-dimensional imagery) for autonomous vehicle AI (usually as deep learning). A handful of Sim2Real models has been extended to the task of training machines for road-crossings, where pedestrian motion and vehicle motion come into contact, for example, within contextual factors of crossing spaces and crossing signals. Haoran et al. (2023) introduced a system for training robot motion control in simulated physical scenes using *Unity 3D*. They used indoor scenes without representation of humans. Ouyang et al. (2018) showed that image processing could be used to insert synthetic images of pedestrians into streetscape scenes, which could be used as training data for vehicle collision detection routines. A system with similar aims was introduced by Strauss et al. (2021). Farley et al. (2023) demonstrated that the infusion of simulated pedestrian data to supervised deep learning could increase vehicle's success in identifying pedestrian collisions by over 27%. They combined deep learning on synthetic static pedestrian scenes from the Multiple Object Tracking Benchmark (MOTSynth) dataset (Fabbri et al., 2021), as well as real roadway scenes of pedestrians as a subset of the Cityscapes dataset (Cordts et al., 2016), "CityPersons" (Zhang et al., 2017). This represents a distinction from the approach of Ouyang et al. (2018), in that simulations are used to generate pedestrians, which are animated in scenes and taken as screengrabs for training.

Nie et al. (2022) introduced a novel twist on this pipeline, inserting virtual pedestrians with unusual, synthetically-generated poses into image-based training data to account for edge cases in pedestrian detection. Typically, the synthetic pedestrian simulations that feed such pipelines are abstract in their representation of locomotion behavior, which is understandable as they focus solely on pedestrian detection, not behavioral inference. We also point out the important distinction that such data are usually static, and do not include movement of pedestrians within the images. Also, synthetic humans are generally generated for pedestrianized environments, without ambient vehicles in the scenes. Pasios and Nikolaidis (2024) recently showed that the CARLA vehicle simulator (Dosovitskiy et al., 2017) can be used to generate vehicle scenes for Sim2Real pipelines, but again this is missing pedestrians and vehicles in the same streetscape. Bucking this trend, however, Vázquez et al. (2014) showed that data from driving scenes of the *Half Life 2* videogame (Valve Corporation, 2004) (which includes procedurally-generated walking avatars) can be used to train deep learning schemes for pedestrian detection.

An approach that sits in the same orbit as the ideas that we propose in this paper was introduced by Yang et al. (2021), taking advantage of high-definition mapping of urban environments (HD maps (Zhang et al., 2023)). They introduce an new twist on Sim2Real, with an interim step of reconstructing synthetic (but frozen) pedestrians from 64-beam LiDAR data (they used the ATG4D data set from Yang et al. (2018)), which they used for subsequent training. Yang et al. (2023) and Wang et al. (2023) demonstrated that this could also be done with LiDAR for reconstruction of vehicles and subsequent insertion of representative (static) models back into driving scenes. In essence, the method distills into a sort of mixed reality approach to generating more useful training scenes for autonomous driving algorithms. Work developed by Priisalu et al. (2021) is closely aligned in objectives to our paper. Priisalu et al. (2021) endeavored to build a "semantic pedestrian locomotion (SPL)" (p. 2) agent that uses reinforcement learning to produce motion control that is semantically related to a given scene. They developed SPL for LiDAR directly: agents appear as point clouds and move through synthetic LiDAR scenes, which matching synthetic vehicles may react to. A critical point, with respect to SPL, is that the agent is supplied with a complete map of the scene a priori. In other words, the trajectories of the agents are reconstructed (using geometry) rather than arriving as the product of behavioral agency. The reinforcement learning policy is derived from *OpenAI*'s Generative Adversarial Imitation Learning (GAIL) (Ho & Ermon, 2016), which has also been used for generating vehicle motion trajectories (Bhattacharyya et al., 2022; Choi et al., 2021). A tangential approach is described in Huynh and Alaghband (2019), but using scene-based Long Short-Term Memory (LSTM) to provide fencing for a trajectory-driven movement LSTM. A salient point that we would like to raise is that these pipelines operate with a great deal of "god-level" information, in the forms of complete views of scenes, geometric details, and geographic information, and sometimes trajectory priors of objects in the scene. Their goals are generally based around "What Happens Next?" type questions, with vector-based answers as a return, rather than striving for novel behavioral agency. This is not intended to be a complaint, rather, we wish to illustrate a distinction relative to our aimed functionality. Recent approaches are advancing toward agency from these principles, for example by building indices of socio-geographic interactions (Bilro et al., 2022) from trajectory-indexed field of view estimates, accessible via Sparse Motion Fields (SMFs) (Barata et al., 2021) or from state data that can be garnered from a given scene (Zhang et al., 2022). Our approach to Sidewalk2Synth, we reason, carries these promising trends forward in some new directions.

There has been quite a large volume of recent activity in developing Vision-Language Models (VLMs) that are trained on streetscape and pedestrian scenes. For example, the Text360Nav system was developed to provide navigation text prompts to users of 360 degree cameras, to assist in mitigating visual challenges and distractions when walking (Nishimura et al., 2024). It was demonstrated on five hours of video data from New York City and Kanto, Japan. A VLM for pedestrian attribute recognition (PAR) was introduced by (Ngo et al., 2024) to essentially marry (through multi-modality) a human-in-image detector with a VLM, using Mixture-of-Experts (Yuksel et al., 2012) to establish consensus (CrossPAR). Wang et al. (2024) developed a related pipeline (spatio-temporal side-tuning) with concatenation of visual and textual tokens in a fusion transformer (Vs et al., 2022). Yang et al. (2024) introduced the Prompt-driven Semantic Guidance (PromptSG) scheme, to make better use of the implied meaning of text prompts to the VLM on input. Side-tuning, PromptSG, CrossPAR, and the related SequencePAR approach developed by (Jin et al., 2023) all make use of the Contrastive Language-Image Pre-Training (CLIP) LLM to associate images with query tokens that can accept text prompts from a traditional LLM framework. For example, Jin et al. (2023) use SequencePAR to generate text data on pedestrian sex, age, and clothing. Generally, VLM approaches share some of the front-end of our Sidewalk2Synth pipeline, but the end product is quite different in form and distinct in its goals. VLMs usually begin with images of naturalistic scenes, either drawn from online video repositories, collected by researchers, or in some cases acquired from streetside robots. Imagery is subjected to convolutional filters and artificial neural networks for deep learning on the features that are extracted by convolution, and image regions that are matched to pertinent urban and pedestrian features. As we will show, we adopt a related approach at the front stage of Sidewalk2Synth. For VLMs, the deeply-learned data are then passed to a large language model (LLM), which is tasked with describing—in text—what elements of the scene could be salient to a user. LLMs such as BLIP (Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation) have been shown to be well suited to pedestrian and street tasks, for which there is often a lot of shifting dynamic detail to consider. We note that, first, the output for VLMs is scene-based: most VLMs aim to describe (by modeling in a linguistic frame) a holistic scene. Usually, they aim to produce contextual understanding of a scene's environment (Song et al., 2025). Any behavior would then be inferred from that environment (e.g., a zebra crossing is for pedestrian movement across roads). This can approximate behavior through tokenization, but it essentially involves matching a text class to an object class. Second, we note that the output of VLMs is usually text, specifically a single text description (e.g., that the image contains "a group of people walking around with umbrellas", (Nishimura et al., 2024, 15,786)). Some innovative work has been accomplished in outputting trajectories, using the LG-Traj approach (LLM-Guided Trajectory) (Chib & Singh, 2024) from VLMs. LG-Traj takes motion cues from deep learning, couples them with learned trajectories from bounding box tracing between images, and predicts future motion in the next few image frames by image extrapolation (not by behavioral modeling we should note). LG-Traj outputs a textual description of what it thinks the trajectory will be, e.g., "According to the given set of coordinates, the motion patterns is a standing still motion" (Chib & Singh, 2024, 6). The VL-TGS approach of (Song et al., 2025) aims to enable maples navigation by robots, using LiDAR data in conjunction with VLMs for scene understanding. In this case, the trajectories that are suggested by LiDAR localization and mapping are down-selected with the assistance of VLM prompts that decide relative traversability as guidance to robot motion planning. A variation of VL-TGS is shown by (Kong et al., 2025), as "autospatial", which includes some contextual details of pedestrian encounters by robots in a similar motion planning pipeline. Autospatial will output pedestrian-relevant navigation details from the image map of the scene, e.g., "The pedestrian is slightly to the right of the robot, at a very close distance. The pedestrian is moving towards west." (Kong et al., 2025, 3). The robot can, then, elect movement rules that relate specifically to pedestrian encounters. In each of the approaches mentioned above, the "model" in a VLM is a language model: it essentially segments a visual image, bounds and frames objects within it, and classifies them to well-trained data sets (or user annotations). The end-sequencing of the relative arrangement of those results is handled by the LLM. LLMs are very different to agent-based behavioral models. Given image-recognized objects, their counts, which side of the image objects appear on, and what their object classification is most probably to be, the LLM outputs some textual description. This is a fantastic accomplishment, which could be useful in providing AI-type assistants for navigation, for assistive technology for pedestrians with visual challenges, and for quickly providing input to driver awareness systems in vehicles. There are, however, some doubts about the utility of VLMs beyond recognition tasks. For example, Wang et al. (2024) investigated whether VLMs could support human gesture recognition and found them to be unsuitable for the task. Huang et al. (2024) cautioned that VLMs have problems in conceptualizing and communicating distance-based and motion-relevant relationships between pedestrians and vehicles: a task that is critical for our applications to streetscape phenomena. What we are trying to accomplish is quite different than recognition, and our approach is thus very different than that of VLM. We aim to build a behavioral model of the scenes that we study, and to translate those behaviors into agent-based rules that can form the basis of what-if simulation. In this sense, the back-end of our pipeline and its outputs depart very significantly from the VLM approach. Our approach is capable of polling objects from images by deep-learning, but we then attempt to simulate (not just describing with text) how they "work together", geographically through space–time action, reaction, interaction, and transactions (social affordances, gestures, body language, etc.). Notably, our approach feeds deeply-learned insight into agents, and not into tokens.

3 Observing and measuring embodied locomotion in real-world bouts of encounter

The Sidewalk2Synth pipeline begins on real streetscapes, with real people, doing real things, in real context (Fig. 2). We are largely interested in the factors that embody people in locomotion on, across, and through streetscapes, and so we established an observational protocol to collect data on embodied streetscape encounters that come into being during pedestrian locomotion. Additionally, the observational protocol is designed to build a typology of those encounters, and then where possible to add measurements and valence to them. We approached this through (1) fixed site observations on streetscapes (Fig. 3), and (2) immersive first-person video diaries of pedestrians while engaged in embodied locomotion. In both cases, the data that were produced were subjected to subsequent analysis to build contextual signals from the encounters. Our data collection replicated several computer vision and deep learning schemes used in autonomous systems, which use sensor modalities that match equipment commonly used by vehicles and robots for artificial perception (e.g., simultaneous localization and mapping (Sun et al., 2023)). Sensed observations and measurements then became inputs to our motion control models and simulations in a subsequent step.

We engaged in a long-term observational study of streetscape sites around New York City (downtown and suburbs) over a period of a 18 months, collecting ~ 1,400 observations of streetscape encounters at 35 different roadside crossings, with different physical environments and activity profiles (see for a sample of the observation areas, which vary by urban design, architecture, urbanity, crowding, lighting, weather, on-street activity, etc.). At each site, we built an atlas of streetscape conditions to include site factors, physical configuration, civic infrastructure, weather and time of day conditions. The atlas was codified to a shared Geographic Information System (GIS). This was supplemented with 3D data acquired by

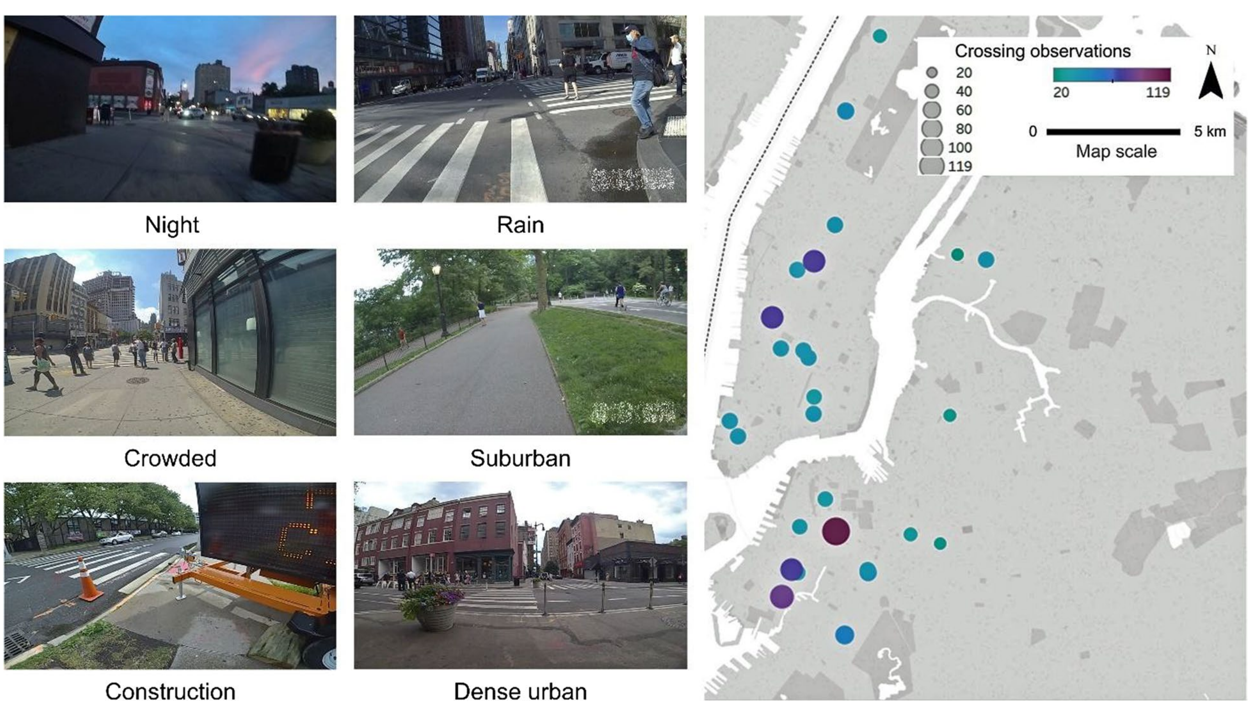


Fig. 2 A selection of observation sites around New York City's dense central city and outlying suburbs

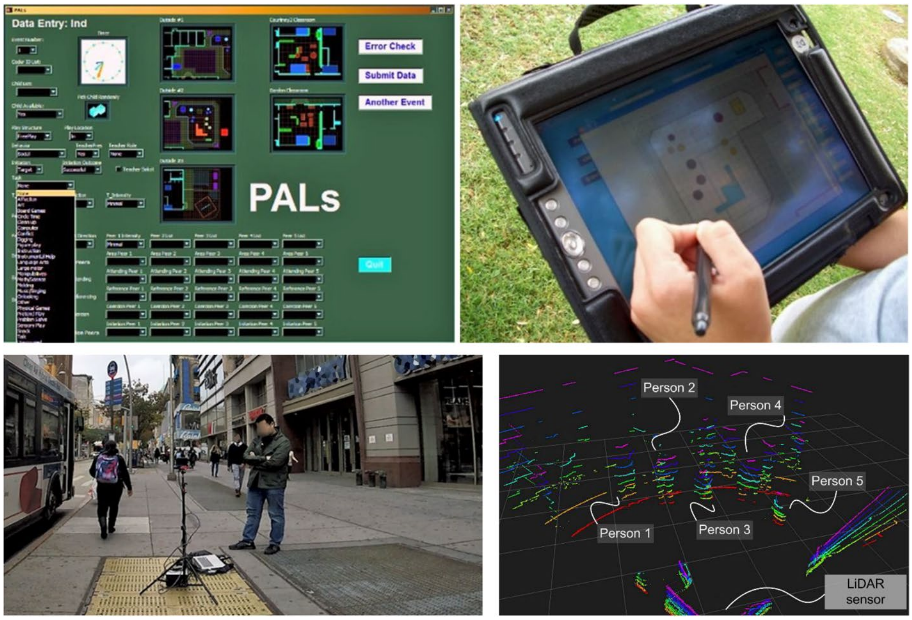


Fig. 3 Top: hand-coded labelling of motion parameters in real-world settings. Bottom: LiDAR measurement of spacings and timings in busy motion scenes

LiDAR for sequences of streetscape activity, which generated allocentric (thing-to-thing) distance measures and timing (to sub-centimeter, sub-second resolution). The census was built for individuals, as well as dyads, groups, and crowds, as well as single vehicles and collective traffic phenomena. To accommodate sensor platforms that are used in vehicle and robot autonomy, we additionally collected RGB video footage from fixed platforms.

To this base, we added hand-coded observation data of streetscape encounters, itemized, classified, and measured using a modification to the Interpersonal Assessment (IPA) framework (Torrens & Griffin, 2013). We first established a set of candidate motion factors in a pre-survey round, followed by a settled set of factors in a survey round, during which we also ascribed valence per-factor. In coding locomotion and embodiment factors, we focused on individual factors of demographics, time geography, ambient awareness, speed, ambulation, object use (including phone use), and risk-taking (when crossing roads). We also coded for site-specific factors of weather, crossing signals, vehicle lanes, and pedestrian crowding. At each observation site, a trained team of socio-behavioral observers marked-up encounters using a modified Interpersonal Process Code (IPC) on a tablet device (Griffin, 2018). These data were fused to the common GIS (Griffin et al., 2007; Torrens et al., 2011, 2012).

First-person (immersive and ego-embodied) data were collected by human participants that were recruited to wear a chest-mounted camera with a high-resolution GPS and networked smart watch while engaging in routine streetscape activities. We collected data over a period of 18 months, for a total of 242 h of immersive video data (Torrens & Kim, 2024b). Taken in totality, the immersive observation data presents a first person record of embodied locomotion encounters on streetscapes around New York City (both downtown areas of New York as well as outlying suburban locations), across a range of encounters across dimensions of place, built environment, physical setting, time of day, urban activity, season, and events. We also collected participants' GPS traces (and sub-trajectories) for these encounters while they were in locomotion.

Collectively, then, the fixed observations (atlas, census, coded encounter) yielded a set of initial conditions for establishing streetscapes by site type, by time of day, by crossing scenario, and by social environment. To sample locomotion through sites, we relied on immersive data. Together, this constitutes the first "sidewalk" portion of Sidewalk2Synth, essentially derived from raw data.

To supplement this, we then added machine-interpreted information on the same scenes. This added value to the encounter data but also had the advantage of allowing us to look at the sorts of sensed realities that might be interpreted by embodied and mobile machines in streetscapes (vehicles and robots). We ran the first-person and third-person video footage through machine learning routines to automatically label scene objects (Fig. 4). We used *YOLO* object detection (Redmon et al.,

Fig. 4 Machine learning motion details from first-person video footage

2016), and *OpenPose* (Cao et al., 2018) and *Detectron2* (Wu et al., 2019) pose detection and motion skeleton extraction, to detect and frame poses for humans that were segmented and identified in the scene. We then used a customized scheme to estimate actions from those poses and to ascribe three-dimensional bounding boxes for pedestrians and vehicles. We also deployed *DensePose* (Güler et al., 2018) to build pseudo-meshes for pedestrians that resulted. To optimize machine learning, we ran the analyses on a customized *Singularity* (Kurtzer et al., 2017) container on our local high-performance computing cluster.

Our observations of real-world motion on streetscapes revealed the key factors that we considered in our model and simulations. We noted key differences in locomotion relative to different configurations of group behavior, and people's embodiment to other pedestrians in very close proximity ahead of crossing.

1. People moving as part of a deliberative dyad or group of more than two were much more likely to either not check their surroundings (particularly when crossing the road), or to rely on incorrect checks (e.g., looking in the wrong direction when facing oncoming traffic).
2. Those in dyads and groups also tended to move more slowly than people moving as individuals. People moving in groups were also less likely to be using a phone than those moving as individuals.
3. We also observed differences in motion due to ambient pedestrian and vehicle traffic. For relatively placid motion scenes, for which there was relatively low foot and vehicle traffic, we saw that people moved more slowly when crossing roadways and that they were more apt to obey crossing signals compared to pedestrians moving in comparatively busy streetscape scenes. Colloquially, this suggests a measure of relative care in motion in low-density scenes. In busier settings, people were observed to move more quickly and with less care to crossing signals. We attribute this to pedestrians either conforming to peer norms in a temporary crossing group (Pfeffer & Hunter, 2013), switching from individual locomotion to engaging in group movement (Coleman & James, 1961; Fernandez & Deneubourg, 2011; Sperber et al., 2019) behavior by temporary "flocking" (matching of velocity, heading relative to nearest-neighbors) (Bikhchandani et al., 1998; Lukeman et al., 2010; Reynolds, 1993), or pedestrians feeling that they had the cover

of the crossing peloton while moving (Das et al., 2005; Faria et al., 2010; Harrell, 1991; Rosenbloom, 2009).

Based on these observations, we resolved to establish experimental simulation scenarios that would permit us (1) to vary the number of crossing pedestrians to include dyads and groups; (2) to vary the ambient pedestrian density; and (3) endow agents with varying care and risk-taking strategies. As we will discuss later in the paper (Sect. 5.2.1), these factors converge conceptually to topics of peer effects (Pfeffer & Hunter, 2013) in locomotion (Gorrini et al., 2014) and social embodiment (Meier et al., 2012; Niedenthal et al., 2005).

4 High-fidelity motion reconstruction by motion capture

In outdoor observation, we learned that people's mannerisms while in embodied locomotion can provide information of their behavior: indicators of motion intent, decision-making, and actions that connect to "enactable" attributes of that motion as future steering, velocity, and stopping rules. In our coded observation, we also noted particular mannerisms that we were able to associate with embodied locomotion.

- Head-checking as a signal of people's interest in and awareness of ambient conditions,
- Hesitation as an indicator of reassessment of locomotion once its embodied effects were confirmed or challenged by next steps,
- Ambulation (of body extremities, usually swung in a particular style) as a motif of locomotion,
- Encumbrance when pedestrians were carrying or pushing/pulling objects, and
- Leaning as a mark of impending locomotion and desired or considered embodiment (Fig. 5).

We coded our observations directly for these mannerisms when surveying natural locomotion scenes. This is exactly the sort of high-resolution detail that we hope to add to the simulations. We reason that body language and other nonverbal communications (NVC) signals (Andersen, 2008; Collett & Marsh, 1981; De Gelder, 2006; Ekman & Friesen, 1981; Kudoh & Matsumoto, 1985; Marques et al., 2025; Mehrabian, 1968; Scheflen, 1972), specifically, can serve as strong (and reliable) indicators of underlying behaviors (Torrens, 2014a, 2014b, 2016b, 2018a; Torrens & Gu, 2023). Further, we propose that NVCs can help to fine-tune agency in our models and simulations in ways that allow embodied locomotion to be embedded synthetically in simulation, especially in mechanisms that human users of the pipeline can interact with in virtual reality, with verisimilitude that matches their embodiment to real world streetscape encounters.

To examine NVC factors further, we examined the nature and use of mannerisms with motion capture in two methodologies. First, we processed first-person and third-person videos that were collected during our observational fieldwork through a deep-learning scheme for pose detection. We used *OpenPose* (Cao et al., 2017, 2018) to build skeletons of human motion (body language as well as gaze direction) from partial affinity fields. The procedure involves segmenting video for the presence of humans against a given background scene, then extracting partial affinity fields, and using those fields to estimate body parts ahead of building a graph-based skeleton of those parts as an estimated pose (Fig. 6). Deep-learned poses were used to index general sets of motion behavior to match the hand-coded observations from our fieldwork. In several instances, we were also able to reconcile distances and timing for these motions using cross-indexed LiDAR observations.

Second, we recruited human participants to engage in motion exercises in a studio mock-up of a physical streetscape (Fig. 7), using marker-based motion capture to collect data on the positioning and timing of their

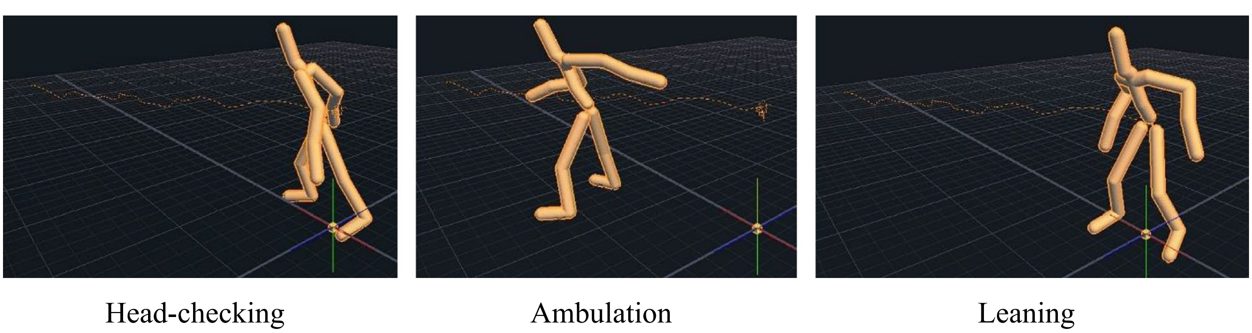


Fig. 5 Examples of mannerisms for different locomotion actions

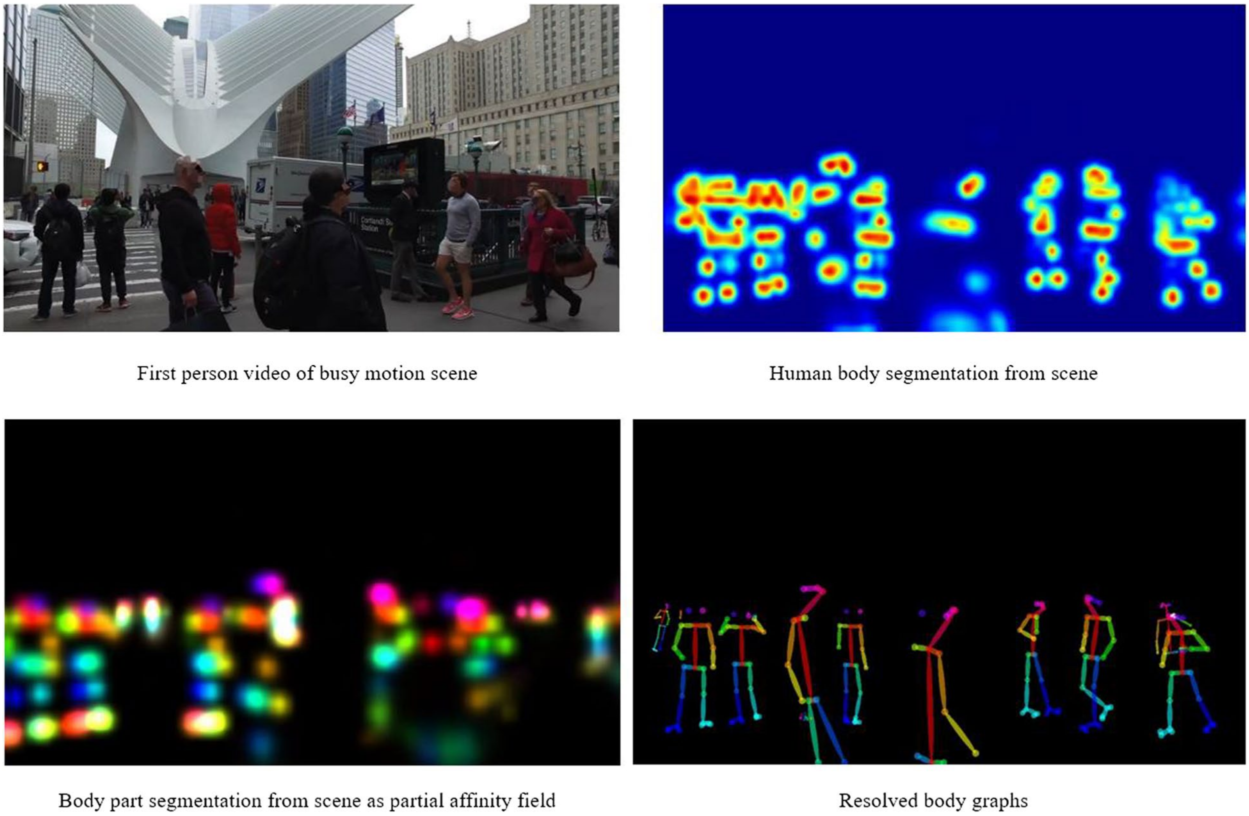


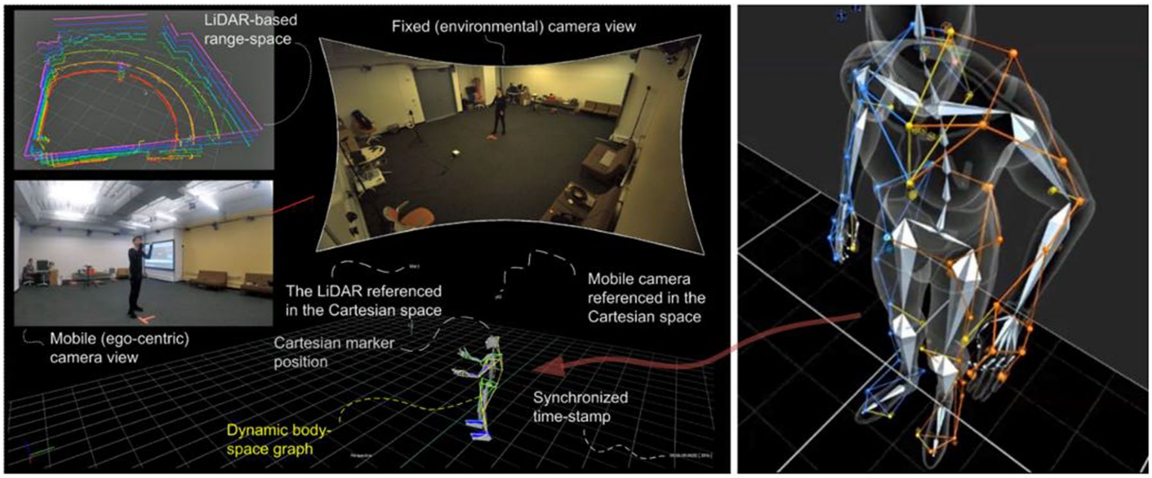
Fig. 6 Machine learning poses from first-person video footage

body movement at very high resolution (centimeter-level localization and up to 800 Hz in timing), each indexed to action states as well as to motion velocities. Moreover, we were able to collect motion data for different dyads and groups (up to ten people sharing a session) and for different motion contexts (crossing a road, using a phone, navigating through a crowd, participating in unidirectional flow, interrupting predominant pedestrian flow, etc.).

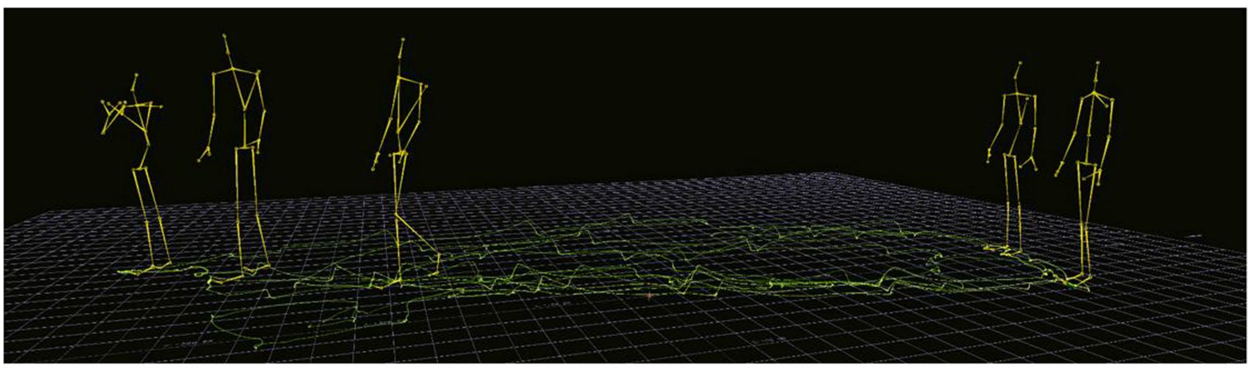
Motion capture by pose detection on video is useful in isolating and delineating rough approximations of human poses in locomotion, and these can further be classified into coarse locomotion actions by automated means. Deep learning to accomplish this can be run for large crowds of people, although occlusion problems will persist for humans that are partially or fully occluded in the scene. Video-based motion capture is relatively "cheap" in researcher involvement. In section 11, we show that it can be fully automated on edge devices which can be placed close to motion scenes such as sidewalks (Potdar & Torrens, 2019).

Our studio-based motion capture produces very high-resolution and high-fidelity motion data. However, the behaviors that we can capture are limited to indoor scenarios. We approximated outdoor locomotion by asking participants in the studio to move as they would normally do in the real world, and by inviting them to move in reaction to scenes of real-world urban scenes that were projected onto a large screen next to them. Ultimately indoor behavior is a proxy for what would really occur in the messy context of an urban scene. Nonetheless, studio-based motion capture produces real locomotion data, and we were able to collect studio data for groups of interacting people (simultaneously, ten at a time at ~ 400 Hz) to reproduce dyadic and group scenarios, as well as small bouts of pedestrian flow.

As we will discuss (Sect. 5), we transferred the motion capture data directly to agent-automata in simulation, so that agent-based pedestrians would move with realistic mannerisms and body language. Real human user-participants can then be immersed in the simulation (see Sect. 7), which may then present as open to social embodiment with these mannerisms on agents. Studio-based motion capture data can be used directly to represent agent motion. The deep-learned (video-sourced) pose skeletons are lower in resolution than the (motion-captured and resolved) skeletal rigs that we use in our agent-automata model are, and so we do not transfer video-learned poses directly into our simulation



Studio-based motion capture using markers to produce detailed motion skeletons.



Simultaneous motion capture for a group of five people, walking as individuals (left) and as a dyad (right).

Fig. 7 Studio marker-based capture of high-resolution and high-fidelity motion

scenarios, although, as we will discuss in Section, these video poses can be used in end-to-end versions of the pipeline that are designed to run outdoors in real-time.

5 Simulating pedestrians as automata with strong-agency AI

The "Sidewalk" components of Sidewalk2Synth are treated using observation and reconstruction as discussed in Sects. 3 and 4. These are designed to feed insight to a second half of the pipeline, the "Synth" component, which is tasked with modeling those observations and producing synthetic representations of streetscapes (both physical and social) in the form of simulations. Users of the system will directly interface with the synthetic simulation, embodying themselves to the streetscape encounters and environments that are represented. Our intent, in designing the synthetic components, is to produce high-fidelity parity with our real-world observations, such that the system can evoke real embodied locomotion and related behavior from users. A first step is to generate synthetic pedestrians, which will provide both physical (collision objects) and social (NVC counterfoils) environments. In Sect. 6, we describe AI for vehicles and traffic. The integration of the human (pedestrian) automata and vehicle (driver) automata in a decision tree framework is shown in Fig. 8. The data architecture that feeds the decision tree is illustrated in Fig. 9.

Embodied locomotion is intertwined with lived experiences that are acted out with sensation and physical and social contact, with the exchange of mannerisms and depth of feeling that implies. Sidewalk2Synth would therefore benefit from fidelity in its congruence to reality: the things in the simulation should behave as they would in their real-world counterparts. Synthetic vehicles should move as they do on real roads, around real crossings, and with faithful response to pedestrians. Similarly, human automata must take on agency that is faithful with respect to real behavior, including both rational and often irrational action to dynamic conditions as they unfold around them. There are cases when substitutes for motion (such as using particle physics in lieu of human

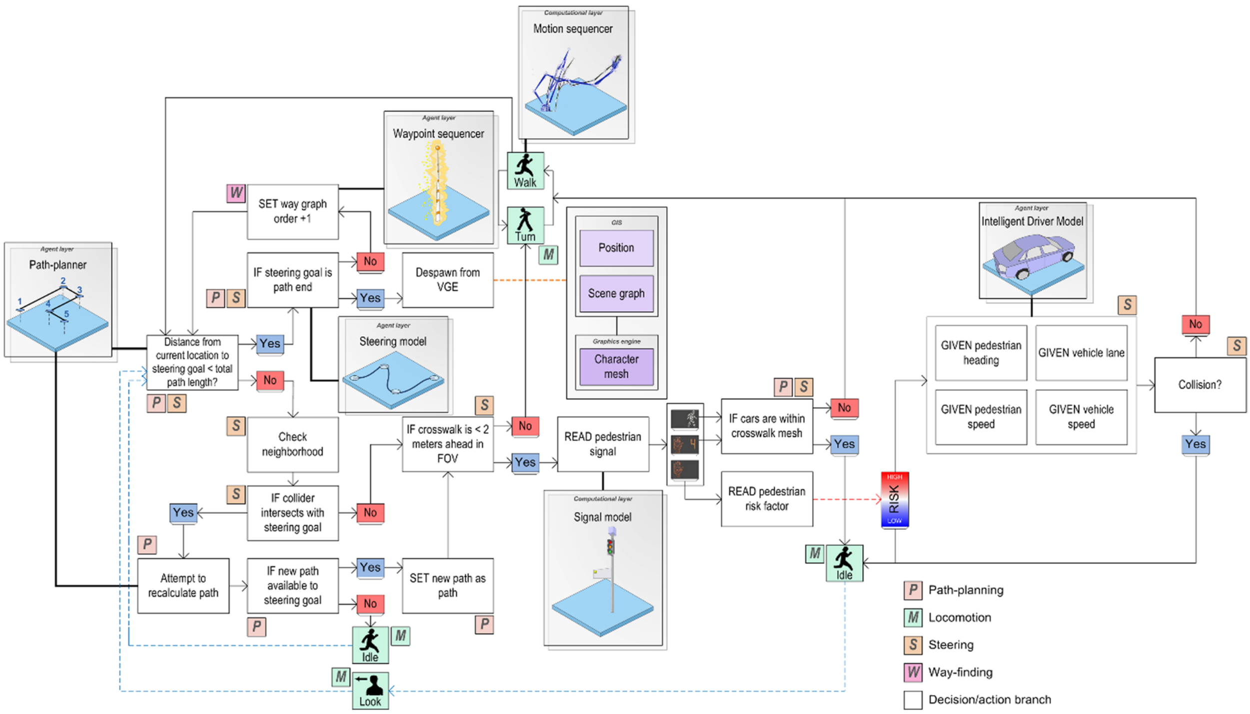


Fig. 8 Agent-automata components come together from six interworking models: a path-planner, a waypoint sequencer, a steering model, and a driver model. These are integrated with computational models (i.e., algorithmic and heuristic rather than behavioral) for motion sequencing and traffic and crosswalk signaling (Torrens & Kim, 2024b)

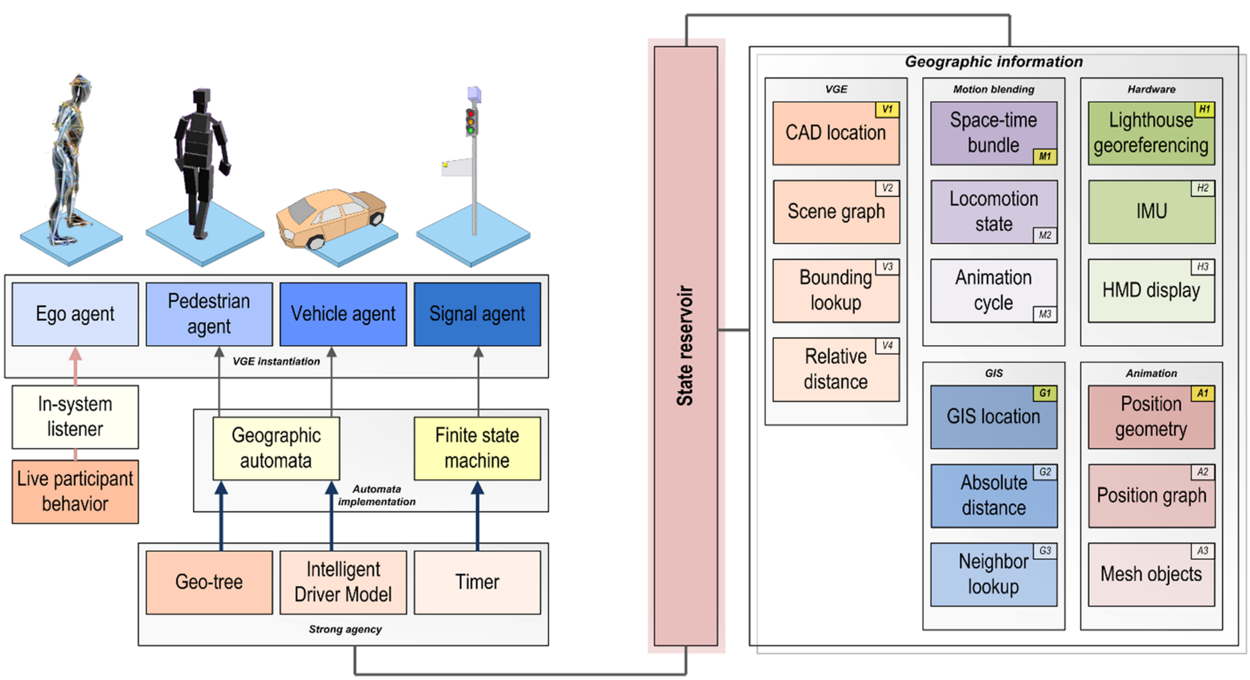


Fig. 9 The data architecture that feeds Sidewalk2Synth in run-time, derived from real-time and run-time feeds sourced in the virtual geographic environment (VGE), GIS, animation, motion-blending and sequencing, and virtual reality hardware. These feeds are managed in a state reservoir before being passed to agent automata and other simulation computation (which are detailed in Fig. 8) (Torrens & Kim, 2024b)

locomotion) could be useful in a simulation effort, e.g., if movement is just an input to a particular simulation system as flow or occupancy counts (Johansson et al., 2012; Treuille et al., 2006). However, for our purposes, *locomotion is the output*, and so we need realistic behavior rather than abstractions. Sidewalk2Synth would also benefit from maintaining congruence to reality in its verisimilitude. Because it is designed to be used by human users as a way to experiment with their (real) embodiment, Sidewalk2Synth should evoke realistic behaviors from its users. Specifically, it should conjure realistic embodiment to both social and physical context. Users of the simulation should feel compelled to engage a full set of their motion skills, mannerisms, and habits, and they should react with distance and timing that matches counterpart context in reality. If the simulation is verisimilar, it could be useful as a means for testing against real-world parameters.

We approached the first topic of congruence—by fidelity—using strong agents. The second facet of congruence—by verisimilitude—is tackled using mobile VR technology (mVR) designed to transpose real human users directly into the simulation as mobile bodies. In the text that follows, we discuss the design of agent-automata; transposition of real human users into the simulation is detailed in Sect. 7.

5.1 Weak or strong agency for motion control?

The use of weak agents is a commonplace approach in applied simulation, where the intent is to produce either visual or statistical movement patterns, with validity of their flow usually assessed, in aggregate (crowd volume) form (Torrens, 2004). Examples include the use of physics-based schemes to generate crowd flow dynamics based on force-based heuristics (Helbing, 1992; Helbing & Molnár, 1995) or continuum mechanics (Henderson, 1971), information search schemes that can generate dynamic density patterns on gridded spaces (Blue & Adler, 2001; Galea et al., 1996), or graph and roadmap structures that can generate localized activity patterns around particular geometries (Sud et al., 2007, 2008). Dedicated motion controllers are also widely used, including vector-based schemes (Reynolds, 1993; Sun et al., 2023) and reverse velocity approaches (Guy et al., 2009) that can be particularly helpful in producing localized (but procedural) motion for collision avoidance. Similarly, a range of inverse and forward kinematics schemes are available for generating realistic-appearing ambulation over short bursts of stride-scale movement (Badler et al., 1987, 1991, 1994), as well as patch-based motion capture recall schemes (Hyun et al., 2013; Lee et al., 2006), and these techniques are usefully deployed in computer animation and special effects. Using combinations of these schemes—usually handovers among algorithms and heuristics at particular scales of motion—can produce very realistic-appearing motion and movement, with pattern dynamics that hold validity against real-world motifs. Moreover, several of these approaches can be run with significant scale advantages so that large mobile crowds of walkers can be simulated (Torrens, 2014b), often with individual variance provided by parameter files that can be paired to real-world censuses and contexts.

Although there are many exceptions to the following statement, we would assert that for the most part such schemes are weak in their fit to actual locomotion behavior. They *look* realistic, but they do not arrive at that realism from fidelity to the real world. This is appropriate for the usual applications to transport systems or planning and policy support for crowd flow management (Johansson et al., 2012; Pauls, 1984; Sime, 1995; Tubbs & Meacham, 2007). But it is problematic if you ask human participants to embody themselves to the simulations. People can easily spot fake behavior, and when they do, it does little to evoke commensurate embodied interaction. In other words, many existing approaches (which we reiterate are never intended for virtual simulation, or for testing embodiment, and so this criticism is just a scientific one) often focus on generating patterns and processes of applied motion and movement scenarios, rather than attempting to replicate real motion behavior with high-fidelity. This is fine: real behavior is not in their remit. They only ever claim weak agency.

An alternative approach could be termed, by contrast, as strong agency, dealing with generalized AI as a way to build agency that is responsive to simulation conditions in ways that authentically and faithfully match to known locomotion conditions and behaviors in the real world. Strong agency is quite difficult to accomplish, as a lot is unknown about how people control their motion in reality, largely because building that understanding can quickly lead to problems of infinite variation and overwhelming context dependency. Simulations can help to build this understanding, but again they must be realistic and authentic to their real-world, physical companions to yield answers relative to that reality, which leads to a circularity problem. We do not claim to have a profound discovery in this regard: here, we simply state that we are trying to advance the science at least a little bit in the direction of strong AI.

5.2 Human automata model

We used automata-based models as the AI driver for synthetic pedestrians. We modeled human automata on a Geographic Automata System (GAS) structure

(Benenson & Torrens, 2003; Torrens & Benenson, 2003), which we index as "G" in what follows:

$$G = K,\ S,\ \mathbf{R}_S,\ L,\ \mathbf{R}_L\ ,\ N,\ \mathbf{R}_N;\ \text{where}:\ \mathbf{R}_S\ \mathbf{:}\ S_{\mathbf{t}} \rightarrow S_{t+1},\ \mathbf{R}_L : L_t \rightarrow L_{t+1},\ \mathbf{R}_N :\ N_t \rightarrow N_{t+1}$$

A state descriptor, K, is used to index whether entities and objects in the model are able to engage in locomotion. Elsewhere, we have experimented with rule sets $\mathbf{R}_K$ that govern whether something that is usually immovable can be brought into motion, e.g., through a forceful collision (Torrens, 2014b), but here we simply rely on K as an index state for features of the built environment as distinct from those that could be mobile. Motion-capable human automata can read the states of fixed automata for ease in data-processing, but in the examples that we show here the K = [fixed] automata were cloistered from the parts of the state transition tree that permit locomotion. As each GAS is a finite state machine (Sipper, 1997), our approach of enveloping information to pertinent subsets of automata-to-automata interactions reduces the computation required to resolve states across the system in a given state transition update epoch.

The main "work" of G is then performed among routines {$\mathbf{R}$}, which are used to control locomotion and its (state) information context, per automaton. Transition functions $\mathbf{R}_S$, $\mathbf{R}_L$, and $\mathbf{R}_N$ govern transition between discrete states in packets of change from $t \rightarrow t+1$. We distinguish between three overriding functions. Under $\mathbf{R}_S$, agent-automata can change general agency states. In this application, we use states {S} to index pedestrians' attitudes as we will discuss shortly. Transition functions {$\mathbf{R}_N$} are used to control different aspects of the human automata's perception and awareness. We consider N as defining different neighborhoods of encounter and information-gathering that automata engage with (and embody to) within the spaces that they progress. In this sense, then, N determines the ego-centric embodiment space of the pedestrian agent, while streetscape spatial and space–time progression is governed by $\mathbf{R}_L$, such that different encounters are continually coming into being in N. N therefore constitutes a shifting set of dynamic information that unfolds around automata G. Aspects of {$\mathbf{R}_N$} are designed to provide system-wide access to information, as in the case of path-planning. However, we also focus dedicated R_N functions on enabling hyper-local awareness for automata, specifically intended to provide information in fleeting moments of space and time that can inform simulated pedestrians' snap judgment based on limited (often proximal and partial) appreciation for simulation conditions. This model function follows the idea of primacy in transient encounter over fixture in geographic representation that is described in NRT concepts, especially rhythmanalysis (Lefebvre, 1992/2004), from the conceptual literature base (Bower, 2014; Cadman, 2009; Sheller, 2017).

Automata control their movement routines via rule-set $\mathbf{R}_L$. Given the significance of locomotion for the application sets of Sidewalk2Synth, an expansive set of rules {$\mathbf{R}_L$} are used (these are discussed in detail in Sect. 5.2.1). We provide ego-location by slipstreaming, i.e., by allowing human automata to cross-register (slip) their position across a range of different geographies in the simulation, on-the-fly as information flows through their state transition schemes (i.e., by streaming) (Torrens, 2015c). Slipstreaming is used to practically embed (and to conceptually embody) automata to 3D built-space representations (Torrens, 2015a), which we handle in model form as Virtual Geographic Environments (VGEs) (Chen et al., 2008; Lin et al., 2015; Torrens, 2015a; Zhang et al., 2007). VGEs provide cross-functionality between geometry and GIS, which we then further expand as Virtual Reality Environments (Torrens & Gu, 2021, 2023). VRE serve as special modalities of VGEs that are delivered to head-mounted displays (HMDs) with 3D rendering and spatial audio that users can immerse themselves into.

We consider a wide range of movement and motion conditions through varied localization states {L}. These hyper-spaces support localization across different interpretations of geography. These include in a traversal graph for pathfinding, in a hierarchical waypoint list for navigation and wayfinding, in vector space for steering and collision detection, in geometric space for collision avoidance and reconciliation, in body-graphs for localization of NVCs such as mannerisms and ambulation, and in relational space for gaze dynamics. Given our intended applications to streetscapes, we based the static built environment for movement on urban scenes, which we provided to human automata as GIS data in planar view. However, we also provided the same data as geometry (CAD and mesh), as well as graph data (navigation meshes). For automata–automata interactions, we supplied vector-space data for velocity determination. Human automata are allowed to geo-position themselves within any one or all of these "geographies" as needed, with the result that they avail of a flexible sense of automata neighborhood input to inform their transition rules.

We designed the system's human automata with tiered spatial agency to represent strong agency at different locomotion scales. At scales of a city-block, we used transition rules that account for agent-pedestrians' gross movement by path-planning as an $\mathbf{R}_L = [A^*]$ heuristic (Hart et al., 1968). At street scale (four sidewalk segments within a given city block), we relied on navigation meshes to provide human automata with wayfinding capabilities between given waypoints. Following the approach shown in Torrens (2018b), we used scaled waypoints with long-term, medium-term, and short-term goal-locations that correspond to navigational features of the built environment (Costa et al., 2011; Cutting et al., 1992; Devlin & Bernstein, 1995; Evans et al., 1984; Gärling & Gärling, 1988; Raubal & Worboys, 1999; Wang & Cutting, 1999). For intra-street movement, we first introduced steering behavior, using combinations of the steering routines from Reynolds (1999), relying on the pedestrian adaptations shown in (Torrens, 2012). However, for collision detection and avoidance movements that were not resolvable by steering, we used reciprocal velocity obstacle algorithms (Fiorini & Shiller, 1998; Guy et al., 2009; Snape et al., 2011; van den Berg et al., 2008; Wilkie et al., 2009).

5.2.1 Human automata movement routines

We modeled movement on a hierarchical motion scheme. For each of the (slipstreamed) spaces that have representation in the model as {L}, we developed a matching locomotion rule, $\mathbf{R}_{\mathrm{L}}$ to govern automata's polling of hyper-local information (N), the behavior that they marshal to that information, and the spatial structures to frame this in GIS, VGE, and VRE. In entwining {L} and N, we build embodiment methodologically in the model. Conceptually, this follows well-researched concepts from behavioral geography (Downs & Stea, 1974; Golledge & Stimson, 1997; Hart, 1987), which indicate that pedestrians generally map motion planning to a set of layered scales (Devlin & Bernstein, 1995; Golledge, 1999; Mark & Frank, 1996; Siegel & White, 1975). Operationally in software, it roughly follows the data flow model shown in (Torrens, 2015c).

Motion in trip space by path-planning At trip scale, pedestrians settle on a route or a route-finding strategy between an origin and a destination: usually a source and sink for activity, akin to the anchor point hypothesis in cognitive mapping (Couclelis et al., 1987; Kuipers, 1982; Kuipers & Levitt, 1988). In real situations, this may take place over a large geographic area, for example, if somebody is walking around a downtown for leisure, or the trip could be comparatively short-lived, say from an establishment to a road crossing. The important point for locomotion over trip-spaces is that a pedestrian will settle on a destination and will devise a path-planning strategy to get there (Hartley et al., 2003; Huber et al., 2014; Lenntorp, 1977; Nabbe et al., 2006; Sun et al., 2021). Usually, this involves a shortest path or some variant, possibly with additional weightings for paths that preserve viewsheds of the streetscape (Hillier & Hanson, 1984; Penn, 2003), or for paths with least right-angle turns (Foltête & Piombini, 2007; Omer & Kaplan, 2019), for example. We handle path-planning using the $A*$ algorithm (Hart et al., 1968) for identifying paths with a minimal traversal cost (in space and time to destination), which trades off shortest path (graph) distance from a locomotion origin and straight-line distance to a destination, and which parsimoniously ekes efficiency by planning toward a specific goal rather than solving for all possible goals in the traversal space. We regard this as realistic because pedestrians on a relatively short stretch of streetscape will generally have just a single goal in mind (Patla & Vickers, 2003) and will usually engage in a combined strategy of minimizing space–time distance to that destination, while preserving streetscape viewshed (Batty, 1997). This latter point, of preserving visibility of the built environment is treated in space syntax concepts (Omer & Goldblatt, 2017; Omer et al., 2015; Ryu et al., 2021), for example. Again, $A*$ yields a (desirable in this regard) combination of both destination-oriented goals in locomotion.

Motion in waypoint space by wayfinding Within a trip, we allow human automata to identify waypoints as motion-relevant interim features along a planned path. Again, we follow concepts from behavioral geography and from psychology that have revealed that walkers make use of waypoints as interim goals to track progression along trip paths (Kato & Takeuchi, 2003; Raubal, 2001b; Spiers & Maguire, 2008; Wang & Cutting, 1999). We relied upon built environment features for our experiments: streetscape curbs, pedestrian crossing locations, and pedestrian crossing lights. Essentially, these built features serve as hyper-local landmarks for waypoint-driven navigation and wayfinding (Caduff & Timpf, 2008; Evans et al., 1984; Kamil & Cheng, 2001; Omer & Goldblatt, 2007; Ruddle et al., 2011). Waypoints were straightforwardly represented in VR simulations as observable features. We also note that built-feature waypoints on streetscapes serve as important boundaries between different activity spaces on the streetscape (chiefly vehicle areas, pedestrian areas, and mixtures of the two such as crosswalks) (Raubal, 2001a). Critically, passage from one of these activity spaces to another may invoke halting states for locomotion (Tabibi & Pfeffer, 2003; Torrens & Kim, 2024b; Zeedyk et al., 2002). For example, Goldhammer et al. (2014) showed that specific termination gaits are used at crosswalks. These locomotion phase shifts provide (visual) representation of what Schmidt and Färber (2009) referred to as “action intention” (p. 300). With animation cycles form mocap as NVCs, user-participants of our system can then (and do) react to these action intentions as visual signals cast by automata-avatars in VRE as ambulation, mannerisms, and body language signals of movement intent. Importantly, several of these NVCs such as milling behavior, gaze, and waiting can be used to represent embodiment to the streetscape. (We detail this shortly.)

Motion in cluttered space by steering While engaged in traversing paths and progressing through waypoints, pedestrians will pursue a tradeoff between a given and a desired locomotion, as an individualized (perspective-based) affordance of their localized embodiment to the streetscape (Raubal, 2008; Withagen & Chemero, 2012). This matches experimental evidence from psychology that pedestrians follow retinal and/or optical flow when moving (Cutting et al., 1992; Matthis et al., 2022). That is, unless they detect an interfering physical obstacle, vehicle obstacle, or human obstacle, pedestrians will tend to follow a preferred velocity (Hurt & Kram, 2006). (This could be leisurely or hurried, and we can code agents to adopt these policies on a hyper-individualized basis using {**S**}.) When faced with the novelty of an imposing obstacle, pedestrians will correct their locomotion, usually temporarily, to first prioritize negotiation around that obstacle and second to return to their desired locomotion (Cutting et al., 1995). In some cases, the originally desired locomotion (by wayfinding and path-planning) may no longer be viable given a pedestrian's destination goals, in which case they may recalculate their "up-scale" locomotion, i.e., they may engage in motion control at scales of space and time that are coarser than the immediacy of the steering decision (Garbrecht, 1971; Gärling & Gärling, 1988; Hillier & Hanson, 1984). Importantly, we handle most collision situations primarily through steering. In other words, pedestrians will try to avoid getting into a situation in which they must engage collision avoidance routines (Kitazawa & Fujiyama, 2010; Patla & Vickers, 2003). This is because, first, pedestrians generally avoid collisions in the real-world as they are socially and physically unacceptable in most situations (Olivier et al., 2013), and second, because collision detection in motion control algorithms is generally more costly (and non-faithful to real-world motion) in simulation (van den Berg et al., 2008), it is a quality to circumnavigate in simulation. We address steering with algorithms that allow pedestrians to take a planned path and its waypoints and to project space–time progression along an intervening path, with the ability to estimate progress between waypoints and with leeway to speed-up to reclaim lost time should they need to steer away from an otherwise desired path (Torrens et al., 2012). Steering itself is handled using a variant of Reynolds's (1999) steering behaviors for autonomous characters, which allow for very parsimonious resolution of steering by seeking and fleeing routines. Space–time projection is handled using a modification to space–time paths from time geography (Lenntorp, 1977), adapted to work with steering behaviors (Torrens, 2012). The geometric hand-off between source and sink nodes for path-planning, a graph-space for path planning, embedding of waypoints as nodes in that space, and the overlay of a relativistic steering space with distance and time look-ups is handled by using slipstreaming in a GIS (Torrens, 2015c). In GAS terms, interchanges in state data are interoperable across transition rules $\{\mathbf{R}_S\}$ and localization data {L} can be co-registered (and translated) between different layers of $L_1, L_2, \ldots, L_n \in \{L\}$. Values of {L} may be stored with diverse spatial data structures, and spatial data lookups may provide data access schemes for the location and (location convention) rule-sets $\{\mathbf{R}_L\}$ (Fayyad et al., 1996; Fujimura & Samet, 1993; Samet, 1989, 1990; Torrens et al., 2011).

Motion in collision space by detection and avoidance When different pedestrians employ different paths, drawing them to the same or to varying destinations, and when pedestrians use individualized locomotion to articulate through space and time, it is inevitable that they will come into potential collisions (Harrigan, 2005). Generally, steering is enough to resolve these collisions while absorbing a small decrease to space–time progression. However, there are cases in which density of activity on a streetscape or multiply-conflicting steering maneuvers among pedestrians in close proximity could produce an impending physical collision (Gérin-Lajoie et al., 2008; Hayduk, 1983). Usually, real people will detect such collisions and if they are unable to steer around them, they will engage in collision-avoidance routines (Basili et al., 2013; Collett & Marsh, 1974; Huber et al., 2014; Kitazawa & Fujiyama, 2010; Knorr et al., 2016; Lynch et al., 2018; Olivier et al., 2012, 2013). In psychology, there is evidence that this is an active brain process for most walkers (Kennedy et al., 2009), which also has secondary import as social psychology (Aiello & Thompson, 1980). In brief, this commonly invokes people's vision, working in tandem with small adjustments to locomotion in very small bundles of space and time (Cutting et al., 1995), as brushing motions, twisting of the body, and sidestepping (Ciolek, 1983), i.e., strong physical embodiment that directly impacts locomotion response. In each case, it is usually necessary for a pedestrian to come to a stop or near to a stop in their desired locomotion to resolve the collision. Because there are two parties involved in the collision, the maneuvers must be balanced in a dialectic (Olivier et al., 2012). This is well-treated algorithmically by Velocity Obstacles (VOs) (Fiorini & Shiller, 1998; Wilkie et al., 2009) and by Reciprocal Velocity Obstacles (RVOs) (van den Berg et al., 2008). For Sidewalk2Synth we use RVOs for collision avoidance; we show this implementation in more detail in Torrens and Gu (2021).

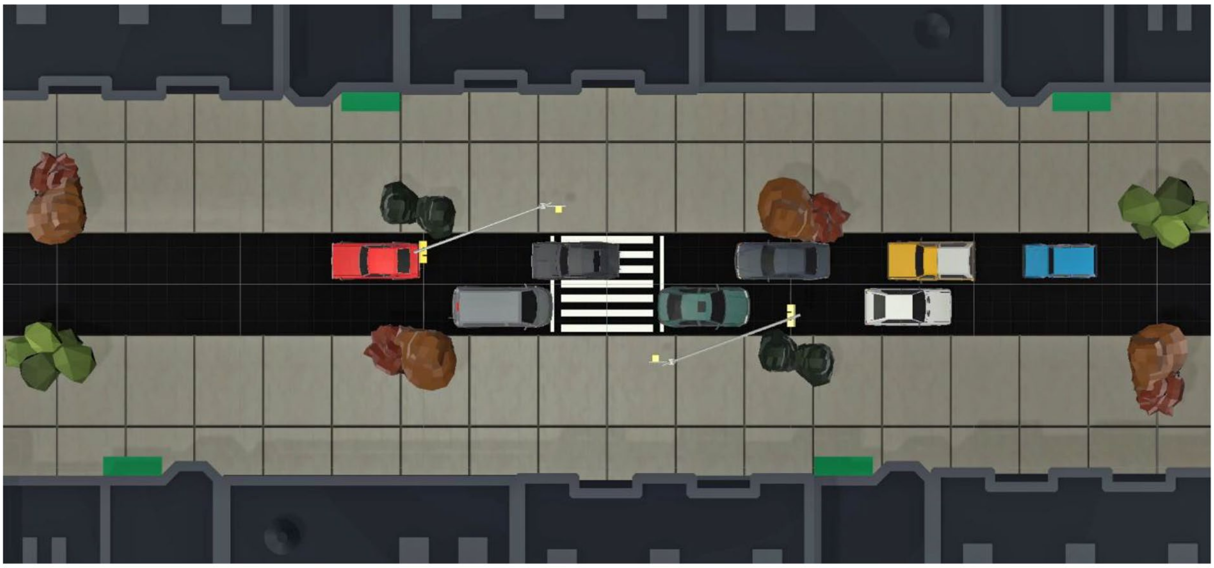
Motion in articulation space by animation cycling Given a motion plan, put into effect as $\mathbf{R}_L$ simultaneously across multiple scales of the streetscape, agent-pedestrians must then generate realistic body motions to both match and satisfy that motion. In real life, such motion is highly individual as gait (Baker & Hart, 2013) and other ambulatory factors that shift based on walking speed (Jordan et al., 2007), ability, and inclination, including many dimensions of desired walking behavior, skill, energy, and effort (Chung & Wang, 2010). Evidence from psychology points to a close coupling between perception (which we can regard as being situational and embodied) and the body motion that produces physical locomotion (Konczak, 1994; Pailhous et al., 1990; Pearson, 2004; Salinas et al., 2017) (which we can regard as sensory). Walking behavior may also be contextually dependent upon interpersonal factors (Cutting & Kozlowski, 1977; Montepare et al., 1987), the given crowd density in a particular part of a streetscape (Hans & Hans, 2015), norms of particular social surroundings (people do not generally sprint through culturally sensitive areas and sacred spaces), civic expectations (waiting at a road crossing when the pedestrian signal indicates "do not walk"), peer pressure between socially-influenceable pedestrians (Pfeffer & Hunter, 2013), activity purpose (tourism vs commuting to school, for example), as well as public rituals such as yielding behavior in a potential collision, entering queues, taking turns at doors to establishments, and so on (Mondada & Tekin, 2023). The physical and built conditions of a streetscape also have obvious impact on gait and other forms of ambulation that tangibly produce locomotion (Franěk, 2013; Thies et al., 2005). At the boundary between different expected locomotion domains (roads as a space where one would expect vehicles to dominate; sidewalks as a space where pedestrians would usually have leeway), action intentions may be visible in relevant non-verbal signals. We represent these with specific animation cycles (waiting by idling at a crosswalk, for example as a signal to users that agent-automata are waiting for a pedestrian signal) (Carol & Roslyn, 2007; Geruschat et al., 2003; Harrell, 1991; Oudejans et al., 1996). The number of potential factors at play in reproducing these behaviors in simulation are massive, and realistically beyond reach in current computer science outside of inverse and forward kinematics (Tolani et al., 2000; Zhao & Badler, 1994). In our pipeline, we sidestepped their implementation with authenticity as transition rules, and instead we opted for realistic-appearing behaviors by using animation cycling (Arikan & Forsyth, 2002; Safonova et al., 2004), but in real motion capture data. Importantly, we stress that animation is only used to produce the kinematics of walking; all other behaviors that lead to those steps are from automata AI. Specifically, we recorded motion capture data of real people's movement behavior for different speeds of locomotion, different beginning and halting conditions, and different steering maneuvers. We then indexed these animations to agents' velocity in simulation, as well as to particular action conditions (waiting at a roadside by using milling behavior or idle swaying, for example). We used motion retargeting (Gleicher, 1998) to adapt motion capture data to pedestrian representations of differing sex and height.

6 Vehicle automata model

To investigate motion control at the roadside curb and during crossing, we also built a vehicle driver model. This was established using a geographic automata version of the Intelligent Driver Model (IDM). The IDM (Kesting et al., 2010) provides microscopic driving dynamics, which collectively can also yield macroscopic phenomena of traffic. However, our model is open to the introduction of driver, vehicle, and traffic models and we have also experimented with the CARLA driving model (Dosovitskiy et al., 2017; Torrens & Gu, 2023) as well as purely macroscopic traffic flow models (Torrens & Gu, 2021). We do not go into significant detail of the driving automata modifications here and instead we refer the reader to the original IDM methodology in Kesting et al. (2010). Our extension of the IDM is discussed in more detail in Kim and Torrens (2024); Torrens and Kim (2024a, 2024b). For the pipeline experiments that we show here, we included vehicle models for coupes, sedans, sports utility vehicles, vans, and trucks, each with different driving norms and with acceleration profiles that match the vehicle type (Fig. 10).

7 Instantiating Sidewalk2Synth as an immersive, traversable virtual reality environment

The final piece of the Sidewalk2Synth pipeline involves opening-up the system to live embodied locomotion form real human users. To support this, we first built a set of VGEs to depict different streetscapes, which we made accessible via mobile Head Mounted Displays (mHMD) to user participants (see a viewscreen from the mHMD as well as a user-participant in the studio in Fig. 10, at the bottom). The mHMD allowed users to view streetscapes in three dimensions and to listen to vehicles with spatial audio. Importantly, the mHMD enabled users to engage their own natural, physical, tangible locomotion control to advance through the VGE and to look around and gaze within the simulated scenes. We enabled this by launching the mHMD



Traffic patterns (space-time bunching) in the IDM as implemented in our model.

Integration of the IDM with other agent components and simulated streetscape.

Fig. 10 Vehicle agents in the intelligent driver model (IDM) along a model streetscape

in a studio space that users could physically traverse. Indeed, this locomotion takes place in a physical studio with one-to-one distance mapping to the VGE streetscape that is represented in the simulation. Second, we coupled the agent-pedestrian and agent-driver models to the VGE, populating the VR streetscapes with dynamic (but individual-based), crowd patterns and traffic patterns. We additionally enabled pedestrian crossing signals and traffic signals within the simulated streetscapes. Third, we recruited cohorts of real human users to engage in a series of motion control trials within the VR system. While in the simulation trials, we recorded participants' visual information (piped directly from the mHMD), their geographic information (polled from wireless georeferencing of the mHMD to lighthouse base stations), key simulation events (available from automata state data), as well as users' gaze behaviors (calculated in a post-processing step using ray-tracing). Each of these four pieces of in-simulation information were fused to a common GIS infrastructure for subsequent analysis. For geographic information, we collected in-simulation data in ways that would permit direct comparisons to GPS data from our real-world observations and diaries of embodied locomotion so that counterfoil measurements could be made. To evoke goal-driven behavior, we tasked human users with approaching a signalized crossing in the VGE and crossing through traffic. Users could choose to do so by adhering to crossing signals, ignoring signals and choosing gaps in moving traffic, or jaywalking outside crossing areas. We did not instruct users as to which strategy to select.

8 Experimental analysis of motion control in the Sidewalk2Synth pipeline

With these functionalities in the VGE, then, the pipeline for Sidewalk2Synth is in place. To test it in use-case scenarios for embodied locomotion, we built a series of experiments with live participant-users to evaluate how they would embody themselves to the VR representation of the model. This test was evaluated as the robustness of motion control in use, gauged against real-world locomotion data form our observation sets described in Sect. 3.

The experiments tested a range of streetscape scenarios for encounters that take place in small encounter-based moments of space and time. These included tasks to evoke embodied locomotion relative to:

- The urban geography and urban design of the built environment;
- Ambient pedestrian-pedestrian interactions as steering and collision avoidance;
- Social dynamics of collective assembly and crossing behavior at the crossing curbside;
- Indicators from signalized crossing infrastructure; and
- Traffic dynamics and traffic gaps.

8.1 Real-world experiments

Over an 18-month duration, we recruited human participants to wear a chest-mounted video camera, GPS, and a smart watch and asked them to record data as they engaged in their day-to-day streetscape activities around New York City and its suburbs. From this corpus of data, we extracted a series of road-crossing epochs form the broader activity-set, and we used these as validation cases for the Sidewalk2Synth pipeline. Specifically, we used crossing trajectories for urban and suburban areas, four-way crossings, end of block and mid-block crossings, signalized and unsignalized crossings, and a variety of road lanes. We also used data for different times of the week and day and night across different seasons with different streetscape physical conditions (puddles, sidewalk obstacles) and crowd conditions.

8.2 Heuristic counterfoil experiments

As a control, we implemented a series of motion control algorithms and heuristics that are popularly used in agent-based modeling of pedestrian locomotion (Fig. 11). Several of these were designed to produce mathematical motion, as Lévy flights (Bartumeus et al., 2005; Brockman et al., 2006; Viswanathan et al., 1996) and Brownian motion (Schweitzer, 1997). Others were coded to produce path-planning behavior by shortest path traversal (A* (Hart et al., 1968) and Dijkstra (1959)) with collision detection and avoidance by Moore and von Neumann cellular automata neighborhoods (Blue & Adler, 2001; von Neumann, 1951). We also implemented a social force model (Helbing & Molnár, 1995) of crowded movement through a sidewalk corridor (with walls as a bounding condition for the physical repulsion factor).

In heuristic experiments, a variety of agent populations were used, depending on the heuristic, ranging

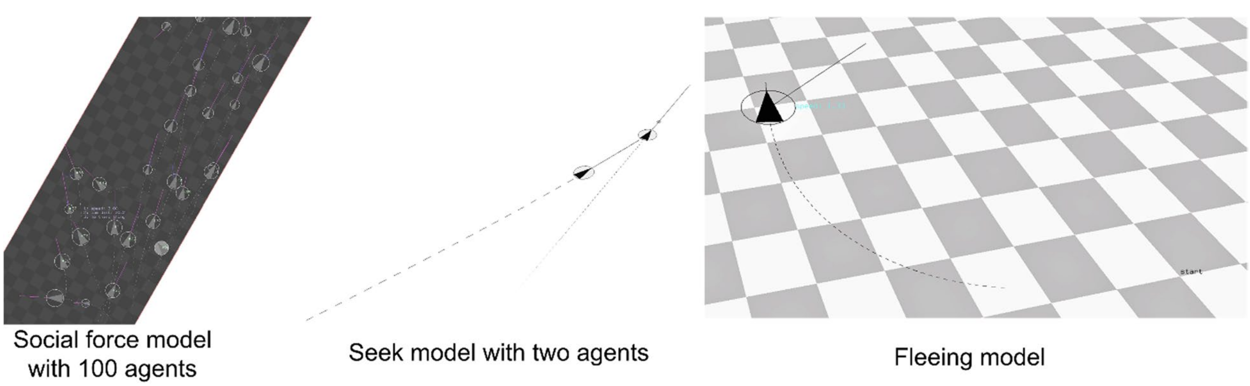


Fig. 11 Samples of the movement generated by the purely algorithmic models

from 1 ≤ n ≤ 100, depending on the simulation scenario. We use the data (spatial and temporal position for locomotion trajectories) as a point of quantitative comparison for our human trials. Primarily, the heuristics show results for specific locomotion behaviors (e.g., path pursuit, repulsion effects, collision avoidance, etc.) that are useful as benchmarks.

8.3 VR experiments

8.3.1 Streetscape geography and urban design

To support experimentation with physical embodied locomotion relative to built environment context, we developed VREs with one-to-one mapping to real world counterparts. Users of the pipeline, moving simultaneously through a VR representation of the geographic environment and through a tactile physical studio space therefore have an opportunity to engage their natural perception, cognition, action, and locomotion skills to plan paths, encounter waypoints, deploy wayfinding and navigation, steer, see and respond to collisions, and engage in ambulation. Critically, the egocentric (user-to-environment) and allocentric (environment-to-environment) distances in the VR environment were set to match those of real-world streetscapes. Specifically, we set aside a physical area of 82.72 square meters (890.34 square feet) that user-participants could move through, while rendering the same space in VR and adding synthetic space covering several city blocks in the virtual surroundings (Fig. 12).

Within the VR space, we included physical features of streetscapes that play into embodiment (Granie et al., 2014), including building façades, window fronts, doorways, awnings, vegetation, sidewalks with varying texture, marked pedestrian crossings (Havard & Willis, 2012), roadways with lane markings (Kadali & Vedagiri, 2013), traffic lights (Yang et al., 2016), pedestrian crossing lights (Lipovac et al., 2013), as well as physical lighting and shading effects (Choi et al., 2006) (Figs. 10 and 12). A unit of distance between these features in the VR representation was matched to a unit distance in the physical traversal space, and we built the scenes using LiDAR measurements from our observations of real crossings in and around New York City. These features can be swapped easily, or they may be drawn procedurally from GIS (Torrens, 2015a, 2015b). In the examples that follow we will discuss experiments for a suburban type streetscape with a signalized pedestrian light-controlled (PELICAN) mid-block crossing and two lanes of traffic. Elsewhere (Torrens & Gu, 2023), we have shown a VR environment that was built for a dense downtown streetscape with a four-way signalized crossing, designed to mimic a counterpart site in Brooklyn, NY, USA. The built setting can also be rendered in very high-resolution or at lower resolutions. In the examples that follow, we show results for a "good-enough" resolution rendering, which user experiments showed to be useful for motion control, and not too distracting vis-à-vis uncanny valley artifacts (Kim & Torrens, 2024).

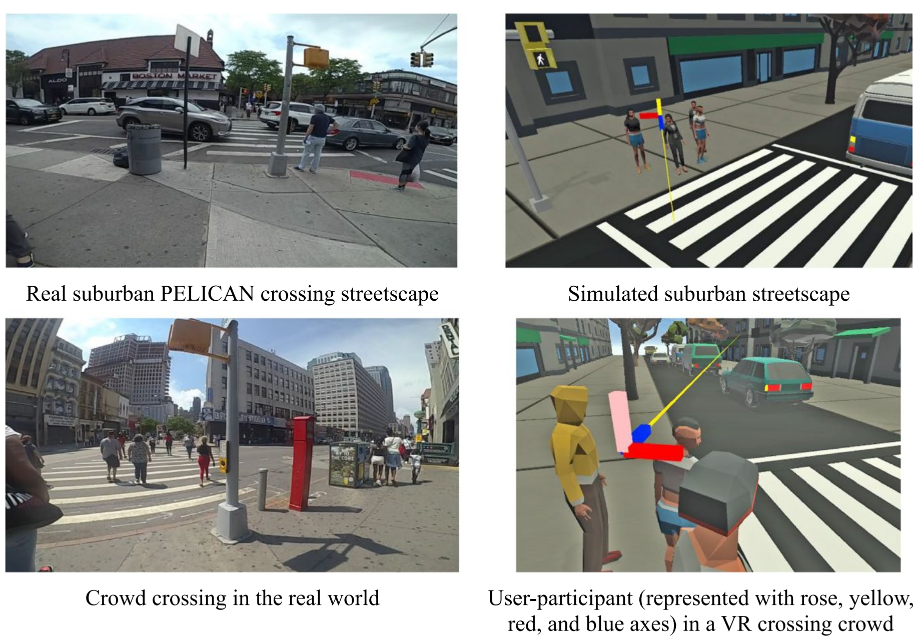


Fig. 12 The urban design for our real observations and counterfoil simulation

8.3.2 Streetscape crowding

Our observations (Sect. 3) showed that crowding at crosswalks in particular was a factor in shaping possible pedestrian embodiment to a given crossing scene. We experimented with varying levels of pedestrian crowding on sidewalks, crosswalk assembly areas at the curbside, and in crosswalks. Only a single user-participant was entered into the experiments per trial, but varying numbers of agent-pedestrians were included to create a diversity of crowding scenarios (Fig. 12). For the experiments that we report here, in suburban settings, our observational data implied that a crossing crowd of one to four pedestrians (five when adding a user-participant) matches the real world (although the size of these crowds could be varied to very large dimensions if desired).

8.3.3 Peer effects on streetscapes

To experiment with peer effects (Pfeffer & Hunter, 2013), e.g., from social influence (Faria et al., 2010; Mirzaei-Alavijeh et al., 2019), group motion (Rosenbloom, 2009), authority effects (Gutierrez et al., 2014), or inter-personal biases (Collett & Marsh, 1974), we varied the demographic makeup and appearances of the avatar representations of the agent-pedestrians. Further, we provided different profiles for motion behavior (and animation cycles) to represent hurried, slow-moving, patient, impatient, signal-abiding, and jaywalking behaviors (Figueroa-Medina et al., 2023; Sueur et al., 2013). Our experimental axes for variation were (1) risk profile (on a continuum from safe motion control by waiting for a pedestrian "walk" signal, to risky motion by crossing or attempting to cross through moving traffic during a "don't walk" signal); (2) sex profile (male or female avatar appearance); and (3) social appearance (wearing business attire or casual clothing). Specifically, risky behavior was weighted to instruct the agent to cross only when there was a traffic gap, but completely ignoring the crossing signal. This resulted in preemptive crossing epochs of up to 15 s before the "walk" signal was illuminated.

8.3.4 Streetscape signals

Our central experimental lever for the mediated (i.e., signs) space of streetscapes centered around streetscape signals: both traffic lights and pedestrian crossing lights (Lipovac et al., 2013). Observational research from other authors has shown significant variation in the cues and actions that pedestrians take in response to signalized features on streetscapes, including age and sex differences in adherence to the motion rules that they advertise (Tom & Granié, 2011), as well as variation in expected norms across different countries (Gang et al., 2011). Similarly, our own observations showed varying propensity among pedestrians to adhere to crossing signals, including alternating checks between signals and reactions of peer pedestrians in different crowd densities at crossing sites (section 11). Pedestrians' propensity to violate crossing rules and lights is particularly well-studied in the literature, and observations point to demographic variation as well as peer effects (Onelcin & Alver, 2015).

8.3.5 Streetscape traffic gaps

Pedestrians and vehicles come into perceptual contact as well as potential physical contact when walkers enter the roadway through jaywalking or at signalized crossings. When doing so, pedestrians usually make a quick judgement of the spacing and timing in traffic interactions that they may have available to effect collision-free locomotion with other pedestrians and with vehicles, which they then map to their own assessment of their capabilities to produce matching movement. This can happen in very small windows of space and time (Kadali & Vedagiri, 2013; Onelcin & Alver, 2015), or it may involve a deliberative assessment of a streetscape scene (Nesoff et al., 2018; Viola et al., 2010). The means by which pedestrians assess and decide on available gaps in traffic is well-covered in the observational literature, with results suggesting that there is significant variation based on age (Dommes & Cavallo, 2011; Zivotofsky et al., 2012), sex (Underwood et al., 2007), and skill (Liu & Tung, 2014; Oxley et al., 2005). In our simulation experiments, we simulated different traffic patterns with different gaps (Liu & Tung, 2014; Plumert et al., 2004). We also programmed different risk and decision profiles (Sueur et al., 2013) into agent locomotion to produce a range of gap acceptance behaviors for users to embody themselves to in simulation. We experimented with a range of different vehicles, acceleration profiles, traffic density, and traffic patterns (free-flow, bunching, gap-closing).

9 Analyzing motion

We considered whether we could validate the motion of users in our system relative to the observed data that we obtained via fieldwork. Doing so required that we build a common ground truth between the streetscape and simulation. Much more experimental analysis is available in our controlled studio settings than in real-world scenarios. Indeed, our controlled studio experiments widely permit motion capture, motion tracking, gaze detection and tracing, gaze fixation, timings, and ego-centric and allo-centric distance calculation. For live scenarios on real streetscapes, we were more limited. However, we can build motion trajectories and full movement paths from differentially-corrected GPS data that we have from real-world scenes and match these to commensurate data in our virtual experiments. For our VR trials, we relied on wireless location-awareness provided by range-finding between users' mHMD and a set of lighthouse base stations distributed around our studio space. Our wireless positioning has a locational accuracy of millimeters, which we verified through motion capture. GPS data, with differential correction, has a below-meter accuracy as we have (almost equidistant) proximity to continuous operating reference stations (CORS) in our GPS trace in nearby Newark, NJ and Long Island, NY.

9.1 Quantitative analysis of locomotion

To evaluate whether the Sidewalk2Synth pipeline was able to create locomotion scenarios with parity to real-world streetscape embodiment, we analyzed motion trajectories directly from our experiments with real walkers on-the-ground in New York City and its suburbs. We also measured counterfoil trajectories for user-participants in our studio-based VR experiments. Additionally, we measured heuristic motion from parallel simulations, and then compared them empirically. We note that the resolution of our VR experiments within the Sidewalk2Synth pipeline is very high (sub-centimeter in space and 120 Hz in time), while GPS signals in the real world were relatively coarse (at best, sub-meter in space and 10 Hz in time). To establish data parity ahead of analysis, we up-sampled the GPS trajectories in Hz, while preserving the original frames. We subsampled crossing epochs from our real-world trajectory corpus that had equivalent lengths that matched movement trials in the VR/studio space.

We used a variety of motion statistics to generate measurements. For step-by-step motion, we studied *correlation of adjacent turning angles* (Benhamou, 2004), as an indicator of step-by-step directional persistence over a trajectory. Low correlation values (near zero in value) may indicate a lack of statistical association between one turning angle and the next (as in Brownian motion, for example). High positive correlation values show codependence between larger step-by-step turning angles, as would occur if a pedestrian is consistently turning as they move (near + 1). High negative correlation is indicative of inverse stepping, as in cases of relatively high sinuosity along a trajectory (near –1) by shuffling or stutter-stepping. We analyzed the general directional trend of motion using *mean cosine of turning angle* (Codling et al., 2008) (p. 823). Higher cosine values indicate relatively straight movement, while lower values may be associated with sinuosity. We calculated cosine of turning angle between successive fixes on trajectories and we then averaged (mean) across the trajectory (so, directional preponderance on a per-trajectory basis). Note that generally turning angle is related to speed of movement: pedestrians likely have more capability to (instantaneously) turn at low speeds than they would have while running, although pedestrians can usually exercise considerable ability to shift direction at most speeds due to their ambulatory dexterity and ability to pivot through weight shifting, twisting, heel maneuvers, and so on, as in dance (Ada et al., 2003). We additionally estimated the *probability of turning in the same direction*, averaged on a per-step basis over the movement sample. Our movement data were collected at relatively high resolutions of space and time, with the implication that movement statistics could be calculated over several scales. To account for this, we also performed *fractal analysis* of trajectories (Nams, 2006), estimating a general linear fractal as well as a mean fractal. The mean fractal is calculated in a forward direction from the first location fix in the motion sequence, then again in a backward direction from the last location fix, taking the average of the two (Nams, 2006; With, 1994). Fractal analysis has the added benefit of allowing us to compare trajectories in our experiments with those of other real-world studies and simulation studies. Generally, the fractal dimension of a trajectory sample would near a value of + 1 for purely straight movement (Torrens et al., 2012), and tend toward a value of + 2 for (infinitely) sinuous movement (purely random walks, for example, would generally cast values toward + 2) (Bartumeus et al., 2005; Batty, 1997). A set of movement-based building blocks were also calculated as part of these compound analyses, including *step count, path length, number of steps per movement trip, steps per unit length of motion*, and *step size*, which we additionally used to assess parity of structure between simulation outputs and real-world measures. We also mention that because our virtual experiments were performed in a studio, we have access to unit distance values that map to real space. In aggregate, the measures of turning angle and of fractality yield a relative measure of (1) directional preponderance, and of (2) relative sinuosity, between motion sequences. In each case, movement statistics were implemented following (Nams, 1996).

9.2 Qualitative analysis of embodiment

We administered a set of questionnaires to the human user-participants to evaluate their personal sense of embodiment while engaged in the simulation trials. The questions were tasked with uncovering two principal factors. The first set (P-questions) evaluated how users felt embodied to the simulation as a virtual medium. The second set (R-questions) evaluated users' embodiment to the simulation as a streetscape, asking specifically how they engaged with dynamic elements of the synthetic environment. The general tone of the questions were as follows (the exact wording of the questions is shown in full in Appendix A, Figure A4).

Embodiment to a virtual medium.

- P1: Whether users felt connected to the synthetic streetscape
- P2: Whether users felt surrounded by the virtual streetscape
- P3: Whether users felt the virtual streetscape simply looked like pictures
- P4: Whether the users felt absent from the virtual streetscape
- P5: Whether the users felt that they had agency in the virtual streetscape
- P6: Whether the users had a realistic sense of navigation in the virtual streetscape
- P7: Whether the users felt compelled to pay attention in the virtual streetscape
- P8: Whether the virtual streetscape held users'captivation
- P9: Whether the virtual streetscape felt real to users
- P10: Whether the virtual streetscape held consistency with users'real world experiences
- P11: Whether users felt they could distinguish the virtual streetscape from the real world

Embodiment to streetscapes

- R1: Whether users crossed synthetic roads with behavior that they felt matched their real-world behavior
- R2: Whether users sought to avoid physical collisions with vehicles
- R3: Whether users sought to avoid physical collisions with pedestrians
- R4: Whether users obeyed crossing signals

10 Results

Our analysis of the performance of the Sidewalk2Synth pipeline with real immersed users had two principal aims. First, we sought to establish whether the model could evoke realistic embodied locomotion in applied and embodied user-participating simulation. Second, if the pipeline is indeed a useful counterfoil to real streetscape dynamics, we aimed to use it to evaluate embodied locomotion scenarios. In short, our results show that the pipeline is a close fit (but not a completely convincing fit) to reality, and that it can support experimental analysis with human users in hard-to-test scenarios with bearing to real streetscapes. Detailed empirical results are available in Appendices. An illustrative example of the results is shown in.

10.1 Can Sidewalk2Synth evoke realistic locomotion from immersed human users?

Our findings demonstrate that the Sidewalk2Synth pipeline is indeed capable of evoking locomotion from users with embodied characteristics that match to real-world counterpart streetscape encounters. Results for real-world movement ($n=20$ counterpart comparisons) are detailed in Appendix A section 11. Results for immersed user locomotion ($n=24$) on Sidewalk2Synth are provided in Appendix A section 11. Results for other agent-based locomotion heuristics ($n=$up to 100) from the movement modeling literature are shown in Appendix A section 11.

An "apples to apples" comparison shows that Sidewalk2Synth produces motion that is markedly distinct from our comparison set of motion control algorithms and heuristics. Essentially, the motion produced in Sidewalk2Synth is much more organic than would be generated by heuristic agent-based routines. The closest empirical fit was between user locomotion in Sidewalk2Synth and crowd steering behavior from the Reynolds model (Reynolds, 1987, 1999), which produced a fractal dimension of ~ 1.01 for generalized steering, and ~ 1.024 for steering by wandering behavior. The averaged fractal dimension for our real-world observations was 1.023 and for simulation it was 1.04. The fractal dimension results for other heuristics were wildly different than our observation data and our human trials in Sidewalk2Synth and we thus conclude that our pipeline is closer to reality than to generalized computational heuristics.

Nevertheless, the results from our comparisons between Sidewalk2Synth and our observation data from real streetscapes indicate that we have some work to do to get a tight fit between reality and our simulation, as metrics are closely matched in some parameters, but in discord for others. Human participants in the Sidewalk2Synth simulations tended, on average, to move with less steps per unit length than their counterparts moving through real urban spaces (an average of 35.478 through the simulation, and 46.482 in real spaces). Average mean cosine results were relatively comparable between reality and simulated environments (average of 0.941 and 0.946 respectively). However, the average probability of turning in the same direction was much higher in simulation than in the real world (0.765, compared to 0.205 for the real world). The average correlation for adjacent turning angles was higher for reality than in the simulation, and the correlation in reality was negative (which is indicative of peripatetic turning on a step-by-step basis). The average fractal dimension for simulation was also higher than we recorded for movement in real spaces: 1.023 for reality and 1.04 for simulation, average across all paths. This means that movement in the simulation is more sinuous than in reality, per-trip: even if the step-by-step results show relative persistence on locomotion, small adjustments can add-up to a larger fractal dimension for the longer trajectory of locomotion. Although, we note that compared to our agent heuristic comparisons, Sidewalk-2Synth was a much closer fit to reality.

Here, we offer some caveats. First, the simulation scenarios contain some relatively extreme test cases of risky behavior, with many events of user jaywalking and dashing between unsafe traffic gaps. The real-world metrics are therefore actually a better fit to the safe scenarios in our simulations than they are too risky scenarios (for which we do not have observation data, due to the hazards involved in real life). Second, there are some resolution disparities between simulation data and real-world GPS data that reduce the efficacy of some of the trajectory statistics. This is a by-product of the synthetic resolution of our model output, which could theoretically approach infinite fineness. Third, we point out to the reader that our simulation scenarios are a much closer fit to reality than the companion heuristics from other models that we tested. Fourth, we note that there is a broad variation in results of movement statistics for different streetscape types. Our real-world experiments showed a shift across results when compared for residential and non-residential streetscapes, for example.

10.2 Did users feel embodied to Sidewalk2Synth?

Our evaluation of user embodiment while engaging in locomotion trials in the simulation showed that users did indeed seem to be convinced by Sidewalk2Synth's ability to support embodied locomotion. In other words, Sidewalk2Synth supported verisimilitude in users' embodied locomotion: against the caveat that of course the simulation is synthetic, the combination of virtual embodiment scenarios delivered against real physical locomotion in a studio setting allowed users to behave naturally with a convincing connection to the real world.

Results from user questionnaires are shown in aggregate form in Appendix A, section 11. Users were invited to score their responses with negative, neutral, or positive valence on a Likert scale. These valences are color-coded in Appendix A, section 11 and the total number of responses per score are listed as integers. Users'answers to the presence questions indicated that they felt "there" in the synthetic streetscape (P1), and indicated that the felt surrounded by the streetscape (P2) (i.e., they had strong ego-centric presence) and that they lost sense of the studio environment while engaged with the mHMDs (P6). Not surprisingly, users were aware that the simulation was virtual (P11). Users did not indicate a negative sentiment toward the pipeline's presence factors when directly questioned on that topic (P3 and P4). Users also felt that they could act-out their behavior in the simulated streetscape (P5) and they felt engaged by the simulation to pursue their behavior (P7 and P8).

10.3 Using Sidewalk2Synth to test different scenarios for embodied locomotion

If the pipeline is indeed trustworthy as a useful counterfoil to reality, can we then test different motion control scenarios of real, live, behaving participant-pedestrians in it? There is a huge swath of potential experimental levers that we could "pull" in virtual form, to evaluate what their impact might be on streetscape policies in reality. Here, we examine two such levers, which have currency in the theoretical and case study literature for streetscape science: crowd size and peer effects. Both come into convergence in the form of crossing safety, which the conceptual literature has associated as being a phenomenon that could potentially frame and explain recent increases in road-crossing harm that have steadily been growing despite dedicated efforts to safeguarding sites on streetscapes via Vision Zero (City of New York 2019) and other campaigns. Experimentation with road-crossing is all but infeasible in reality due to the dangers involved. As such, virtual streetscapes have several experimental value platforms, but need to be realistic at micro-scales of streetscape environments, individual behaviors, and interpersonal dynamics of pedestrians to be reliable as a platform for testing.

10.3.1 Crowd size

In simulation, we established a range of scenarios to vary the crowd size at crossings. These ranged from dyads to larger groups ("crossing pelotons"). The crowding scenarios showed that immersed user-participants in relatively dense pelotons tended to move in largely the same manner as those in dyads. Average mean cosine for low density (0.951) and higher density (0.941) were a close match, as were the probabilities of turning in the same direction (0.773 for dyads, 0.758 for pelotons), average correlation of adjacent turning angles (0.565 for dyads, 0.521 for pelotons), and average fractal dimension (1.044 for dyads, 1.046 for pelotons). However, human participants that crossed with pelotons used more steps per unit length (40.642 on average) than those in dyads (38.848 on average). One explanation is that dense pelotons create more potential collisions, and participant stutter-stepping is used in locomotion as a response.

10.3.2 Peer effects

We established a set of simulation scenarios with varying mixtures of risk-taking among agents. Risk-averse agents would wait to cross a road only at a signaled crossing site, and they would adhere to crossing signals for "walk" and "don't walk" signals (green pedestrian icon and red hand icon respectively). Risk-taking agents were apt to ignore these signals, would move quickly to cross a road, and would also accept very tight gaps in moving traffic while jaywalking. These groupings were introduced as scenarios to test whether human participants would peer-adopt risk-taking behavior or risk-aversion as a norm in their own crossing decisions.

For human participants immersed in the simulation, risky-behaving agent peer groupings were associated with a lower number of steps per unit length of locomotion taken by immersed human participants in the simulation scenarios. Average steps per length for risk-averse peer environments were 47.601, while they were on average 32.512 and 37.802 respectively for mixed-risk and risk-taking peer group crossing contexts. Participants moved faster and hesitated more in risky peer groups than they did among risk-averse peers. Average mean cosine was higher when exposed to risk-taking peers (averaged as 0.935 for safe peers and climbing to 0.939 for mixed-risk peers and then again to an average of 0.95 for risky peer contexts). The probability of turning in the same direction on a step-by-step basis increased as human participants were exposed to risky peers: averages were 0.751 for risk-averse peer context, 0.76 for mixed-risk contexts, and 0.776 around risk-taking peers. Fractal dimension results for different scenarios of risk context were relatively stable across experiments (averaging 1.041 to 1.046). These results show that human participants adopted straighter movement paths when with risk-taking peers than they did around risk-averse agents. We interpreted this as indicating that human participants (perhaps blindly) followed risk-taking peers at crossings. This peer effect was also apparent when we reviewed playback animations of the participant trials.

11 Automating Sidewalk2Synth as an end-to-end pipeline

Development of Sidewalk2Synth from first principles involves a concerted effort to sweep through data collection, model development, user studies, and simulation experiments. To examine whether we might be able to encapsulate Sidewalk2Synth in a portable and generalizable pipeline, we developed a preliminary automated version. This is designed to run, end-to-end (i.e., from live data feeds, through modeling, via simulation, to output) in real-world contact scenarios. Our prototype automation takes on two forms: a container-based solution that is designed to run on edge devices, and a portable augmented reality version that is intended to run on tablets and phones.

11.1 Edge Sidewalk2Synth

Edge devices are growing in popular use for smart city applications. Generally, they consist of relatively light computer hardware (typically, system-on-chip (SoC) devices) that contain a power source, a small motherboard, a low-power CPU or GPU, and wireless networking. Edge devices typically run containerized software and firmware that can be tailored to particular applications, such as computer vision. Usually, edge devices are federated into an array of near-phenomena computing and they are tasked to perform preliminary analyses near streetscapes, for example, and to pass results (but not necessarily full video streams) to a centralized computer for down-stream activity such as multi-site simulation. In this configuration, they are referred to as Wi-Edge (Torrens, 2022a, 2022b), working wirelessly in concert to perform edge computing or edge artificial intelligence, typically over high-bandwidth and low-latency (HB/LL) network architectures.

For our prototype, we set up the front-end of Sidewalk2Synth (video input and machine-learning) and partially implemented the back-end (trajectory prediction from learned bounding boxes and video frame association) on Wi-Edge. In this configuration, the agent-based pedestrian and driver behavior are sidestepped completely. Our results show that Edge Sidewalk2Synth can work in real-time on busy streetscapes (Fig. 13, section 11). Detected crossing signals, pedestrians, pedestrian pose-graphs, pedestrian trajectories, and streetscape atlases could potentially be passed on to centralized simulation running on a CPU or cluster, feasibly with edge input across multiple sites in a city. (We note that we only tested edge Sidewalk2Synth on one edge device for this paper, but elsewhere (Vyas & Torrens, 2024) we have shown it working across many devices by federated learning.)

11.2 Augmented Sidewalk2Synth

Our second automation prototype tackles the missing components of Edge Sidewalk2Synth, chiefly the lack of a live run-time agent-based simulation. We implemented an augmented reality (AR) version of the pipeline, Augmented Sidewalk2Synth, running completely in real-time via Unity on Apple iOS (Torrens & Gu, 2023). In the AR implementation, a light version of the pedestrian agents and driver agents runs. Computer vision and manual geo-fencing is used to interpret the VGE directly from a tablet camera. We ran detections for sidewalks, roads, pedestrian crossings, street furniture, lampposts, traffic signs, and buildings. Agents will steer to avoid collision objects, other agents, and agent-driven vehicles. They will also cross only at crosswalks. The agent system is rendered via Unity over the live camera video feed,

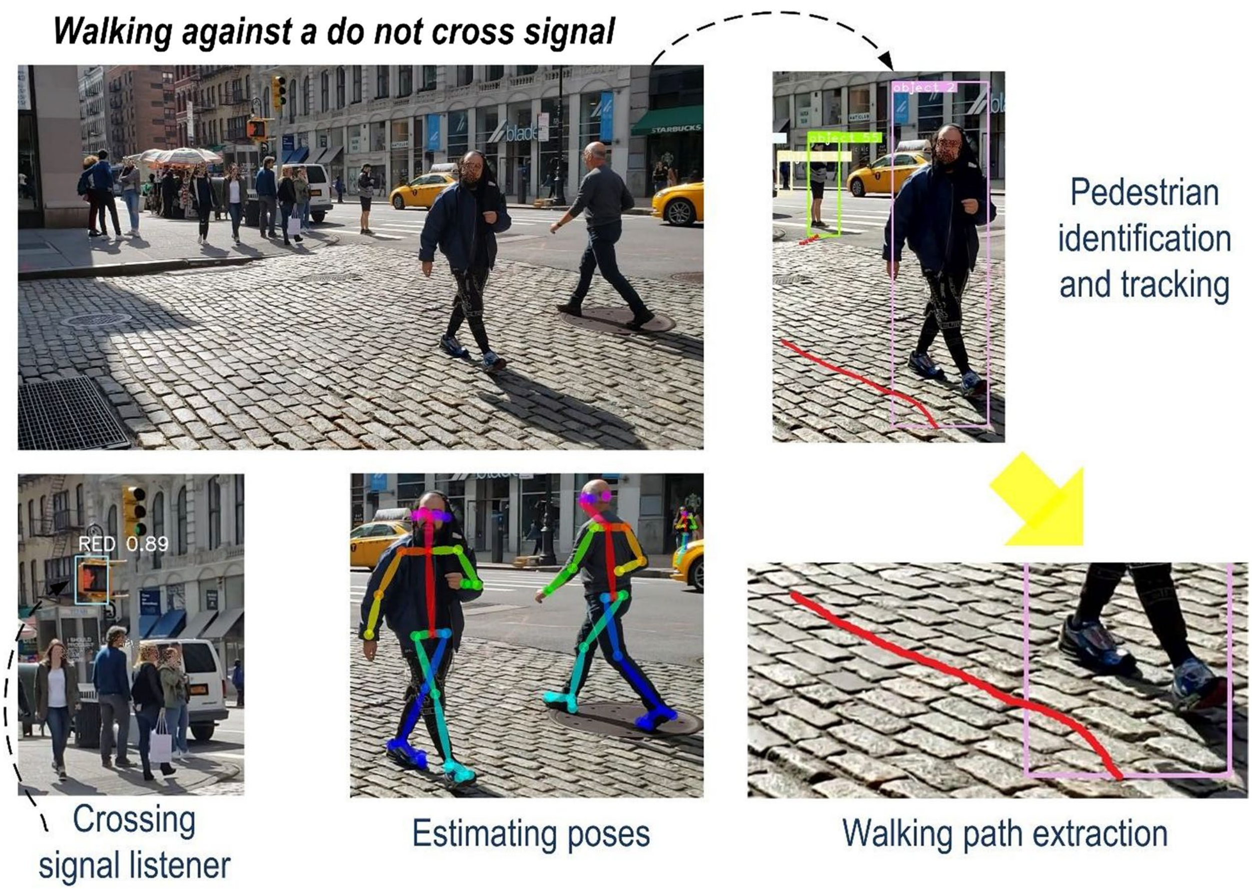


Fig. 13 Edge AI deployment (via Wi-Edge) of deep learning for detection of locomotion during road-crossing, and embodiment to crossing signals (Potdar & Torrens, 2019; Torrens, 2022b)

Fig. 14 Augmented reality version of Sidewalk2Synth running on a nighttime streetscape in downtown Brooklyn, NY, USA. Images are direct screenshots from an Apple tablet

providing a mixed-reality (MR) mash-up of Sidewalk2Synth and the live scene. In this way, users can embody themselves directly in the pipeline, running what-if scenarios on top of live streetscape dynamics (Fig. 14).

12 Conclusions and future work

In this paper, we have evaluated whether simulations might be drawn to closer parity with real-world dynamics for streetscape applications. Our approach leans on two dimensions of this query: whether human movement through streetscapes might be represented in higher-fidelity, and whether one might be able to represent embodiment in simulation. We consider both factors as convergent around the topic of embodied locomotion, and we established an experimental protocol to collect data on real-world embodied locomotion, to pipe those data into model form, and to generate dynamic simulations with the products. We also examined whether real human users could be brought "into" such simulations, via embodied locomotion, for the purposes of experimentation with virtual scenarios. We approached user embodiment in two fused formats: first, we allowed users to move with real locomotion through a studio setting, where they could walk, wander, and look around with their natural behavior; and second, we developed a high-fidelity streetscape simulation that we could deliver to users as embodied VR. An obvious question is whether this works with a reasonable match to real-world embodied locomotion, and we evaluated this match by testing simulated and real locomotion against real-world trajectory samples from streetscapes around New York City and its suburbs. We also engaged users with questionnaires to evaluate their sense of embodiment in the simulation and its authenticity relative to their real world encounters.

A limitation of our approach is that it only considers one-way pedestrian movement (a human user and its surrounding peer dyad and group crossing together without ongoing pedestrian traffic from the other side of the street). This approach was used to control for influence on gap determination, gap acceptance, and gap action. However, it is unrealistic in dense urban areas. In a related paper we have implemented a different version of the simulation with two-way crossing and crowd formation along sidewalk segments (Torrens & Gu, 2021, 2023). Integrating that crowd-based approach into Sidewalk2Synth will be a task for future development. Critically, we need more physical, tangible studio floor space to build user experiences that can stretch over two sidewalk segments and a roadway. Our proposed solution is to use redirected walking (Razzaque et al., 2002; Sun et al., 2016, 2018) as a way to fold the virtual space of the simulation to fit within a smaller physical space.

Our results show that a pipeline that moves from observation, through models, to simulation and user interaction is feasible and we introduced a detailed methodology for how that can be accomplished using mixtures of sensing technology, deep learning, agent AI, and VR graphics, with GIS providing support. Our analysis points to quantitative and qualitative evidence that users engage simulated streetscapes with embodied locomotion that matches reasonably and sensibly to the real world. We also showed a preliminary proof-of-concept using Sidewalk2Synth to evaluate road-crossing scenarios with simulated dangers.

Several future questions remain open to exploration around the ideas that we have advocated for in this paper, and we discuss them both briefly and partially here with hope that readers may be interested in interpreting them in applications that intrigue them. The first relates to the concept of embodiment, which is broadly interpreted in conceptual form in the existing literature (see Sect. 2). We focused our development efforts on embodied locomotion, but other facets of embodiment could be studied with similar observational and simulation-based protocols, particularly social embodiment (which we reason could be very usefully examined through NVCs as motion capture data or deep learning on poses). A second promising vector for future study could focus around automating the sorts of pipelines that we have shown in Sidewalk2Synth. Our early prototyping has shown that components of the Sidewalk2Synth technology stack can be ported to edge devices as a form of Edge AI and of mixed-reality AR. One could feasibly imagine federations of these interoperating as Wi-Edge (Torrens, 2022a, 2022b, 2023) over federated learning (Vyas & Torrens, 2024), with the ability to generate widespread streetscape insight from hyper-local epochs of embodiment. In ongoing work, we are developing a fused system that will run Sidewalk2Synth seamlessly over edge devices and AR, via a federated simulation. Third, it is feasible that aspects of the pipeline that we demonstrated here could be ported to wearable technologies, particularly to AR glasses. We have shown a prototype of the agent AI simulation stack that works as AR in (Torrens & Gu, 2021, 2023). The prospect of turning AR-based cameras outward to the embodied streetscape, while an edge device, on-device computing, or off-site computing reasons on the ego-centric viewshed of an individual pedestrian, and returns personalized and localized insight that is both context-aware and situation-aware is potentially a very promising area of development and inquiry, with significant potential returns to the conceptual literature in behavioral geography in particular. Fourth, it is feasible that pipelines that flit easily from reality to simulation and back, involving real human users at parity to the space–time scales of their lived experiences could suggest a new era for geosimulation (Benenson & Torrens, 2004; Torrens, 2004) in particular. In this paper, we aimed to develop simulations with high realism. Our intended output was therefore not an abstraction of reality, but rather an *alternative reality*. We note that this is an approach that differs from deep learning (LeCun et al., 2015) and existing forms of Sim2Real modeling ((Doersch & Zisserman, 2019)which is by contrast focused on pattern-matching as an outcome, rather than fidelity of behavior). Approaches such as Sidewalk2Synth, instead, aim toward new veins of modeling that are more aligned with artificial general intelligence (AGI) (LeCun, 2022). An obvious question, perhaps, is whether a foundation model for embodied locomotion might be in reach and if so whether it could inform human applications, or even applications in robotics.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s44212-025-00081-z.

Additional file 1.

Authors' contributions
PT and RK contributed equally to all components of the research and analysis. PT wrote the manuscript.

Funding
No funding was involved in this project.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Consent to participate
Data collection and human subjects experiments were conducted under an approved protocol from the institution's Institutional Review Board. All participants provided informed consent to participate in data collection and experiments. Some video and image data were collected in public spaces: in New York City, such recording is permitted.

Competing interests
The authors declare no competing interests.

Received: 6 April 2025 Revised: 5 June 2025 Accepted: 28 June 2025
Published online: 04 August 2025

References
Ada, M., Suda, K., & Ishii, M. (2003). Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and Motor Skills, 97*(3), 697–708.
Adams, F. (2010). Embodied cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 619–628.
Aiello, J.R., and Thompson, D.E. (1980). Personal space, crowding, and spatial behavior in a cultural context. In Human Behavior and Environment, Volume 4. Environment and Culture, eds. I. Altman, J. F. Wohlwill and A. Rapoport, 107-178. New York: Pienum Press.
Andersen, P. A. (2008). *Nonverbal Communication: Forms and functions* (2nd ed.). Waveland Press.
Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149*(1), 91–130.
Arikan, O., & Forsyth, D. A. (2002). Interactive motion generation from examples. *ACM Transactions on Graphics, 21*(3), 483–490.
Ashby, W. R. (1958). Requisite variety and its implications for the control of complex systems. *Cybernetica, 1*(2), 83–99.
Badler, N. I., Barsky, B. A., & Zeltzer, D. (1991). *Making them move: Mechanics, control, and animation of articulated figures*. Morgan Kauffman.
Badler, N.I., Bindiganavale, R., Granieri, J.P., Wei, S., and Zhao, Z. 1994. Posture interpolation with collision avoidance. In *Proceedings of Computer Animation'94, Geneva, Switzerland, May 25–28, 1994*, ed. L. O'Connor, 13–20. Geneva: IEEE.
Badler, N. I., Manoochehri, K. H., & Walters, G. (1987). Articulated figure positioning by multiple constraints. *Computer Graphics and Applications, 7*(6), 28–38.
Baker, R., & Hart, H. M. (2013). *Measuring Walking: A Handbook of Clinical Gait Analysis*. Mac Keith Press.
Barata, C., Nascimento, J. C., Lemos, J. M., & Marques, J. S. (2021). Sparse motion fields for trajectory prediction. *Pattern Recognition, 110*(1), Article 107631.
Bartumeus, F., da Luz, M. G. E., Viswanathan, G. M., & Catalan, J. (2005). Animal search strategies: A quantitative random-walk analysis. *Ecology, 86*(11), 3078–3087.
Basili, P., Sağlam, M., Kruse, T., Huber, M., Kirsch, A., & Glasauer, S. (2013). Strategies of locomotor collision avoidance. *Gait & Posture, 37*(3), 385–390.
Batty, M. (1997). Predicting where we walk. *Nature, 388*(6637), 19–20.
Batty, M., Desyllas, J., & Duxbury, E. (2003). Safety in numbers? Modelling crowds and designing control for the Notting Hill carnival. *Urban Studies, 40*(8), 1573–1590.
Benenson, I., and Torrens, P.M. 2003. Geographic Automata Systems: A new paradigm for integrating GIS and geographic simulation. In *Proceedings of the Seventh International Conference on Geocomputation, University of Southampton, Southampton, UK, 8–10 September, 2003*, ed. D. Martin, 367–370. Southampton: Geocomputation CD-ROM.
Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automata-Based Modeling of Urban Phenomena*. John Wiley & Sons.
Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology, 229*(2), 209–220.
Bhargava, M., & Donthu, N. (1999). Sales response to outdoor advertising. *Journal of Advertising Research, 39*(4), 7–18.
Bhattacharyya, R., Wulfe, B., Phillips, D. J., Kuefler, A., Morton, J., Senanayake, R., & Kochenderfer, M. J. (2022). Modeling human driving behavior through generative adversarial imitation learning. *IEEE Transactions on Intelligent Transportation Systems, 24*(3), 2874–2887.
Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives, 12*(3), 151–170.
Bilro, P., Barata, C., and Marques, J.S. 2022. Pedestrian Trajectory Prediction Using LSTM and Sparse Motion Fields. In *Proceedings of the 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, August 21–25, 2022*, eds. M. Jenkin, H. I. Christensen and C.-L. Liu, 3529–3535. Los Alamitos, CA: IEEE.
Blue, V., & Adler, J. (2001). Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B, 35*, 293–312.
Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., and Zhang, J. 2016. End to end learning for self-driving cars. *arXiv preprint* arXiv:1604.07316.
Bower, M. (2014). Affectively driven perception: Toward a non-representational phenomenology. *Husserl Studies, 30*(1), 225–245.
Brockman, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature, 439*, 462–465.
Cadman, L. (2009). Autonomic non-representational theory/non-representational geographies. In N. Thrift & R. Kitchin (Eds.), *The International Encyclopedia of Human Geography* (Vol. 7, pp. 456–463). Elsevier.
Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing, 9*(4), 249–267.

Camara, F., Bellotto, N., Cosar, S., Nathanael, D., Althoff, M., Wu, J., Ruenz, J., Dietrich, A., & Fox, C. W. (2020). Pedestrian models for autonomous driving Part I: Low-level models, from sensing to tracking. *IEEE Transactions on Intelligent Transportation Systems, 22*(10), 6131–6151.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint* arXiv:1812.08008.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. 2017. Realtime multi-person 2D pose estimation using Part Affinity Fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 17), Honolulu, HI, July 21–26, 2017*, eds. R. Chellappa, Z. Zhang and A. Hoogs, 7291–7299. Los Alamitos, CA: IEEE.

Carol, H., and Roslyn, H. 2007. The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations. *Accident Analysis \& Prevention* 39 (2):224–237.

Chen, M., Claramunt, C., Çöltekin, A., Liu, X., Peng, P., Robinson, A. C., Wang, D., Strobl, J., Wilson, J. P., Batty, M., Kwan, M.-P., Lotfian, M., Golay, F., Joost, S., Ingensand, J., Senousi, A. M., Cheng, T., Bandrova, T., Konecny, M., … Lü, G. (2023). Artificial intelligence and visual analytics in geographical space and cyberspace: Research opportunities and challenges. *Earth-Science Reviews, 241*, Article 104438.

Chen, M., Sheng, Y., Wen, Y., Sheng, J., & Su, H. (2008). Virtual geographic environments oriented 3D visualization system. *Journal of System Simulation, 20*(19), 5105–5108.

Chen, Z., Jingjing, Y., Xintao, L., & Guo, Z. (2024). Reinterpreting activity space in tourism by mapping tourist-resident interactions in populated cities. *Tourism Recreation Research, 49*(4), 797–811.

Chib, P.S., and Singh, P. 2024. Lg-traj: Llm guided pedestrian trajectory prediction. *arXiv preprint* arXiv:2403.08032.

Choi, A.-S., Kim, Y.-O., Oh, E.-S., & Kim, Y.-S. (2006). Application of the space syntax theory to quantitative street lighting design. *Building and Environment, 41*(3), 355–366.

Choi, S., Kim, J., & Yeo, H. (2021). TrajGAIL: Generating urban vehicle trajectories using generative adversarial imitation learning. *Transportation Research Part c: Emerging Technologies, 128*(1), Article 103091.

Chrisley, R. (2003). Embodied artificial intelligence. *Artificial Intelligence, 149*(1), 131–150.

Chung, M.-J., & Wang, M.-J.J. (2010). The change of gait parameters during walking at different percentage of preferred walking speed for healthy adults aged 20–60 years. *Gait & Posture, 31*(1), 131–135.

Ciolek, T. M. (1983). The proxemics lexicon: A first approximation. *Journal of Nonverbal Behavior, 8*(1), 55–79.

City of New York 2019. Vision Zero. *City of New York*, June 18, 2019. Last accessed June 18 2019. Available from https://www1.nyc.gov/site/visionzero/index.page.

Clark, A. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press (USA).

Clough, P. T. (2008). The affective turn: Political economy, biomedia and bodies. *Theory, Culture & Society, 25*(1), 1–22.

Codling, E. A., Plank, M. J., & Benhamou, S. (2008). Random walk models in biology. *Journal of the Royal Society Interface, 5*(25), 813–834.

Coleman, J. S., & James, J. (1961). The equilibrium size distribution of freely-forming groups. *Sociometry, 24*(1), 36–45.

Collett, P., & Marsh, P. (1974). Patterns of public behavior: Collision avoidance on a pedestrian crossing. *Semiotica, 12*(4), 281–300.

Collett, P., and Marsh, P. 1981. Patterns of public behavior: Collision avoidance on a pedestrian crossing. In *Nonverbal Communication, Interaction and Gesture. Selections from Semiotica. (Approaches to Semiotics; 41)*, ed. A. Kendon, 199–217. The Hague: Mouton.

Çöltekin, A., Lochhead, I., Madden, M., Christophe, S., Devaux, A., Pettit, C., Lock, O., Shukla, S., Herman, L., and Stachoň, Z. 2020. Extended reality in spatial sciences: a review of research challenges and future directions. *ISPRS International Journal of Geo-Information* 9 (7):Online 439.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 26-July 1, 2016*, eds. R. Bajcsy, F.-F. Li and T. Tuytelaars, 3213–3223. Los Alamitos, CA: IEEE.

Corporation, V. (2004). *HλLF-LIFE2 (Half Life 2)*. Valve Corporation.

Costa, P., Fernandes, H., Vasconcelos, V., Coelho, P., Barroso, J., and Hadjileontiadis, L. 2011. Landmarks detection to assist the navigation of visually impaired people. In *Human-Computer Interaction: Towards Mobile and Intelligent Interaction Environments: 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9–14, 2011, Proceedings, Part III*, ed. J. A. Jacko, 293–300. Berlin, Heidelberg: Springer Berlin Heidelberg.

Couclelis, H., Golledge, R., Gale, N., & Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology, 7*(2), 99–122.

Cresswell, T. (1999). Embodiment, power and the politics of mobility: The case of female tramps and hobos. *Transactions of the Institute of British Geographers, 24*(2), 175–192.

Crouch, D. (2000). Places around us: Embodied lay geographies in leisure and tourism. *Leisure Studies, 19*(2), 63–76.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*(5), 353–356.

Cutting, J. E., Springer, K., Braren, P. A., & Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology: General, 121*(1), 41–72.

Cutting, J. E., Vishton, P. M., & Braren, P. A. (1995). How we avoid collisions with stationary and moving obstacles. *Psychological Review, 102*(4), 627–651.

Das, S., Manski, C. F., & Manuszak, M. D. (2005). Walk or wait? An empirical analysis of street crossing decisions. *Journal of Applied Econometrics, 20*(4), 529–548.

Davidson, J., & Milligan, C. (2004). Embodying emotion sensing space: Introducing emotional geographies. *Social & Cultural Geography, 5*(4), 523–532.

De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7*(3), 242–249.

de Valera, T. (1986). Grafton Street: A collage of time and people. *Dublin Historical Record, 39*(4), 122–131.

Devlin, A. S., & Bernstein, J. (1995). Interactive wayfinding: Use of cues by men and women. *Journal of Environmental Psychology, 15*, 23–38.

Dijkstra, E. W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik, 1*, 269–271.

Dodge, M., & Kitchin, R. M. (2005). Code and the transduction of space. *Annals of the Association of American Geographers, 95*(1), 162–180.

Doersch, C., and Zisserman, A. 2019. Sim2real transfer learning for 3D human pose estimation: motion to the rescue. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, Canada, December 8–14, 2019*, eds. H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc and E. B. Fox, 12949–12961. Red Hook, NY, USA: Curran Associates Inc.

Dommes, A., & Cavallo, V. (2011). The role of perceptual, cognitive, and motor abilities in street-crossing decisions of young and older pedestrians. *Ophthalmic and Physiological Optics, 31*(3), 292–301.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. 2017. CARLA: An open urban driving simulator. *arXiv preprint* arXiv:1711.03938.

Dourish, P. (2001). *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press.

Downs, R. M., & Stea, D. (1974). *Image and Environment: Cognitive Mapping and Spatial Behavior*. Aldine Publishing Company.

Duan, J., Yu, S., Tan, H. L., Zhu, H., & Tan, C. (2022). A survey of embodied AI: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence, 6*(2), 230–244.

Edensor, T. (2012). *Geographies of Rhythm: Nature, Place, Mobilities and Bodies*. Ashgate Publishing Ltd.

Ekman, P., & Friesen, W. V. (1981). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. In A. Kendon, T. A. Sebeok, & J. Umiker-Sebeok (Eds.), *Nonverbal Communication, Interaction, and Gesture: Selections from SEMIOTICA* (pp. 57–106). Mouton Publishers.

Evans, G. W., Skorpanich, M. A., Gärling, T., Bryant, K. J., & Bresolin, B. (1984). The effects of pathway configuration, landmarks and stress on environmental cognition. *Journal of Environmental Psychology, 4*(4), 323–335.

Fabbri, M., Brasó, G., Maugeri, G., Cetintas, O., Gasparini, R., Ošep, A., Calderara, S., Leal-Taixé, L., and Cucchiara, R. 2021. MOTSynth: How can synthetic data help pedestrian detection and tracking? In *Proceedings of the 18th IEEE/CVF International Conference on Computer Vision (ICCV), (virtual), October 10–17, 2021*, eds. T. Berg, J. Clark, Y. Matsushita and C. J. Taylor, 10849–10859. Los Alamitos, CA, USA: IEEE.

Faria, J. J., Krause, S., & Krause, J. (2010). Collective behavior in road crossing pedestrians: The role of social information. *Behavioral Ecology, 21*(6), 1236–1242.
Farley, A., Zand, M., and Greenspan, M. 2023. 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, June 6–8, 2023*, eds. K. Li and H.-C. Lin, 169–176. Los Alamitos, CA, USA: IEEE.
Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27–34.
Fernandez, A. A., & Deneubourg, J. L. (2011). On following behaviour as a mechanism for collective movement. *Journal of Theoretical Biology, 284*(1), 7–15.
Figueroa-Medina, A. M., Valdés-Díaz, D., Colucci-Ríos, B., Cardona-Rodríguez, N., & Chamorro-Parejo, A. (2023). Analysis of walking speeds and success rates on mid-block crossings using virtual reality simulation. *Accident Analysis & Prevention, 183*, Article 106987.
Fiorini, P., & Shiller, Z. (1998). Motion planning in dynamic environments using velocity obstacles. *The International Journal of Robotics Research, 17*(7), 760–772.
Foltête, J.-C., & Piombini, A. (2007). Urban layout, landscape features and pedestrian usage. *Landscape and Urban Planning, 81*(3), 225–234.
Franěk, M. (2013). Environmental factors influencing pedestrian walking speed. *Perceptual and Motor Skills, 116*(3), 992–1019.
Franklin, S., and Graesser, A. 1997. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent Agents III: Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages, (ATAL'96), Lecture Notes in AI*, eds. J. P. Mller, M. J. Wooldridge and N. R. Jennings: Springer-Verlag.
Fujimura, K., & Samet, H. (1993). Planning a time-minimal motion among moving obstacles. *Algorithmica, 10*(1), 41–63.
Galea, E.R., Owen, M., and Lawrence, P. 1996. The EXODUS Model. *Fire Engineers Journal* 56 (183 (July)):26–30.
Gang, R., Zhuping, Z., Wei, W., Yong, Z., & Weijie, W. (2011). Crossing behaviors of pedestrians at signalized intersections: Observational study and survey in China. *Transportation Research Record, 2264*(1), 65–73.
Garbrecht, D. (1971). Pedestrian paths through a uniform environment. *Town Planning Review, 42*(1), 71–85.
Gärling, T., & Gärling, E. (1988). Distance minimization in downtown pedestrian shopping. *Environment and Planning A, 20*(4), 547–554.
Gérin-Lajoie, M., Richards, C. L., Fung, J., & McFadyen, B. J. (2008). Characteristics of personal space during obstacle circumvention in physical and virtual environments. *Gait & Posture, 27*(2), 239–247.
Geruschat, D. R., Hassan, S. E., & Turano, K. A. (2003). Gaze behavior while crossing complex intersections. *Optometry and Vision Science, 80*(7), 515–528.
Gibson, J. J. (1950). *The perception of the visual world*. Houghton-Mifflin.
Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton-Mifflin.
Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton, Mifflin and Company.
Gleicher, M. 1998. Retargetting motion to new characters. In *Proceedings of SIGGRAPH 98: The 25th Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, July 19–24, 1998*, eds. C. C. Yang and T. M. Murali, 33–42. New York: Association for Computing Machinery.
Goldhammer, M., Hubert, A., Koehler, S., Zindler, K., Brunsmann, U., Doll, K., and Sick, B. 2014. Analysis on termination of pedestrians' gait at urban intersections. In *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), October 8–11, 2014, Qingdao, China*, ed. ITSC, 1758–1763. Los Alamitos, CA, USA: IEEE.
Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences, 13*(4), 154–159.
Golledge, R. A. (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. The Johns Hopkins Press.
Golledge, R., & Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. The Guilford Press.
Gorrini, A., Bandini, S., & Sarvi, M. (2014). Group dynamics in pedestrian crowds. *Transportation Research Record: Journal of the Transportation Research Board, 2421*(1), 51–56.
Granie, M.-A., Brenac, T., Montel, M.-C., Millot, M., & Coquelet, C. (2014). Influence of built environment on pedestrian's crossing decision. *Accident Analysis & Prevention, 67*(1), 75–85.
Griffin, W. A. (2018). pyObs: Open-source software for computer-assisted behavioral observation coding. *Journal of Open Research Software, 6*(1), 5.
Griffin, W.A., Schmidt, S.K., Nara, A., Torrens, P.M., and Sechler, C. 2007. Integrating ABM and GIS to model typologies of playgroup dynamics in preschool children. In *Proceedings of Agent 2007 (North American Association for Computational Social and Organizational Sciences), Evanstaon, IL, USA, November 12–14, 2007*, eds. M. North, C. Macal and D. Sallach, 17–24. Evanston, IL: Argonne National Labs and the University of Chicago.
Güler, R.A., Neverova, N., and Kokkinos, I. 2018. Densepose: dense human pose estimation in the wild. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, June 18–22, 2018*, eds. M. Brown, B. Morse and S. Peleg, 7297–7306. Los Alamitos, CA: IEEE.
Guo, Z., & Liu, X. (2024). How artificial intelligence cooperating with agent-based modeling for urban studies: A systematic review. *Transactions in GIS, 28*(3), 654–674.
Gutierrez, C. M., Slagle, D., Figueras, K., Anon, A., Huggins, A. C., & Hotz, G. (2014). Crossing guard presence: Impact on active transportation and injury prevention. *Journal of Transport & Health, 1*(2), 116–123.
Guy, S.J., Chhugani, J., Kim, C., Satish, N., Lin, M., Manocha, D., and Dubey, P. 2009. ClearPath: highly parallel collision avoidance for multi-agent simulation. E. Grinspun and J. Hodgins. Paper read at ACM SIGGRAPH/Eurographics Symposium on Computer Animation, August 1–2, 2009, at New Orleans.
Hans, A., and Hans, E. 2015. Kinesics, haptics and proxemics: Aspects of non-verbal communication. *IOSR Journal of Humanities and Social Science (IOSR-JHSS)* 20 (2):47–52.
Haoran, L., Shasha, L., Mingjun, M., Guangzheng, H., Yaran, C., and Dongbin, Z. 2023. NeuronsGym: A Hybrid Framework and Benchmark for Robot Tasks with Sim2Real Policy Learning. *arXiv preprint* arXiv:2302.03385.
Harrell, W. A. (1991). Factors influencing pedestrian cautiousness in crossing streets. *The Journal of Social Psychology, 131*(3), 367–372.
Harrigan, J. A. (2005). Proxemics, kinesics, and gaze. In J. A. Harrigan, R. Rosenthal, & K. Scherer (Eds.), *The New Handbook of Methods in Nonverbal Behavior Research* (pp. 137–198). Oxford University Press.
Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics, 4*(2), 100–107.
Hart, R. A. (1987). Environmental psychology or behavioral geography? Either way it was a good start. *Journal of Environmental Psychology, 7*(4), 321–329.
Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron, 37*(5), 877–888.
Havard, C., & Willis, A. (2012). Effects of installing a marked crosswalk on road crossing behaviour and perceptions of the environment. *Transportation Research Part f: Traffic Psychology and Behaviour, 15*(3), 249–260.
Hayduk, L. A. (1983). Personal space: Where we now stand. *Psychological Bulletin, 94*(2), 293–335.
He, J., Hong, L., Frías-Martinez, V., and Torrens, P.M. 2015. Uncovering social media reaction pattern to protest events: a spatiotemporal dynamics perspective of Ferguson unrest. In *Lecture Notes in Computer Science 9471. Social Informatics. Proceedings of the 7th International Conference, SocInfo 2015, Beijing, China, December 9–12, 2015*, eds. T.-Y. Liu, N. C. Scollon and W. Zhu, 67–81. Berlin: Springer.
Helbing, D. (1992). A fluid-dynamic model for the movement of pedestrians. *Complex Systems, 6*(5), 391–415.
Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*, 4282–4286.
Henderson, L. F. (1971). The statistics of crowd fluids. *Nature, 229*(5284), 381–383.
Hillier, B., & Hanson, J. (1984). *The Social Logic of Space*. Cambridge University Press.
Ho, J., and Ermon, S. 2016. Generative adversarial imitation learning. In *30th Conference on Advances in Neural Information Processing Systems*

(NeurIPS), Barcelona, Spain, December 5–10, 2016, 6391. New York, NY, UA: NeurIPS.
Hong, L., Fu, C., Torrens, P.M., and Frías-Martinez, V. 2017. Understanding citizens' and local governments' digital communications during natural disasters: the case of snowstorms. In *Proceedings of the 2017 Association for Computing Machinery Web Science Conference, Troy, NY, June 26–28, 2017*, 141–150. Troy, NY: ACM.
Huang, J., Jiang, P., Gautam, A., & Saripalli, S. (2024). GPT-4V Takes the Wheel: Promises and Challenges for Pedestrian Behavior Prediction. *Proceedings of the AAAI Symposium Series, 3*(1), 134–142.
Huber, M., Su, Y.-H., Krüger, M., Faschian, K., Glasauer, S., & Hermsdörfer, J. (2014). Adjustments of speed and path when avoiding collisions with another pedestrian. *PLoS ONE, 9*(2), Article e89589.
Hurt, C. P., & Kram, R. (2006). Conscious control of preferred walking speed: Are we paying attention? *Medicine & Science in Sports & Exercise, 38*(5), S441–S442.
Huynh, M., and Alaghband, G. 2019. Trajectory prediction by coupling Scene-LSTM with human movement LSTM. In *Advances in Visual Computing. 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7–9, 2019, Proceedings, Part I*, eds. G. Bebis, R. Boyle, B. Parvin, D. Koracin, D. Ushizima, S. Chai, S. Sueda, X. Lin, A. Lu, D. Thalmann, C. Wang and P. Xu, 244–259. Cham: Springer International Publishing.
Hyun, K., Kim, M., Hwang, Y., & Lee, J. (2013). Tiling Motion Patches. *IEEE Transactions on Visualization and Computer Graphics, 19*(11), 1923–1934.
Jin, J., Wang, X., Li, C., Huang, L., and Tang, J. 2023. SequencePAR: Understanding pedestrian attributes via a sequence generation paradigm. *arXiv preprint* arXiv:2312.01640.
Johansson, A., Batty, M., Hayashi, K., Al Bar, O., Marcozzi, D., & Memish, Z. A. (2012). Crowd and environmental management during mass gatherings. *The Lancet Infectious Diseases, 12*(2), 150–156.
Johansson, A., Helbing, D., Al-Abideen, H., & Al-Bosta, S. (2008). From crowd dynamics to crowd safety: A video-based analysis. *Advances in Complex Systems, 11*(4), 497–527.
Jordan, K., Challis, J. H., & Newell, K. M. (2007). Walking speed influences on gait cycle variability. *Gait & Posture, 26*(1), 128–134.
Kadali, B. R., & Vedagiri, P. (2013). Effect of vehicular lanes on pedestrian gap acceptance behaviour. *Procedia-Social and Behavioral Sciences, 104*(1), 678–687.
Kamil, A. C., & Cheng, K. (2001). Way-finding and landmarks: The multiple-bearings hypothesis. *Journal of Experimental Biology, 204*(1), 103–113.
Kato, Y., & Takeuchi, Y. (2003). Individual differences in wayfinding strategies. *Journal of Environmental Psychology, 23*(2), 171–188.
Kennedy, D. P., Glascher, J., Tyszka, J. M., & Adolphs, R. (2009). Personal space regulation by the human amygdala. *Nature Neuroscience, 12*(10), 1226–1227.
Kesting, A., Treiber, M., & Helbing, D. (2010). Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences, 368*(1928), 4585–4605.
Kim, R., and Torrens, P.M. 2024. 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, June 24–26, 2024, Atlanta, GA, USA*, eds. M. Loper and A. Pellegrini, 119–130. New York, NY, USA: Association for Computing Machinery.
Kitazawa, K., & Fujiyama, T. (2010). Pedestrian vision and collision avoidance behavior: Investigation of the information process space of pedestrians using an eye tracker. In W. W. F. Klingsch, C. Rogsch, A. Schadschneider, & M. Schreckenberg (Eds.), *Pedestrian and Evacuation Dynamics 2008* (pp. 95–108). Springer.
Kiverstein, J. (2012). The meaning of embodiment. *Topics in Cognitive Science, 4*(4), 740–758.
Knorr, A. G., Willacker, L., Hermsdörfer, J., Glasauer, S., & Krüger, M. (2016). Influence of person- and situation-specific characteristics on collision avoidance behavior in human locomotion. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1332–1343.
Konczak, J. (1994). Effects of optic flow on the kinematics of human gait: A comparison of young and older adults. *Journal of Motor Behavior, 26*(3), 225–236.
Kong, Y., Song, D., Liang, J., Manocha, D., Yao, Z., and Xiao, X. 2025. Autospatial: Visual-language reasoning for social robot navigation through efficient spatial reasoning learning. *arXiv preprint* arXiv:2503.07557.
Krishna, A., & Schwarz, N. (2014). Sensory marketing, embodiment, and grounded cognition: A review and introduction. *Journal of Consumer Psychology, 24*(2), 159–168.
Krumm, J., & Horvitz, E. (2007). Predestination: Where do you want to go today? *IEEE Computer, 40*(4), 105–107.
Kudoh, T., & Matsumoto, D. (1985). Cross-cultural examination of the semantic dimensions of body postures. *Journal of Personality and Social Psychology, 48*(6), 1440.
Kuipers, B. (1982). The "Map in the Head" Metaphor. *Environment and Behavior, 14*(2), 202–220.
Kuipers, B. J., & Levitt, T. S. (1988). Navigation and mapping in large scale space. *AI Magazine, 9*(2), 25.
Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute. *PLoS ONE, 12*(5), Article e0177459.
LeCun, Y. 2022. A path towards autonomous machine intelligence version 0.9. 2, 2022–06–27. *Open Review* 62 (1):1–62.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
LeCun, Y., Muller, U., Ben, J., Cosatto, E., and Flepp, B. 2005. Off-road obstacle avoidance through end-to-end learning. In *Proceedings of the 18th International Conference on Neural Information Processing Systems, December 5–8, 2005, Vancouver, British Columbia, Canada*, eds. Y. Weiss, B. Schölkopf and J. C. Platt, 739–746. Cambridge, MA, USA: MIT Press.
Lee, K. H., Choi, M. G., & Lee, J. (2006). Motion patches: Building blocks for virtual environments annotated with motion data. *ACM Transactions on Graphics, 25*(3), 898–906.
Lefebvre, H. 1992/2004. *Rhythmanalysis: Space, Time and Everyday Life (Translated by Stuart Elden and Gerald Moore)*. London: Continuum.
Lenntorp, B. (1977). Paths in space-time environments: A time-geographic study of movement possibilities of individuals. *Environment and Planning A, 9*(8), 961–972.
Liao, M., Mei-Po, K., & Liu, X. (2024). Exploration of human activity fragmentation in cyber and physical spaces using massive mobile phone data. *Annals of GIS, 30*(4), 417–434.
Lin, H., Batty, M., Jørgensen, S. E., Fu, B., Konecny, M., Voinov, A., Torrens, P. M., Lu, G., Zhu, A.-X., Wilson, J. P., Gong, J., Kolditz, O., Bandrova, T., & Chen, M. (2015). Virtual environments begin to embrace process-based geographic analysis. *Transactions in GIS, 19*(4), 1–5.
Lindblom, J. (2015). *Embodied Social Cognition*. Springer.
Lipovac, K., Vujanic, M., Maric, B., & Nesic, M. (2013). The influence of a pedestrian countdown display on pedestrian behavior at signalized pedestrian crossings. *Transportation Research Part f: Traffic Psychology and Behaviour, 20*(1), 121–134.
Liu, Y.-C., & Tung, Y.-C. (2014). Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed. *Safety Science, 63*, 77–82.
Loukaitou-Sideris, A., & Ehrenfeucht, R. (2011). *Sidewalks: Conflict and Negotiation Over Public Space*. MIT Press.
Lukeman, R., Li, Y.-X., & Edelstein-Keshet, L. (2010). Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences, 107*(28), 12576–12580.
Lynch, S. D., Kulpa, R., Meerhoff, L. A., Pettré, J., Crétual, A., & Olivier, A.-H. (2018). Collision avoidance behavior between walkers: Global and local motion cues. *IEEE Transactions on Visualization and Computer Graphics, 24*(7), 2078–2088.
Mark, D. M., & Frank, A. U. (1996). Experiential and formal models of geographic space. *Environment and Planning b: Planning and Design, 23*(1), 3–24.
Marques, P., Váz, P., Silva, J., Martins, P., & Abbasi, M. (2025). Real-time gesture-based hand landmark detection for optimized mobile photo capture and synchronization. *Electronics, 14*(4), 704.
Matthis, J. S., Muller, K. S., Bonnen, K. L., & Hayhoe, M. M. (2022). Retinal optic flow during natural locomotion. *PLoS Computational Biology, 18*(2), Article e1009575.
Mehrabian, A. (1968). Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation, 1*(6), 203–207.
Meier, B. P., Schnall, S., Schwarz, N., & Bargh, J. A. (2012). Embodiment in social psychology. *Topics in Cognitive Science, 4*(4), 705–716.

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of sadness and depression—gait patterns associated with dysphoric mood. *Psychosomatic Medicine, 71*(5), 580–587.
Middleton, J. (2010). Sense and the city: Exploring the embodied geographies of urban walking. *Social & Cultural Geography, 11*(6), 575–596.
Mirzaei-Alavijeh, M., Babakhani, M., Jalilian, F., Vaezi, M., Jalilian, F., Khashij, S., & Hamzeh, B. (2019). Socio-cognitive determinants of safe road-crossing behaviors: An application of the prototype willingness model. *Journal of Injury and Violence Research, 11*(1), 93.
Mishra, S., Welch, T. F., Torrens, P. M., Fu, C., Zhu, H., & Knaap, E. (2015). A tool for measuring and visualizing connectivity of transit stop, route and transfer center in a multimodal transportation network. *Public Transport, 7*(1), 77–99.
Mondada, L., & Tekin, B. S. (2023). Joining the queue as a newcomer: The instructably visible order of queuing. In M. Lynch & O. Lindwall (Eds.), *Instructed and Instructive Actions* (pp. 73–99). Routledge.
Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior, 11*(1), 33–42.
Nabbe, B., Hoeim, D., Efros, A.A., and Herbert, M.I. 2006. Opportunistic use of vision to push back the path-planning horizon. Y. Liu. Paper read at 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 9–15, at Beijing.
Nams, V. O. (1996). The VFractal: A new estimator for fractal dimension of animal movement paths. *Landscape Ecology, 11*(5), 289–297.
Nams, V. O. (2006). Improving accuracy and precision in estimating fractal dimension of animal movement paths. *Acta Biotheoretica, 54*(1), 1–11.
Nesoff, E. D., Milam, A. J., Pollack, K. M., Curriero, F. C., Bowie, J. V., Gielen, A. C., & Furr-Holden, D. M. (2018). Novel Methods for Environmental Assessment of Pedestrian Injury: Creation and Validation of the Inventory for Pedestrian Safety Infrastructure. *Journal of Urban Health, 95*(2), 208–221.
Ngo, B.-H., Ngo, S.-T., Le, P.-D., Phan, Q.-M., Tran, M.-T., and Le, T.-N. 2024. Cross-PAR: Enhancing Pedestrian Attribute Recognition with Vision-Language Fusion and Human-Centric Pre-training. Paper read at Proceedings of the Asian Conference on Computer Vision.
Nie, Y., Lu, B., Chen, Q., Miao, Q., & Lv, Y. (2022). SynPoses: Generating Virtual Dataset for Pedestrian Detection in Corner Cases. *IEEE Journal of Radio Frequency Identification, 6*(1), 801–804.
Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184–211.
Nine Inch Nails. 2016. *The Idea of You*. Not the Actual Events (Halo 29). Los Angeles, CA: The Null Corporation. CD.
Nishimura, C., Kurita, S., and Seki, Y. 2024. Text360Nav: 360-degree image captioning dataset for urban pedestrians navigation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), May 20–25, 2024, Torino, Italy*, 15783–15788. Torino, Italy: ACL Anthology.
Olivier, A.-H., Marin, A., Crétual, A., Berthoz, A., & Pettré, J. (2013). Collision avoidance between two walkers: Role-dependent strategies. *Gait & Posture, 38*(4), 751–756.
Olivier, A.-H., Marin, A., Crétual, A., & Pettré, J. (2012). Minimal predicted distance: A common metric for collision avoidance during pairwise interactions between walkers. *Gait & Posture, 36*(3), 399–404.
Omer, I., & Goldblatt, R. (2007). The implications of inter-visibility between landmarks on wayfinding performance: An investigation using a virtual urban environment. *Computers, Environment and Urban Systems, 31*(5), 520–534.
Omer, I., & Goldblatt, R. (2017). Using space syntax and Q-analysis for investigating movement patterns in buildings: The case of shopping malls. *Environment and Planning b: Urban Analytics and City Science, 44*(3), 504–530.
Omer, I., & Kaplan, N. (2019). Structural properties of the angular and metric street network's centralities and their implications for movement flows. *Environment and Planning b: Urban Analytics and City Science, 46*(6), 1182–1200.
Omer, I., Rofè, Y., & Lerman, Y. (2015). The impact of planning on pedestrian movement: Contrasting pedestrian movement models in pre-modern and modern neighborhoods in Israel. *International Journal of Geographical Information Science, 29*(12), 2121–2142.
Onelcin, P., & Alver, Y. (2015). Illegal crossing behavior of pedestrians at signalized intersections: Factors affecting the gap acceptance. *Transportation Research Part f: Traffic Psychology and Behaviour, 31*, 124–132.
Openshaw, S. (1983). *The Modifiable Areal Unit Problem*. GeoBooks.
Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A, 16*(1), 17–31.
Oudejans, R. R., Michaels, C. F., van Dort, B., & Frissen, E. J. (1996). To cross or not to cross: The effect of locomotion on street-crossing behavior. *Ecological Psychology, 8*(3), 259–267.
Ouyang, X., Cheng, Y., Jiang, Y., Li, C.-L., and Zhou, P. 2018. Pedestrian-synthesis-gan: Generating pedestrian data in real scene and beyond. *arXiv preprint* arXiv:1804.02047.
Oxley, J., Charlton, J., and Fildes, B. 2005. *The effect of cognitive impairment on older pedestrian behaviour and crash risk*. Melbourne: Monash University Accident Research Centre. MUARC Report #244. Available from. Last accessed.
Pailhous, J., Ferrandez, A.-M., Flückiger, M., & Baumberger, B. (1990). Unintentional modulations of human gait by optical flow. *Behavioural Brain Research, 38*(3), 275–281.
Pasios, S., and Nikolaidis, N. 2024. CARLA2Real: a tool for reducing the sim2real gap in CARLA simulator. *arXiv preprint* arXiv:2410.18238.
Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research, 148*(1), 133–138.
Pauls, J. (1984). The movement of people in buildings and design solutions for means of egress. *Fire Technology, 20*(1), 27–47.
Pearson, K. G. (2004). Generating the walking gait: Role of sensory feedback. *Progress in Brain Research, 143*(1), 123–129.
Penn, A. (2003). Space syntax and spatial cognition: Or why the axial line? *Environment and Behavior, 35*(1), 30–65.
Pfeffer, K., & Hunter, E. (2013). The effects of peer influence on adolescent pedestrian road-crossing decisions. *Traffic Injury Prevention, 14*(4), 434–440.
Picard, R.W. 2000. *Affective Computing*: MIT press.
Plumert, J. M., Kearney, J. K., & Cremer, J. F. (2004). Children's perception of gap affordances: Bicycling across traffic-filled intersections in an immersive virtual environment. *Child Development, 75*(4), 1243–1253.
Potdar, K., and Torrens, P.M. 2019. Modelling spatio-temporal patterns in pedestrian behavior at the edge with Jetson SOMs. *NVIDIA GPU Technology Conference (GTC), Washington, D.C., November 4–6*:DCP91291.
Priisalu, M., Paduraru, C., Pirinen, A., & Sminchisescu, C. (2021). Semantic synthesis of pedestrian locomotion. In H. Ishikawa, C.-L. Liu, T. Pajdla, & J. Shi (Eds.), *Computer Vision – ACCV 2020* (pp. 470–487). Springer International Publishing.
Raubal, M. (2001a). Human wayfinding in unfamiliar buildings: A simulation with a cognizing agent. *Cognitive Processing, 2*(3), 363–388.
Raubal, M. (2001b). Ontology and epistemology for agent-based wayfinding simulation. *International Journal of Geographic Information Science, 15*(7), 653–665.
Raubal, M. 2008. Wayfinding: affordances and agent simulation. In *Encyclopedia of GIS*, eds. S. Shekhar and H. Xiong, 1243–1246. New York: Springer US.
Raubal, M., and Worboys, M. 1999. A formal model of the process of wayfinding in built environments. In *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, eds. C. Freksa and D. M. Mark, 748–748. Berlin: Springer-Verlag.
Razzaque, S., Swapp, D., Slater, M., Whitton, M.C., and Steed, A. 2002. Redirected walking in place. In *Proceedings of EGVE '02: Eurographics Workshop on Virtual Environments, Barcelona, Spain May 30 - 31, 2002*, eds. W. Stürzlinger and S. Müller, 123–130. New York, NY, USA: ACM.
Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. 2016. 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, June 26-July 1, 2016*, eds. F.-F. Li, R. Bajcsy and T. Tuytelaars, 779–788. Los Alamitos, CA: IEEE.
Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*(4), 25–34.
Reynolds, C. W. (1993). An evolved, vision-based behavioral model of coordinated group motion. In J.-A. Meyer, H. L. Roitblat, & S. W. Wilson (Eds.), *From Animals to Animats 2* (pp. 384–392). The MIT Press.

Reynolds, C.W. 1999. Steering behaviors for autonomous characters. In *Proceedings of Game Developers Conference, San Jose, CA*, ed. Miller Freeman Game Group, 763–782. San Francisco, CA: Miller Freeman Game Group.
Rosenbloom, T. (2009). Crossing at a red light: Behaviour of individuals and groups. *Transportation Research Part f: Traffic Psychology and Behaviour, 12*(5), 389–394.
Ruddle, R. A., Volkova, E., Mohler, B., & Bülthoff, H. H. (2011). The effect of landmark and body-based sensory information on route knowledge. *Memory & Cognition, 39*(4), 686–699.
Ryu, S., Chen, A., Su, J., Liu, X., & Yu, J. (2021). Considering Space Syntax in Bicycle Traffic Assignment with One or More User Classes. *Sustainability, 13*(19), 11078.
Safonova, A., Hodgins, J. K., & Pollard, N. S. (2004). Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Transactions on Graphics, 23*(3), 514–521.
Salinas, M. M., Wilken, J. M., & Dingwell, J. B. (2017). How humans use visual optic flow to regulate stepping during walking. *Gait & Posture, 57*, 15–20.
Samet, H. (1989). *Applications of Spatial Data Structures*. Addison-Wesley.
Samet, H. (1990). *The Design and Analysis of Spatial Data Structures*. Addison-Wesley.
Sarker, A.A., Mishra, S., Welch, T.F., Golias, M.M., and Torrens, P.M. 2015. A model framework for analyzing public transit connectivity and its application in a large-scale multi-modal transit network (15–3279). In *Proceedings of the Transportation Research Board 94th Annual Meeting, January 11–15, 2015, Washington D.C.*, 13–20. Washington, D.C.: Transportation Research Board of the National Academies.
Scheflen, A. E. (1972). *Body Language and the Social Order: Communication as Behavioral Control*. Prentice-Hall.
Schick, L., & Malmborg, L. (2010). Bodies, embodiment and ubiquitous computing. *Digital Creativity, 21*(1), 63–69.
Schmidt, S., & Färber, B. (2009). Pedestrians at the kerb – Recognising the action intentions of humans. *Transportation Research Part f: Traffic Psychology and Behaviour, 12*(4), 300–310.
Schweitzer, F. (1997). Active brownian particles: Artificial agents in physics. In T. Pöschel & L. Schimansky-Geier (Eds.), *Stochastic Dynamics* (pp. 358–371). Springer-Verlag.
Shaw, D. B. (2015). Streets for cyborgs: The electronic flâneur and the posthuman city. *Space and Culture, 18*(3), 230–242.
Sheller, M. (2017). From spatial turn to mobilities turn. *Current Sociology, 65*(4), 623–639.
Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A, 38*(2), 207–226.
Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in Child Development and Behaviour, 10*(1), 9–55.
Sime, J. D. (1995). *Crowd Psychology and Engineering. Safety Science, 21*(1), 1–14.
Simonsen, K. (2013). In quest of a new humanism: Embodiment, experience and phenomenology as critical geography. *Progress in Human Geography, 37*(1), 10–26.
Sipper, M. (1997). *Evolution of Parallel Cellular Machines: The Cellular Programming Approach*. Springer.
Snape, J., Van Den Berg, J., Guy, S. J., & Manocha, D. (2011). The hybrid reciprocal velocity obstacle. *IEEE Transactions on Robotics, 27*(4), 696–706.
Song, D., Liang, J., Xiao, X., and Manocha, D. 2025. VL-TGS: trajectory generation and selection using vision language models in mapless outdoor environments. *IEEE Robotics and Automation Letters* 10 (6).
Sperber, A. L., Kappeler, P. M., & Fichtel, C. (2019). Should I stay or should I go? Individual movement decisions during group departures in red-fronted lemurs. *Royal Society Open Science, 6*(3), Article 180991.
Spiers, H. J., & Maguire, E. A. (2008). The dynamic nature of cognition during wayfinding. *Journal of Environmental Psychology, 28*, 232–249.
Strauss, K., Savkin, A., and Tombari, F. 2021. Attention-based adversarial appearance learning of augmented pedestrians. *arXiv preprint* arXiv:2107.02673.
Sud, A., Andersen, E., Curtis, S., Lin, M., C., and Manocha, D. 2008. Real-time path planning in dynamic virtual environments using multiagent navigation graphs. *IEEE Transactions on Visualization and Computer Graphics* 14 (3):526–538.
Sud, A., Gayle, R., Anderson, E., Guy, S., Lin, M., and Manocha, D. 2007. Real-time navigation of independent agents using adaptive roadmaps. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '07), Irivine, CA, November 5 to 7*, eds. A. Majumder, L. Hodges and D. Cohen-Or, 99–106. New York, NY: Association for Computing Machinery.
Sueur, C., Class, B., Hamm, C., Meyer, X., & Pelé, M. (2013). Different risk thresholds in pedestrian road crossing behaviour: A comparison of French and Japanese approaches. *Accident Analysis & Prevention, 58*, 59–63.
Sun, Q., Patney, A., Wei, L.-Y., Shapira, O., Lu, J., Asente, P., Zhu, S., Mcguire, M., Luebke, D., and Kaufman, A. 2018. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Trans. Graph.* 37 (4):Article 67.
Sun, Q., Wei, L.-Y., and Kaufman, A. 2016. Mapping virtual and physical reality. *ACM Trans. Graph.* 35 (4):Article 64.
Sun, S., Sun, C., Duives, D.C., and Hoogendoorn, S.P. 2021. Deviation of pedestrian path due to the presence of building entrances. *Journal of Advanced Transportation* 2021 (Online):5594738.
Sun, Y., Huang, F., Wen, W., Hsu, L.T., and Liu, X. 2023. Multi-robot cooperative LiDAR SLAM for efficienct mapping in urban sciences. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-1/W1–2023:473–478.
Tabibi, Z., and Pfeffer, K. 2003. Choosing a safe place to cross the road: the relationship between attention and identification of safe and dangerous road-crossing sites. *Child: care, health and development* 29 (4):237–244.
Thies, S. B., Richardson, J. K., & Ashton-Miller, J. A. (2005). Effects of surface irregularity and lighting on step variability during gait: A study in healthy young and older women. *Gait & Posture, 22*(1), 26–31.
Thrift, N. (2008). *Non-Representational Theory: Space, Politics, Affect*. Routledge.
Tolani, D., Goswami, A., & Badler, N. I. (2000). Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models, 62*, 353–388.
Tom, A., & Granié, M.-A. (2011). Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads. *Accident Analysis & Prevention, 43*(5), 1794–1801.
Torrens, P. M. (2004). Geosimulation, automata, and traffic modeling. In P. Stopher, K. Button, K. Haynes, & D. Hensher (Eds.), *Handbook of Transport Geography and Spatial Systems* (Vol. 5, pp. 549–565). Emerald Publishing Group Ltd.
Torrens, P. M. (2010a). Agent-based modeling and the spatial sciences. *Geography Compass, 4*(5), 428–448.
Torrens, P. M. (2010b). Geography and computational social science. *GeoJournal, 75*(2), 133–148.
Torrens, P. M. (2012). Moving agent pedestrians through space and time. *Annals of the Association of American Geographers, 102*(1), 35–66.
Torrens, P. M. (2014a). High-fidelity behaviours for model people on model streetscapes. *Annals of GIS, 20*(3), 139–157.
Torrens, P. M. (2014b). High-resolution space–time processes for agents at the built–human interface of urban earthquakes. *International Journal of Geographical Information Science, 28*(5), 964–986.
Torrens, P. M. (2015a). Geographical agents in three dimensions. In A. Singleton & C. Brunsdon (Eds.), *Geocomputation: A Practical Primer* (pp. 40–62). Sage.
Torrens, P. M. (2015b). Intertwining agents and environments. *Environmental Earth Sciences, 74*(10), 7117–7131.
Torrens, P. M. (2015c). Slipstreaming human geosimulation in virtual geographic environments. *Annals of GIS, 21*(4), 325–344.
Torrens, P. M. (2016a). Computational Streetscapes. *Computation*, *4*(3), 37.
Torrens, P. M. (2016b). Exploring behavioral regions in agents' mental maps. *The Annals of Regional Science, 57*(2–3), 309–334.
Torrens, P. M. (2018a). Artificial intelligence and behavioral geography. In D. R. Montello (Ed.), *Handbook of Behavioral and Cognitive Geography* (pp. 357–372). Edward Elgar Publishing.
Torrens, P. M. (2018b). A computational sandbox with human automata for exploring perceived egress safety in urban damage scenarios. *International Journal of Digital Earth, 11*(4), 369–396.
Torrens, P. M. (2022a). Data science for pedestrian and high street retailing as a framework for advancing urban informatics to individual scales. *Urban Informatics, 1*(1), 1–33.
Torrens, P. M. (2022b). Smart and sentient retail high streets. *Smart Cities, 5*(4), 1670–1720.
Torrens, P. M. (2023). Agent models of customer journeys on retail high streets. *Journal of Economic Interaction and Coordination, 18*(1), 87–128.
Torrens, P. M. (2024). Ten traps for non-representational theory in human geography. *Geographies, 4*(2), 253–286.

Torrens, P.M., and Benenson, I. 2003. Geographic Automata Systems. Paper read at Special session: Geographical Perspectives on Complexity Theory and Complex Systems 2, 99th Annual Meeting of the Association of American Geographers, March, 2003, at New Orleans.
Torrens, P. M., & Griffin, W. A. (2013). Exploring the micro-social geography of children's interactions in preschool: A long-term observational study and analysis using Geographic Information Technologies. *Environment and Behavior, 45*(5), 584–614.
Torrens, P.M., and Gu, S. 2021. Real-time experiential geosimulation in virtual reality with immersion-emission. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, 19–28. Beijing, China: Association for Computing Machinery.
Torrens, P. M., & Gu, S. (2023). Inverse augmentation: Transposing real people into pedestrian models. *Computers, Environment and Urban Systems, 100*, Article 101923.
Torrens, P. M., Kevrekidis, I., Ghanem, R., & Zou, Y. (2013). Simple urban simulation atop complicated models: Multi-scale equation-free computing of sprawl using geographic automata. *Entropy, 15*(7), 2606–2634.
Torrens, P. M., & Kim, R. (2024a). Evoking embodiment in immersive geosimulation environments. *Annals of GIS, 30*(1), 35–66.
Torrens, P. M., & Kim, R. (2024b). Using immersive virtual reality to study road-crossing sustainability in fleeting moments of space and time. *Sustainability, 16*(3), 1327.
Torrens, P. M., Li, X., & Griffin, W. A. (2011). Building agent-based walking models by machine-learning on diverse databases of space-time trajectory samples. *Transactions in Geographic Information Science, 15*(s1), 67–94.
Torrens, P. M., & Nara, A. (2013). Polyspatial agents for multi-scale urban simulation and regional policy analysis. *Regional Science Policy and Practice, 44*(4), 419–445.
Torrens, P. M., Nara, A., Li, X., Zhu, H., Griffin, W. A., & Brown, S. B. (2012). An extensible simulation environment and movement metrics for testing walking behavior in agent-based models. *Computers, Environment and Urban Systems, 36*(1), 1–17.
Treuille, A., Cooper, S., and Popovi. (2006). Continuum Crowds. ACM Transactions on Graphics 25 (3):1160–1168.
Tuan, Y.-F. 1975. Place: an experiential perspective. *Geographical Review*:151–165.
Tuan, Y.-F. (1979). Space and place: Humanistic perspective. In S. Gale & G. Olsson (Eds.), *Philosophy in Geography (Theory and Decision Library, Volume 20)* (pp. 387–427). Springer.
Tubbs, J., & Meacham, B. (2007). *Egress Design Solutions: A Guide to Evacuation and Crowd Management Planning*. John Wiley & Sons.
Underwood, J., Dillon, G., Farnsworth, B., & Twiner, A. (2007). Reading the road: The influence of age and sex on child pedestrians' perceptions of road risk. *British Journal of Psychology, 98*(1), 93–110.
Urry, J. (2002). *The Tourist Gaze*. Sage.
van den Berg, J., Lin, M., and Manocha, D. 2008. Reciprocal Velocity Obstacles for real-time multi-agent navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), May 19–23, 2008, Pasadena, CA, USA*, ed. J. Triesch, 1928–1935. Los Alamitos, CA, USA: IEEE.
Vázquez, D., López, A. M., Marín, J., Ponsa, D., & Gerónimo, D. (2014). Virtual and Real World Adaptation for Pedestrian Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(4), 797–809.
Viola, R., Roe, M., & Shin, H. (2010). *The New York City Pedestrian Safety Study & Action Plan, August 2010*. New York City Department of Transportation.
Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., & Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature, 381*(6581), 413–415.
von Neumann, J. (1951). The General and Logical Theory of Automata. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 1–41). Wiley.
Vs, V., Valanarasu, J.M.J., Oza, P., and Patel, V.M. 2022. Image fusion transformer. In *Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), October 16–19, 2022, Bordeaux, France*, eds. A. Cavallaro, J.-C. Pesquet and R. Willet, 3566–3570. Los Alamitos, CA: IEEE.
Vyas, H., and Torrens, P.M. 2024. Tackling "leaky" facial and object recognition operations over federated learning. In *Proceedings of the 2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC), Chicago, IL, USA, June 11–13, 2024*, eds. N. Bizon, J. M. Lopez-Guede and B. Selami, 1–5. Los Alamitos, CA: IEEE.
Wang, J., Manivasagam, S., Chen, Y., Yang, Z., Bârsan, I.A., Yang, A.J., Ma, W.-C., and Urtasun, R. 2023. Cadsim: Robust and scalable in-the-wild 3d reconstruction for controllable sensor simulation. *arXiv preprint* arXiv:2311.01447.
Wang, R. F., & Cutting, J. E. (1999). Where we go with a little good information. *Psychological Science, 10*(1), 71–75.
Wang, X., Zhu, Q., Jin, J., Zhu, J., Wang, F., Jiang, B., Wang, Y., and Tian, Y. 2024. Spatio-temporal side tuning pre-trained foundation models for video-based pedestrian attribute recognition. *arXiv preprint* arXiv:2404.17929.
Wilkie, D., van den Berg, J., and Manocha, D. 2009. Generalized velocity obstacles. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, October 10–15, 2009*, eds. N. Xi and W. R. Hamel, 5573–5578. Los Alamitos, CA: IEEE.
With, K. (1994). Using fractal analysis to assess how species perceive landscape structure. *Landscape Ecology, 9*(1), 25–36.
Withagen, R., & Chemero, A. (2012). Affordances and classification: On the significance of a sidebar in James Gibson's last book. *Philosophical Psychology, 25*(4), 521–537.
Wrigley, N., Holt, T., Steel, D., & Tranmer, M. (1996). Analysing, modelling, and resolving the ecological fallacy. In P. A. Longley & M. Batty (Eds.), *Spatial Analysis: Modelling in a GIS Environment* (pp. 25–40). GeoInformation International.
Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. 2019. Detectron2. *GitHub*, January 7, 2023. Last accessed January 1 2020. Available from https://github.com/facebookresearch/detectron2.
Yang, B., Luo, W., and Urtasun, R. 2018. Pixor: Real-time 3d object detection from point clouds. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, June 18–23, 2018*, 7652–7660. Los Alamitos, CA, USA: IEEE.
Yang, X., Abdel-Aty, M., Huan, M., Jia, B., & Peng, Y. (2016). The effects of traffic wardens on the red-light infringement behavior of vulnerable road users. *Transportation Research Part f: Traffic Psychology and Behaviour, 37*(1), 52–63.
Yang, Z., Manivasagam, S., Chen, Y., Wang, J., Hu, R., and Urtasun, R. 2023. Reconstructing objects in-the-wild for realistic sensor simulation. In *Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, Innit, UK, May 29-June 2, 2023*, ed. K. Althoefer, 11661–11668. Los Alamitos, CA, USA: IEEE.
Yang, Z., Manivasagam, S., Liang, M., Yang, B., Ma, W.-C., and Urtasun, R. 2021. Recovering and simulating pedestrians in the wild. In *Proceedings of the 2020 Conference on Robot Learning (PMLR 155), Cambridge, MA, USA, November 16–18, 2020*, 419–431. New York, NY, USA: MLR Press.
Yang, Z., Wu, D., Wu, C., Lin, Z., Gu, J., and Wang, W. 2024. A pedestrian is worth one prompt: Towards language guidance person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 16–22, 2024, Seattle, WA, USA*, eds. W. Scheirer, O. Camps, R. Zabih, R. Cucchiara and S. Sarkar, 17343–17353. Los Alamitos, CA: IEEE.
Yuksel, S. E., Wilson, J. N., & Gader, P. D. (2012). Twenty years of mixture of experts. *IEEE Transactions on Neural Networks and Learning Systems, 23*(8), 1177–1193.
Zeedyk, M. S., Wallace, L., & Spry, L. (2002). Stop, look, listen, and think?: What young children really do when crossing the road. *Accident Analysis & Prevention, 34*(1), 43–50.
Zhang, F., Wenzhong, S., Min, C., Wei, H., & Liu, X. (2023). Open HD map service model: An interoperable high-Definition map data model for autonomous driving. *International Journal of Digital Earth, 16*(1), 2089–2110.
Zhang, J., Gong, J., Lin, H., Wang, G., Huang, J., Zhu, J., Xu, B., & Teng, J. (2007). Design and development of distributed virtual geographic environment system based on web services. *Information Sciences, 177*(19), 3968–3980.
Zhang, P., Xue, J., Zhang, P., Zheng, N., & Ouyang, W. (2022). Social-aware pedestrian trajectory prediction via states refinement LSTM. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(5), 2742–2759.
Zhang, S., Benenson, R., and Schiele, B. 2017. CityPersons: A diverse dataset for pedestrian detection. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21–26, 2017*, eds. R. Chellappa, A. Hoogs and Z. Zhang, 3213–3221. Los Alamitos, CA, USA: IEEE.
Zhao, J., & Badler, N. I. (1994). Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics, 13*(4), 313–336.

Ziemke, T. (2013). What's that thing called embodiment? In Proceedings of the 25th Annual Conference of the Cognitive Science Society: Volume I, eds. R. Alterman and D. Kirsch, 1305-1310. New York: Psychology Press.
Zivotofsky, A. Z., Eldror, E., Mandel, R., & Rosenbloom, T. (2012). Misjudging their own steps: Why elderly people have trouble crossing the road. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(4), 600–607.

Publisher's Note