Building Component Segmentation-oriented Indoor Localization using BIM-based Synthetic Data Generation and Deep Learning

Haomiao Zhao[1], Yifeng Sun[2], Chi Chiu Lam[2], Junxiang Zhu[3], Mi Pan[2], Mun On Wong[2, *]

[1] School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
[2] Department of Civil and Environmental Engineering, University of Macau, Macao SAR, China
[3]Department of Engineering, University of Cambridge, Cambridge, United Kingdom
[*]Corresponding author.
z5521569@ad.unsw.edu.au, mc35239@um.edu.mo, fstccl@um.edu.mo, jz652@cam.ac.uk, mipan@um.edu.mo, mowong@um.edu.mo

Abstract

This paper presents a novel image-based indoor localization approach that leverages semantic segmentation techniques by combining synthetic images generated from building information modelling (BIM) and deep learning (DL) methods. Unlike traditional image-based methods that compare overall visual features, our approach focuses on stable building components—such as floors, ceilings, walls, doors, and windows—that remain largely consistent over time. To enhance localization accuracy, we introduce a new similarity metric based on the intersection-over-union (IoU) ratio for each building component. This metric accounts for the spatial distribution of building elements, allowing for more robust comparisons between semantic masks derived from real-captured images and those generated from synthetic BIM models. The method reduces the dependence on detailed BIM models, enhancing reliability by focusing on the geometric and semantic consistency of key building elements. The feasibility of the approach is demonstrated through a case study with an institutional building, showcasing promising localization performance despite low-fidelity BIM models.

Keywords –

Indoor Localization; Semantic Segmentation; Building Information Modeling; Deep Learning; Image Retrieval; Synthetic Data Generation

1 Introduction and Related Works

Accurate indoor localization plays a crucial role in enabling a wide range of applications such as facility management, disaster response, and navigational assistance in complex indoor environments [1]. In recent years, various indoor localization methods have been developed, such as dead reckoning approaches with inertial sensors and piloting approaches using range-based signals [2-3]. However, these methods have their own constraints. For instance, range-based signals such as Wi-Fi, Bluetooth, and ultra-wideband signals are highly susceptible to environmental interference (e.g., walls) and require additional deployment of hardware infrastructure within buildings. Inertial measurement unit (IMU)-based methods suffer from cumulative drift errors, leading to decreased positioning accuracy over time [4]. In contrast, image-based localization methods offer a promising alternative, as they have a lower reliance on specialized hardware and can provide discrete, location-referenced measurements [5].

Traditional image-based indoor localization methods typically incorporate an offline database collection and an online image retrieval process [6]. During offline collection, various views of indoor environments need to be captured to establish an image database along with the associated locations and orientations. Subsequently, the online image retrieval is performed to leverage a real-captured query image to search the database for the most visually similar match, from which the position can be inferred. Nevertheless, the database collection process demands extensive preparatory work, and the effectiveness of the database may be affected over time when the indoor environment evolves, and the visual characteristics of certain areas change significantly [7].

To reduce the manual effort in data collection, recent research has explored building information modelling (BIM) to generate synthetic images in an automated and controllable manner. For instance, Ha et al. [8] generated the BIM-rendered images and then utilized a convolutional neural network (i.e., VGG 16 [9]) to measure the similarity between the real-captured images and the BIM-rendered images for indoor localization.

Similarly, in the study by Acharya et al. [10], the authors used 3D indoor models to generate rendered synthetic images and perform similarity comparison on a modified GoogLeNet-based deep convolutional neural network architecture (i.e., PoseNet [11]). These studies showcase the benefits of BIMs for image database collection, such as rapid image generation and customizable camera configurations like positions, orientation, and focal lengths. Despite these advantages, these BIM-derived methods have their limitations. To ensure the effectiveness of the similarity calculation between the real-captured and BIM-rendered images, these methods rely heavily on the high fidelity of BIM models, such as the inclusion of furniture, fixtures, and decorations. Such a prerequisite, in turn, requires certain human efforts to maintain the similarity of visual appearance continuously and meticulously. Any discrepancy, caused by insufficient detail in modelling or significant changes in actual environments, can severely degrade the positioning accuracy.

To address these challenges, this study proposes a semantic segmentation-oriented indoor localization approach by leveraging synthetic BIM image generation and deep learning (DL) techniques. Instead of comparing the holistic visual similarity of environments, the proposed approach concentrates on fundamental building components, i.e., floors, ceilings, walls, doors, and windows, since these components usually remain stable regardless of the interior changes and thus provide robust referencing patterns for similarity computation. In this way, the approach reduces the dependency on the highly detailed BIM models. The objectives of this study are: (1) to create an automated approach to generate the semantic masks of BIM-rendered images and real-captured images for similarity computation; (2) to develop a semantic segmentation-oriented similarity metric to implement image retrieval; and (3) to investigate the feasibility and quantify the performance of the proposed approach using a real-world case study. The detailed method development and validation are discussed in the following sections.

2 Methodology

The proposed approach consists of three key stages, including BIM-based synthetic image generation, DL-based image segmentation for real captures, and similarity-based image retrieval. Fig. 1 illustrates the architecture of the proposed approach.

2.1 BIM-based Synthetic Mask

In this study, fundamental building components (i.e., floors, ceilings, walls, doors, and windows) are prioritized since they generally maintain stable geometric and spatial relationships over time. In contrast, furniture and decorative elements may sometimes change and thus offer less reliable features for image matching. Specifically, the relative positions and proportions of these building elements can provide persistent features to enhance similarity measurements. Therefore, this study intends to generate semantic masks to exploit both the semantic and geometric information contained in each captured view.

To construct a database of semantic masks, BIM models are selected as they offer accurate geometries of the target environment and allow flexible configurations for camera deployment. A custom plugin is developed via the Revit API [12] to automate the placement of virtual cameras within the BIM environment for capturing various interior views. Specifically, the cameras are positioned at predefined intervals along two

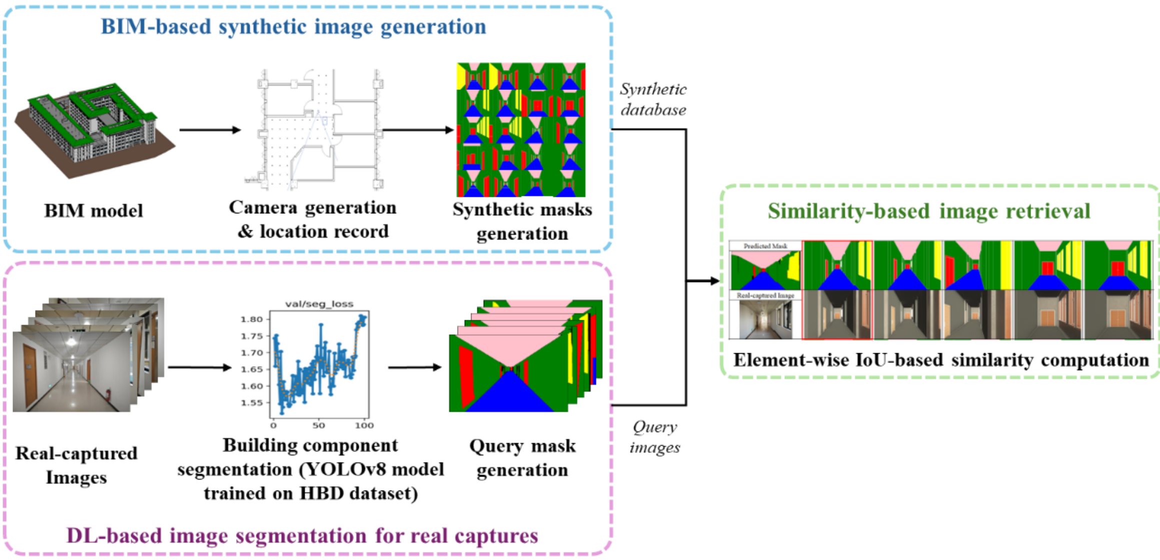

Figure 1. The architecture of the proposed approach

axes, and the boundaries of interior spaces are used to exclude any out-of-bound camera placements. From each camera's position, four axis-aligned orientations are recorded to ensure comprehensive interior coverage. Moreover, camera parameters such as height, image resolution, and dots-per-inch (DPI) can be further customized to produce images with suitable sizes and qualities.

Once the virtual cameras are configured, each building component type within the BIM model is assigned a distinct rendered color to generate component-specific semantic masks. This process involves filtering each target component into a collection and then overriding its graphical properties to establish consistent color coding. Through these steps, synthetic semantic masks from various camera views are generated, along with the corresponding positions and orientations, together constituting the reference database for image retrieval.

During the large-scale semantic mask generation, certain images may lack sufficient spatial features (such as the cases where cameras are placed too close to walls) and provide limited values for image matching. To address this, a filtering process is introduced to enhance the quality of the synthetic BIM-based mask database. The filtering criterion is to retain images that simultaneously include both the ceiling and the floor classes to ensure sufficient semantic and geometric patterns are covered in the masks across various capturing views. In this way, the less distinguishing semantic masks are removed by filtering to enhance the referential quality of the database.

2.2 DL-based Image Segmentation for Real Captures

After preparing the offline database, it is also necessary to process real-captured images and create corresponding masks that have consistent classes with the synthetic BIM masks. To do so, this study employs supervised learning techniques to classify and delineate the building components within images. Specifically, an off-the-shelf image segmentation model, YOLOv8 [13], is adopted as it enables rapid inference speed while maintaining good recognition performance. To align with the classes in the synthetic BIM masks, this study selects the HKU building image dataset (HBD) as the training material since it comprises large-scale annotated images of both building interiors and exteriors, covering all the interested building components [14]. Given the scope of this study, a subset of HBD containing only the interior scenes is utilized, yielding 1,780, 224, and 231 images for training, validation, and testing.

The training process is performed using a computer with configurations of Intel i9-14900HX CPU, NVIDIA GeForce RTX 4060 laptop GPU, and 8 GB GPU memory. After 300 epochs, the training and validation losses plateau, and the training process is terminated. As a result, the best-performing model with the highest validation is subsequently deployed for real-world predictions. The segmentation performance on the test set is summarized in Table 1, which indicates the model obtains moderately high segmentation accuracy. The segmentation of ceilings and floors performs relatively well, probably due to their fixed positions and shapes within images. In contrast, the segmentation result of windows is less effective because of the limited training materials (731 instances in total in the interior scenes of the HBD dataset [14]) and their various appearances.

Table 1. Building component segmentation results on the test set

Class	Images	Instances	mAP^{50}	$mAP^{50\text{-}95}$
Wall	228	554	0.829	0.670
Door	183	315	0.938	0.868
Window	34	94	0.741	0.540
Ceiling	190	213	0.915	0.776
Floor	215	237	0.945	0.866

2.3 Similarity-based Image Retrieval

Last, similarity-based image retrieval is implemented to match the real query images and the BIM-based image database to calculate the most similar pair and acquire the camera positions for indoor localization. Specifically, both the BIM-derived synthetic masks and the real-captured semantic mask are utilized for comparison. These semantic masks effectively capture the composition of the architectural elements, where their spatial arrangement and proportions help characterize the interior layouts. In general, two images deemed similar would exhibit analogous distribution and counts of building elements, occupying comparable proportions and locations within each image. Based on this pattern, this study designs a similarity metric using the element-wise intersection-over-union (IoU) ratio to reflect the overlaps between the real and synthetic masks. The similarity metric focuses on the five fundamental building components to compute the class-balanced IoU values. For each class c_i in the class set C, the pre-class binary masks are generated for the real and synthetic masks, as shown in Equations (1) – (2). Next, a pre-class IoU (IoU_{c_i}) is calculated to quantify the overlap between these corresponding binary masks. Given that the spatial distribution of door and window elements usually provides more distinguishable cues to describe the location-aware context than the other classes, this study assigns different weights w_{c_i} to each class. As a result, the overall similarity metric s is derived by aggregating these weighted IoU values into a single score, as shown in Equations (3) – (4).

$$R_{c_i} = \{(x,y)|R(x,y) = c_i\}, \quad c_i \in C \tag{1}$$

$$S_{c_i} = \{(x,y)|S(x,y) = c_i\}, \quad c_i \in C \tag{2}$$

where $C = \{c_1, c_2, \ldots, c_n\}$ is the set of semantic classes, R and S is the real mask and the synthetic mask, and R_{c_i} and S_{c_i} are the pre-class binary real and synthetic masks that include the pixel locations labelled as a class c_i.

$$IoU_{c_i} = |R_{c_i} \cap S_{c_i}| / |R_{c_i} \cup S_{c_i}| \tag{3}$$

$$s = \sum_{c_i \in C} w_{c_i} IoU_{c_i} \Big/ \sum_{c_i \in C} w_{c_i} \tag{4}$$

where the default weights of the door, window, wall, floor, and ceiling classes are empirically set to 0.325, 0.325, 0.15, 0.1, and 0.1.

By comparing the real and synthetic masks, the similarity metric identifies the image pairs whose semantic layouts most closely resemble one another, thereby determining the best candidate for indoor localization.

3 Case Study

3.1 Experiment Setup

To validate the feasibility and localization performance of the proposed approach, a real-world case study was carried out. The experiment was conducted in an institutional building at the University of Macau, China, where a publicly accessible corridor area was selected as the testbed. The corridor area is an L-shaped space with a length of approximately 175 meters. Both the fundamental building elements and the furniture and decorative elements (e.g., garbage bin, water dispenser, and poster board) exist inside the corridor area. The floorplan layout of the testbed is displayed in Fig. 2.

The proposed approach was applied to the corridor area for indoor localization. First, the LOD 200 BIM model of the institutional buildings was employed to construct the offline database. This BIM model contained fundamental building components, while other furniture and decorative components were not created due to the considerable efforts required. This study then leveraged the custom plugin to generate a set of virtual cameras, where the camera interval was set to 0.914 meters (3 feet), resolution was set to 640×480 pixels, and the DPI was configured at 600. The camera interval was determined through an iterative testing process. Initial experiments with smaller intervals yielded excessive redundancy in the image database, as numerous virtual cameras generated highly similar images with negligible variation. In contrast, larger intervals substantially reduced the size of the reference database, but this came at the cost of increased localization errors due to inadequate image coverage. Therefore, a moderate interval was determined via trial-and-errors to balance the database size and the localization accuracy.

Subsequently, 616 virtual cameras were generated, covering 2,464 capturing views, with cameras oriented

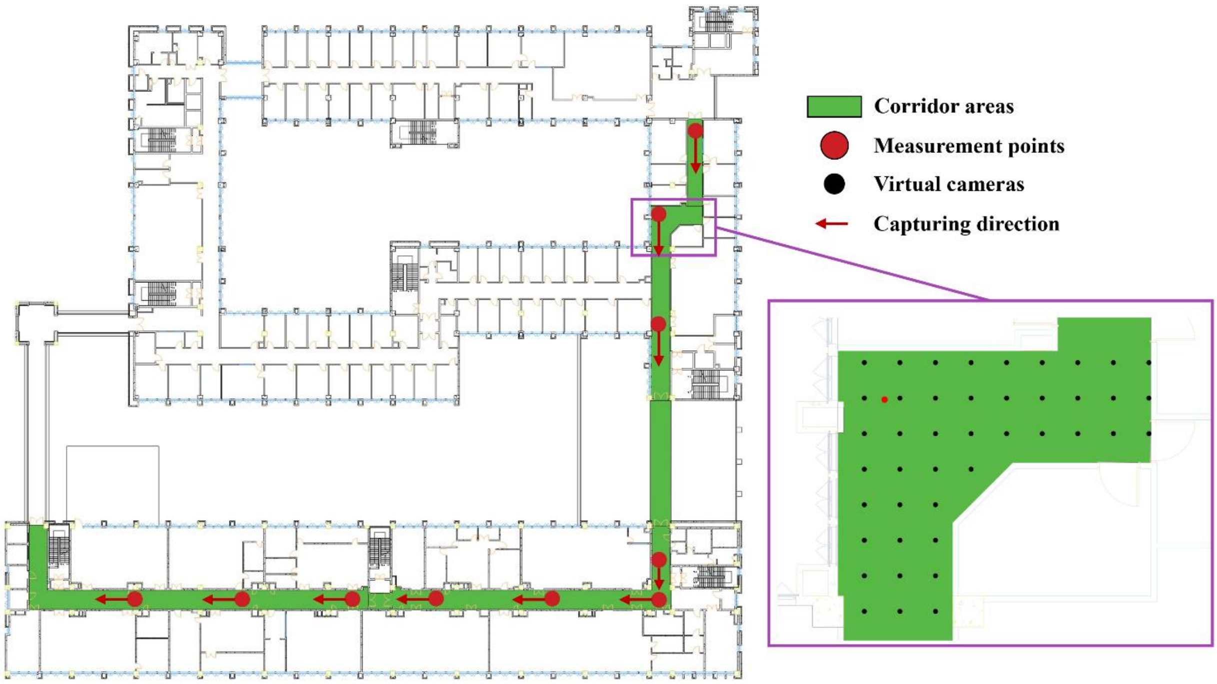

Figure 2. Virtual camera deployment and the measurement points in the testbed BIM model

along the four directions of the X and Y axes. After filtering, 1,005 semantic masks remained as the reference data in the offline database. Fig. 2 shows the deployment of the virtual cameras in the BIM model.

Next, this study conducted a pedestrian walk inside the indoor environment. During the walk, ten measurement points were selected to capture real photos along the walking direction. Additionally, their relative locations were recorded manually using a tape to form the ground truth data. In this study, a digital camera equipped with a mobile device (Xiaomi Mi 13 [15]) was utilized to capture the real photos. Fig. 2 shows the locations of the measurement points in the BIM model. These photos were then performed segmentation predictions to generate the masks of building components. Last, the mask of each measurement point severed as the query images to match with the offline database. The most similar pair was calculated based on the class-balanced IoU metric.

For comparison, this study selected the method proposed by Ha et al. [8] as the baseline approach. The baseline approach leveraged BIM models to generate realistic rendered images and then utilized convolutional neural networks to extract visual features from images for similarity computation. Accordingly, we generated the BIM-rendered images of the corridor areas and took the fourth convolutional block of the ResNet50 model [16] built on the Keras framework [17] to characterize the image features. Subsequently, the cosine similarity metric was employed for image retrieval.

To perform quantitative analysis, this study introduced two indicators to compare the accuracy of the proposed and baseline approaches. The first indicator is the system error, which measures the deviation between the actual positions and its nearest reference points in the offline database. This system error reveals the inherent impact of the interval spacing of the virtual cameras in BIM on localization accuracy. The second indicator is the top10 range accuracy, which comprehensively quantifies the image retrieval results in terms of both location proximity and mask similarity. The traditional top10 accuracy only considers whether the nearest reference point of the query image exists in the first ten search results or not, without accounting for other reference points that also lie near the actual location. Hence, the top10 range accuracy addresses this limitation by incorporating two parts, i.e., a positioning score and a sequencing score.

The positioning score (p_i) gauges whether the matched reference lies within an acceptable distance (i.e., a specified radius) from the actual position and further quantifies their proximity, as shown in Equation (5). The positioning score ranges from 0 to 1, where a value closer to 1 indicates that the matched reference point is closer to the actual location. The sequencing score (q_i) calculates the relative similarity of each matched result using the IoU metric, as shown in Equation (6). Consequently, the top10 range accuracy ($t_{10}\mathrm{Acc}$) is determined by selecting the maximum product between the position score and the sequencing score among the first ten retrieved results (Equation (7)).

$$p_i = \begin{cases} (r - d_i)/r, & \text{when } d_i \le r \\ 0, & \text{when } d_i > r \end{cases} \quad (5)$$

$$q_i = s_i / s_1 \quad (6)$$

$$\mathrm{t_{10}Acc} = \max_{i \in [1,10]} \{p_i \times q_i\} \quad (7)$$

where d_i is the distance between the i-th matched reference point and the true location, r is the radius of the acceptable range (i.e., 2 meters in this study), and s_i is the IoU metric obtained from Equation (4).

3.2 Experimental Results and Performance Evaluation

After the experiment, the results of the case study are calculated accordingly. For the system error, the results indicate that an average deviation of 0.38 meters exists between the actual locations and their nearest reference points. This error indicates that due to the camera interval configuration, the proposed approach reaches approximately decimetre-level accuracy in the best scenarios, which should meet the requirements of most pedestrian indoor navigation applications.

Next, the results of the top10-range accuracy for the two approaches are summarized in Table 2.

Table 2. Top10 range accuracy of the two approaches

Measurement point	$\mathrm{t_{10}Acc}$	
	The proposed approach	The baseline approach
1	0.88	0
2	1	0
3	0.77	0.95
4	0	0
5	0.54	0.54
6	0.54	0
7	0.51	0
8	0	0
9	0.33	0.34
10	1	0
Average	0.57	0.18

Specifically, the results demonstrate that the proposed approach achieves an average $\mathrm{t_{10}Acc}$ of 0.57, showing a moderate level of performance for indoor localization. Fig. 3 illustrates more details about the image retrieval results, revealing that the IoU metric can identify highly similar mask images for all measurement points. However, the range accuracy is sometimes constrained when multiple images exhibit high resemblance across different locations. Comparatively, the first three measurement points which contain all five building components would result in relatively higher accuracy, given that these images provide richer semantic and spatial details. On the other hand, the rest measurement points only incorporate a reduced set of components (i.e., floors, ceilings, walls, and doors), which may result in similar patterns across various locations and thus sometimes yield suboptimal matches.

By contrast, the baseline approach produces less satisfactory results, leading to an average t_{10}Acc of 0.18. The relatively low accuracy could be attributed to the inconsistent visual features between the BIM-rendered images and the real-captured images. Particularly, the BIM models include only fundamental building components painted in default colours, which significantly differ from the actual indoor environment in terms of visual fidelity. Therefore, the baseline approach may become ineffective when applied to low LOD BIM requirements.

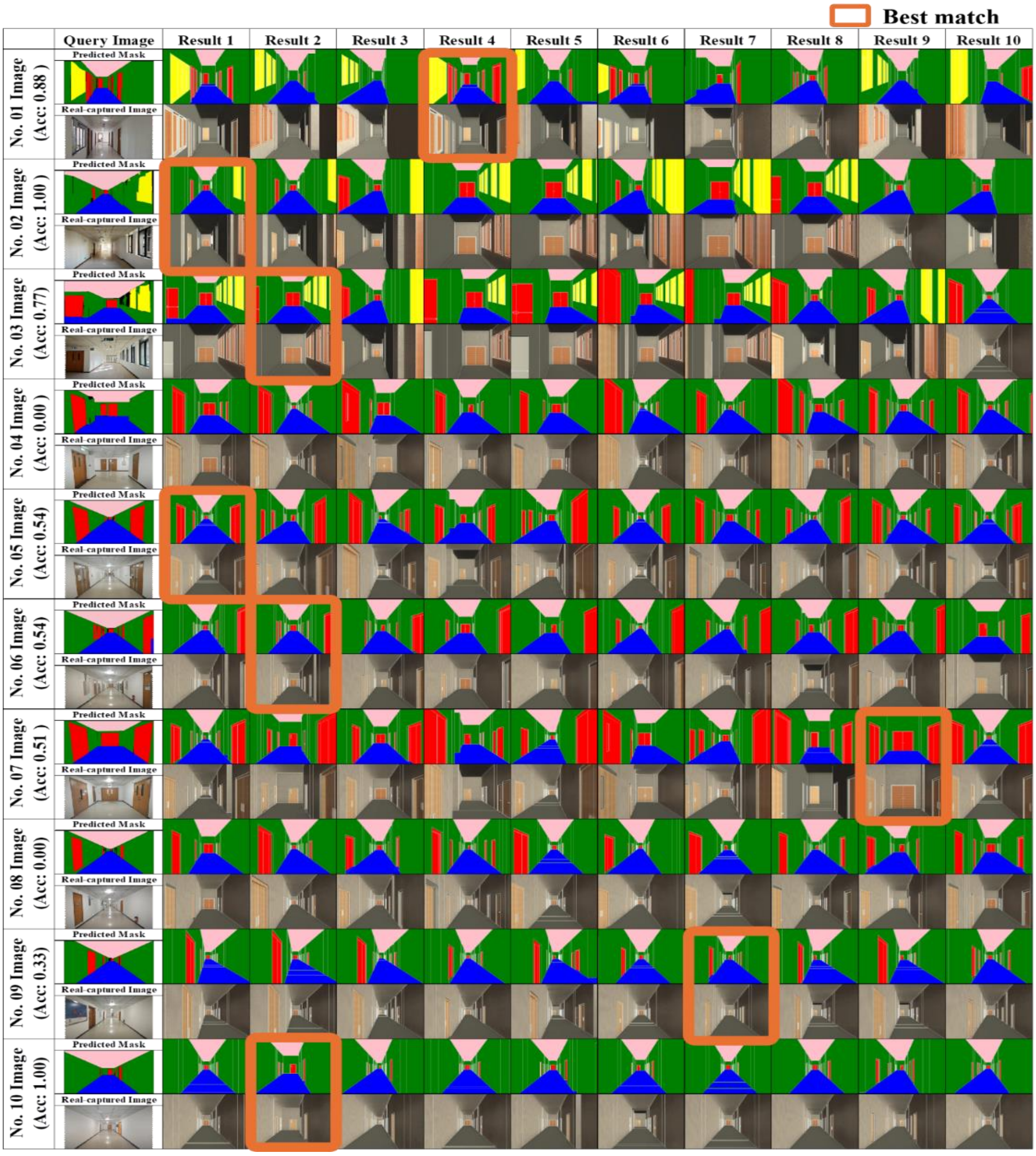

Figure 3. Image retrieval results of the proposed approach.
Note: Queries with no box-selected images indicate that the best match is absent from the top10 results

4 Discussion and Conclusion

This study presents a novel image-based indoor localization approach by integrating BIM synthetic data generation and building component semantic segmentation. Through the development and evaluation of the proposed approach, the benefits and limitations are identified. Specifically, the approach concentrates on the fundamental building components and generates semantic masks for both offline BIM database construction and online image querying. The benefits of this approach are multi-fold. First, it operates effectively with relatively low-fidelity BIM models, thereby minimizing the manual effort required for detailed modelling or continuous updates. Only essential building components are required, which enables greater applicability for real-world uses. Second, it provides a more robust performance against environmental changes. Prior methods often rely heavily on visual fidelity between BIM models and real-world environments, making them vulnerable to discrepancies in actual conditions. By contrast, the proposed approach makes use of the spatial and semantic information of those static building components, which offers consistent patterns for feature matching over time.

Compared to traditional localization methods such as IMU-based tracking or Wi-Fi fingerprinting, the proposed image-based approach offers advantages in measurement mechanisms and ease of deployment. The accuracy of IMU-based methods is gradually affected by cumulative drift errors, which require additional recalibration. Range-based approaches (e.g., Wi-Fi signals) may suffer from environmental interference and often rely on additional infrastructure, such as multiple access points, beacons, or specialized sensors, which may not be readily available in all environments. By contrast, the proposed approach determines the user's locations using independent capturing and similarity matching, which does not incur any accumulated errors. In addition, the approach provides position references using stable architectural features extracted from images without the need for extensive sensor deployments. This makes the approach scalable to large-scale indoor environments.

The result of the case study indicates that the top 10-range accuracy of the proposed approach achieves moderate performance (0.57), which significantly outperforms the conventional baseline approach (0.18). Moreover, the result also discloses that the top 10-range accuracy varies across different measurement points. When positioning with various surrounding features like distinct combinations of windows and doors, the proposed approach results in a better performance. It is because these divergent environmental features provide a more distinguishable image pattern in the segmentation mask to separate it from others in the image database. However, the proposed approach may struggle in common corridor scenes, where repetitive features like similar doors and walls occur and confuse the image retrieval results.

Several limitations are also identified and require further investigation. First, the parameters of virtual camera deployment, such as orientation directions, camera spacing intervals, and focal lengths, are set empirically in the present study, which needs to be optimized by systematic experiments to better understand how they affect localization accuracy and computational efficiency. Second, the current approach only considers discrete measurements using individual image captures, which makes localization accuracy highly susceptible to segmentation errors. Incorporating continuous measurements with probabilistic models could mitigate this issue and offer a more robust solution. Third, the case study reveals that the proposed approach may become less effective when applied to locations with similar spatial patterns of building components. To address this, additional essential elements (e.g., mechanical, electrical, and plumbing components) could be included in the segmentation process to enhance feature distinguishability. Finally, future work should validate the method across diverse interior environments to confirm its generalizability and practical applicability. Specifically, the present study is still in the proof-of-concept stage to investigate the feasibility and performance of the proposed approach for BIM-based indoor localization. In the future, the approach should be further integrated into self-developed prototypes or existing software to empower different applications such as facility management and emergency response. Such integration has the potential to assist various practitioners, including facility managers, construction engineers, and emergency responders, in enhancing operational efficiency and decision-making. Besides, systematic comparison to existing approaches using open datasets is needed to quantify its performance and identify its utilization and potential challenges.

To conclude, this study demonstrates the potential of leveraging spatial and semantic details of fundamental building components in capturing views to develop an efficient and reliable approach for indoor localization. At the same time, follow-up research is still needed to fully exploit the advantages of the approach and broaden its adoption in more complex indoor environments.

Acknowledgements

The authors acknowledge funding support from The Science and Technology Development Fund, Macao S.A.R (FDCT project no. 0034/2024/RIB1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101034337.

References

[1] P. Roy and C. Chowdhury. A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent & Robotic Systems*, 101(63), 2021. https://doi.org/10.1007/s10846-021-01327-z.

[2] M.O. Wong, H. Zhou, H. Ying, and S. Lee. A voice-driven IMU-enabled BIM-based multi-user system for indoor navigation in fire emergencies. *Automation in Construction*, 135, 104137, 2022. https://doi.org/10.1016/j.autcon.2022.104137.

[3] M. O. Wong and S. Lee. Indoor navigation and information sharing for collaborative fire emergency response with BIM and multi-user networking. *Automation in Construction*, 148:104781, 2023. https://doi.org/10.1016/j.autcon.2023.104781.

[4] J. Kunhoth, A.G. Karkar, S. Al-Maadeed, and A. Al-Ali. Indoor positioning and wayfinding systems: a survey. *Human-centric Computing and Information Sciences*, 10(1):1–10, 2020. https://doi.org/10.1186/s13673-020-00222-0.

[5] I. Ha, H. Kim, S. Park, and H. Kim. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, 140:23–31, 2018. https://doi.org/10.1016/j.buildenv.2018.05.026.

[6] H. Kawaji, K. Hatada, T. Yamasaki, and K. Aizawa. Image-based indoor positioning system: fast image matching using omnidirectional panoramic images. *In Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis*, pages 1–4, October 2010. https://doi.org/10.1145/1878039.1878041.

[7] X. Sun, Y. Xie, P. Luo, and L. Wang. A dataset for benchmarking image-based localization. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7436–7444, 2017.https://doi.org/10.1109/CVPR.2017.598

[8] I. Ha, H. Kim, S. Park, and H. Kim. Image-based indoor localization using BIM and features of CNN. *In Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1–4, 2018. https://doi.org/10.22260/ISARC2018/0107.

[9] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. https://doi.org/10.48550/arXiv.1409.1556.

[10] D. Acharya, K. Khoshelham, and S. Winter. BIM-PoseNet: indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:245–258, 2019. https://doi.org/10.1016/j.isprsjprs.2019.02.020

[11] A. Kendall, M. Grimes, and R. Cipolla. Posenet: A convolutional network for real-time 6-DOF camera relocalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2938–2946, 2015. https://doi.org/10.1109/iccv.2015.336.

[12] Autodesk. Revit 2024 API Documentation. On-line: https://www.revitapidocs.com/, Accessed: 24/12/2024.

[13] G. Jocher, J. Qiu, and A. Chaurasia. Ultralytics YOLO (Version 8.0.0) [Computer software]. On-line: https://github.com/ultralytics/ultralytics, Accessed: 24/12/2024.

[14] M. O. Wong, H. Ying, M. Yin, X. Yi, L. Xiao, W. Duan, and L. Tang. Semantic 3D reconstruction-oriented image dataset for building component segmentation. *Automation in Construction*, 165:105558, 2024. https://doi.org/10.1016/j.autcon.2024.105558.

[15] Xiaomi. Xiaomi 13 - Specifications. On-line: https://www.mi.com/global/product/xiaomi-13/specs, Accessed: 01/13/2025.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. https://doi.org/10.1109/cvpr.2016.90.

[17] Chollet, F., & others. (2015). *Keras*. Retrieved from https://keras.io