INDOOR TRAJECTORY RECONSTRUCTION USING BUILDING INFORMATION MODELING AND GRAPH NEURAL NETWORKS

Mingkai Li, Peter Kok-Yiu Wong, Cong Huang & Jack Chin-Pang Cheng
The Hong Kong University of Science and Technology, Hong Kong

ABSTRACT: *Trajectory reconstruction of pedestrian is of paramount importance to understand crowd dynamics and human movement pattern, which will provide insights to improve building design, facility management and route planning. Camera-based tracking methods have been widely explored with the rapid development of deep learning techniques. When moving to indoor environment, many challenges occur, including occlusions, complex environments and limited camera placement and coverage. Therefore, we propose a novel indoor trajectory reconstruction method using building information modeling (BIM) and graph neural network (GNN). A spatial graph representation is proposed for indoor environment to capture the spatial relationships of indoor areas and monitoring points. Closed circuit television (CCTV) system is integrated with BIM model through camera registration. Pedestrian simulation is conducted based on the BIM model to simulate the pedestrian movement in the considered indoor environment. The simulation results are embedded into the spatial graph for training of GNN. The indoor trajectory reconstruction is implemented as GNN conducts edge classification on the spatial graph.*

KEYWORDS: *Indoor trajectory reconstruction; Graph neural network; Building information modeling; Camera-based tracking; Spatial graph; Pedestrian simulation*

1. INTRODUCTION

Indoor trajectory reconstruction refers to the process of estimating the path or trajectory followed by a moving object or person within an indoor environment. This can be useful in various applications, such as indoor navigation, activity recognition, or monitoring systems. There are different approaches to indoor trajectory reconstruction, including sensor-based methods and computer vision (CV) techniques. Sensor-based methods rely on sensors, such as accelerometers, gyroscopes, magnetometers, or depth sensors, to track the movement of an object or person. The sensor data is processed using techniques like sensor fusion or Kalman filtering to estimate the trajectory (Patron-Perez et al. 2015). This approach is commonly used in devices like smartphones or wearable devices. Wi-Fi or Bluetooth signals can also be used to estimate the location of a device within an indoor environment (Traunmueller et al. 2018). By measuring the signal strength from different access points or beacons, it is possible to determine the approximate position of the device. Trajectory reconstruction can be achieved by tracking the device's movements over time using signal strength variations. However, as people pay more attention to privacy, these methods have become more controversial and inconvenient, since it requires pedestrians to actively upload signals. CV techniques can be employed to reconstruct trajectories using visual information captured by cameras or depth sensors (Wong et al. 2022). These methods may involve object detection, tracking, and motion estimation algorithms. For example, by tracking the position of a person in multiple frames of a video, it is possible to reconstruct their trajectory within the indoor environment. In this regard, CV techniques are more acceptable for public as it required less information exposure.

Person re-identification (ReID) is a CV task that involves identifying and tracking individuals across different cameras or video frames (Zheng et al. 2015). The goal of ReID is to match a person's identity across non-overlapping camera views or at different points in time within a video sequence. In scenarios such as closed circuit television (CCTV) systems, where multiple cameras are installed in an area, ReID can help track individuals as they move between camera views. It is particularly useful in crowded or complex environments where traditional tracking methods may fail due to occlusion or changes in appearance. ReID applies deep learning techniques such as convolutional neural networks (CNNs) to extract discriminative features from the person's appearance (Cheng et al. 2016), which are further compared among different individuals to match the same person's features while differentiating them from others. However, ReID performs differently indoors and outdoors, as they differ significantly in terms of lighting conditions, camera placement, and occlusion. Indoor environments often have controlled lighting and less occlusion of pedestrian, which can result in more consistent appearance of individuals, thereby ReID algorithms usually could achieve better performance. However, indoor cameras are often installed at fixed positions with controlled angles and have narrow fields of views and limited camera coverage due to the narrow space of indoor environment and occlusion of building elements. It is not realistic to install cameras to

cover all indoor space as it requires large investment and maintenance cost in CCTV system. Hence, other techniques are required to enable indoor trajectory reconstruction.

Building information modeling (BIM) is a powerful tool for building management and provides a common data environment to connect different platforms to support various applications (Song et al. 2022; Cheng et al. 2022). Integration of BIM and CV techniques has emerged as a hot topic in recent years and unlocks a lot of applications. For example, construction activities can be recognized and analyzed by CV algorithms, and the relevant information can be extracted and intergraded into BIM model to improve the construction progress monitoring (Deng et al. 2020; Braun et al. 2020). Furthermore, information about the building components can also be intergraded in the digital representation of the BIM model for automatic detection and identification (Troncoso-Pastoriza, et al. 2018).

Graph neural network (GNN) is a type of deep learning model that is specifically designed to operate on unstructured data that can be represented as graphs (Zhou et al. 2020). The key idea behind GNNs is to propagate information across the nodes and edges of a graph, allowing each node to gather and update information from its neighbors. GNNs have shown promising results in various domains, including social network analysis, molecular chemistry, recommendation systems, and CV tasks involving graphs or structured data (Wu et al. 2020). In recent years, GNNs have been applied to the building design and management. Nauata et al. (2020) applied GNN to generate house layout following given relational architecture. Cheng et al. (2022) leveraged GNN to conduct crowd prediction in the building. In this regard, GNN is a potential technique to assist the indoor trajectory reconstruction as the building layout can be represented as a graph and can be further processed by GNNs.

This paper proposed an indoor trajectory reconstruction method using BIM and GNN. A spatial graph is proposed to depict the indoor environment. Pedestrian simulation is conducted using the agent-based model established based on BIM model to enrich the spatial graph. The CCTV system is integrated with BIM model by camera registration, so that the information generated by CV algorithms based on the cameras' videos can be related to specific location in the BIM model. With the information from CCTV system, the spatial graph is then processed by a GNN to reconstruct the indoor trajectory of pedestrian. Section 2 introduces the methodology in details, while Section 3 provides an example to illustrate the proposed method.

2. METHODOLOGY

The proposed framework of indoor trajectory reconstruction is shown in Fig. 1. The BIM model is first used to establish a spatial graph to describe the spatial relationship among the indoor spaces including corridors and rooms. To be specific, the floor plan of the Revit model is analyzed by a Dynamo algorithm to identify the entrances, exits, intersections and dead ends. These points would be nodes in the graph and corridors connecting these nodes would be edges. Besides, DWG file is exported from BIM model and imported into a pedestrian simulation software called AnyLogic. The movement of pedestrian inside the building is simulated. The required time for a person to move from one point to others is recorded and embedded into the graph as edge attribute. The CCTV layouts are linked to the BIM model by camera registration so that the identification of some person in the field of view of one camera can provide information of the location of the person in the building. ReID algorithm is adopted to identify a specific person across several cameras. The series of timestamps and positions of one person will be passed to the spatial graph and possessed by a graph neural network to identify the trajectory of this person.

2.1 Integration of BIM and CCTV System

Cameras have been widely used in buildings for safety and efficiency surveillance. Especially with the integration of artificial intelligence and building BIM, many intelligent applications have emerged. To unlock the potential of CCTV-BIM integration, the first step is to localize the cameras in the considered environment, based on which the event or person identified in cameras can be linked to specific position in the BIM model.

2.1.1 Camera Registration with BIM

Camera registration, also known as camera pose estimation, is the process of relating the camera coordinates to the real-world coordinates of objects or scenes. Some previous studies leverage conformity of geometric primitives such as points, lines, and planes to determine the translation, rotation, and scale in reference to as-planned models or real world (Asadi et al. 2019). These methods usually require manual operation and rely on predefined viewpoint assumption including camera position and orientation (Lukins and Trucco 2007; Rebolj et al. 2008). Asadi et al. (2019) automated the registration process by performing an augmented monocular simultaneous localization and

mapping and perspective detecting and matching between the image frames and their corresponding BIM views. Though automated methods are more efficient, manual approaches with low technical threshold are still common as the camera registration process is one-off for a scene as long as the cameras are fixed.

Fig. 2 shows an example of manual camera registration with BIM model. The position and rough orientation of the camera are provided, so that the objects such as doors and walls in the field of view (FOV) can be easily mapped to the building elements in the BIM model. Several characteristic points at the intersection line of wall and floor, such as corner points of walls, will be selected in the FOV, while their correspondences will be identified in the BIM model. The function for transforming the pixel coordinates in camera's FOV to global coordinates in the BIM model can further be established. For example, for the FOV in Fig.2, two characteristic points $C_1(C_1')$ and $C_2(C_2')$ are selected, for any point $C_p(x_p, y_p)$ in the FOV, its corresponding point C_p' can be determined by following equations.

$$x_p' = \frac{x_p - x_1}{x_2 - x_1}(x_2' - x_1') + x_1' \tag{1}$$

$$y_p' = \frac{y_p - y_1}{y_2 - y_1}(y_2' - y_1') + y_1' \tag{2}$$

The above equation can only be used for cameras with no distortion, otherwise some corrections are needed. When more than two characteristic points are identified, the coordinates can take the average of the calculation results of every pair of the points using the above equation. For a FOV where characteristic points could not be found, several markers with known global coordinates can be set on the floor, which can be easily identify in the camera's FOV to establish the transformation function.
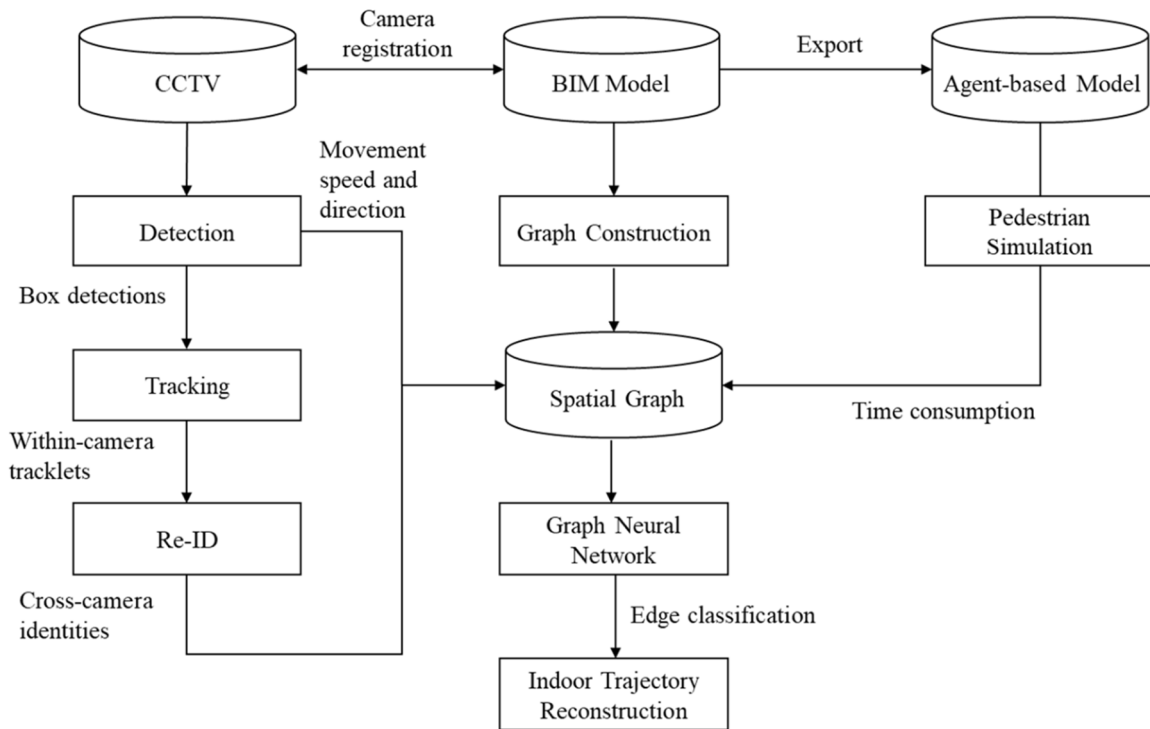

Fig. 1: Proposed framework of indoor trajectory reconstruction

Camera registration allows for accurate transformation between the real-world coordinates and the 2D coordinates in camera. It is crucial in multi-camera systems, where multiple cameras are used to capture a scene from different viewpoints. By accurately registering the cameras, it becomes possible to merge or fuse the information from different cameras and create a consistent and comprehensive representation of the scene. Overall, camera registration is a fundamental step in computer vision applications that involve cameras, enabling precise mapping between the real world and the image plane, and facilitating accurate measurements and analysis of the captured visual data.
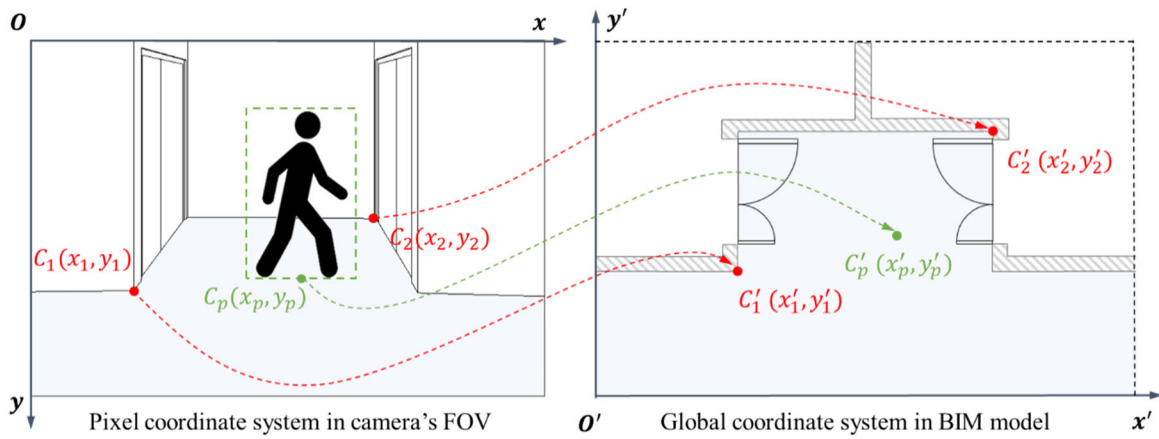
Fig. 2: Example of camera registration with BIM

2.1.2 Real-to-BIM

With camera registration, the information captured by CCTV system can be reflected into BIM. By adopting CV techniques, pedestrian can be detected with a bounding box. It is assumed that the midpoint of the bottom line of the bounding box can roughly represent the location of a person if the detection module in Section 2.3.1 reaches certain accuracy. Based on this, the movement direction of a person can be identified, which can be further matched with the directions of different branches of corridors to estimate the trajectory of the person. Besides, the speed of the person can be further calculated, as the movement distance of the person in real world can be achieved through transformation and the time consumed is also known.

As some cameras will be installed outside some rooms, the videos from cameras can be used to estimate whether a person has entered a room. Firstly, necessary information is extracted from BIM model using Dynamo, including door's dimension, door's location, and room's name. The location of a door is simplified as a segment AB on the 2D plane, while a person is abstracted as a point P. There are three cases based on the relative position of the line segment and the point, shown in Fig. 3. Based on this, we develop the algorithm to detect, when a person disappears from the video, the room he/she enters, or whether the person leaves, as shown in Fig. 4. For each point, we repeat such a process m times to collect the distances between this point and m line segments (doors), then regard the line segment that has the smallest distance as the final result. In other words, the closest door a person nears when he/she disappears is identified as the one the person enters.
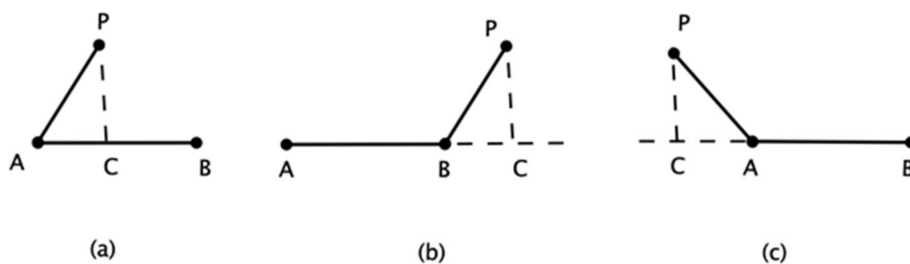

Fig. 3: Three cases for the shortest distance between a point and a line segment

2.2 Spatial Graph

Spatial graph is proposed to represent the indoor environment based on the BIM model. Besides, pedestrian simulation is conducted using agent-based model, which is derived from BIM model. The simulation results are then embedded in the spatial graph. With the information from CCTV system, the indoor trajectory can be reconstructed based on the spatial graph using GNN.

2.2.1 Graph Construction

Spatial graph is proposed by improving the medial axis transform (MAT) (Lee 2004) for indoor trajectory

reconstruction. As shown in Fig. 5, MAT adds nodes at every turning point in the building, while spatial graph skips those nodes that are not fork, because the trajectory of pedestrian will not have multiple possibilities when passing through this kind of nodes. What these two methods have in common is that the edges in the both graphs represent sections of the corridor. Spatial graph includes several kinds of nodes: "entrance/exit" nodes (in green), "dead end" nodes (in dark green), "room" node (in orange), "fork" nodes (in red) and "camera" nodes (in blue). Dividing the nodes into different categories can depict the indoor environment more accurately and provide more information for GNN.

Algorithm 1: DETECT the room in which each person enters

Input: A person's coordinate P
A set of 2D vectors represent m doors $D = \{\overrightarrow{A_1 B_1}, \overrightarrow{A_2 B_2}, \ldots, \overrightarrow{A_m B_m}\}$

1   $rid \leftarrow 0$
2   $dist \leftarrow \infty$
3   **for** $i \leftarrow 1$ **to** m **do**
4     $C_i \leftarrow$ the projection point of P onto $\overrightarrow{A_i B_i}$
5     $r \leftarrow (\overrightarrow{A_i P} \cdot \overrightarrow{A_i B_i}) / |\overrightarrow{A_i B_i}|^2$
6   **if** $r \leq 0$ **then**
7       $d \leftarrow |A_i P|$
8   **if** $r \geq 1$ **then**
9       $d \leftarrow |B_i P|$
10   **if** $0 < r < 1$ **then**
11       $d \leftarrow |C_i P|$
12   **if** $d \leq dist$ **then**
13       $rid \leftarrow j$
14       $dist \leftarrow d$
15   $M_{rid} \leftarrow$ the midpoint of $A_{rid} B_{rid}$
16   **if** $|PM_{rid}| \geq k * |A_{rid} B_{rid}|$ **then**
17     $r \leftarrow 0$
18   **else**
19     $r \leftarrow rid$
20   **return** r

Output: The room in which the person enters $r \in [0, m]$

Fig. 4: Pseudo code to detect the room in which each person enters
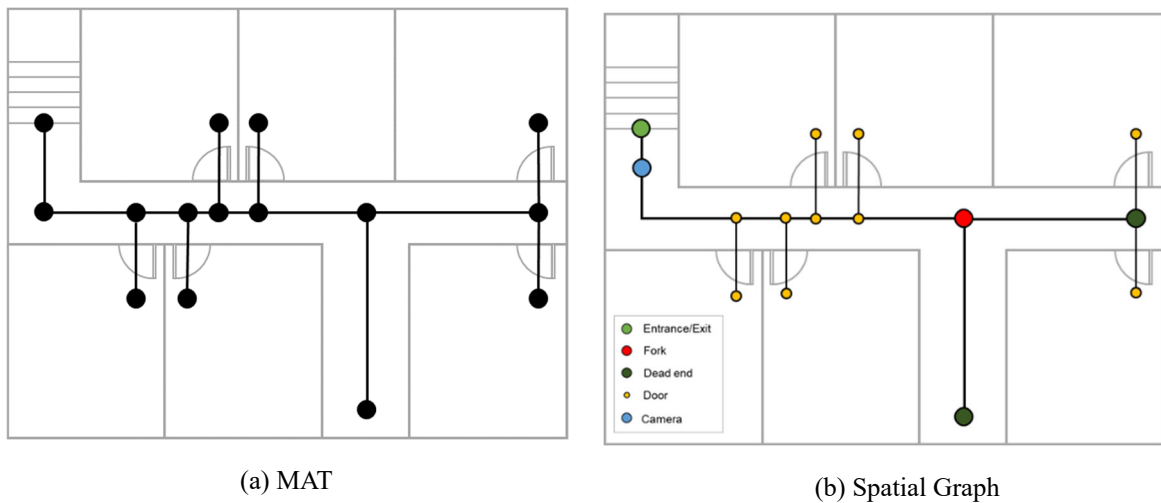

(a) MAT

(b) Spatial Graph

Fig. 5: Spatial graph compared with medial MAT (Lee, 2004)

2.2.2 Pedestrian Simulation

To achieve more information for indoor trajectory reconstruction, pedestrian simulation is conducted to analyze the behavior and the time required for a person to reach a specific position in the considered environment. We adopt agent-based modeling (ABM) for pedestrian simulation. ABM has been widely applied to simulate the real-world operation, particularly pedestrian movement, traffic network and manufacturing chain. The three main elements in ABM are agents, their attributes as well as behavioral rules (Cheng and Gan, 2013). Each agent is an autonomous component having its attributes defined by user-input parameters, such as its size and moving speed. Every agent also behaves according to a set of decision rules, for example, approaching multiple places in a specified sequence. With these characteristics, every agent in an environment persistently interacts with each other, in pursuit of specific objectives.

There have been extensive case studies that were able to simulate human behaviors in various scenario. Said et al. (2012) modeled how occupants in a high-rise building react in case of emergency and find the quickest evacuation route. On the other hand, Liu et al. (2014) demonstrated the functionality of ABM in simulating typical pedestrian flow phenomenon, such as bidirectional flow in corridors and through bottlenecks. Their simulation result consisted of the flow rate, movement velocity and spatial density. Suggested by Seyfried et al. (2005), pedestrian flow is notably consistent with the fundamental traffic theory. This inspired us to analyze the speed-density-flow relationship for pedestrian movement, by adopting traffic flow theories.

One typical ABM engine is AnyLogic, which possesses excellent user-friendliness by allowing graphical drag-drop control and advanced Java codes. Moreover, it supports realistic visualization by either component markup or importing geometric models from external application. In this paper, the BIM model is imported to AnyLogic using DWG file as intermedia. Then the walls of the building can be generated automatically. By setting the source and target of pedestrian flow, the movement of pedestrian in the considered indoor environment can be simulated. Besides, we also added "line service" at the positions of fire doors (which are normally closed) to consider the time delay of passing the doors. The pedestrian queuing behavior and parameter setting are investigated by Kim. et al. (2013), which laid the foundation for establishing a robust framework for our project.

2.2.3 GNN-based Trajectory Reconstruction

With the pedestrian simulation based on ABM, the spatial graph can be further enriched with the time consumption information from one position to the other within the indoor environment. Each edge in the graph has two features: "time consumption" and "pass_or_not". The former is a feature to indicate the time required for a person to reach a specific position from the other. The time is taken as the average value detected from the simulation. The latter is a binary-class feature showing whether a person has passed a specific path, which is represented by an edge. The movement direction mentioned in Section 2.1.2 is used to estimate the path that the person is most likely to travel through, which is achieved by matching the detected direction with the directions of different branched of corridors. Besides, each node will have 5 features: x-coordinate, y-coordinate, category, timestamp when a person is detected, and speed of detected person. "Category" refers to the categories of node according to Section 2.2.1. Speed of detected pedestrian relies on the camera registration and CV techniques to estimate the speed. For those nodes that no pedestrian pass, the value of this feature is set as 0. Base on this graph representation, the indoor trajectory reconstruction can be formulated as an edge classification task on graph, as shown in Fig. 6, aiming to divide all edges into those that pedestrians pass by and those that do not.
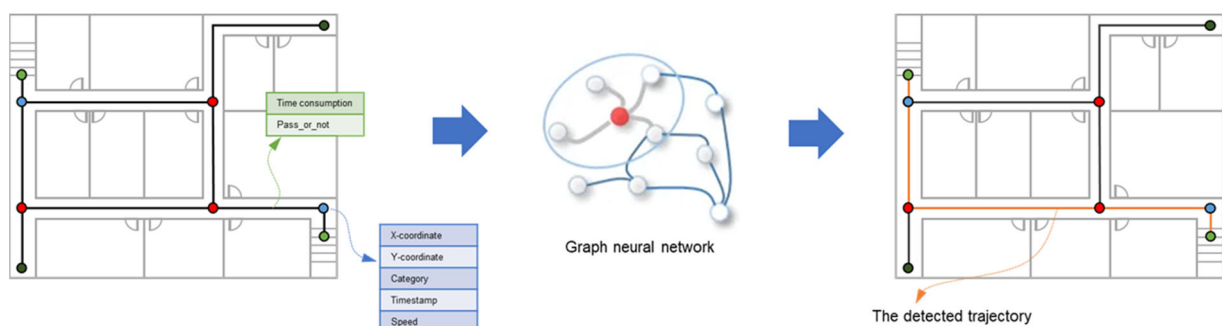

Fig. 6: Edge classification for indoor trajectory reconstruction

2.3 Multi-target Multi-Camera Tracking and Re-Identification

Given a query video, our method targets on detecting, tracking and identifying all people within this video, which is composed of three parts, a detecting module, a tracking module, and a ReID module. The whole framework is shown in Fig. 7.
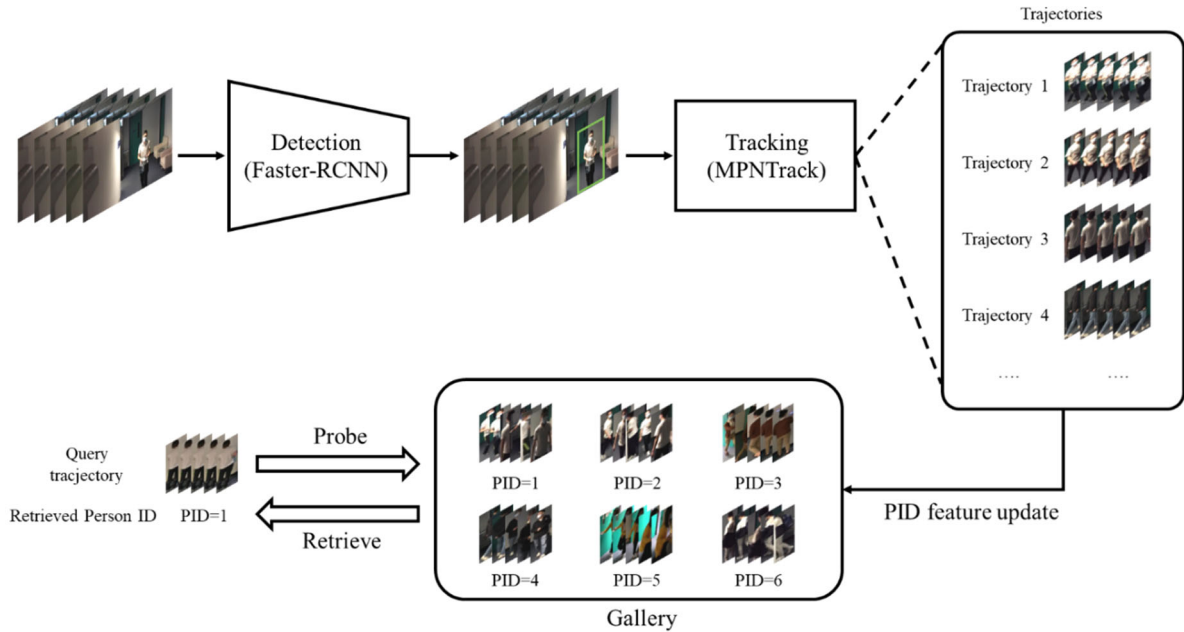

Fig. 7: The framework for multi-target multi-camera tracking and re-identification

2.3.1 Detection

For each frame in the query video, we first apply a Faster-RCNN (Ren et al. 2015) to detect all persons within this frame. In detail, given a frame, a CNN is applied to extract pixel-wise features, which are then sent to a region proposal network (RPN) to generate region-of-interest (RoI) which may contain a person. These regions are fed into another CNN classifier to determine whether they correspond to a person or not. Finally, a non-max-suppression (NMS) method is applied for redundancy removal. NMS first outputs the highest scoring box and then suppresses all overlapping boxes with that box, repeating this process until all boxes are processed.

2.3.2 Tracking

The tracking module works for aligning objects in later frames with those in previous frames. We use a GNN called MPNTrack (Braso and Leal-Taixe, 2020) to achieve this goal. To be specific, for positive predictions in each frame, we first apply a RoIAlign operation to extract their features. We then construct a graph. Features of positive predictions in each frame are treated as nodes, and all prediction pairs across frames form the edges within this graph. For each edge in this graph, we encode its feature as the deviation between the features of its two end nodes. These edge features are then passed through a multi-layer perceptron (MLP) for classification. If two end nodes of an edge correspond to the same person within two frames, we label it as one, otherwise zero. In this way, we associate predictions across frames, which predictions correspond to the same instance and which are not. Therefore, we derive appeared instances in the query video.

2.3.3 Re-ID

Finally, we deploy a Re-ID module to identify these instances. We first forge an instance gallery. Given the training videos, the detection and tracking module was first used to obtain different instances. We then randomly select n (n=10 in our experiments) instances for each person and store them in the instance gallery. We extract feature vector for each instance from the well-trained Re-ID model and use these feature vectors as the high-level semantic representations for persons. During testing, for different instances obtained from the query video, we treat them as probe, extract their feature vectors and retrieve their identifications stored in the instance gallery. Specifically, we compute the cosine distance between the queried feature vectors and all the stored feature vectors, and treat the person with the least feature distance as the output identification. The detailed structure is shown in Fig. 8.
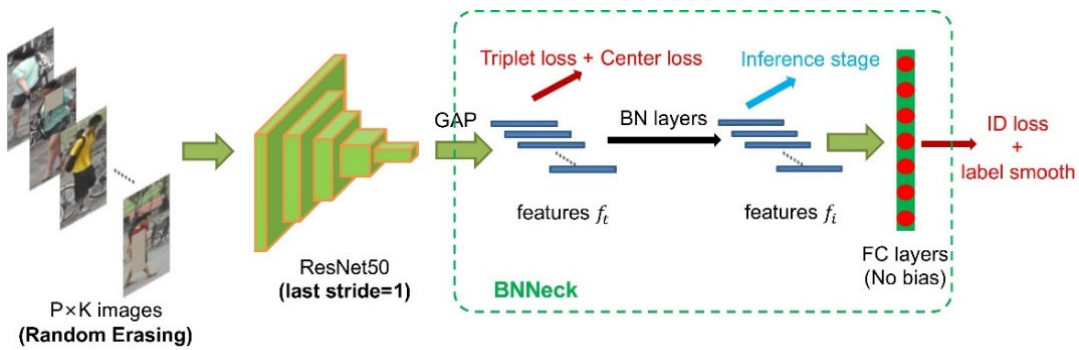
Fig. 8: The detailed structure of the Re-ID module (Luo et al. 2019)

3. ILLUSTRATIVE EXAMPLE

We selected a part of the HKUST campus with more than 200 rooms to demonstrate the method proposed in this paper. Firstly, we developed a Dynamo program to derive the spatial graph based on the floor plan from the BIM model. Fig. 9 shows the process of graph construction. Then DWG file is exported to AnyLogic to support agent-based pedestrian simulation (as shown in Fig. 10) to enrich the spatial graph. Currently, the simulation model is established automatically based on the DWG file, though some manual adjustment is needed. The pedestrian flow logic is established manually, which could be automated with some further developed. We extracted the average time for pedestrian to reach a position from the other as the "time consumption" feature for the corresponding edge. Besides, by selecting different combinations of starting and end points, the movement of pedestrian in the building following designated paths can be simulated, the results can be further used to train the GNN.

(a) BIM model

(b) Floor plan

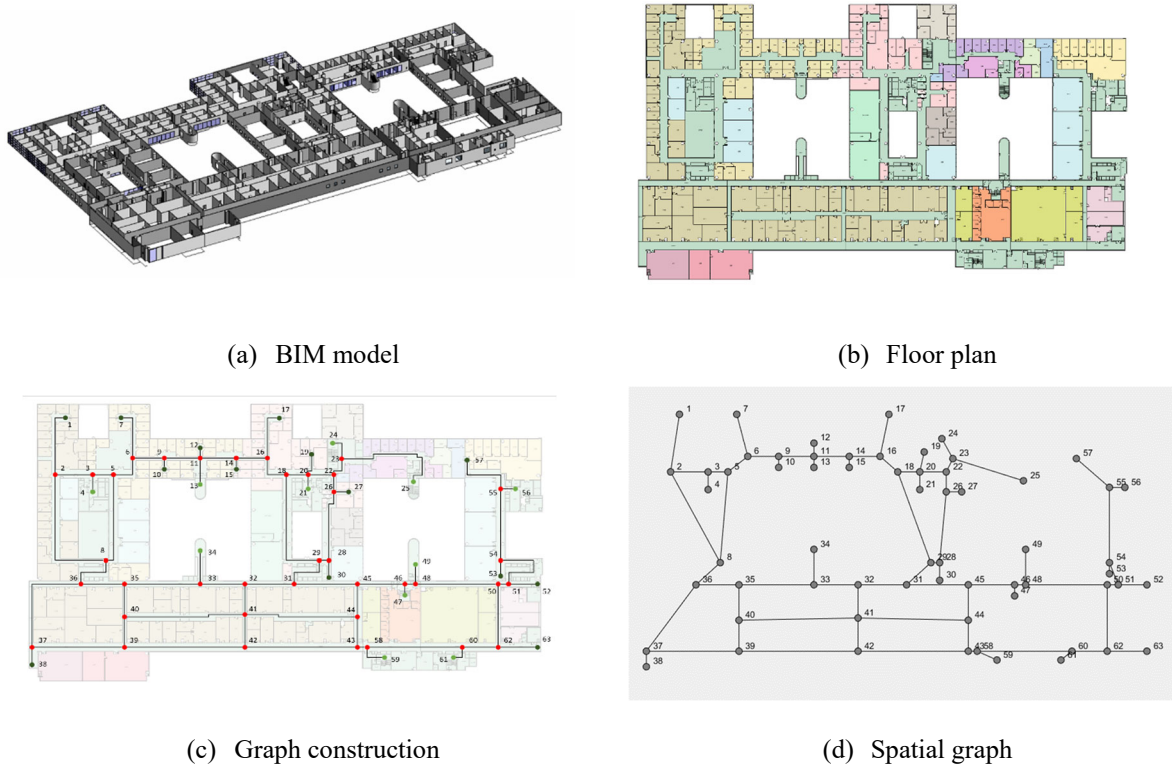(c) Graph construction

(d) Spatial graph

Fig. 9: Graph construction based on BIM model

The CCTV system is integrated with the BIM model through camera registration, so that the information captured by cameras can be linked to the specific location in the BIM model as well as the spatial graph. In our experiment, we conduct camera registration manually for 6 cameras using characteristic points and Equation (1) and (2). With the ReID techniques, the same person appears in different cameras' POV can be identified. The ReID model is pre-trained on a public benchmark -- DukeMTMC-reID (Ristani et al., 2016), and got an accuracy of 100% for the

100 pedestrians we observed in the building. The high accuracy may result from the stable lighting condition and fewer pedestrian appearing in the FOV of each camera, compared to the outdoor environment.

Fig. 11 shows the feature maps and ReID process, while Fig. 12 provides an example of ReID results.
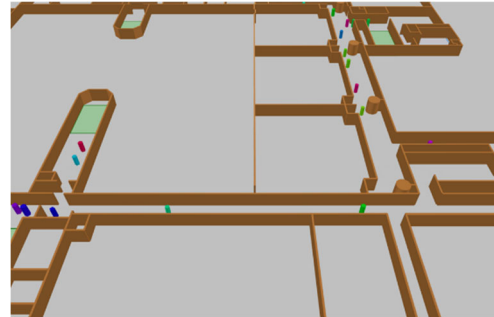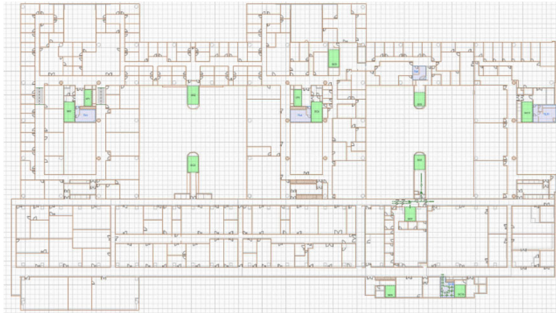
(a) AnyLogic model established based on BIM model

(b) 3D view of the pedestrian simulation
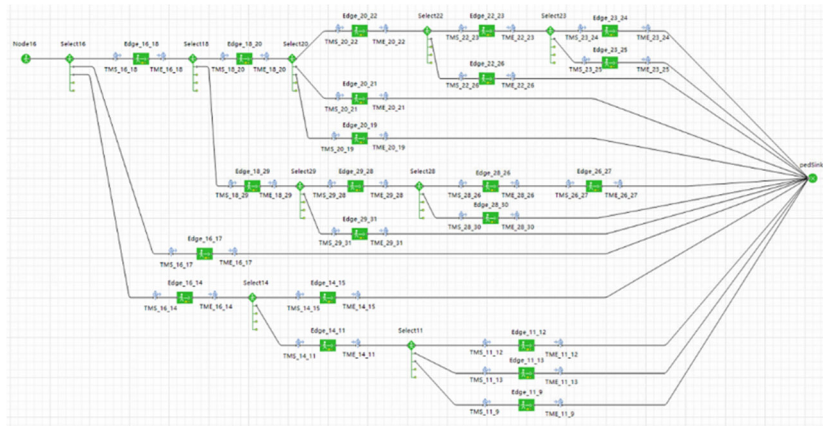
(c) Pedestrian flow logic
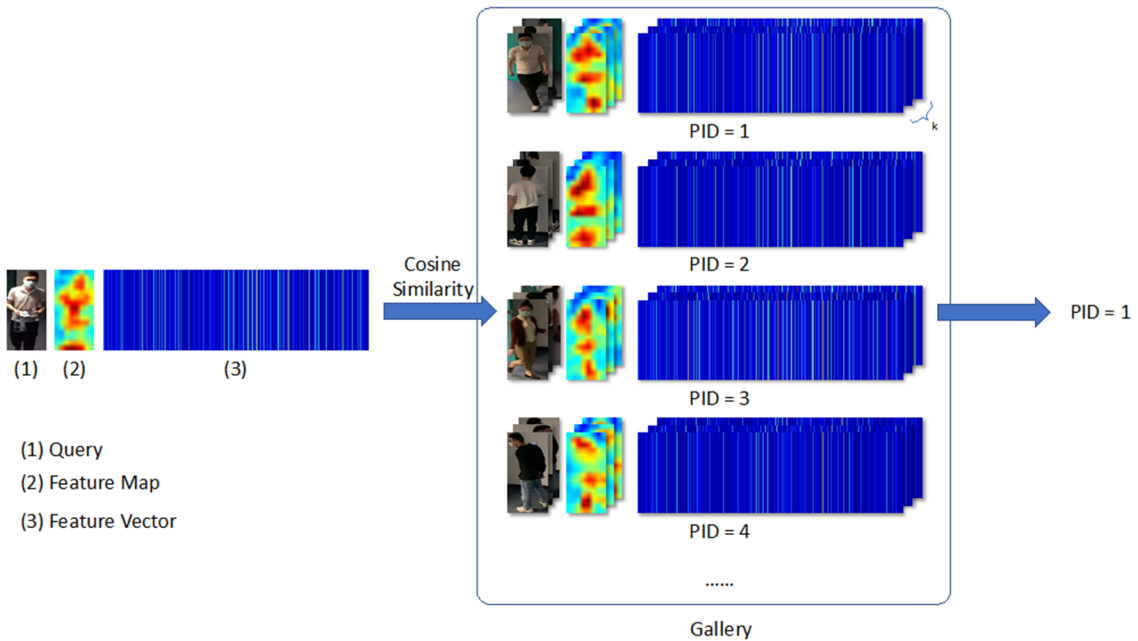
Fig. 10: Pedestrian simulation based on AnyLogic

Fig. 11: Feature maps and Re-ID process
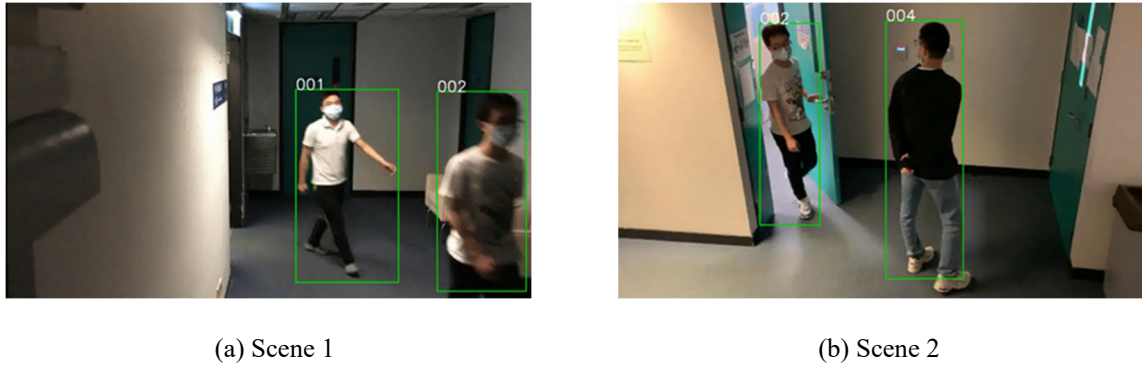
(a) Scene 1   (b) Scene 2

Fig. 12: Examples of the Re-ID results for the two cameras' videos

Information including ReID results as well as the direction and speed of the detected person were embedded into the graph and processed by GNN to make edge classification. The indoor trajectory can be reconstructed by classifying edges into 2 categories: those the person passed, and those did not. The GNN classification achieved 81.2% for the trajectories for the observed 100 pedestrians. We found that for some cases where pedestrians stopped during the process, GNN would produce some wrong classifications, since staying will affect the total time for a person to reach a specific location.

4. CONCLUSION AND DISCUSSION

This paper proposed an indoor trajectory reconstruction method integrating BIM and GNN. A spatial graph is proposed based on BIM to depict the connection of indoor spaces and integrate information captured by CCTV system. The CCTV system is related to the BIM model through camera registration. ABM-based pedestrian simulation is leveraged to simulate the movement of persons within the building, which provides more information to the spatial graph. Trajectory reconstruction is implemented using GNN, which works on spatial graph to aggregate information and classify edges. This study provides an automated approach to trace the pedestrian in the building, which could provide building managers with more insight in the indoor movement pattern and crowd distribution, and thereby could support a lot of smart applications such as indoor navigation, ambient-assisted facility management, precise product delivery, etc.

The proposed approach still has several limitations. For environments with a large number of rooms, the CCTV system usually cannot cover all the entrances of rooms due to limited number of cameras. In this scenario, for those rooms whose doors are not in the FOVs of cameras, we could not achieve the time of staying in the room for a detected person only using cameras, hence it may affect the GNN's performance on edge classification. Other techniques such as Internet of things could be explored to provide supplementary information for those positions that are not covered by cameras. Besides, optimization of camera layout can also be investigated to enlarge the coverage of camera and reduce blind areas. In addition, an automated camera registration method can also be included to improve the convenience of applying the proposed method.

REFERENCES

Asadi, K., Ramshankar, H., Noghabaei, M., & Han, K. (2019). Real-time image localization and registration with BIM using perspective alignment for indoor monitoring of construction. Journal of Computing in civil Engineering, 33(5), 04019031.

Brasó, G., & Leal-Taixé, L. (2020). Learning a neural solver for multiple object tracking. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 6247-6257).

Braun, A., Tuttas, S., Borrmann, A., & Stilla, U. (2020). Improving progress monitoring by fusing point clouds, semantic data and computer vision. Automation in Construction, 116, 103210.

Cheng, D., Gong, Y., Zhou, S., Wang, J., & Zheng, N. (2016). Person re-identification by multi-channel parts-based CNN with improved triplet loss function. *In Proceedings of the iEEE conference on computer vision and pattern recognition* (pp. 1335-1344).

Cheng, J. C., & Gan, J. (2013). Integrating agent-based human behavior simulation with building information modeling for building design. International Journal of Engineering and Technology, 5(4), 473.

Cheng, J. C., Kwok, H. H., Li, A. T., Tong, J. C., & Lau, A. K. (2022). BIM-supported sensor placement optimization based on genetic algorithm for multi-zone thermal comfort and IAQ monitoring. Building and Environment, 216, 108997.

Cheng, J. C., Poon, K. H., & Wong, P. K. Y. (2022). Long-Time gap crowd prediction with a Two-Stage optimized spatiotemporal Hybrid-GCGRU. Advanced Engineering Informatics, 54, 101727.

Deng, H., Hong, H., Luo, D., Deng, Y., & Su, C. (2020). Automatic indoor construction process monitoring for tiles based on BIM and computer vision. Journal of construction engineering and management, 146(1), 04019095.

Kim, I., Galiza, R., & Ferreira, L. (2013). Modeling pedestrian queuing using micro-simulation. Transportation Research Part A: Policy and Practice, 49, 232-240.

Lee, J. (2004). A spatial access-oriented implementation of a 3-D GIS topological data model for urban entities. GeoInformatica, 8, 237-264.

Liu, S., Lo, S., Ma, J., & Wang, W. (2014). An agent-based microscopic pedestrian flow simulation model for pedestrian traffic problems. IEEE Transactions on Intelligent Transportation Systems, 15(3), 992-1001.

Lukins, T. C., & Trucco, E. (2007, September). Towards Automated Visual Assessment of Progress in Construction Projects. In BMVC (pp. 1-10).

Luo, H., Gu, Y., Liao, X., Lai, S., & Jiang, W. (2019). Bag of tricks and a strong baseline for deep person re-identification. In the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 0-0).

Nauata, N., Chang, K. H., Cheng, C. Y., Mori, G., & Furukawa, Y. (2020, Fall). House-gan: Relational generative adversarial networks for graph-constrained house layout generation. In Computer Vision–ECCV 2020: 16th European Conference (Part I 16, pp. 162-177).

Patron-Perez, A., Lovegrove, S., & Sibley, G. (2015). A spline-based trajectory representation for sensor fusion and rolling shutter cameras. International Journal of Computer Vision, 113(3), 208-219.

Rebolj, D., Babič, N. Č., Magdič, A., Podbreznik, P., & Pšunder, M. (2008). Automated construction activity monitoring system. Advanced engineering informatics, 22(4), 493-503.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.

Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016). Performance measures and a data set for multi-target, multicamera tracking. European Conference on Computer Vision Workshops (EECVW), Amsterdam, The Netherlands (pp. 43–51).

Said, H., Kandil, A., & Cai, H. (2012). Agent-based simulation of labour emergency evacuation in high-rise building construction sites. In Construction Research Congress 2012: Construction Challenges in a Flat World (pp. 1104-1113).

Seyfried, A., Steffen, B., Klingsch, W., & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. Journal of Statistical Mechanics: Theory and Experiment, 2005(10), P10002.

Song, C., Chen, Z., Wang, K., Luo, H., & Cheng, J. C. (2022). BIM-supported scan and flight planning for fully autonomous LiDAR-carrying UAVs. Automation in Construction, 142, 104533.

Traunmueller, M. W., Johnson, N., Malik, A., & Kontokosta, C. E. (2018). Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities. Computers, Environment and Urban Systems, 72, 4-12.

Troncoso-Pastoriza, F., López-Gómez, J., & Febrero-Garrido, L. (2018). Generalized vision-based detection, identification and pose estimation of lamps for BIM integration. Sensors, 18(7), 2364.

Wong, P. K. Y., Luo, H., Wang, M., & Cheng, J. C. (2022). Enriched and discriminative convolutional neural

network features for pedestrian re-identification and trajectory modeling. Computer-Aided Civil and Infrastructure Engineering, 37(5), 573-592.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 4-24.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. *In Proceedings of the IEEE international conference on computer vision* (pp. 1116-1124).

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI open, 1, 57-81.