ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Leveraging Reinforcement Learning for Generating Construction Workers' Moving Path: Opportunities and Challenges

Minguk Kim[1]*, Tae Wan Kim[2]

[1] *Department of Architecture & Urban Design,* Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 22012, Korea, *E-mail address:* mingukkim@inu.ac.kr
[2] *Department of Architecture & Urban Design,* Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 22012, Korea, *E-mail address:* taewkim@inu.ac.kr

Abstract: Travel distance is a parameter mainly used in the objective function of Construction Site Layout Planning (CSLP) automation models. To obtain travel distance, common approaches, such as linear distance, shortest-distance algorithm, visibility graph, and access road path, concentrate only on identifying the shortest path. However, humans do not necessarily follow one shortest path but can choose a safer and more comfortable path according to their situation within a reasonable range. Thus, paths generated by these approaches may be different from the actual paths of the workers, which may cause a decrease in the reliability of the optimized construction site layout. To solve this problem, this paper adopts reinforcement learning (RL) inspired by various concepts of cognitive science and behavioral psychology to generate a realistic path that mimics the decision-making and behavioral processes of wayfinding of workers on the construction site. To do so, in this paper, the collection of human wayfinding tendencies and the characteristics of the walking environment of construction sites are investigated and the importance of taking these into account in simulating the actual path of workers is emphasized. Furthermore, a simulation developed by mapping the identified tendencies to the reward design shows that the RL agent behaves like a real construction worker. Based on the research findings, some opportunities and challenges were proposed. This study contributes to simulating the potential path of workers based on deep RL, which can be utilized to calculate the travel distance of CSLP automation models, contributing to providing more reliable solutions.

Key words: site layout planning, reinforcement learning, construction site, pathfinding

1. INTRODUCTION

Construction Site Layout Planning (CSLP) is the task of determining the attributes (e.g., type, size, and rotation), location, and the number of temporary facilities temporarily placed on the construction site. Placing these facilities in an optimal location can reduce project costs, construction duration, and site safety can be improved [1]. Therefore, to find the optimal layout, in many studies, optimization models have been developed through various algorithms such as Genetic Algorithm [2–5], Total Potential Energy [6], Fuzzy logic [7], and Particle swarm optimization [8]. One of the main goals of this optimization problem is to minimize the resources' total travel distance between facilities for the efficiency of resources.

Approaches to compute travel distance between facilities in previous studies focus on the shortest distance between facilities. For example, some studies using the linear distance between facilities for optimization, do not consider obstacles [2,6,7]. Studies using access roads for computing the travel distance require predefined road networks and there is no guarantee that actual workers and equipment follow these networks [3,8]. As another example, there are approaches that only target the shortest distance such as Visibility Graph, Dijkstra, and A* [4,9,10], and most recently, there was an attempt to consider the uncertainty of movement in following the shortest distance by a construction worker in a grid-based site layout through fuzzy graph theory [5].

Due to these limitations, when optimizing a site layout, the travel distance calculated from existing approaches may become an inappropriate input. Therefore, in order to solve this problem, it is necessary to choose a different approach that can describe the movement path of actual workers.

However, in general, if there are several alternatives to the route to the destination, humans do not necessarily follow one shortest route, but can choose a safer and more comfortable route according to the situation within a reasonable range [11]. In fact, it is mentioned that in the study of human wayfinding, various factors are considered in addition to the shortest route in human wayfinding [12–15]. Similarly, in an environment such as a construction site, the route may vary depending on the difference in risk depending on the region, and in fact, among the areas where site workers can cross, there is a tendency to bypass the area where heavy equipment is frequently used [16].

Therefore, in order to obtain reliable construction site layout solutions in the optimization model, a wayfinding approach that mimics the wayfinding behavior of construction workers on a real site is necessary. To this end, reinforcement learning (RL) influenced by neurological learning mechanisms can be used [17]. The reason is that the learning form of RL is similar to the process for the human learners to obtain a cognitive-like map [18]. Generally, it is known that humans use cognitive maps to navigate in a given environment [19]. According to neuroscience research, the human cognitive map, which enables positioning and navigation of the spatial environment, is created by repeating the process of activating certain cells called *grid cells* in the hippocampus region of the brain whenever one visits a location in the environment [20]. In other words, if humans are located in an unknown environment, they will not have spatial information about that environment so, in order to find a path, humans need to search the environment to repeatedly construct spatial information in a given environment.

Since the RL, which learns a policy to take action on each state in the direction of maximizing the cumulative reward in iterative search, has computational similarity to cognitive mapping, the abstract concept of the cognitive map can be mathematically transformed [21]. Thus, RL can be used to create pathfinding similar to the behavior of construction workers. Therefore, for the purpose of using the RL-based pathfinding approach, this paper identifies the tendencies of workers' wayfinding behavior from a literature review on the walking behavior of pedestrians and construction workers. In addition, this paper shows that the identified construction workers' wayfinding tendencies can be designed as a reward, and accordingly, the potential paths of the workers can be simulated through deep RL. Lastly, several opportunities and challenges are proposed to improve the RL-based approach based on the findings of the study.

2. BACKGROUND

2.1. Reinforcement learning

RL, One of the machine learning technology, aims to learn how to solve a specific problem by interacting with the environment surrounding, which is formulated as Markov Decision Process

(MDP). In MDP, which is extended from the concept that the future state is independent of the previous state (also known as Markov property), an agent acts according to the lapse of discrete timesteps. When the agent observes the current state s belonging to a finite set of states in the environment and takes a specific action a belonging to a finite set of actions, it moves to the next state s' which also belongings to a finite set of states according to the state transition probability $P_{ss'}^{a}$ and gets the reward $R_{ss'}^{a}$. This process repeats until a terminal state is reached. In this episodic task, in order for the agent to achieve the desired purpose, it is necessary to receive as many cumulative rewards as possible.

To this end, In this paper, one of the policy-based deep reinforcement learning (DRL) algorithms, Proximal Policy Optimization (PPO) [22], is used. PPO is a stable and appropriate algorithm to learn the optimal policy to maximize the cumulative reward in an environment containing a continuous action space. The loss function defined in PPO to approximate the optimal policy is defined as (1).

$$L^{CLIP+VF+S}(\theta) = \widehat{\mathbb{E}}_t[L_t^{CLIP}(\theta) - c_1 L_t^{VF}(\theta) + c_2 S[\pi_\theta](s_t) \qquad (1)$$

Where $\widehat{\mathbb{E}}_t$ is the expectation operator, $L_t^{CLIP}(\theta)$ is the clipped surrogate objective, $L_t^{VF}(\theta)$ is the squared-error loss of value functions $\left(V_\theta(s_t) - V_t^{targ}\right)^2$, S denotes the entropy loss, π_θ is the policy that depicts the probability of taking a specific action in a specific state, which is approximated by neural networks as an optimal policy, and c_1, c_2 are coefficient. Here, the clipped surrogate objective $L_t^{CLIP}(\theta)$ is defined as:

$$L_t^{CLIP}(\theta) = \widehat{\mathbb{E}}_t[\min\left(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)] \qquad (2)$$

Where $\hat{A}_t$ is an estimator of the advantage function at timestep t, $r_t(\theta)$ is the probability ratio $\pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$, and ϵ is a hyperparameter. $r_t(\theta)$ is clipped at $1-\epsilon$ or $1+\epsilon$ depending on whether the advantage is positive or negative. In this manner, large policy updates can be prevented, stable learning is possible.

Finally, in the PPO algorithm, T trajectory segments are collected by interacting between the agent and the environment during T timesteps using the old policy $\pi_{\theta_{old}}(a_t|s_t)$ from N actors at every iteration. Then, we compute $\hat{A}_t$ using a truncated version of generalized advantage estimation (GAE) (3).

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_t + \cdots + \cdots (\gamma\lambda)^{T-t+1}\delta_{T-1} \qquad (3)$$

Where $\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$, and γ is the discount factor. Then the loss function (1) has constructed on these $N * T$ time steps of data, and optimized it with mini-batch Stochastic Gradient Descent, for K epochs.

2.2. The tendency of human wayfinding behavior

In RL, an agent's specific goal is achieved through reward-based learning. Therefore, it is necessary to identify how construction workers behave in the real world. To this end, it is necessary to identify their tendency from studies dealing with human wayfinding behavior.

Humans seem to subconsciously prefer paths that seek to minimize complexity [12]. For instance, people use the shortest or less congested routes [13,14] and take the straightest possible route to It tries to maintain linearity [12]. Other studies show that humans follow the path with the longest sight of the line first [12] or try to use the path with the least change in direction [15]. In addition, paths may vary depending on individual stress levels [13], and factors such as the attractiveness of the route, sidewalk quality, absence of long waits at traffic lights, etc. affect their paths in addition to safety [14]. Also, there is a tendency to use the first noticed route or to use the usual route [14]. These various characteristics are also shown in construction simulations [23].

Additionally, workers at the construction site tend to change their path to avoid the dangers caused by on-site hazards [24]. Therefore, since the difference in danger level by type of facility or zone is relatively large, the workers' path may also be affected by it (e.g., the area around the tower crane is more dangerous than a temporary office).

In summary, since construction sites are relatively less influenced by others compared to general pedestrian environments, among the identified tendencies, the dominants of workers' wayfinding are the shortest path, maintaining direction, and facility avoidance.

3. FEASIBILITY OF THE RL-BASED PATHFINDING APPROACH

3.1. State and Action space

To model a RL agent that represents construction workers, one must define a state space and an action space. First, state space is defined as information about the environment obtained through spatial perception at every time step. It consists of the following elements: 1) the relative position between the agent and the destination, 2) the agent's velocity, 3) the difference between the agent's moving direction and the direction to the destination, 4) the ratio between the total distance traveled to the destination and the linear distance, and 5) visual perception data by using rays. Second, for the natural movement of the agent, the action space consists of horizontal and vertical movements, which are continuous types, and the agent’s movement speed is set at 1.38m/s, which is the average human walking speed [25].

3.2. Reward design

Although reward design is the most critical part of the success of RL, it is difficult to design a good reward signal [18]. A general reward signal design process is known as an informal trial-and-error search process to find a signal that produces a satisfactory result. In this process, the reward formulation to make the RL agent move like the real construction worker was configured, and the reward signal value with the best learning result was found by selecting whether the average cumulative reward convergence or not and the convergence speed as the evaluation criteria for agent performance.

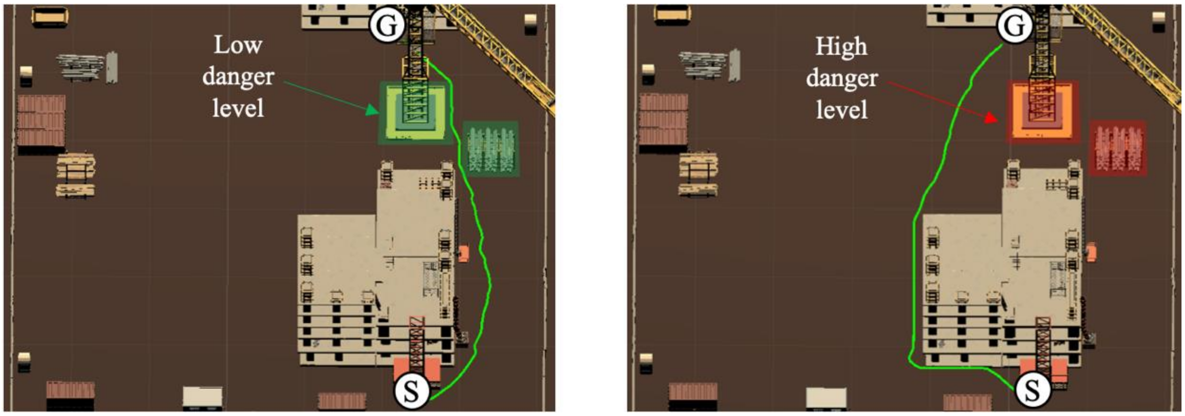
As a result of this, the RL agent is able to keep moving by getting a very small penalty for each timestep. And, in order to guide the agent to the destination, the agent is rewarded as its moving direction is closer to the destination. The identified human tendencies then are able to map as rewards. Accordingly, the agent avoids the facilities while considering the danger level of the facility, and moves to the destination by the shortest path while changing the direction to the minimum.

3.3. Simulation

In this paper, we used the Unity engine and ML-agents tool [26] to show the simulation results. Unity is a game engine that provides an intuitive interface for development close to realistic simulations and provides a machine learning library called ML-agents. For the simulation, we used the PPO algorithm aforementioned in Section 2.1, and Curriculum Learning (CL) [27] to improve the learning performance was also applied. (Table 1) shown the hyperparameters used throughout the simulation. The RL agents were trained on two different start-goal sets in the same construction site environment and the paths generated by these agents are shown similar to actual workers’ pathfinding (Fig 1-2).

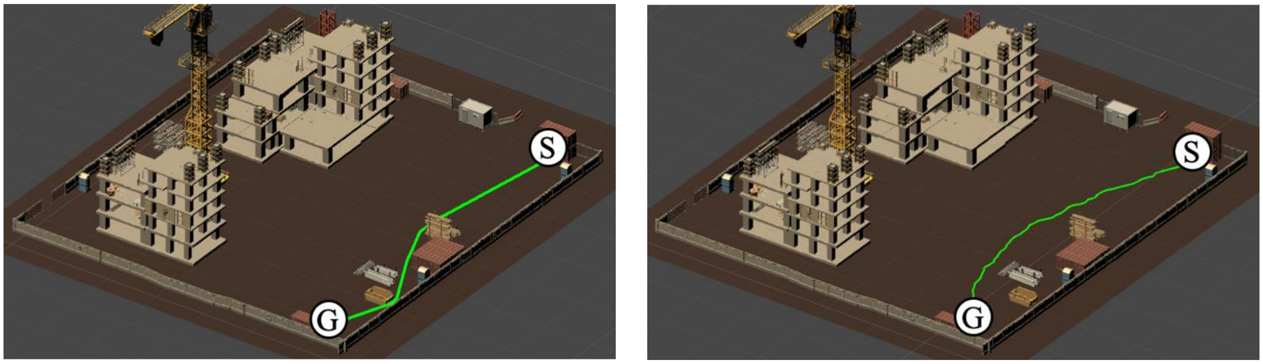
(Fig. 1) shows that the RL agent is able to generate different potential paths according to the danger level of facilities (tower crane and laydown area) that exist between the start and goal point. Because the facilities in (Fig 1. (a)) have a relatively high level of danger compared to those in (Fig

1. (b)), the path seems to detour the dense area of hazardous facilities while being relatively safe. (Fig. 2) shows the path generated by the A* algorithm (Fig. 2(a)) and the RL agent (Fig. 2(b)) for another start-goal set. When compared to the A* algorithm, it can be seen that the RL agent simulated a path with less direction change. These paths are particularly seen in the paths of workers moving loads with equipment such as wheelbarrows.



(a) (b)

Figure 1. RL-based paths with different danger level in a construction site environment



(a) (b)

Figure 2. Paths simulated by (a) A* algorithm, and (b) RL-based approach

4. ELICITED OPPORTUNITIES AND CHALLENGES

The given results confirm that the proposed approach plausibly generates the paths of real construction workers, and reveal the potential of this approach for calculating travel distances in CSLP. Therefore, the RL-based approach has the potential to overcome the limitations of existing approaches, such as not considering the existence of obstacles, only following a predefined network, or generating a path based only on the shortest distance among human wayfinding tendencies. Based on the results of the study, several opportunities and challenges were found.

First, there is still room for improvement to obtain satisfactory results through the reward design. For example, since the interpretation of a facility's danger level may differ from person to person or situation to situation, it may be necessary to examine the danger level of each facility from what construction workers perceived, and how to implement it as an appropriate reward function should also be considered.

Second, the given result only shows that the generated path can be similar to the actual path. However, in order to bring the plausibility of the proposed approach, it is necessary to carry out a face validation of how realistic the path generated by the proposed approach is [28]. Specifically, after generating paths with the existing and proposed approaches in virtual site layout environments, a survey can be conducted asking real workers to evaluate how similar each path is to the trajectories of real workers. These experimental studies can confirm the usefulness and necessity of the proposed approach [29].

Third, one of the disadvantages of reinforcement learning is that the training period is long. In this paper, it took about 24 minutes on average to train the RL agent 1 million times. To solve this problem, it is necessary to investigate the performance of various hyperparameters and perform sensitivity analysis. Through this, not only the simulation time can be shortened but also the computational performance can be improved based on the sensitivity analysis result.

Lastly, a framework may be proposed to facilitate this approach. The framework includes the required functions for implementation, required attributes that define the types and characteristics of site objects required for travel distance computation and Construction Site Layout (CSL) optimization, and may include well-configured CL and learning environment configuration.

Table 1. Simulation hyperparameters

Hyperparameter	**Value**
Batch size	128
Buffer size	2560
Learning rate α	3e-4
Regulation term Beta β	3e-3
Clipping parameter Epsilon ϵ	0.2
GAE parameter Lambda λ	0.95
Number of epochs	3
Discount factor Gamma γ	0.99
Time horizon (trajectory length)	64
Hidden units	256
Number of layers	3

5. CONCLUSION

The pathfinding approaches used in existing CSLP optimization models focus only on the shortest path. However, humans consider factors other than the shortest path and choose a safer and more comfortable path within a reasonable range. In order to simulate a realistic path for workers reflecting these characteristics, a RL-based pathfinding approach that mimics the decision-making and behavioral processes of wayfinding of construction workers were proposed.

To do so, in this paper, the tendency of workers' wayfinding was identified through a literature review on pedestrians and construction workers' walking behavior. Also, the feasibility of the proposed approach was confirmed by using a simulation developed by mapping the identified tendencies to the reward design. The simulation showed that the RL agent behaves like a real construction worker.

The following future works were proposed based on the research findings: 1) In order to obtain satisfactory agent performance, it is necessary to examine the perceived danger level of each facility from construction workers and the way to design a proper reward formulation. 2) Face validation and similarity validation by using surveys and the workers' actual trajectory data should be performed to bring the plausibility of the proposed approach. 3) To improve training performance, hyperparameter performance can be investigated and sensitivity analysis and tuning can be performed. In addition, the similarity of the simulated paths can be improved by coordinating the reward value using the trajectory data of the actual workers. 4) A framework may be proposed to facilitate the proposed approach.

In this paper, the potential of the RL-based pathfinding approach for mimicking the actual construction worker path was confirmed, which can be applied as a new approach to compute the travel distance of the CSLP optimization model in the future. Accordingly, the proposed approach is able to contribute to the reliability of the construction site layout derived from the optimization model in the future.

ACKNOWLEGEMENTS

This work was supported by the National Research Foundation of Korea(NPF) grant funded by the Korea government(MSIT) (No. NRF-2021R1A2C1013188)

REFERENCES

[1] M. Kim, H. Ryu, T.W. Kim, Typology Model of Temporary Facility Constraints for Automated Construction Site Layout Planning, Appl. Sci. 11 (2021) 1027. https://doi.org/https://doi.org/10.3390/app11031027.

[2] M.J. Mawdesley, S.H. Al-jibouri, H. Yang, Genetic algorithms for construction site layout in project planning, J. Constr. Eng. Manag. 128 (2002) 418–426. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:5(418).

[3] H.M. Sanad, M.A. Ammar, M.E. Ibrahim, Optimal construction site layout considering safety and environmental aspects, J. Constr. Eng. Manag. 134 (2008) 536–544. https://doi.org/10.1061/(ASCE)0733-9364(2008)134.

[4] S.S. Kumar, J.C.P. Cheng, A BIM-based automated site layout planning framework for congested construction sites, Autom. Constr. 59 (2015) 24–37. https://doi.org/10.1016/j.autcon.2015.07.008.

[5] M. Rezaee, E. Shakeri, A. Ardeshir, H. Malekitabar, Optimizing travel distance of construction workers considering their behavioral uncertainty using fuzzy graph theory, Autom. Constr. 124 (2021) 103574. https://doi.org/10.1016/j.autcon.2021.103574.

[6] M. Andayesh, F. Sadeghpour, Dynamic site layout planning through minimization of total potential energy, Autom. Constr. 31 (2013) 92–102. https://doi.org/10.1016/j.autcon.2012.11.039.

[7] A.R. Soltani, T. Fernando, A fuzzy based multi-objective path planning of construction sites, Autom. Constr. 13 (2004) 717–734. https://doi.org/10.1016/j.autcon.2004.04.012.

[8] V. Benjaoran, V. Peansupap, Grid-based construction site layout planning with Particle Swarm Optimisation and Travel Path Distance, Constr. Manag. Econ. 0 (2019) 1–16. https://doi.org/10.1080/01446193.2019.1600708.

[9] M. Andayesh, F. Sadeghpour, A comparative study of different approaches for finding the shortest path on construction sites, Procedia Eng. 85 (2014) 33–41. https://doi.org/10.1016/j.proeng.2014.10.526.

[10] I. Abotaleb, K. Nassar, O. Hosny, Layout optimization of construction site facilities with dynamic freeform geometric representations, Autom. Constr. 66 (2016) 15–28.

https://doi.org/10.1016/j.autcon.2016.02.007.

[11] A. Sevtsuk, R. Basu, X. Li, R. Kalvo, A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco, Travel Behav. Soc. 25 (2021) 41–51. https://doi.org/10.1016/j.tbs.2021.05.010.

[12] R.C. Dalton, The secret is to follow your nose: Route path selection and angularity, Environ. Behav. 35 (2003) 107–131. https://doi.org/10.1177/0013916502238867.

[13] H. Li, J. Zhang, L. Xia, W. Song, N.W.F. Bode, Comparing the route-choice behavior of pedestrians around obstacles in a virtual experiment and a field study, Transp. Res. Part C Emerg. Technol. 107 (2019) 120–136. https://doi.org/10.1016/j.trc.2019.08.012.

[14] J.F. Morrall, Analysis of factors affecting the choice of route of pedestrians, Transp. Plan. Technol. 10 (1985) 147–159. https://doi.org/10.1080/03081068508717309.

[15] E.K. Sadalla, D.R. Montello, Remembering Changes in Direction, Environ. Behav. 21 (1989) 346–363. https://doi.org/10.1177/0013916589213006.

[16] T. Cheng, U. Mantripragada, J. Teizer, P.A. Vela, Automated Trajectory and Path Planning Analysis Based on Ultra Wideband Data, J. Comput. Civ. Eng. 26 (2012) 151–160. https://doi.org/10.1061/(asce)cp.1943-5487.0000115.

[17] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015) 436–444. https://doi.org/10.1038/nature14539.

[18] Sutton, Richard S., A.G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[19] T. Wolbers, M. Hegarty, What determines our navigational abilities?, Trends Cogn. Sci. 14 (2010) 138–146. https://doi.org/10.1016/j.tics.2010.01.001.

[20] C.F. Doeller, C. Barry, N. Burgess, Evidence for grid cells in a human memory network, Nature. 463 (2010) 657–661. https://doi.org/10.1038/nature08704.Evidence.

[21] T.E.J. Behrens, T.H. Muller, J.C.R. Whittington, S. Mark, A.B. Baram, K.L. Stachenfeld, Z. Kurth-Nelson, What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior, Neuron. 100 (2018) 490–509. https://doi.org/10.1016/j.neuron.2018.10.002.

[22] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal Policy Optimization Algorithms, (2017) 1–12. http://arxiv.org/abs/1707.06347.

[23] O.S. Binhomaid, Construction Site-Layout Optimization Considering Workers ' Behaviors Around Site Obstacles , Using Agent-Based Simulation, Univ. Waterloo. (2019) 129.

[24] K. Yang, C.R. Ahn, H. Kim, Tracking divergence in workers' trajectory patterns for hazard sensing in construction, Constr. Res. Congr. 2018. (2018) 126–133. https://doi.org/10.1061/9780784481288.013.

[25] K. Teknomo, Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model, ArXiv Prepr. ArXiv1610.00029. (2016). http://arxiv.org/abs/1610.00029.

[26] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, D. Lange, Unity: A General Platform for Intelligent Agents, (2018) 1–28. http://arxiv.org/abs/1809.02627.

[27] Y. Bengio, J. Louradour, R. Collobert, J. Weston, Curriculum learning, Proc. 26th Annu. Int. Conf. Mach. Learn. (2009) 41–48. https://doi.org/10.1017/S1047951100000925.

[28] F. Klügl, A validation methodology for agent-based simulations, Proc. ACM Symp. Appl. Comput. (2008) 39–43. https://doi.org/10.1145/1363686.1363696.

[29] M.H. Dridi, Pedestrian Flow Simulation Validation and Verification Techniques, Curr. Urban Stud. 03 (2015) 119–134. https://doi.org/10.4236/cus.2015.32011.