

# Microscopic modeling of attention-based movement behaviors<sup>☆,☆☆</sup>

Danrui Li<sup>a,\*</sup>, Mathew Schwartz<sup>a</sup>, Samuel S. Sohn<sup>a</sup>, Sejong Yoon<sup>b</sup>,  
Vladimir Pavlovic<sup>a</sup>, Mubbasir Kapadia<sup>a</sup>

<sup>a</sup> Rutgers, the State University of New Jersey, 110 Frelinghuysen Rd, Piscataway, 08854, NJ, USA

<sup>b</sup> The College of New Jersey, 2000 Pennington Road Ewing, 08628, NJ, USA

## ARTICLE INFO

Dataset link: <https://tianchi.aliyun.com/dataset/162948>

### Keywords:

Pedestrian dynamics  
Visual attention  
Retail environment  
Transportation hubs  
Architectural design

## ABSTRACT

For transportation hubs, leveraging pedestrian flows for commercial activities presents an effective strategy for funding maintenance and infrastructure improvements. However, this introduces new challenges, as consumer behaviors can disrupt pedestrian flow and efficiency. To optimize both retail potential and pedestrian efficiency, careful strategic planning in store layout and facility dimensions was done by expert judgement due to the complexity in pedestrian dynamics in the retail areas of transportation hubs. This paper introduces an attention-based movement model to simulate these dynamics. By simulating retail potential of an area through the duration of visual attention it receives, and pedestrian efficiency via speed loss in pedestrian walking behaviors, the study further explores how design features can influence the retail potential and pedestrian efficiency in a bi-directional corridor inside a transportation hub.

## 1. Introduction

Pedestrian dynamics within public spaces have garnered significant attention in recent years due to their implications for the planning of transportation facilities, commercial centers, and other urban infrastructures (Hoogendoorn et al., 2004). While pedestrian safety has been the focus for a long time, recent demands call for explorations about the pedestrian dynamics in retail areas within high-density transportation hubs. As passenger growth increases the expenditure of transportation facility operations, leveraging such pedestrian flows to expand commercial activities becomes a practical way for the management to fund maintenance and asset improvements (Baron, 2019).

However, while increases in pedestrian flows provide financial benefits, research suggests consuming behaviors may introduce slow-down effects or turbulence to pedestrian flows (Zacharias, 2021). These conflicting outcomes create a challenge when incorporating retail areas into high-density transportation hubs. In order to maintain the efficiency requirements of transportation services, architectural designers must strike a balance between retail potential and pedestrian efficiency through strategic store layout and facility dimensions (Hänseler, 2016). This process is currently done through intuition and expert judgement, in part due to the difficulties in simulating the possible outcomes given modifications to the design and layout.

<sup>☆</sup> This material is based upon work supported by the U.S. Department of Homeland Security, United States under Grant Award 22STES00001-03-02. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. The research was also supported in part by U.S. National Science Foundation #1955365, #1955404, and #2324598.

<sup>☆☆</sup> This article belongs to the Virtual Special Issue on IG005588: VSI: Pedestrians & Crowds.

\* Corresponding author.

E-mail addresses: [danrui.li@rutgers.edu](mailto:danrui.li@rutgers.edu) (D. Li), [cadop@umich.edu](mailto:cadop@umich.edu) (M. Schwartz), [samuel.sohn@rutgers.edu](mailto:samuel.sohn@rutgers.edu) (S.S. Sohn), [yoons@tcnj.edu](mailto:yoons@tcnj.edu) (S. Yoon), [vladimir@rutgers.edu](mailto:vladimir@rutgers.edu) (V. Pavlovic), [mubbasir.kapadia@rutgers.edu](mailto:mubbasir.kapadia@rutgers.edu) (M. Kapadia).

<https://doi.org/10.1016/j.trc.2024.104583>

Received 27 October 2023; Received in revised form 12 March 2024; Accepted 19 March 2024

Available online 3 April 2024

0968-090X/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

One of the difficulties is the diversity of pedestrian interactions with environmental objects, which limits the applications of previous studies. Most prior works follow an assumption that pedestrians tend to steer towards environmental objects when they are attracted to them (Kwak et al., 2013; Zhou et al., 2022). However, Li (2023) shows that such a model does not impose the dominant impact on the slowing-down effects in transportation hub retail areas. Instead, pedestrians who slow down to gaze at stores without changing directions actually provide the major contributions (Li, 2023). Such a behavior, which we refer to as *attention-based movement*, differs from endogenous distracted movement in Kremer et al. (2021) because in our context agents will interact with exogenous environmental objects rather than something they hold. Instead, our work is related to the attracted movement behaviors described in Wang (2014). Since the modeling of attention-based movement remains largely unexplored, in this study we focus on this subgroup to pave the way to the design optimization goal above.

Another challenge is to find metrics for pedestrian efficiency and retail potential, and then model them simultaneously. Retail potential, often referred to as the probability of purchasing a product/item, is unable to directly inspect in our study, because we focus on attention-based movements where pedestrians do not enter the store to purchase. However, retail potential is found to be correlated with the time of looking at the products (the duration of visual attention) (Gidlöf et al., 2017). Such influence becomes even stronger in rapid decisions (Milosavljevic et al., 2012), which are often the cases in transportation hubs. Therefore, we consider the average visual attention that pedestrians pay to a retail area as a reasonable measure for the retail potential of it. It can also be seen as the visual saliency of retail areas from an exocentric perspective (Kremer et al., 2021). Pedestrian efficiency, on the other hand, can be measured by the speed loss compared to pedestrians' desired walking speed (Helbing et al., 2000; Kwak et al., 2013). Therefore, we need to simulate not only pedestrians' visual attention but also their changes in walking behaviors when they are interacting with environmental objects.

In this paper, we propose a novel pedestrian model referred to as *attention-based movement behaviors*. Our model is a combination of visual attention and locomotion, designed for simulating attention within retail transportation hubs. Section 2 provides an overview of related work. Section 3 describes the dataset, models, and parameter settings. Section 4 compares its simulation results to a real-world dataset and prior work. Then, our model is applied to a sensitivity analysis with regard to the architectural design features in Section 5 to demonstrate its capability in design optimization. Finally, the paper ends with a discussion and conclusion in Section 6.

## 2. Related work

### 2.1. Visual attention

In this paper, the visual attention of a pedestrian is a status when the pedestrian moves their eyes to a location and the focus of attention is in accordance with eye movements, referred to as *spatial overt attention* (compared to the attention that does not involve eye movements) (Carrasco, 2011). Visual attention can be triggered by not only internal factors (e.g., pedestrians' preferences) (Gidlöf et al., 2017), but also external factors such as the physical feature (color, luminance, etc.) of environmental objects (Saunders and Gero, 2004), the presence of surrounding pedestrians (Gallup et al., 2012), and visual coverage (one of the most explored aspects) (Xie et al., 2007; Nassar, 2011; Chen, 2011; Wang et al., 2014; Kremer et al., 2021; Zhou et al., 2022).

Visual coverage is the state of environmental objects covering pedestrians' visual perception fields. It describes not only the area they cover but also the location (i.e., the object is on the center of our retinal image or the fringe of that). In previous studies, visual coverage has been represented in different ways. Xie et al. (2007) explored the boundary of visual perception fields (referred to as visual catchment area) using angular separation of environmental objects. Chen (2011) and Wang et al. (2014) used distance and proportion of visual coverage, while (Nassar, 2011) used a piece-wise scaling term on the proportion of visual coverage and multiplied it by distance. Zhou et al. (2022) and Kremer et al. (2020) employed observation angles and distances. Some of their representations of visual coverage are illustrated in Fig. 3. We found that not all of these prior works can fully describe the area and the location of visual coverage at the same time (summarized in Table 1).

Additionally, pedestrians continuously shift their visual attention among various objects in retail areas, resulting in a series of attention initiation and termination events. This differs from evacuation scenarios where we care about how people initiate their attention to a sign rather than how they terminate it. In our context, we study the attention transitions in both directions (i.e. the initiation and the termination of attention). Furthermore, the visual attention model should be calibrated by real-world data, and must easily fit to data. Importantly, this model must have fast execution times due to the large number of agents (typically hundreds) involved in transportation hub simulations.

Our work is compared with prior literature in Table 1 across: (1) Visual coverage representation in area and location (2) State transitions (3) Performant execution (4) Limitations specific to the models. Our work addresses gaps in prior works required for evaluating retail in transportation hubs. Specifically, we will build a visual attention model that can fully describe visual coverage, predict attention transitions in both directions calibrated by real-world data, and run fast.

### 2.2. Locomotion in retail areas

Locomotion models describe how pedestrians move in space at a resolution approximately of the human body. In the narrowest applications, they are used in simple environments without consideration for visual attention. The most prolific of which is the Social Force Model (SFM) (Helbing and Molnár, 1995) and its subsequent variations (Johansson et al., 2007; Karamouz et al.,

**Table 1**  
Visual attention module functionality comparison among previous work and ours.

Prior work	Visual coverage representation		All state transitions	Fast running	Other limitations
	Area	Location			
Xie et al. (2007)	✓	–	✓	✓	Not considering pedestrian orientations
Nassar (2011)	✓	Partially	–	✓	Not calibrated by real data
Wang et al. (2014)	✓	Partially	–	✓	Not calibrated by real data
Kremer et al. (2020)	✓	✓	–	–	Not calibrated by real data
Zhou et al. (2022)	Partially	✓	✓	✓	Invariant to attraction sizes
Ours	✓	✓	✓	✓	

2009; Zanlungo et al., 2012). Generally, the SFM determines pedestrian positions using their second derivatives that are summed by various hand-crafted forces. These forces represent the influences from walking goals, other pedestrians, the environment, etc.

Due to the popularity of the SFM, many previous studies that consider visual attention have built their models upon it. In Kwak et al. (2013) and Zhou et al. (2022), additional forces pointing to environmental objects were implemented to reflect the trend of pedestrians walking towards them. Nonetheless, as mentioned in Section 1, such a model design differs from the nature of attention-based movements, where pedestrians maintain their walking direction, but simultaneously draw attention to environmental objects. Such a model is therefore not considered in this paper.

Wang (2014) modeled attention-based movements by simply reducing walking speed to a fixed value when pedestrians draw their visual attention to environmental objects. But such a design is not supported by others' empirical findings, where pedestrians with visual attention showed varying speed, such as a reduction in walking speed as they approached the environmental objects (Li, 2023).

While there is little literature about the varying walking speed when visual attention is drawn, some cognitive science experiments may provide insights. Warren and Fajen (2008) claimed that pedestrian locomotion is guided by optical flow. Specifically, pedestrians try to maintain the angular position of their walking goals in their field of vision. We use this research as a reference point to explore similar rules for attention-based movements.

Therefore, in this paper, we will build a locomotion model for attention-based movements upon the SFM, and reveal how visual attention affects locomotion by the relative angular speed of the environmental objects.

### 2.3. Pedestrian-simulation-based design evaluation

Evaluating architectural design by crowd simulation has been a long-time practice (Hoogendoorn et al., 2004; Feng et al., 2016; Hu et al., 2020). However, the types of models vary in three key aspects. First, while many studies focus on egress behaviors in emergency scenarios (Hu et al., 2020), we consider normal daily scenarios (Feng et al., 2016) in which our evaluation metrics shift from safety (such as pedestrian density) to efficiency and retail potential.

Second, compared to some prior works that evaluate the design on a large scale with a focus on the graphical configurations of entire floor plans (Borgers and Timmermans, 1986; Feng et al., 2016), we adopt a micro-scale perspective focusing on the dimensions of environmental objects, similar to (Shukla, 2009; Hu et al., 2020). This is because, for the majority of transportation hubs, their floor plans are dominated by their main functions, leaving little configurational flexibility to retail spaces.

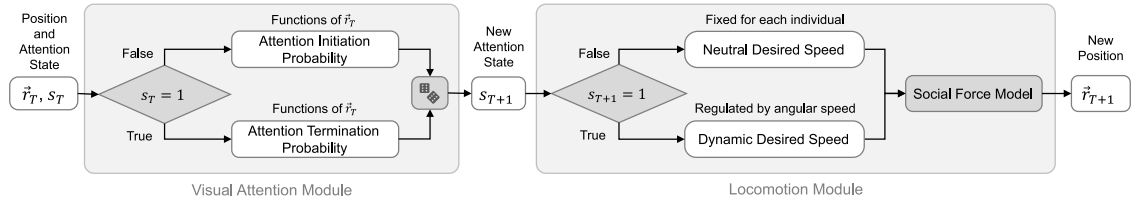
Finally, previous studies vary in the study of independent variables. While some researchers aim to find optimized sets of design variables either by searching through continuous parameter space (summarized in Nguyen et al. (2014)) or by comparing several designs that are drafted by experts (Pantano et al., 2021; Zhou et al., 2022), we want to get a more detailed understanding of the impacts of several variables by local sensitivity analysis or factorial designs, as in the work (Zhang et al., 2017).

## 3. Method

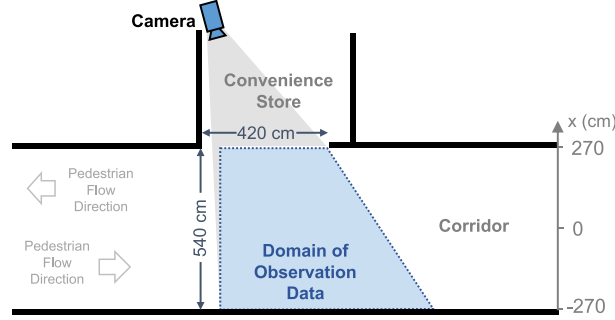
The attention transitions and walking speed variations in attention-based movement behaviors are jointly simulated as follows. For each virtual pedestrian, the simulation framework is a loop consisting of a visual attention module and a locomotion module (see Fig. 1), which produce attention states and positions respectively at every time step. We define the attention state  $s_T$  as a binary value, where  $s_T = 0$  means the pedestrian's attention is not drawn to the object and  $s_T = 1$  means the contrary.

The loop begins with the visual attention module (Section 3.2), where we derive the new visual attention state  $s_{T+1}$  by calculating the probability of state transition from its previous state  $s_T$ . In other words, if the visual attention is not initiated ( $s_T = 0$ ), we calculate the probability of initiating it. Otherwise, we calculate the probability of terminating it. Both probabilities are functions of position  $\vec{r}_T$ .

In the locomotion module (Section 3.3), we first calculate the desired walking speed of the pedestrian based on the visual attention state. If the pedestrian is not drawing attention to the environmental object ( $s_{T+1} = 0$ ), we set the desired walking speed to a neutral value, which is fixed for each pedestrian. Otherwise, the desired walking speed is determined by the angular speed to the environmental object (detailed in Section 3.3.2). Finally, the desired walking speed is used as an input of a Social Force Model, where the new position  $\vec{r}_{T+1}$  is calculated (see Section 3.3.3).



**Fig. 1. Model Architecture:** The simulation framework is a loop composed of two modules: visual attention and locomotion. The loop begins with a position  $\vec{r}_T$  and a visual attention state  $s_T$  at time step  $T$ . Based on them, the probability of initiating or terminating the attention is calculated. From the probability, we get the attention state at the next time step  $s_{T+1}$ . Then the new attention state determines the desired speed of a pedestrian, which is fed into a Social Force Model to get a new position  $\vec{r}_{T+1}$ .



**Fig. 2. Field observation site and the camera installation.** The study area covers the front of a convenience store. And the origin of the x-axis is set to the middle point of the corridor section line.

### 3.1. Data collection

Our model is fit to empirical data by [Tianchi Platform \(2023\)](#). For clarity, we briefly describe the data collection process. Data was collected from an underground bi-directional corridor at a metro station. A solitary convenience store is situated within the corridor, devoid of additional stores, obstacles, or environmental objects that may get pedestrians' visual attention. A camera was positioned at the store entrance (see [Fig. 2](#)), unobtrusively recording 45-minute video segments during off-peak daytime hours. After removing 97 pedestrians who went into the store, we captured the movements of 1,153 pedestrians. Due to the prevalence of pedestrian occlusions, pedestrian positions were manually labeled every 15 frames (0.5 s), drawing upon methodologies employed in prior research ([Gallup et al., 2012](#)). Then pedestrian trajectories were smoothed by cubic interpolation.

In this paper, a pedestrian's head turning towards the store is treated as a surrogate of their visual attention to it. The dependent variable – the visual attention state to the store – was manually labeled from video recordings every 0.5 s as:

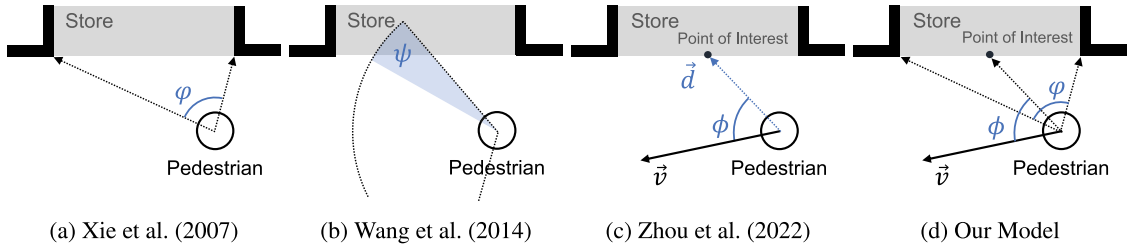
$$s_T = \begin{cases} 1, & \text{if pedestrian turns head towards store,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

### 3.2. Visual attention module

In this section, we firstly explain why we derive the visual attention state  $s_{T+1}$  by calculating the probability of attention state transition from its previous state  $s_T$  in Section 3.2.1. Then, we show how the position  $\vec{r}_T$  is converted to two variables that fully describe the visual coverage information, and how the variables constitute functions that represent the probabilities of attention state transitions in Section 3.2.2.

#### 3.2.1. Markov chain structure

Predicting attention states over time can be treated as a sequence prediction task, where some previous studies ([Zhou et al., 2022](#); [Wang et al., 2014](#)) do not consider temporal dependency between prediction results. In those cases, the likelihood of a given sequence was formulated as  $p(s_{1:T+1}) = p(s_1)p(s_2) \cdots p(s_T)p(s_{T+1})$ . While such models can successfully estimate the probability of attention status at time  $T$ , they fail to simulate the duration of behaviors as Markov chain does ([Haldi and Robinson, 2009](#)). To overcome the limitation, a first-order Markov chain is applied in this paper as:  $p(s_{1:T+1}) = p(s_1)p(s_2|s_1) \cdots p(s_T|s_{T-1})p(s_{T+1}|s_T)$ , where we predict attention state sequences by modeling the probabilities of the state transitions. The transition probability is only conditioned on the previous state.



**Fig. 3.** Input variables in prior and our work. Generally, our model combines the observation angle  $\phi$  in Zhou et al. (2022), with the angular separation  $\phi$  in Xie et al. (2007), resulting in a representation that can deal with varying environmental object sizes and orientations.

When fitting this model, we only need to model two probabilities: the probability of initiating visual attention when a pedestrian draws no attention to the object  $p(s_{T+1} = 1 | s_T = 0)$ , and that of the opposite event  $p(s_{T+1} = 0 | s_T = 1)$ . In simulations, the attention state of a pedestrian at the next time step is determined by sampling from a single-trial binomial distribution:

$$s_{T+1} \sim \mathcal{B}(n = 1, p = p(s_{T+1} = 1 | s_T)) \quad (2)$$

### 3.2.2. The probability of attention state transition

To implement the relations between visual coverage information and visual attention, we convert the position  $\vec{r}_T$  into two input variables below. Then, the probabilities of visual attention transitions are defined as functions of them.

**Angular Separation  $\phi$**  Similar to the definition in Xie et al. (2007), the angular separation describes the size of coverage that the environmental object applies to the visual field of a pedestrian. From a top-down view of the walking environment, If two tangent lines are drawn from the pedestrian position to environmental object boundaries, the angle between the two is the angular separation (Fig. 3).

**Observation Angle  $\phi$**  Similar to the definition in Zhou et al. (2022), it is the angle between the walking direction and the direction pointing from the pedestrian to the environmental object. In Zhou's work, a point of interest represented an individual object (e.g., a sign). In contrast, our use case is an entire store. To reduce the large footprint of a store to a representative point, we select the mid-point of the store's entryway as the point-of-interest (see Fig. 3(d)). We use such a representation in both our model and our implementation of Zhou et al. (2022) in Section 3.4.

To keep the model in its possible simplest form but include the possible interaction effects between inputs, we choose polynomial logistic regressions to predict the binary output (attention state). To avoid over-fitting, we limit the highest order to 2 and conduct step-wise backward feature exclusion at the final step of the modeling. That is, starting from modeling visual attention with all variables and their high-order combinations, we remove terms one by one until the prediction performance begins to drop. So we have the probability of attention initiation (written in a form prior to step-wise backward feature exclusion):

$$p_\theta(s_T = 1 | s_{T-1} = 0) = \sigma(\theta_0\phi + \theta_1\phi + \theta_2\phi^2 + \theta_3\phi^2 + \theta_4\phi\phi + \theta_5) \quad (3)$$

and the probability of attention termination:

$$p_\theta(s_T = 0 | s_{T-1} = 1) = 1 - p_\theta(s_T = 1 | s_{T-1} = 1) = 1 - \sigma(\theta_6\phi + \theta_7\phi + \theta_8\phi^2 + \theta_9\phi^2 + \theta_{10}\phi\phi + \theta_{11}) \quad (4)$$

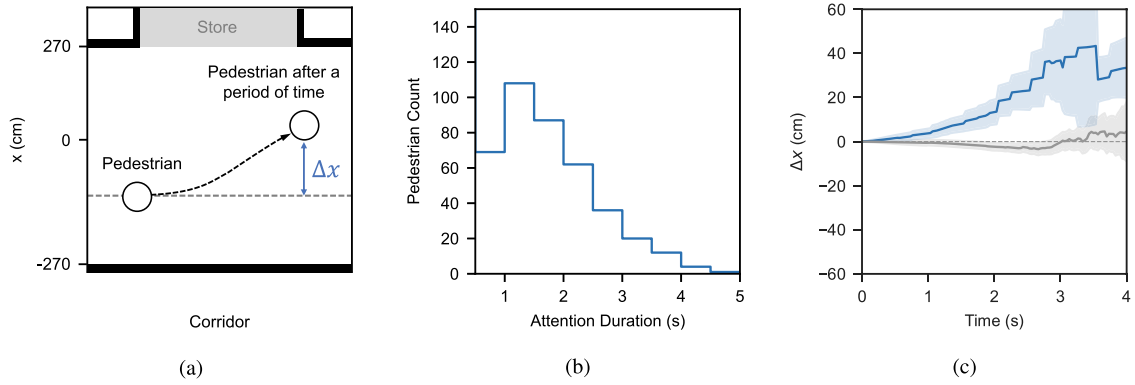
where  $\sigma(\cdot)$  denotes the sigmoid function. And all  $\theta$ s ranging from  $\theta_0$  to  $\theta_{11}$  are parameters to be fit. By plugging Eqs. (3) and (4) into Eq. (2), the attention state is simulated as:

$$s_{T+1} \sim \mathcal{B}(1, p), \text{ where } p = \begin{cases} \sigma(\theta_0\phi + \theta_1\phi + \theta_2\phi^2 + \theta_3\phi^2 + \theta_4\phi\phi + \theta_5), & \text{if } s_T = 0, \\ \sigma(\theta_6\phi + \theta_7\phi + \theta_8\phi^2 + \theta_9\phi^2 + \theta_{10}\phi\phi + \theta_{11}), & \text{if } s_T = 1. \end{cases} \quad (5)$$

Finally, to speed up the simulation and align with prior observation, we cut  $s_{T+1}$  down to zero when the angular separation  $\phi$  is less than a small threshold. It means pedestrians will never draw attention to an object when it occupies a too-small area of their vision fields. The threshold is chosen as 0.29 (in radian) from the empirical study in Xie et al. (2007). It should be noted that such a cut-off is only applied in simulation. And it imposes little impact on model fitting since the  $\phi$  in most datapoints are larger than the threshold.

### 3.3. Locomotion module

This section describes the locomotion of attention-based movements. In the first two subsections, we analyze our empirical data with regard to direction and speed changes. In the third subsection, we demonstrate how the empirical findings are integrated into the Social Force Model. Finally, we show the parameters, specifications, and boundary conditions in the simulation.



**Fig. 4. Walking direction change (lateral positional deviation) for pedestrians with visual attention.** (a) The illustration of walking direction change (lateral positional deviation)  $\Delta x$ . (b) The distribution of visual attention duration, shown in a histogram. Pedestrians without visual attention are not shown here. (c) The average deviation of lateral position after the initiation of attention. The gray solid line represents the deviation of pedestrians without visual attention after entering the camera.

### 3.3.1. Weak direction changes

Given the pedestrians with attention movement behaviors will finally continue their original routes rather than walking into the store, we prefer measuring the accumulated impact of direction change instead of the difference between the first and the last time step. Therefore, we define walking direction change as the lateral positional deviation  $\Delta x$  of a pedestrian after the initiation of visual attention. Fig. 4(a) depicts the definition.

We firstly explore how  $\Delta x$  changes over time. Here we study all pedestrians who initiate their visual attention inside our camera view with a duration  $\geq 1.5$  s. We set this threshold to lower the impacts of mislabeling attention states. As shown in Fig. 4(b), after 2.5 s of the attention initiation, most pedestrians have terminated their attention. For remaining pedestrians, their average  $\Delta x$  is significantly more than zero, but the value is only around 30 cm (see Fig. 4(c)). Therefore, for the majority of pedestrians with visual attention, their lateral position deviation is small.

Secondly, we explore whether  $\Delta x$  is affected by store proximity. We divide all pedestrians into 5 groups by their lateral positions when attention is initiated. Then we compare how average  $x$  changes across the groups. We observe no significant  $\Delta x$  differences between groups after 1.5 s of the attention initiation (F-test statistics=0.32 with the degrees of freedom (4, 112), p-value=0.866). Also, we do not observe significant differences after 2.5 s of the attention initiation (F-test statistics=1.24 with the degrees of freedom (3, 25), p-value=0.317). We are unable to compare groups with longer attention duration due to lack of data. While the insignificant results do not suggest store proximity applies no impact on  $\Delta x$  in the real world, they indicate that the impact is small when attention duration is less than 2.5 s.

Such an observation demonstrates that direction changes are very weak in attention-based movement behaviors. However, prior works (Kwak et al., 2013; Zhou et al., 2022) use additional SFM forces to describe similar behaviors. Additional forces are usually applied when pedestrians are assumed to steer towards environmental objects, and the steering effects increase (or decrease) with proximity. Since it is not aligned with our observation, we conclude that attention-based movements cannot be well explained or simulated by additional SFM forces in prior works.

### 3.3.2. Capped walking speed

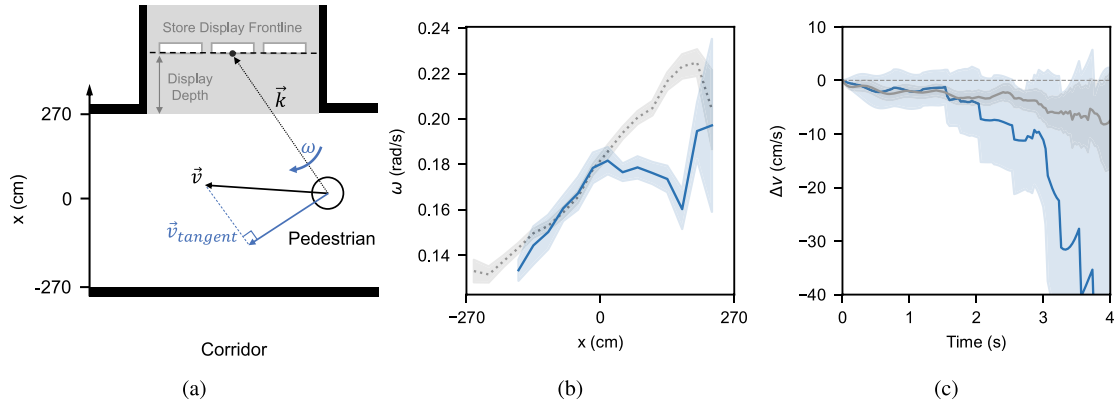
As direction changes do not play a key role in attention-based movement behaviors, speed changes are studied instead. Based on the idea in Warren and Fajen (2008), the angular speed with regard to the store display is studied.

Here the store display refers to goods that are visible to pedestrians outside the store. Such items would be clothes in the window display or food on shelves. While the layouts of store displays vary, the convenience store in our study adopts a simple strategy, where items are stacked on rows of shelves that form a frontline to pedestrians outside (see Fig. 5(a)). In this way, we represent the store display by a frontline. This representation can be extended to other displays such as window displays, where the display depth (the distance from the frontline to the store entrance) is small. Next, the position of the store display is further defined as the midpoint of the store display frontline (see Fig. 5(a)). It should be noted that this midpoint for store display is different from that for the store in Fig. 3(d).

Then the angular speed to the store display is calculated in this way: firstly we construct a vector  $\vec{k}$  pointing from the pedestrian to the midpoint of the store display. Then a tangential velocity  $\vec{v}_{tangent}$  is derived from projecting walking velocity  $\vec{v}$  onto the direction that is perpendicular to  $\vec{k}$ . Finally, the angular speed  $\omega$  is the  $\vec{v}_{tangent}$  divided by the length of  $\vec{k}$ .

To explore whether the angular speed is related to the proximity to the store, the average angular speeds stratified by pedestrian lateral positions are compared. The average angular speed is calculated in this way: for a given lateral position stratum, we query all datapoints falling into the stratum. Then the datapoints are grouped by pedestrian IDs to get the average angular speeds for each pedestrian. Finally, the average angular speeds are averaged across all pedestrians.





**Fig. 5. Angular speed  $\omega$  to store display.** (a) The illustration of angular speed  $\omega$  and tangential velocity  $\vec{v}_{tangent}$ . (b) How average angular speed  $\omega$  varies with lateral position ( $x$  coordinate).  $x = 0$  represents the middle of the corridor section.  $x = 270$  represents the wall boundary of the corridor on the store side and  $x = -270$  represents the opposite boundary. The gray dotted line represents the angular speed of pedestrians without visual attention to the store. And the shaded areas represent 95% CIs of the metrics. (c) The average walking speed change after the initiation of attention. The gray solid line represents the speed change of pedestrians without visual attention after entering the camera. The shaded areas represent 95% CIs of the metrics.

The result shows that pedestrians with visual attention to the store differ from those without attention in their angular speed. As shown in Fig. 5(b), while the average angular speed for pedestrians without attention increases with the proximity to the store entrance (gray dotted line), the angular speed for pedestrians with visual attention (blue line) is capped by a threshold value (its mean  $\mu_\omega$  and standard deviation  $\sigma_\omega$  are included in Table 3b). As a result, pedestrians whose attention is drawn to the store will walk more slowly, with the speed decreasing more as their distance to the store reduces.

We also explore how walking speed varies over time after the attention initiation. However, we do not observe significant speed differences for pedestrians who initiate visual attention inside camera view (see Fig. 5(c)) due to lack of data. Although we cannot reach a thorough conclusion on speed change over time, we can still use the current observation on the angular speed to build a locomotion model that approximates the slowing dynamics.

### 3.3.3. Social force model integration

The general idea of the locomotion module is: if pedestrians draw their visual attention to the store, they will adjust their desired speed to ensure the angular speed towards the store display will not exceed a certain value. We firstly explain how the desired speed regulates locomotion using Social Force Model. Then we demonstrate how to design the dynamic desired speed function.

In the Social Force Model (using the notation in Eqs. (2) and (3) in Johansson et al., (2007), for a pedestrian  $\alpha$ , its velocity  $\vec{v}_\alpha$  is calculated by:

$$\frac{d\vec{v}_\alpha(t)}{dt} = \frac{1}{\tau_\alpha} (v_\alpha^0 \vec{e}_\alpha - \vec{v}_\alpha) + \sum_{\beta(\alpha)} \vec{f}_{\alpha\beta}(t) + \sum_i \vec{f}_{\alpha i}(t) + \vec{\xi}_\alpha(t) \quad (6)$$

where  $f_{\alpha\beta}(t)$  and  $\vec{f}_{\alpha i}$  denote the repulsive effects from other pedestrians and the environment, respectively.  $\vec{\xi}_\alpha(t)$  is a random noise attribute. The first term (the goal force), describes the tendency to adjust the current velocity  $\vec{v}_\alpha$  to a desired one  $v_\alpha^0 \vec{e}_\alpha$ . It should be noted that  $v_\alpha^0$  is the desired speed in scalar and  $\vec{e}_\alpha$  is the desired direction (vector) pointing to the pedestrian's goal.

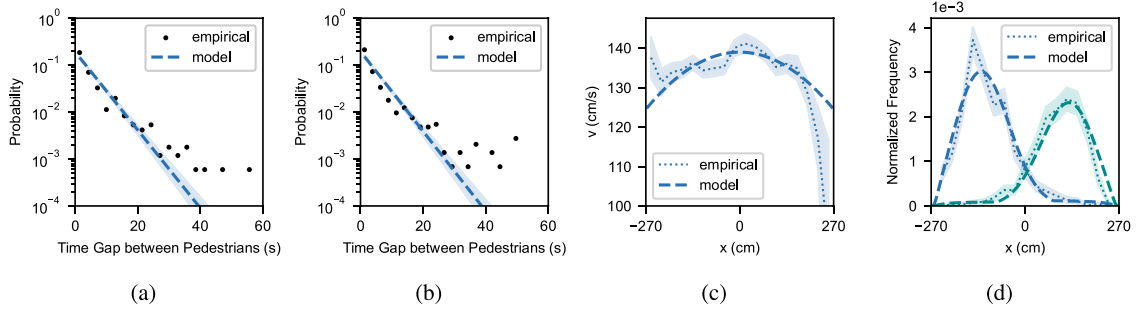
In Helbing and Molnár (1995) and Johansson et al. (2007), while the desired speed  $v_\alpha^0$  varies among pedestrians, it is a constant for a single pedestrian over time. In our paper, we turn it into a dynamic one. We update the desired speed at every time step by scaling a neutral speed with the ratio between an ideal angular speed and the current angular speed. And the ratio is capped by 1. As a result, if the current angular speed is greater than the ideal angular speed, the ratio will be less than 1. Then a smaller desired speed will guide a pedestrian to slow down. If the current angular speed is smaller than the ideal, the ratio will still be 1 so no influence is applied. The calculation steps are detailed below.

We first calculate the current angular speed by dividing the magnitude of tangential speed  $\vec{v}_{tangent}$  with the length of  $\vec{k}$  (illustrated in Fig. 5(a)):

$$\omega_{current} = \frac{\|\vec{v}_{tangent}\|}{\|\vec{k}\|} = \left(1 - \frac{\vec{v} \cdot \vec{k}}{\|\vec{v}\| \cdot \|\vec{k}\|}\right) \cdot \frac{\|\vec{v}\|}{\|\vec{k}\|} \quad (7)$$

Then we calculate the ratio  $\zeta$  between the ideal angular speed  $\omega_{ideal} \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$  (detailed in Section 3.3.2) and the current angular speed.  $\omega_{ideal}$  is sampled once when a pedestrian is initiated in the simulation. The ratio is capped by 1.

$$\zeta = \min(\omega_{ideal} / \omega_{current}, 1) \quad (8)$$



**Fig. 6. Boundary conditions:** (a) and (b) shows the probability distributions of the time gaps between pedestrians in two directions. The dots are empirical measurements. The dashed lines are exponential fitting results. The y-axes are shown in log scales. (c) shows how the average walking speed varies with lateral coordinates for pedestrians without visual attention, suggesting that pedestrians close to walls tend to walk more slowly. (d) shows the density distributions of pedestrians as a function of lateral coordinates. The two distributions in two colors represent the pedestrian flows in two directions. The shaded areas represent the 95% CIs for the metrics.

Next, the desired speed is scaled by the ratio. Besides, such a scaling only happens when the visual attention has been initiated (i.e.  $s_{T+1} = 1$ ):

$$v_{\alpha}^0 = v_{neutral} \cdot [\zeta \cdot s_{T+1} + (1 - s_{T+1})] \quad (9)$$

where  $v_{neutral}$  represents the desired speed without visual attention. It is fixed for a single pedestrian over time. But it varies among pedestrians, which is elaborated in Section 3.3.4. In this way, the desired speed controlled by attention states acts as a dynamic term in the goal force component of the Social Force Model.

### 3.3.4. Simulation configurations

To demonstrate that the success of our model relies on the design of visual attention module and its locomotion regulation (rather than the overfitted parameters in SFM and boundary condition), we configure SFM and the boundary condition only using pedestrians that do not show visual attention. In other words, we split all pedestrians in the empirical dataset into 3 disjoint groups: (1) pedestrians that never initialize visual attention or walk into the store, i.e. commuters. (2) pedestrians that initialize visual attention, that is, having attention-based movement behaviors. (3) pedestrians that enter or exit the store, i.e. consumers. We only use the first group for SFM and boundary condition (except flow rates) configuration. And we use first two groups to fit visual attention module and the flow rates. Then, we run two simulations with visual attention module activated and deactivated respectively. We expect that the simulation without visual attention module should produce results that resemble the behaviors of commuters (the first group). And the one with visual attention should be similar to the union of first two groups. The third group is not in our scope, as we clarified in Section 1. The results are shown in Figs. 9(c) and 9(d). And the configuration details are provided as below:

**Social force model.** We do not fit the Social Force Model (locomotion module) ourselves because there are few neighboring pedestrians for a large proportion of our trajectories, which lead to insufficient data for a traditional model fitting. Instead, we adapt existing models and specifications in prior work. Given that prior models were fitted using dataset with different pedestrian flow densities, we assign different parameters and specifications for different flow density scenarios. In this paper, we consider the cases where the pedestrian Level-of-Service is no more than Level C ( $1.4 \text{ m}^2$  per person, equivalent to a flow rate at around  $0.5 \text{ people}/(\text{m s})$  in all directions in our paper) as low-density scenarios, otherwise as high-density.

For low-density simulations that appear in the model validation Section 3.4.2 and the first case study Section 5.1, we adapt the parameters and specifications in Zanlungo et al. (2012), since it is fitted for such scenarios and is capable of replicating the speed distribution along the walking facility section (Zanlungo et al., 2012). Based on its original parameter values, we change two of them ( $A$  and  $\tau^{-1}$ , see Table 3c) to let the walking speed distribution to be aligned with the empirical data (see Fig. 9(c)). For high-density simulations that appear in the second case study Section 5.2, we adopt (Johansson et al., 2007), which does better in maintaining pedestrian flow stability in high-density scenes.

**Boundary condition.** We set the boundary conditions of pedestrian flows to resemble the original dataset in (Li, 2023). The boundary conditions are set by three key attributes: pedestrian flow rate, neutral speed, and density distribution. They are analyzed and modeled as below:

The pedestrian flow rate, referring to the number of pedestrians moving through a designated area over a period of time, is represented by time gaps between two pedestrians crossing the same vertical plane. After removing the gaps over 60 s, the time gaps follow an exponential distribution where the mean values are detailed in Table 3d (distributions are illustrated in Figs. 6(a) and 6(b)).

For pedestrians who never have attention drawn to the store, their walking speed distribution is used as the mean of the neutral walking speed  $\mu_v$  for all agents. Fig. 6(c) shows that pedestrians tend to walk faster at the center of the corridor. Such a trend is approximated by a second-degree polynomial function, where the mean of the neutral speed is affected by the initial lateral



**Table 2**

Variables for Representation Evaluation. It should be noted that some variables ( $\psi$ ,  $d$ ,  $n_g$ ,  $d_{lr}$ ,  $\rho$ ,  $t_g$ ,  $a_{max}$ ) are neither parts of our visual attention module nor locomotion module. They are only used for representation evaluation that is shown in 4.1.

Variable	Notion	Definition
<i>Intrinsic variable</i>		
Angular separation	$\varphi$	Xie et al. (2007), also see Section 3.2.2
Observation angle	$\phi$	Zhou et al. (2022), also see Section 3.2.2
Visibility	$\psi$	Wang et al. (2014)
Observation distance	$d$	Zhou et al. (2022)
Walking speed	$v$	
Angular speed	$\omega$	Angular speed with regard to environmental objects, see Section 3.3.2 for details
Cumulative gazing time	$t_g$	The cumulative time of gazing for a pedestrian
Walking direction	$d_{lr}$	A binary variable (0 or 1) for walking directions in bi-directional corridors
<i>Extrinsic variable</i>		
Gazing pedestrian number	$n_g$	The number of gazing surrounding pedestrians whose observation angle is less than 120 degrees and the distance is less than 3 m
Pedestrian density	$\rho$	The X-T definition (Duives, 2016) adapted from (Edie, 1963)
Maximum surrounding acceleration	$a_{max}$	The maximum acceleration of surrounding pedestrians. The scope of “surrounding” is same as that in gazing pedestrian number $n_g$

position:  $\mu_v(x) = a_v x^2 + b_v$ , with  $x$  in centimeters and  $\mu_v$  in centimeters per second. When agents are initialized, their desired walking speeds without visual attention  $v_{neutral}$  are sampled from  $\mathcal{N}(\mu_v(x), \sigma_v^2)$ , where the standard deviation  $\sigma_v$  is also calculated from the observation data.

Finally, the density function of pedestrians’ initial lateral position is modeled the same as a prior work (Zanlungo et al., 2014). In Zanlungo et al. (2014), the density function is proportional to a Boltzmann factor:  $p(x) \propto e^{U(x)}$ . And the energy function  $U(x)$  is designed to resemble a Gaussian at its maximum, drop to zeros at wall boundaries, and maintain a constant value far enough from the walls and its peak:

$$U(x) = \frac{a_\rho}{x} + \frac{a_\rho}{L_c - x} + \left(\frac{\delta}{b_\rho L_c}\right)^2$$

$$\delta = \begin{cases} x - c_\rho L_c & \text{if } |x - c_\rho L_c| \leq d_\rho L_c \\ d_\rho L_c & \text{if } |x - c_\rho L_c| \geq d_\rho L_c \end{cases}$$

where  $L_c = 540$  cm is the width of the corridor.  $a_\rho$  is fixed to 30 cm, as indicated in Zanlungo et al. (2014). The remaining parameters are calibrated using grid search with our dataset. The definition and the calibration results of these parameters are summarized in Table 3d. The fitted distribution and empirical measurements are visualized in Fig. 6(d).

### 3.4. Evaluation protocols

Our model is evaluated in two steps: firstly we compare the attention prediction results to ground truth, where we measure the performance of the visual attention module solely. Then we compare the simulation results to ground truth to know the performance of the whole model.

#### 3.4.1. Attention metrics

We evaluate the visual attention module by three aspects. First, to prove that we chose powerful variables as model inputs, our representations are compared to other possible combinations of variables in previous literature. Second, our model is compared to visual-attention models in selected previous work (Wang et al., 2014; Zhou et al., 2022; Kremer et al., 2020) to prove its performance in attention transition events (initiation and termination). Finally, as prediction errors for individual time steps may accumulate over time, good prediction performance on transition events may not necessarily mean good performance on sequential predictions. Therefore, to evaluate such accumulated errors, we compare the probability distributions of attention duration among all models.

**Representation Evaluation** A list of independent variables (intrinsic and extrinsic) is derived from a literature review. They are defined and calculated as in previous studies (see Table 2). To provide a general impression of variable impacts, we enumerate all uni-variate and bi-variate combinations for all listed variables. And we use those combinations to construct logistic regression models with the highest order up to 2. Then we compare the prediction performances of all models.

**Evaluation on transition events** Since previous work is not constrained to logistic regression, our model is compared to the works of (Wang et al., 2014; Zhou et al., 2022; Kremer et al., 2020) by fitting and validating with our dataset. All models are fit using gradient descent with an Adam optimizer. However, for (Wang et al., 2014) and Kremer et al. (2020), we also apply grid-search for comparison as their model structures are not designed for gradient descent and suffer exploding gradients in certain cases. Although we have applied smoothing terms to mitigate their numerical instabilities in gradient-descent, we believe their models may exhibit larger potentials if they are fit by grid-search.

Both evaluation tasks follow the same procedure: the datapoints in the original empirical dataset are firstly down-sampled at a sampling rate of 6 Hz. After the standardization, the sampled data is split into a training set and a testing set at a ratio of 3:1. Next, models are fit to the training set. Evaluation metrics are derived by running prediction tasks on the testing set. All procedures above are repeated (100 times for representation evaluation and 30 times for the other) to get the distribution of the metrics. ROC Curves and AUC scores are provided to describe the true positive rate when the false positive rate is controlled. The results using this protocol are shown in Section 4.1.

**Evaluation on attention duration** As we assume a model with fewer accumulated errors produces an attention duration distribution that is closer to empirical observation, we calculate the difference between the attention duration distribution of empirical data and that of simulation for each model. The duration distribution for each model is derived by simulating all attention states of each pedestrian trajectory frame-by-frame where the previous attention state for each frame is the previous simulation result. On the contrary, in the previous two evaluation tasks, previous attention states were always the ground truth. The difference between two duration distributions is shown by the difference of their cumulative distribution functions (CDFs). Wasserstein distance, which sums up the absolute of these differences, was chosen as the quantitative metric. The results are shown in Section 4.1.

### 3.4.2. Locomotion metrics

To demonstrate that our model can jointly predict visual attention and speed variations, which are proxies of retail potential and pedestrian efficiency, the proportion of long-attention pedestrians and the average speed are compared to empirical datasets. In addition, to demonstrate our model's ability to show the spatial variance of these metrics, the metrics are calculated as functions of lateral position. They are defined as below:

**Proportion of Long-attention Pedestrians**  $P_{long}(X)$  Pedestrians who draw their attention to the store for no less than 2.5 s are defined as long-attention pedestrians. The duration threshold here is chosen based on the empirical observation in [Gidlöf et al. \(2017\)](#), where the probability of buying starts to rapidly increase when pedestrians spend more than 2.5 s looking at a product. Then, given a lateral coordinate stratum  $X \in [a, b)$ , we count the number of pedestrians whose datapoints fallen into this stratum as  $N_{all}$ . Next, we count the number of long-attention pedestrians in this stratum as  $N_{long}$ . Finally we have  $P_{long}(X) = N_{long}/N_{all}$ .

**Walking Speed**  $v(X)$  For each pedestrian  $p$  in the population  $S$ , for all its datapoints falling into the lateral coordinate stratum  $X \in [a, b)$ , average the walking speed as  $v_{p,X}$ . After we get all  $v_{p,X}$  for every  $p$ , the final result  $v(X)$  is the average of all  $v_{p,X} > 0$ .

## 4. Results

In this section we firstly show the prediction performance of our visual attention module. Then we demonstrate that our entire framework can simulate the attention-based movement with regard to both visual attention and locomotion.

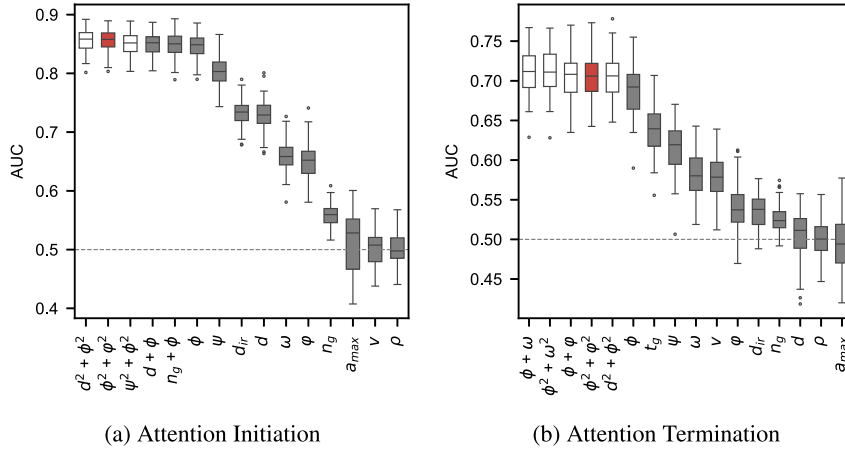
### 4.1. Visual attention modeling

In the representation evaluation, our model representation was among the top candidates with regard to prediction performance, whereas other top candidates showed no significant difference from ours (see [Fig. 7](#)).

Among all uni-variate logistic regressions, observation angle  $\phi$  was the most influential variable in both attention state transition events. On the contrary, some variables (mostly extrinsic ones) imposed little impact on the dataset. For instance, the AUC scores of pedestrian density  $\rho$  and maximum surrounding acceleration  $a_{max}$  on both attention transition events were close to 0.5 (see [Fig. 7](#)). Also, gazing pedestrian number  $n_g$  (defined in [Table 2](#)) played little role in attention termination predictions. However, the results do not indicate the extrinsic variables are not influential in all real-world cases. Because our dataset did not capture pedestrian flows with extreme densities.

Next, we compare our model with prior work in visual attention modeling. For attention initiation events (see [Fig. 8\(a\)](#)), our model was on par with ([Zhou et al., 2022](#)), a reasonable result given our representation is a coordinate transform of [Zhou et al. \(2022\)](#) for the attention initiation part. For attention termination events (see [Fig. 8\(b\)](#)), our model achieved 11% higher AUC score than the best prior work. With regard to attention duration, our model achieved the closest distribution to empirical observation (40% less Wasserstein distance than the best prior work, see [Fig. 8\(d\)](#)). While ([Zhou et al., 2022](#)) and [Kremer et al. \(2020\)](#) tend to overestimate the proportion of pedestrian with short attention duration, ours excels in predicting long-attention pedestrians, which increases the overall performance. Although we introduce more parameters than some previous studies ([Wang et al., 2014](#); [Kremer et al., 2020](#)), it achieves prediction improvements without significant additional computational cost.

Finally, using the whole dataset, our final model fitting results are shown in [Table 3a](#). It should be noted that some variable combination terms (such as  $\theta_8$  for  $\varphi^2$  in attention termination predictions) were removed from Eq. (5) by step-wise backward feature selections.



**Fig. 7.** AUC scores of visual attention predictions with different representations. Boxplots are drawn based on 100 trials with randomized samplings of our dataset. An AUC score equal to 1 indicates the model is always correct in the test dataset. And a score close to 0.5 indicates that the model performance is the same as random choices. The red boxes denote the performance of our representation. The gray boxes are representation performances that have significantly different mean values from ours. The x-axis labels are abbreviations of representations, where a single letter means a uni-variate linear logistic regression. And two letters connected by a plus (such as  $\phi + \omega$ ) mean a bi-variate linear logistic regression. A squared letter means the logistic regression is polynomial up to 2 degrees. It should be noted that we only display the top five results and all uni-variate linear fitting results here.

#### 4.2. Locomotion modeling

We firstly show that attention-based movement pedestrians make a major contribution to the slowing effects. In Fig. 9(a), the difference between the mean walking speed function of all pedestrians (group 1+2+3, see group definitions in Section 3.3.4) and that of group 1+2, is smaller than the difference between group 1+2+3 and group 1. It shows the removal of group 3 (consumers) does not impose a major impact to pedestrian flows, especially for  $x$  coordinate intervals where most pedestrians appear. But group 2 does so, which makes itself as our focus in this paper.

Our simulation was generally in accordance with the empirical data. In Fig. 9(b), the proportion of long-attention pedestrians  $P_{long}$  decreased with the distance to the store entrance. In Fig. 9(d), walking speed function  $v$  showed a decaying slowing effect of the store when the distance increases.

We also show our simulation results without visual attention module in Fig. 9(c) (in magenta), where we observe the walking speed distribution aligns with empirical data where we only keep commuters in the empirical data (in green). After we include pedestrians with visual attention into the empirical data (in blue, see Fig. 9(d)), we observe the slowing effect that can only be captured when we enable the visual attention module in our simulation (in red). Therefore, we prove that the slowing effect stems from visual attention module and its regulation on locomotion, rather than the over-fitting from SFM or boundary condition parameters.

#### 5. Case study

In this section, we explore how architectural design features may influence pedestrian efficiency and retail potential using our model. The first application aims to reveal the influence of three key design feature variables (corridor width, store entrance width, and display depth) qualitatively. The second one shows the process of striking a balance between efficiency and retail potential by changing the corridor width incrementally in simulations.

##### 5.1. Qualitative analysis on design feature impacts

To reduce the computation complexity in simulation, discretizations are applied to design feature variables: We pick two values for each variable. Then we compare the simulation results between all combinations of the values. The values are discretized as: corridor width  $L_c = \{3.5 \text{ m}, 8 \text{ m}\}$ , store entrance width  $L_s = \{4.2 \text{ m}, 6 \text{ m}\}$ , and display depth (defined in Section 3.3.2)  $L_d = \{0.5 \text{ m}, 5 \text{ m}\}$ . The combination of these values provide 8 test conditions (illustrated in Fig. 10).

The pedestrian flows are generated in two directions. For each direction, the flow rate for a unit of corridor width is fixed to  $0.08 \text{ people m}^{-1} \text{ s}^{-1}$ . Other parameters are set as Table 3.

In the analysis, the average speed loss  $\Delta v$  and long-attention-pedestrian proportion  $P_{long}$  are compared between trials. The speed loss here is defined as the speed difference between the simulation and an ideal simulation, which has the same corridor width but has no stores. The two metrics above are shown in their spatial distribution form. In other words, rather than analyzing the corridor as a whole, the corridor is divided into rectangular cells. And the metrics are calculated for each cell using the pedestrian trajectory segments inside it.

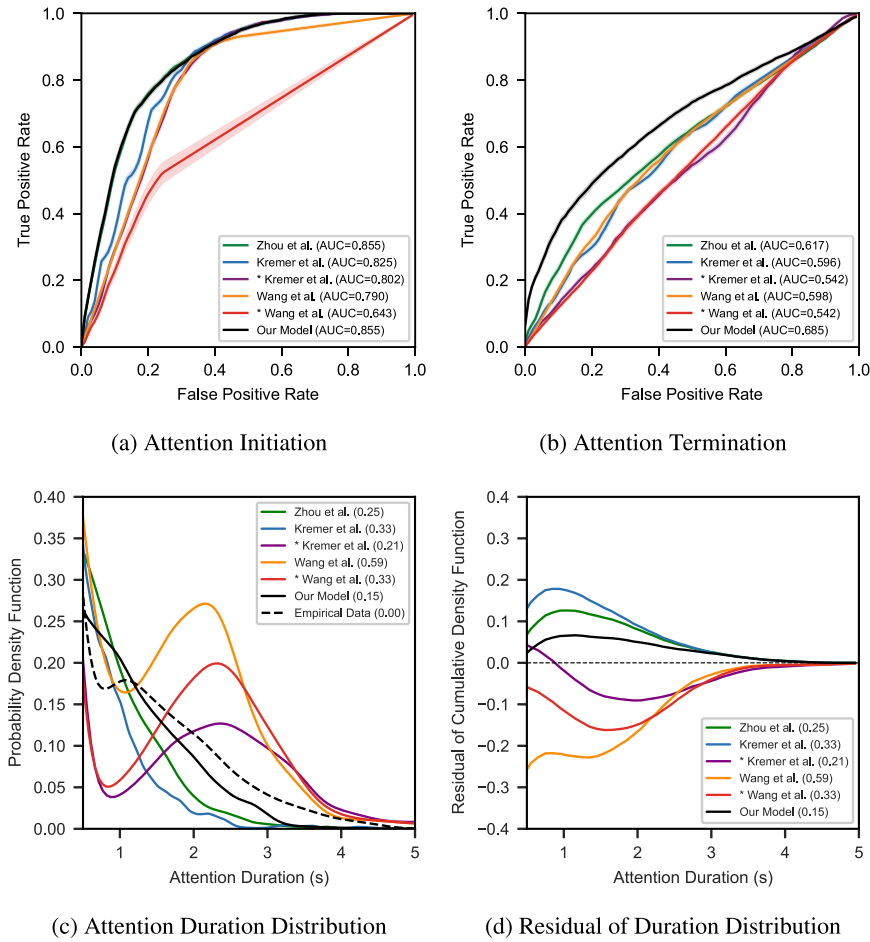
**Table 3**

All model parameters and simulation specifications. The first table is the final model fitting for visual attention modules. Parameters are fitted using standardized input variables (the means and the standard deviations are shown in the “Value” column). All parameter estimates were significantly different from zero ( $p$ -value < 0.05). The second table includes parameters in Section 3.3.3 that calculate desired speed. The third table covers all parameters in the SFM for low density scenarios, where asterisks denote the parameter values that differ from the original prior work (Zanlungo et al., 2012). The last table lists all parameters in the boundary condition configuration.

Variable Notation	Value	Parameter Notation	Estimate	Std. errors	ValueZ scores
<i>Attention initiation events</i>					
$\varphi$	$0.981 \pm 0.433$	$\theta_0$	3.167	0.351	9.031
$\phi$	$1.797 \pm 0.558$	$\theta_1$	-1.542	0.056	-27.308
$\varphi^2$	$1.151 \pm 1.008$	$\theta_2$	-2.359	0.306	-7.709
(intercept)	-	$\theta_5$	-4.683	0.107	-43.682
<i>Attention termination events</i>					
$\varphi$	$1.366 \pm 0.383$	$\theta_6$	-0.804	0.160	-5.009
$\phi$	$1.350 \pm 0.504$	$\theta_7$	-2.510	0.363	-6.918
$\phi^2$	$2.076 \pm 1.472$	$\theta_9$	1.060	0.266	3.980
$\phi\varphi$	$1.806 \pm 0.776$	$\theta_{10}$	0.828	0.193	4.284
(intercept)	-	$\theta_{11}$	1.177	0.048	24.560
(a) Visual attention module					
Parameter	Description		Value	Unit	
$\mu_{\omega}$	Mean of ideal angular speed towards the store		0.18	rad/s	
$\sigma_{\omega}$	Standard deviation of ideal angular speed towards the store		0.04	rad/s	
(b) Attention-based locomotion regulation					
Parameter	Description		Value	Unit	
$A$	Weight of pedestrian repulsive effect		1.13*	m/s <sup>2</sup>	
$B$	Distance discounting factor of pedestrian repulsive effect		1	m	
$r_v$	Pedestrian perception range		5.6	m	
$t_{max}$	Maximum predicted collision time		6.1	s	
$A^w$	Weight of environment repulsive effect		0.9	m/s <sup>2</sup>	
$B^w$	Distance discounting factor of environment repulsive effect		1.0	m	
$\theta_v$	Velocity bias		0.16	rad	
$\lambda$	Asymmetry parameter		0.95		
$\tau^{-1}$	Inverse of relaxation time		2*	s <sup>-1</sup>	
$\kappa$	Weight of physical collision		5000		
(c) Social Force Model in low density					
Parameter	Description		Value	Unit	
$t_+$	Mean time gap between pedestrians (direction 1)		5.11	s	
$t_-$	Mean time gap between pedestrians (direction 2)		5.22	s	
$a_v$	Parameter in neutral speed function		-1.3	10 <sup>-4</sup>	
$b_v$	Intercept of neutral speed function		139	cm/s	
$\sigma_v$	Standard deviation of desired walking speed		30	cm/s	
$a_p$	Distance regulation to the wall		24.78	cm	
$b_p$	Width factor of pedestrian flow		0.2		
$c_p$	Positional factor of maximum density		0.24		
$d_p$	Probability of walking on the wrong side		0.36		
$L_c$	Width of the corridor		540	cm	
(d) Boundary condition					

As illustrated in Fig. 11(a), design feature variables imposed complex impacts on the distribution of  $P_{long}$ .  $P_{long}$  did not always decrease with the distance to the store. The trend was conditioned on corridor widths. For very narrow corridors, a larger proportion of pedestrians drew their attention to the store when they were located further away from it. This can be attributed to the constrained visual coverage in narrow corridors. For pedestrians who walk on the store side, they will suffer a larger observation angle in a narrow corridor, which discourages them to draw their attention for a long time. For the other two variables, store widths were positively correlated to visual attention while display depths played little role.

While not influential to  $P_{long}$ , display depths greatly affected  $\Delta v$ . As shown in Fig. 11(b), shallower storefront displays led to larger speed decreases, which resulted from the angular speed adaptations. Since store displays are much closer, pedestrians have to decrease speed more to keep the angular speed at a low level. We also observed such impacts interacting with the other two variables. For instance, the speed decrease got smaller when the store became wider. That can be explained by the spread-out of visual attention distribution, which makes pedestrians get attracted when they are further away from the store, lowering their need to maintain the angular speed.



**Fig. 8.** Compare our visual attention module with prior work. (a) and (b) are ROC curves with regard to attention initiation and termination predictions. When the curve gets closer to the top-left corner, the model achieves a higher true positive rate when the false positive rate is fixed. Shaded areas represent the standard errors of true positive rates. Our model is on par with Zhou et al. (2022) in attention initiation modeling, shown by their overlapping curves. The model names with asterisks (e.g., Wang et al.) refer to the modified models fitted by gradient descent. (c) compares the attention duration distributions of model results and empirical data by probability density functions. For visualization quality, the lines are smoothed by a Gaussian kernel. In the legend, the numbers in the parenthesis are the Wasserstein distances to the empirical distribution. (d) shows the differences between the attention duration distribution of model simulation and that of empirical data, represented in the residuals of cumulative distribution function. When the line is closer to  $y=0$ , the simulation result is closer to empirical data. The legend is the same as the previous sub-figure, and lines are also smoothed by a Gaussian kernel.

## 5.2. Optimization: strike the balance

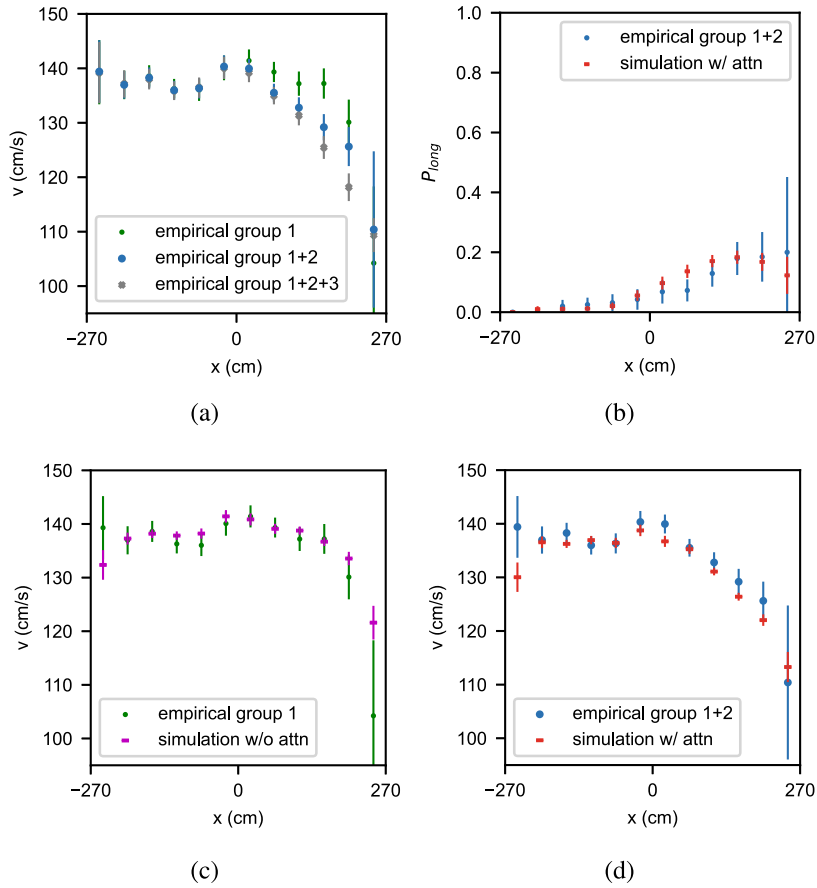
Finally, to demonstrate our model can help designers find the balance between efficiency and retail potential, we used our model to run a local sensitivity analysis on corridor width in two store display depth settings. These options resemble decisions in real practice where designers need to find an ideal facility dimension when other external factors are fixed.

In the simulation, display depths are set as 5 m and 0.5 m respectively. In both cases, the flow rates for each direction are fixed to 2 people/s, and the entrance widths are fixed to 4.2 m. Other parameters are set as Table 3. We incrementally change the corridor width (illustrated in Fig. 12) to evaluate its impact on two metrics (long-attention pedestrian proportion  $P_{long}$  and average walking speed  $v$ ). Here the corridor is analyzed as a whole without spatial discretizations.

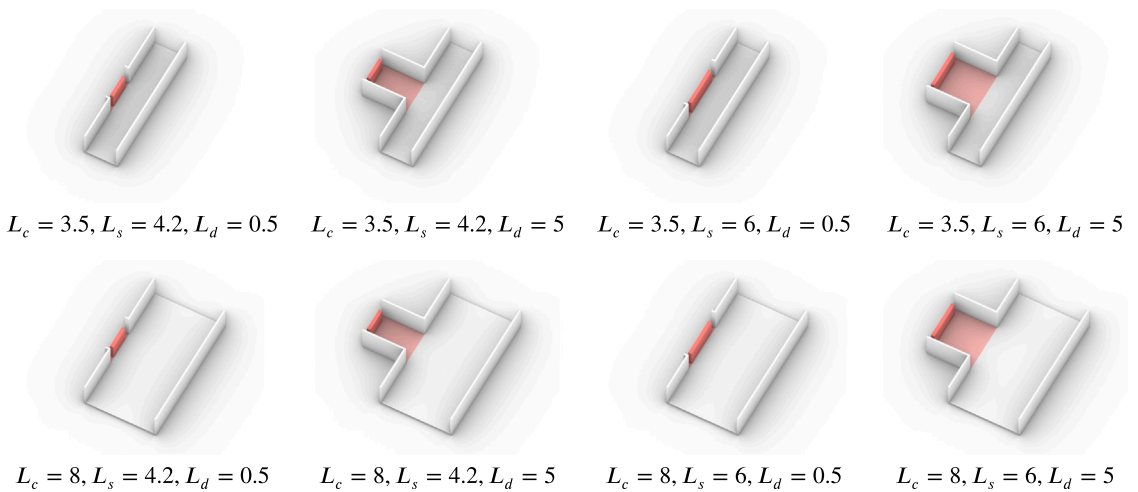
As shown in Fig. 13, we observed the trade-offs between pedestrian flow efficiency (average walking speed) and retail potential (long-attention pedestrian) in both scenarios. Although the actual balance point depends on the designers' judgments, the figures help designers to make data-informed decisions by providing the possible outcomes in different scenarios.

## 6. Conclusion

In this paper, we propose a novel pedestrian simulation framework focused on providing feedback on architectural design decisions to increase retail potential and pedestrian flow efficiency for common retail spaces similar to transportation hubs. Our

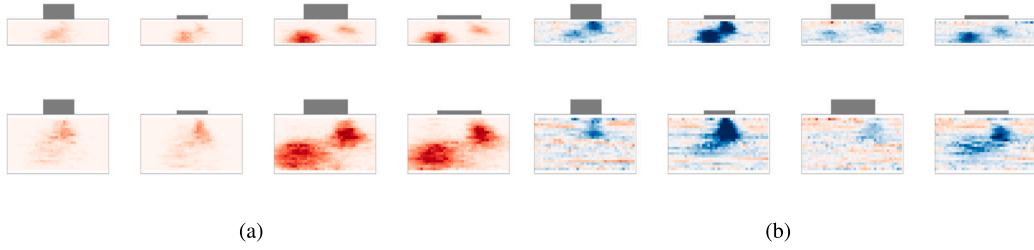


**Fig. 9. Simulation results compared to empirical data.** To better visualize the contribution of our model, we split all pedestrians in the empirical dataset into 3 disjoint groups: (1) commuters; (2) pedestrians with attention-based movement behaviors; (3) consumers. (see definitions in Section 3.3.4) (a) Mean walking speed  $v$  as a function of  $x$  coordinate in empirical data. The error bars represent the 95% CIs for the metrics. We compare the slowing effects when different groups are included in the population. (b) The proportion of long-attention pedestrians  $P_{long}$  as a function of  $x$  coordinate. (c) Mean walking speed  $v$  as a function of  $x$  coordinate for commuters (group 1). Green: empirical measurement. Magenta: simulation data with visual attention module disabled (simulation of group 1). (d) Mean walking speed  $v$  as a function of  $x$  coordinate for commuters and pedestrians with attention-based movement (group 1+2). Blue: empirical measurement. Red: simulation data with visual attention module enabled (simulation of group 1+2).

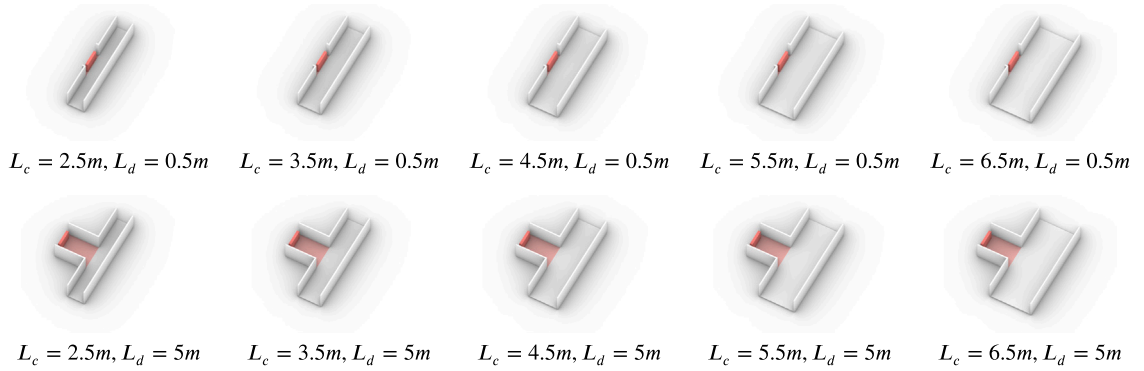


**Fig. 10. Design settings with varying features in isometric views.** White floors represent the corridor segment. Light red floors are store areas. And the red volumes represent store displays. Corridor width  $L_c$ , store width  $L_s$ , and store depth  $L_d$  are all in the unit of meter here.

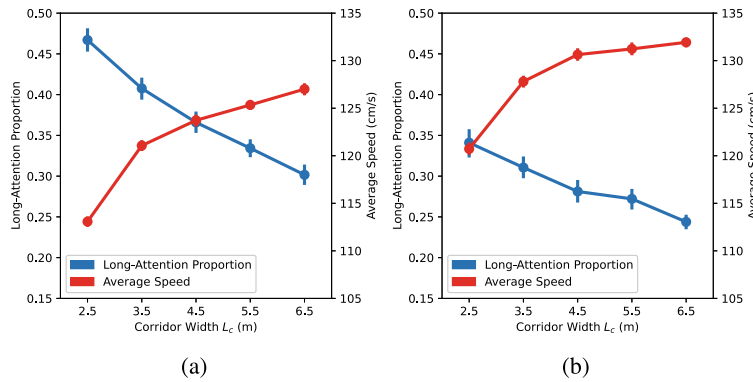




**Fig. 11. Simulations with varying design features** Refer to Fig. 10 for corresponding 3D representation. (a) The distribution map of  $P_{long}$ , which means the probability of observing a pedestrian with long-time visual attention to the store; (b) The distribution map of  $\Delta v$ , which means the average speed loss compared to the same corridor without a store. Maps are top-down views of the corridor in real-world dimensions. The gray boxes represent the dimensions of the store width and display depth.



**Fig. 12. Design settings with varying corridor width in isometric views.** White floors represent the corridor segment. Light red floors are store areas. And the red volumes represent store displays.



**Fig. 13. Simulations with varying corridor widths.** The proportion of long-attention pedestrians and average walking speed for all pedestrians are plotted in the same diagram with different scales. There are two scenarios where (a) store display depth is 0.5 m, and (b) depth is 5 m. The error bars represent the 95% CIs for the metrics.

model uses visual attention to introduce the modification of the Social Force Models desired velocity coefficient for the goal force. The model that combines visual attention and locomotion, is compared to prior works by fitting to an existing dataset. Finally, we demonstrate qualitative results through a simple case study of a corridor and a single storefront.

In general, the model evaluates retail potential and pedestrian flow efficiency by jointly predicting the visual attention and locomotion of the pedestrians at the same time. The visual attention, which is the proxy to retail potential, is able to simulate the transitions between the dual states of visual attention, making it adaptable to multi-store simulations. As for the model representation, we extend the model in Zhou et al. (2022) to the varying sizes of attractions by changing the distance factor in the model to angular separation in Xie et al. (2007). In this way, the attention prediction is purely guided by ego-centric visual information, which maximizes its potential in generalizations. Also, Our model outperforms prior work in the prediction of attention

terminations. While previous work by Zhou et al. (2022) chose a time-decaying function to represent how visual attentions fade, our study indicates such transitions can be better explained by pedestrian visual fields.

Then the locomotion model is conditioned on visual attention prediction. Compared to prior work that chose a fixed walking speed for pedestrians with visual attention (Wang, 2014), we propose a dynamic walking speed scheme. We found that when pedestrians are attracted by environmental objects, their walking speed is regulated by the angular speed with regard to the objects. Such a setting can be related to experimental findings by Warren and Fajen (2008), where the author concluded that human locomotion is guided by optical flows.

Additionally, we show that our model is responsive to design feature variations, showing its potential in design optimizations. By configuring store widths and display depths, our model can be applied to various environmental objects such as stores, banners, and window displays. By adjusting corridor widths, architects can compare different design proposals with regard to facility dimensioning. Although the model in this paper is built upon a dataset of attention-based movement behaviors, its findings may give insights into similar topics such as signage detections or evacuation wayfinding.

Our work has some limitations that should be addressed in future work: First, in visual attention modeling, we only discuss a limited number of extrinsic factors in diluted pedestrian flows due to data availability. Future works should cover more extrinsic factors. Second, the empirical data we use is limited in the diversity of location and in pedestrian flow density. Our finding about factor influences on visual attention does not apply to pedestrian flows with extreme high densities. Third, while attention-based movement behaviors play a dominant role in our context, integration of other response modes would be beneficial to create more realistic simulations. For example, the pedestrians who walk into the store. Fourth, although our visual attention model is compared to the original forms of prior models, a comparison that considers the varying parameter sizes of the models may better reveal the advantage of ours. Since we have not found a way to fairly compare models of which parameter size cannot be easily scaled, we consider this topic as a potential future work. Finally, since the overall performance of our simulation framework is closely related to that of the Social Force Model, the limitations inherit in this model propagate, and such a base model can be improved to explore our attention-based model input.

## CRedit authorship contribution statement

**Danrui Li:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Mathew Schwartz:** Writing – review & editing, Writing – original draft, Visualization, Methodology. **Samuel S. Sohn:** Writing – review & editing, Formal analysis. **Sejong Yoon:** Writing – review & editing, Funding acquisition, Formal analysis. **Vladimir Pavlovic:** Writing – review & editing, Funding acquisition, Formal analysis. **Mubbasir Kapadia:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis.

## Data availability

The empirical data can be found at <https://tianchi.aliyun.com/dataset/162948>.

## Acknowledgments

This material is based upon work supported by the U.S. Department of Homeland Security, United States under Grant Award 22STESE00001-03-02. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. The research was also supported in part by U.S. National Science Foundation #1955365, #1955404, and #2324598.

Our study (ID: Pro2024000416) is identified as non-human subject research by Human Research Protection Program (HRPP) Institutional Review Board (IRB) at Rutgers University. The research activities do not meet the regulatory definition of human subjects research provided in 45 CFR 46.102. This project does not involve interaction or intervention with living individuals, nor does this project involve the use of identifiable private information or bio-specimens.

The dataset used in the research originates from Tianchi Platform (<https://tianchi.aliyun.com/dataset/162948>).

## References

- Baron, N., 2019. Designing Paris Gare du Nord for pedestrians or for clients? New retail patterns as flow optimization strategies. *Eur. Plan. Stud.* 27 (3), 618–637. <http://dx.doi.org/10.1080/09654313.2018.1562651>, Publisher: Routledge.
- Borgers, A., Timmermans, H., 1986. A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geogr. Anal.* 18 (2), 115–128. <http://dx.doi.org/10.1111/j.1538-4632.1986.tb00086.x>.
- Carrasco, M., 2011. Visual attention: The past 25 years. *Vis. Res.* 51 (13), 1484–1525. <http://dx.doi.org/10.1016/j.visres.2011.04.012>.
- Chen, C.H., 2011. Attention theory-based agent system: Using shopping street design simulation as an example. *J. Chinese Inst. Eng., Trans. Chinese Inst. Eng., Ser. A* 34 (1), 155–168. <http://dx.doi.org/10.1080/02533839.2011.553035>.
- Duives, D., 2016. Analysis and Modelling of Pedestrian Movement Dynamics at Large-scale Events (Ph.D. thesis). Delft University of Technology, <http://dx.doi.org/10.4233/UUID:08831F69-9B8E-44CF-8AFE-F4A3E7BC9A9C>.
- Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. Port of New York Authority. Google-Books-ID: 3uBmYgEACAAJ.
- Feng, T., Yu, L.-F., Yeung, S.-K., Yin, K., Zhou, K., 2016. Crowd-driven mid-scale layout design. *ACM Trans. Graph.* 35 (4), 132:1–132:14. <http://dx.doi.org/10.1145/2897824.2925894>.
- Gallup, A.C., Hale, J.J., Sumpter, D.J., Garnier, S., Kacelnik, A., Krebs, J.R., Couzin, I.D., 2012. Visual attention and the acquisition of information in human crowds. *Proc. Natl. Acad. Sci. USA* 109 (19), 7245–7250. <http://dx.doi.org/10.1073/pnas.1116141109>, Publisher: National Academy of Sciences.

- Gidlöf, K., Anikin, A., Lingonblad, M., Wallin, A., 2017. Looking is buying. How visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. *Appetite* 116, 29–38. <http://dx.doi.org/10.1016/j.appet.2017.04.020>.
- Haldi, F., Robinson, D., 2009. Interactions with window openings by office occupants. *Build. Environ.* 44 (12), 2378–2395. <http://dx.doi.org/10.1016/j.buildenv.2009.03.025>.
- Hänseler, F.S., 2016. Modeling and Estimation of Pedestrian Flows in Train Stations (Ph.D. thesis). INTER, EPFL, Lausanne, p. 152. <http://dx.doi.org/10.5075/epfl-thesis-6876>.
- Helbing, D., Farkas, I.J., Vicsek, T., 2000. Freezing by heating in a driven mesoscopic system. *Phys. Rev. Lett.* 84 (6), 1240–1243. <http://dx.doi.org/10.1103/PhysRevLett.84.1240>, Publisher: American Physical Society.
- Helbing, D., Molnár, P., 1995. Social force model for pedestrian dynamics. *Phys. Rev. E* 51 (5), 4282–4286. <http://dx.doi.org/10.1103/PhysRevE.51.4282>, Publisher: APS.
- Hoogendoorn, S.P., Hauser, M., Rodrigues, N., 2004. Applying microscopic pedestrian flow simulation to railway station design evaluation in Lisbon, Portugal. *Transp. Res. Rec.* 1878 (1), 83–94, Publisher: SAGE Publications Sage CA: Los Angeles, CA.
- Hu, K., Yoon, S., Pavlovic, V., Faloutsos, P., Kapadia, M., 2020. Predicting crowd egress and environment relationships to support building design optimization. *Comput. Graph.* 88, 83–96. <http://dx.doi.org/10.1016/j.cag.2020.03.005>.
- Johansson, A., Helbing, D., Shukla, P.K., 2007. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Adv. Complex Syst.* 10, 271–288. <http://dx.doi.org/10.1142/S0219525907001355>.
- Karamouzas, I., Heil, P., van Beek, P., Overmars, M.H., 2009. A predictive collision avoidance model for pedestrian simulation. In: Egges, A., Geraerts, R., Overmars, M. (Eds.), *Motion in Games*. In: *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, pp. 41–52. [http://dx.doi.org/10.1007/978-3-642-10347-6\\_4](http://dx.doi.org/10.1007/978-3-642-10347-6_4).
- Kremer, M., Haworth, B., Kapadia, M., Faloutsos, P., 2020. Watch out! modelling pedestrians with egocentric distractions. In: *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games. MIG '20*, Association for Computing Machinery, New York, NY, USA, pp. 1–10. <http://dx.doi.org/10.1145/3424636.3426910>.
- Kremer, M., Haworth, B., Kapadia, M., Faloutsos, P., 2021. Modelling distracted agents in crowd simulations. *Vis. Comput.* 37 (1), 107–118. <http://dx.doi.org/10.1007/s00371-020-01969-4>.
- Kwak, J., Jo, H.-H., Luttinen, T., Kosonen, I., 2013. Collective dynamics of pedestrians interacting with attractions. *Phys. Rev. E* 88 (6), 062810. <http://dx.doi.org/10.1103/PhysRevE.88.062810>.
- Li, D., 2023. How do retail stores affect pedestrian walking speed: An empirical observation. [arXiv:2310.04818](https://arxiv.org/abs/2310.04818).
- Milosavljevic, M., Navalpakkam, V., Koch, C., Rangel, A., 2012. Relative visual saliency differences induce sizable bias in consumer choice. *J. Consum. Psychol.* 22 (1), 67–74. <http://dx.doi.org/10.1016/j.jcps.2011.10.002>.
- Nassar, K., 2011. Sign visibility for pedestrians assessed with agent-based simulation. *Transp. Res. Rec.* 2264 (1), 18–26. <http://dx.doi.org/10.3141/2264-03>, Publisher: SAGE Publications Inc.
- Nguyen, A.-T., Reiter, S., Rigo, P., 2014. A review on simulation-based optimization methods applied to building performance analysis. *Appl. Energy* 113, 1043–1058. <http://dx.doi.org/10.1016/j.apenergy.2013.08.061>.
- Pantano, E., Pizzi, G., Bilotta, E., Pantano, P., 2021. Enhancing store layout decision with agent-based simulations of consumers' density. *Expert Syst. Appl.* 182, 115231. <http://dx.doi.org/10.1016/j.eswa.2021.115231>, Publisher: Pergamon.
- Saunders, R., Gero, J.S., 2004. Curious agents and situated design evaluations. *AI EDAM* 18 (2), 153–161. <http://dx.doi.org/10.1017/S0890060404040119>, Publisher: Cambridge University Press.
- Shukla, P.K., 2009. Genetically optimized architectural designs for control of pedestrian crowds. In: Korb, K., Randall, M., Hendtlass, T. (Eds.), *Artificial Life: Borrowing from Biology*. In: *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, pp. 22–31. [http://dx.doi.org/10.1007/978-3-642-10427-5\\_3](http://dx.doi.org/10.1007/978-3-642-10427-5_3).
- Tianchi Platform, 2023. Position and visual attention trajectories in underground retail area. URL: <https://tianchi.aliyun.com/dataset/162948>.
- Wang, W., 2014. Modeling of Pedestrian Behavioral Dynamics in Complex Urban Settings and Built Environment (Ph.D. thesis). City University of Hong Kong, URL: [https://scholars.cityu.edu.hk/en/theses/theses\(6294b498-c4ce-40a7-9865-61096a81e805\).html](https://scholars.cityu.edu.hk/en/theses/theses(6294b498-c4ce-40a7-9865-61096a81e805).html).
- Wang, W.L., Lo, S.M., Liu, S.B., Kuang, H., 2014. Microscopic modeling of pedestrian movement behavior: Interacting with visual attractors in the environment. *Transp. Res. C* 44, 21–33. <http://dx.doi.org/10.1016/j.trc.2014.03.009>, Publisher: Elsevier Ltd.
- Warren, W.H., Fajen, B.R., 2008. Behavioral dynamics of visually guided locomotion. In: *Coordination: Neural, Behavioral and Social Dynamics*. Vol. 2008, Springer, Berlin, Heidelberg, pp. 45–75. [http://dx.doi.org/10.1007/978-3-540-74479-5\\_3](http://dx.doi.org/10.1007/978-3-540-74479-5_3), ISSN: 18600832.
- Xie, H., Filippidis, L., Gwynne, S., Galea, E.R., Blackshields, D., Lawrence, P.J., 2007. Signage legibility distances as a function of observation angle. *J. Fire Protect. Eng.* 17 (1), 41–64. <http://dx.doi.org/10.1177/1042391507064025>.
- Zacharias, J., 2021. Pedestrian dynamics on narrow pavements in high-density Hong Kong. *J. Urban Manag.* 10 (4), 409–418. <http://dx.doi.org/10.1016/j.jum.2021.05.003>.
- Zanlungo, F., Chigodo, Y., Ikeda, T., Kanda, T., 2014. Experimental study and modelling of pedestrian space occupation and motion pattern in a real world environment. In: *Pedestrian and Evacuation Dynamics 2012*. Springer International Publishing, Cham, pp. 289–304. [http://dx.doi.org/10.1007/978-3-319-02447-9\\_24](http://dx.doi.org/10.1007/978-3-319-02447-9_24).
- Zanlungo, F., Ikeda, T., Kanda, T., 2012. A microscopic “social norm” model to obtain realistic macroscopic velocity and density pedestrian distributions. In: Perc, M. (Ed.), *PLoS One* 7 (12), e50720. <http://dx.doi.org/10.1371/journal.pone.0050720>.
- Zhang, Z., Jia, L., Qin, Y., 2017. Optimal number and location planning of evacuation signage in public space. *Saf. Sci.* 91, 132–147. <http://dx.doi.org/10.1016/j.ssci.2016.07.021>.
- Zhou, M., Dong, H., Wang, X., Hu, X., Ge, S., 2022. Modeling and simulation of crowd evacuation with signs at subway platform: A case study of Beijing subway stations. *IEEE Trans. Intell. Transp. Syst.* 23 (2), 1492–1504. <http://dx.doi.org/10.1109/TITS.2020.3027542>, Conference Name: IEEE Transactions on Intelligent Transportation Systems.