A SAM-Based Approach for Automatic Indoor Point Cloud Segmentation

Muataz S. A. Albadri[1], Patricia González-Cabaleiro[1], Rosa M. Túñez-Alcalde[1], Antonio Fernández[1], Lucía Díaz-Vilariño[1]

[1]CINTECX, Universidade de Vigo, GeoTECH group, 36310 Vigo, Spain (muatazsafaaabed.albadri, patricia.gonzalez.cabaleiro, rosamaria.tunez, antfdez, lucia)@uvigo.gal

KEYWORDS: Scan-to-BIM, as-built BIM, segment anything model, building reconstruction, 3D indoor modelling

ABSTRACT: Foundation models in computer vision, such as the Segment Anything Model (SAM), have demonstrated remarkable zero-shot performance in image segmentation. Leveraging these models for automated building segmentation can contribute to the efficiency of Scan-to-BIM workflows. Automatic 3D modelling has become widely relied on point cloud data; however, the nature of this data hinders the direct application of the foundation models. This study explores the potential use of SAM for automatic point cloud segmentation, proposing a SAM-based approach for segmenting building components, such as rooms, doors, and windows. The proposed method employs SAM to generate masks for an image that represents projected point clouds. Point clouds are then retrieved for each mask, which are further classified to identify building components. Room segmentation starts with the extraction of a section that defines the room boundary, followed by horizontal projection of the section. In contrast, door and window segmentation starts by projecting planes containing wall points onto their normal vectors. The experiments have been performed using three real case studies. The findings demonstrate the method's effectiveness without requiring any pretraining process, highlighting that the application of the foundation models in point cloud segmentation is a promising direction.

1. Introduction

Scan-to-BIM is a process that converts point cloud data into detailed 3D models and typically comprises several stages including point cloud data acquisition, pre-processing, and processing. However, Scan-to-BIM remains a time-consuming and labour-intensive process, primarily relying on manual efforts. Advances in the acquisition system, such as laser scanners, have boosted interest in automated indoor modelling. A key challenge in automating the Scan-to-BIM workflows is achieving automatic and accurate point cloud segmentation. Automatic segmentation still faces challenges due to the massive raw data, the variable quality dependent on acquisition systems, and the occlusions that clutter indoor scans (Sun et al., 2024). The raw data are unstructured, presenting significant challenges for the direct application of advanced 2D segmentation techniques to 3D point clouds. In recent years, 2D and 3D-based methods have been presented for indoor point cloud segmentation. Though these methods have demonstrated their effectiveness in recent studies, more robust and general approaches are still needed.

Recent advances in Artificial Intelligence (AI) have introduced successful foundation models in the Large Language Models (LLM) such as Generative Pre-trained Transformers (GPT) (Radford et al., 2019) and DeepSeek (Bi et al., 2024). The foundation models are being extended to the field of vision such as the Segment Anything Model (SAM) (Kirillov et al., 2023). SAM was trained on extensive datasets of images and masks and delivers exceptional zero-shot performance without requiring task-specific training. This model is based on automatically generating masks for image segmentation and is a highly versatile model in fields like healthcare, robotics, augmented reality, autonomous systems, and construction, offering a promising alternative to traditional methods by reducing computational and data-intensive demands.

This study investigates SAM's potential for segmenting indoor point clouds to facilitate the automatic 3D reconstruction of existing buildings. The study proposes a SAM-based approach to segment building components such as rooms, doors, and windows. The proposed approach aims to use SAM to create 2D masks of images created from rasterized point clouds, followed by retrieving points belonging to each mask from the rasterized point cloud and finally classifying these retrieved points. The method focuses on single-storey indoor point clouds.

The rest of the paper is organized as follows. Section 2 reviews indoor point cloud segmentation methods and the recent use of SAM in the construction sector. The proposed method is then detailed in Section 3 and evaluated with different case studies in Section 4. Finally, section 5 presents the conclusions of this work.

2. Related Work

2.1 Room segmentation

Room segmentation is the process of identifying and isolating distinct surfaces within an indoor environment. Room segmentation methods are categorized into two main approaches: 2D-based and 3D-based approaches. Each category leverages different assumptions and computational strategies to achieve room segmentation. While 2D methods typically rely on projecting point cloud data, 3D segmentation methods process the data directly in its three-dimensional space. In 3D methods, Frías et al. (2020) and (Túñez-Alcalde et al., 2024) developed a 3D mathematical morphology strategy to segment rooms based on the study of the continuity of the empty space. Ochmann et al. (2016) proposed a segmentation technique that relies on prior knowledge of scanning positions to segment rooms. In contrast, 2D segmentation methods simplify the problem by assuming that walls are vertical. These methods segment rooms by analysing empty spaces in 2D projections of point clouds. The transition back to 3D is achieved by identifying projected points that define each 2D room area. Macher et al. (2017) presented a projection-based method to create a floor map. They projected a slice of point clouds above the doors including the ceiling. They analysed projected images to generate the map using a region-growing algorithm. Gourguechon et al. (2023) proposed a methodology composing image analysis processing and the trajectory data, along with refinement steps to produce planimetric masks for room segmentation. More recently, Tang et al. (2024) introduced a hybrid method that integrates both 2D and 3D geometric constraints. By leveraging the strengths of both 2D projections and 3D spatial analysis, the proposed method addressed challenges posed by complex indoor environments.

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-131-2025

2.2 Door and window segmentation

Doors and windows are building elements embedded within walls and normally have regular shapes. The literature has different techniques to segment open and closed doors and windows. Díaz-Vilariño et al., (2017) and Staats et al. (2017) presented methods to detect open doors based on the trajectory. The closed doors and windows can be also detected based on finding parametric shapes in images as presented by (Díaz-Vilariño et al., 2015). They introduced an approach to detect closed and open doors based on images and point clouds. In the realm of AI, Mehranfar et al. (2024) proposed a projection-based method to detect doors and windows in point clouds using the YOLOv8 object detection network. Points belonging to walls were projected to x-z and x-y planes in the proposed method. The YOLOv8 network was trained on image datasets from various buildings.

2.3 SAM

SAM generates masks for image segmentation and the process is controlled by six parameters. The first parameter is `points_`, which controls the number of equidistant points placed along each side of the input image to create a uniform grid of prompts that guide the mask image generation process. Increasing the value of `points_` will increase the density of points, potentially improving mask generation accuracy for small details. Additionally, `iou_`, `stability`, and `min_mask` parameters control the Intersection over Union, stability score, and minimum area (in pixels) for a mask to be retained respectively. Finally, the `n_layers` and `down_factor` parameters are the crop layers and point sampling respectively. SAM has been reported across a wide range of applications in different sectors (Hazzaa et al., 2024). In the construction sector, SAM has been implemented for structural corrosion inspection (Chai et al. 2024), and structural surface damage detection (Ye et al., 2024) based on image datasets. SAM was also employed to segment point clouds of the underground tunnels (Ye et al. 2025). Wang et al. (2024) proposed an approach to segment mechanical, electrical, and plumbing (MEP) systems using two data sources: camera data and LiDAR point cloud data. The presented approach was based on different vision foundation models including SAM. Yang et al. (2023) introduced a method that segments 3D point clouds using SAM in RGB-D images. The proposed method used SAM to generate masks in RGB images and then project the 2D masks into 3D point clouds based on the depth information provided by the RGB-D images. To the best of our knowledge, the capabilities of SAM for segmenting building indoor elements have yet to be explored.

3. Method

Our approach encompasses four key stages: occupancy image creation, 2D mask image generation, point cloud segmentation, and classification. The first stage involves creating an occupancy image by rasterizing point clouds. The second stage applies SAM to generate a 2D mask image for the occupancy image. The third stage consists of retrieving point clouds based on the 2D mask image. Finally, the retrieved point clouds are classified during the fourth stage. The following subsections will explain the specific application of this workflow to room segmentation, as well as window and opening segmentation.

3.1 Room segmentation

Room segmentation is performed based on the assumption that rooms are horizontally bounded by vertical walls, regardless of their shape and orientation. Consequently, rooms are defined by the boundaries formed by their walls in the horizontal projection of a storey's point cloud. The indoor point clouds typically include floor, furniture, beams, and ceiling, which cause occlusions that hinder the definition of room boundaries. A well-defined room boundary is obtained at a section above the furniture and below the beams.

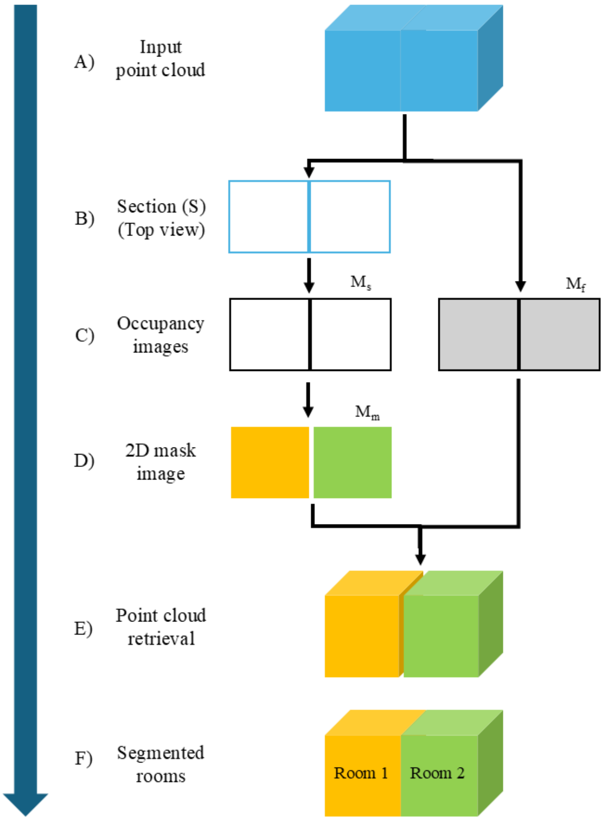

Figure 1 Room segmentation process. A) 3D point cloud, B) C) Section extraction, C) Occupancy image creation, D) 2D mask image generation using SAM, E) Point cloud retrieval, and F) Segmented rooms.

The SAM-based approach is applied to segment rooms within a single-story point cloud data (PCD), summarized in Figure 1. The input data is 3D point clouds (Figure 1A). Firstly, RANSAC detects a plane containing the ceiling from the PCD. The method then extracts a section (S) from the input PCD, excluding furniture, beams, and ceiling (Figure 1B). This section is situated below the ceiling (or the beams) by a specified distance (d_s) and has a defined thickness (t). The section (S) is then horizontally projected to create an occupancy image (M_s), which is used in the mask generation step. Similarly, the input PCD is also projected into the x-y plane to create another occupancy image (M_f), which is used in the point cloud retrieval step. These images are created using the same pixel size (l) and represent the presence of points in the vertical columns above the pixels. Pixels containing point cloud data are assigned a black colour, indicating occupied regions (Figure 1C).

In the mask generation step, the method uses SAM to create 2D masks for the M_s image by adjusting SAM's parameters to control the mask creation process. SAM generates a mask for each area, with each mask assigned a distinct colour corresponding to a specific room area within the M_s image. An image (M_m) is then created containing the generated 2D masks

(Figure 1D). The M_m image matches the size of the M_f image to facilitate the retrieval of point clouds for the masks from the PCD. During point cloud retrieval, RGB values and pixel indices are obtained from the M_f image. Pixels sharing the same RGB values are grouped. Corresponding points for each RGB group are then extracted from the PCD (Figure 1E). Furthermore, SAM generates masks for not-room areas; therefore, a room or not-room classification is performed. The retrieved point clouds with a cross-sectional area greater than a minimum room area threshold (A) are classified as rooms. The A threshold is identified based on the architectural standards of the buildings.

Points representing walls potentially are not fully included in the classified rooms, since the proposed method segments the rooms based on indoor room areas. This can result in an incomplete representation of walls. To address this problem, a buffer is applied to enhance room segmentation. This involves expanding the horizontal room boundary outward by a distance (d_o) equal to half of the wall thickness. The d_o is adjusted based on prior knowledge of the average thickness of walls. The updated boundaries are then used to re-retrieve the corresponding point clouds for the rooms from the input PCD (Figure 1F). The remaining points (rest), which may include points corresponding to doors or windows, are preserved for use in subsequent processing.

3.2 Door and window segmentation

Doors and windows are embedded into walls and typically have regular shapes (e.g., doors are rectangular vertically oriented, while windows can be rectangular, squared, etc). In addition, doors usually have standard sizes and are installed at floor level while windows are typically installed at a certain distance above the floor level. Our method considers these general assumptions to explore the capabilities of SAM to detect doors and windows as shown in Figure 2. First, the method segments planes ($\pi_{i,j}$) containing walls of each segmented room; $\pi_{j,i}$ for the j-th wall plane in the i-th room, $1 \leq j \leq n_i$, where n_i is the number of walls in room i. RANSAC is used in the plane segmentation process. The segmented wall planes are refined by extracting the points aligned with each plane from the remaining points (rest). Planes corresponding to ceilings and floors are also segmented. Point clouds that do not correspond to the floor, ceiling, and wall planes are classified as others, which may include points corresponding to furniture or similar. Each $\pi_{i,j}$ plane is then rasterized based on its normal vector, and an occupancy image (W_i) is created (Figure 2A).

Second, 2D masks are generated using SAM for areas that represent doors, windows, and others. Closed doors and windows are typically represented in the wall planes, while open or uninstall doors and windows are not. To detect the open or uninstall doors and windows, the method expands the generated masks to include boundary points for their openings. A dilation method is used to extend the masks. Then, a 2D mask image (W_m) is created containing the generated 2D masks (Figure 2B). Third, point clouds belonging to each mask are retrieved from the projected $\pi_{j,i}$ plane (Figure 2C).

Finally, the retrieved point clouds are classified into doors, windows, or others by analysing their shapes and positions (Figure 2D). A retrieved point cloud is classified as a door if its 2D boundary has dimensions within standard door size thresholds, and the boundary's bottom edge aligns with the floor level. On the other hand, windows are classified using the shape compactness presented in Equation (1) (Montero and Bribiesca; 2009). Shape compactness is a measure widely used to analyse and compare the geometric properties of objects.

$$C = \frac{4\pi\, Area}{(perimeter)^2} \tag{1}$$

Where C is the value of shape compactness, calculated by the shape's area and perimeter. C value ranges from 0 to 1, with 1 corresponding to a perfect circle, 0.7854 corresponding to a square, and 0.6046 corresponding to an equilateral triangle, while $C \rightarrow 0$ corresponds to shapes with negligible area compared to their perimeter. Consequently, a retrieved point cloud is classified as a window when it has a significant C and the boundary's bottom edge does not align with the floor level. A threshold (c) of shape compactness is used in the windows classification process. Since windows are most commonly rectangular, the threshold (c) is set as higher than the compactness (C) of an equilateral triangle.

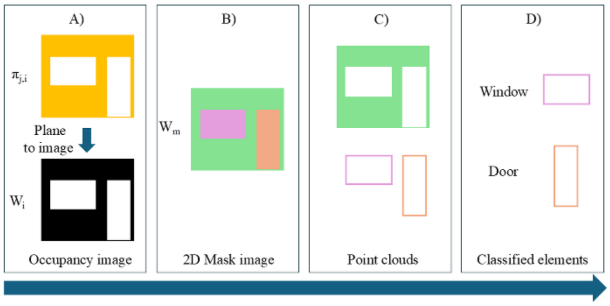

Figure 2. Door and window segmentation process.
A) Occupancy image generation, B) 2D mask image generation, C) Point cloud retravel, and D) classification step.

4. Results and discussion

4.1 Case studies

The proposed approach was tested on three real datasets (CS1, CS2, and CS3), containing 5.3 million, 8.2 million, and 24.3 million 3D points, respectively. Each dataset represents a one-storey, multi-room case study as shown in Figure 3. The CS1 and CS3 correspond to two storeys of different real residential buildings and CS2 is one storey from the ISPRS Benchmark (Khoshelham et al., 2017). The case studies vary in terms of rooms' size and shape, walls' thickness, ceilings' heights, and clutter, and include both open and closed doors and windows.

4.2 Room Segmentation Results

The experiment results at different stages are presented in Figure 4. Figure 4A shows the ground truth which was created manually. The input data consisted of XYZ coordinates (Figure 4B). The method first detected the ceiling using a RANSAC threshold of 0.07 m and extracted the section (S) at d_s equal to 0.5 m and had t equal to 0.3 m across all case studies as shown in Figure 4C. A pixel size (l) of 0.01 m was selected to create the occupancy images (M_s and M_f). This pixel size was determined by exploring the impact of different pixel sizes. Subsequently, the mask image (M_m) was generated using SAM's parameters, as presented in Table 1. Five parameters were assigned the same values, while the parameter (`points_`) was changed across all experiments. The `points_` parameter was tuned based on the quality of the generated masks.

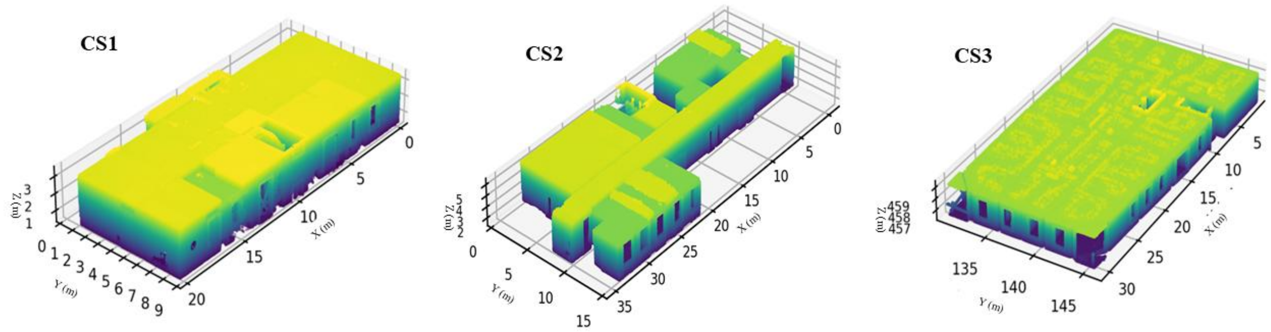

Figure 3. Overview of the case studies.

The retrieved point clouds are presented in Figure 4D. Room or not classification was performed by setting the *A* threshold as 1.5 m^2. The classified rooms were refined by buffering their boundaries by 0.07, 0.1, and 0.12 m of CS1, CS2, and CS2 respectively. The new boundaries were used to re-retrieve the corresponding points for the rooms from the input data. The segmented rooms of all case studies are presented in Figure 4E.

SAM parameters \ Case studies	CS1	CS2	CS3
`points_`	11	15	30
`iou_`	0.85		
`stability_`	0.95		
`n_layers`	1		
`down_factor`	2		
`min_mask`	100		

Table 1. SAM’s parameters for all case studies.

As shown in Figure 4E, corridors in CS2 and CS3 were not segmented because SAM did not generate masks for the corridors during the mask generation step. Consequently, a reprocessing was conducted to segment the corridors. This reprocessing was applied only to the remaining points, which contain points corresponding to the corridors. The same parameter values were applied except the `points_` parameter was set to 5 for both case studies. Figure 4F shows all segmented rooms of CS2 and CS3 after the reprocessing. Points not corresponding to the segmented rooms were preserved as remaining points (rest) (Figure 4G).

To evaluate the performance of the proposed method, Intersection over Union (IoU) was used. IoU is widely used as an evaluation metric in 2D segmentation. IoU quantifies the overlap between segmented and ground truth regions after the point clouds have been projected onto a 2D plane. Firstly, the segmented rooms were classified as corresponding to their respective ground truth rooms. Let P_i represent a set of 2D points for a segmented room (in X-Y projection) obtained from the proposed approach, and G_i for a ground truth room. The intersection I (Equation 2) represents the number of common points between P_i and G_i. Then, the intersection ratios p_i and g_i are calculated as expressed in Equations 3 and 4. p_i represents the proportion of intersected points relative to the total points in the segmented room. g_i represents the proportion of intersected points relative to the total points in the ground truth room. These ratios provide insight into the correspondence between the segmented and ground truth rooms. Moreover, they allow over-segmented and under-segmented rooms to correctly match their respective ground truth rooms.

$$I = P_i \cap G_i \tag{2}$$

$$p_i = \frac{|I|}{|P_i|} \tag{3}$$

$$g_i = \frac{|I|}{|G_i|} \tag{4}$$

Where $|\cdot|$ denotes the cardinality of a set. The P_i and G_i are classified as corresponding rooms if at least one of these ratios exceeds a predefined threshold th of 75%, as expressed by:

$$\text{If } p_i \geq th \text{ or } g_i \geq th, \quad \text{then } P_i \text{ and } G_i \text{ are corresponding rooms.} \tag{5}$$

Once the ground truth rooms were matched with the segmented rooms, the IoU was calculated for the corresponding rooms as expressed in Equation 6. The mean of IoU for all rooms in each case study was also computed by Equation 7.

$$\text{IoU} = \frac{|I|}{|P_i \cup G_i|} \tag{6}$$

$$\text{Mean IoU} = \frac{\sum_{i=1}^{N} \text{IoU}_i}{N} \tag{7}$$

Where IoU_i represents the IoU for the *i*-th room and N denotes the total number of rooms. Figure 5 presents the IoU values for all case studies.

The results presented in Table 2 show that the method correctly segmented the 17 rooms for CS1. Most of the segmented rooms in CS1 had IoUs exceeding 90% (Figure 5). Regarding CS2 results, seven rooms were accurately segmented, with IoU exceeding 90% (Figure 5). However, the method did not accurately segment the two corridors at the centre of CS2, in Figure 6, because SAM did not create accurate masks for these corridors. This inadequate mask generation resulted in over and under-segmentation as shown in Figure 6 CS2.

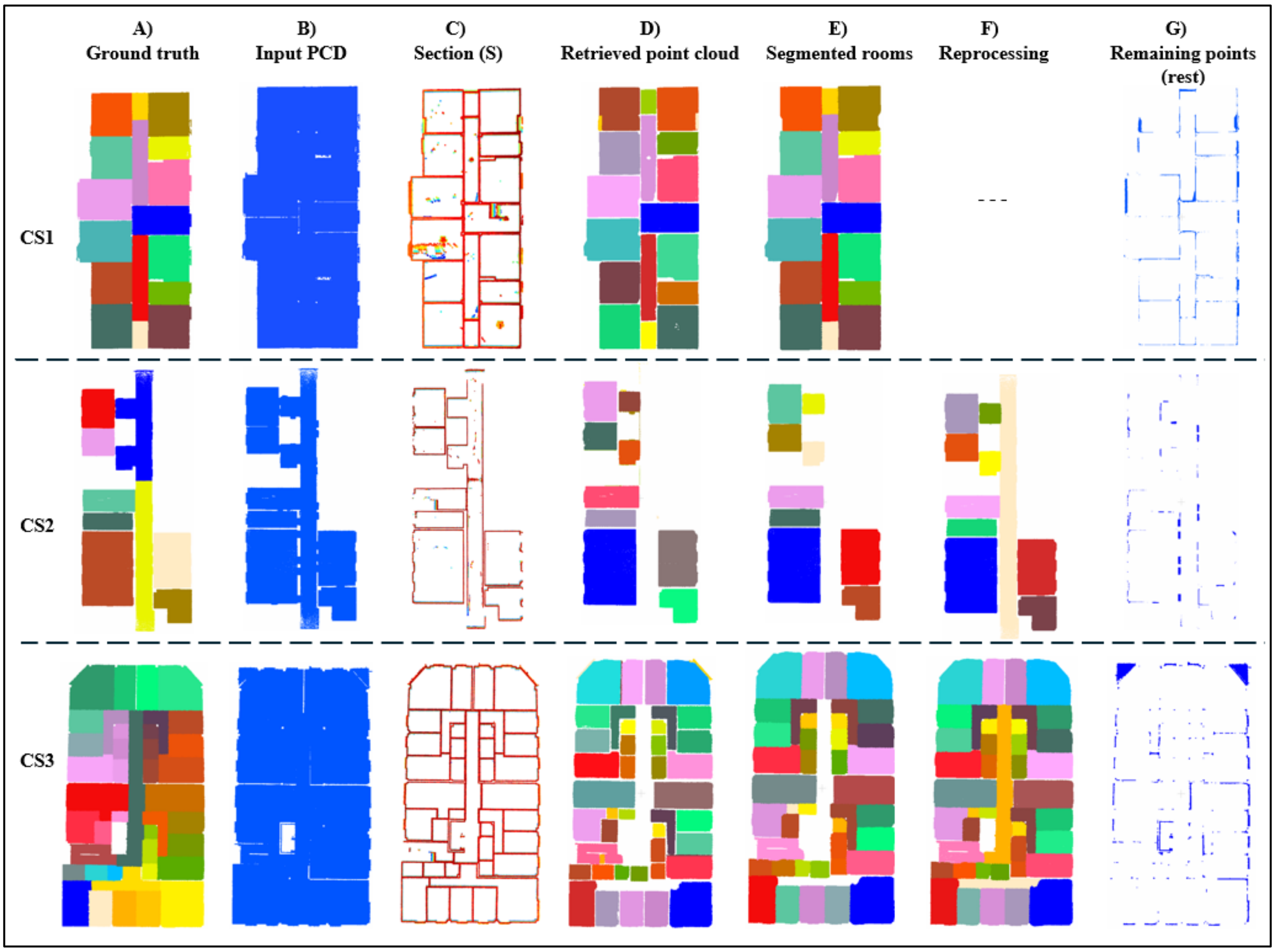

Figure 4. Room segmentation results in three dimensions (top view) for all case studies. A) Ground truth, B) Input point cloud, C) Section extraction, D) Point cloud retrieval, E) The segmented rooms, F) The results after reprocessing for CS2 and CS3, G) Remaining points.

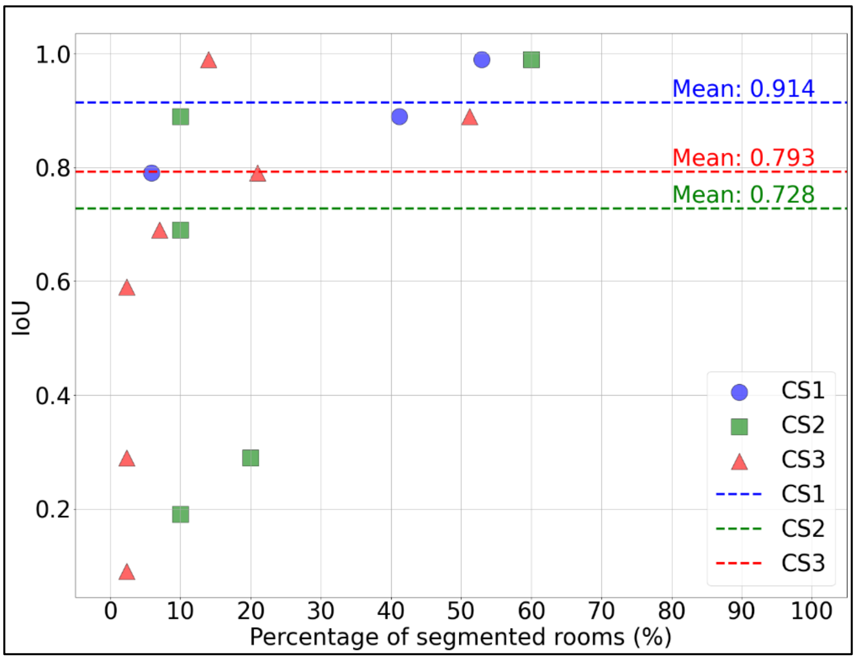

Figure 5. IoU values for room segmentation across the three case studies. The markers represent the IoU values for rooms; blue circular markers to CS1, green square markers to CS2, and red triangular markers to CS3. Each marker illustrates the percentage of rooms that achieved specific IoU values.

	No. of rooms			
	Ground truth	Segmented rooms	Over-segmentation	Under-segmentation
CS1	17	17	0	0
CS2	9	10	2	1
CS3	41	43	2	0
All	67	70	4	1

Table 2. Room segmentation results.

In CS3, the method segmented 43 rooms; only two were Over-segmented (Figure 6 CS3). This over-segmentation occurred because SAM had created separate masks for the entrances in these two rooms and these entrances had areas more than the *A* threshold. Additionally, the corridor and the adjacent room in CS3 (Figure 6b) were not correctly separated because SAM did not exactly recognize the corridor and room boundaries, which resulted in inaccurately generated masks. Generally, most of the segmented rooms in CS3 had an IoU of more than 80% (Figure 5).

Overall, the results indicate that the method can successfully handle room segmentation with minimal errors, though improvements are needed for more complex areas like long corridors and room entrances. The method showed significant performance and correctly segmented 65 out of 67 rooms across the three case studies. Despite some challenges, most of the segmented rooms achieved IoU exceeding 85%, highlighting the method's effectiveness. Furthermore, the `points_` parameter plays a crucial role in the method, and selecting an appropriate value can further enhance the method's efficiency.

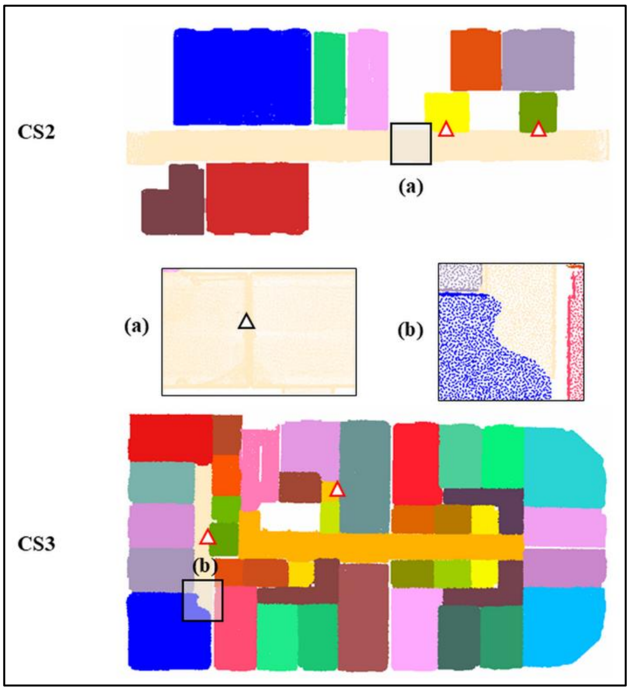

Figure7. CS2 and CS3 room segmentation results. Red triangles localize over-segmentations and black triangles for under-segmentations, (a) focus on under-segmentation, and (b) on inaccurate segmentation of the corridor and adjacent room.

4.3 Door and window segmentation

The door and window segmentation experiments were conducted on the CS1 and CS2. A RANSAC threshold of 0.07 m was used to segment planes containing floor, ceiling, and walls for each room. Planes containing walls were enhanced from the remaining points (rest) shown in Figure 4G. The remaining points aligned with a plane were added to this plane. A pixel size (l) of 0.01 m was used for occupancy image generation. SAM's parameter values presented in Table 1 were applied for 2D mask image generation, except the parameter (`points_`) was set as 5.
After the point cloud retrieval, the door classification was performed using thresholds of door dimensions, with widths ranging from 0.7 to 1 meter and door heights ranging from 1.9 to 2.5 meters. Retrieved point clouds within these thresholds and aligned with the floor levels were classified as doors. On the other hand, retrieved point clouds with shape compactness (C) of more than 0.65, and did not align with the floor level, were classified as windows.

		CS1	CS2
Doors	Ground truth	16	42
	Segmented	16	38
	Precision %	100	94.74
	Recall %	100	85.71
Windows	Ground truth	16	13
	Segmented	14	12
	Precision %	85.7	100
	Recall %	75	92.31

Table 3. Door and window segmentation results.

The door and window segmentation results for CS1 and CS2 are presented in Table 3. The main challenges encountered across the experiments were (1) detecting multiple doors embedded into a single wall, (2) detecting doors and windows in the case of occlusions and low-resolution data, and (3) detecting closed doors and windows. Figure 7 shows the results for six selected rooms, summarizing the outcomes. All doors and windows were correctly detected in Rooms 1, 2, and 3, while the challenges arose in Rooms 4, 5, and 6 (see Figure 7). To address these challenges, the segmentation process was reprocessed for rooms facing these issues. The pixel size (l) and the `points_` parameters were adjusted during the reprocessing. This reprocessing successfully resolved the first and second challenges as the method segmented the multiple doors embedded into a single wall and doors with partially captured data, see Figure 8 Rooms 4 and 5. Only two doors and one window in CS2 were undetected due to these two challenges. The third challenge persisted because SAM did not recognize the regions of the closed doors and windows in the occupancy images. Figure 8 shows an undetected window in Room 6 (red plane) because the window was closed.

The method performance was evaluated using precision and recall metrics as presented in Equations 8 and 9. The metrics were calculated by analysing the doors and windows classified as True Positive (TP), False Positive (FP), and False Negative (FN).

$$\text{Precision} = \left(\frac{\text{TP}}{\text{TP+FP}}\right) \times 100 \tag{8}$$

$$\text{Recall} = \left(\frac{\text{TP}}{\text{TP+FN}}\right) \times 100 \tag{9}$$

Where TP refers to doors and windows presented in both the segmentation results and the ground truth, while FP refers to those presented in the segmentation results, that do not actually exist in the ground truth. FN refers to doors and windows presented in the ground truth, and they are not detected in the segmentation results.

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-131-2025 | © Author(s) 2025. CC BY 4.0 License.

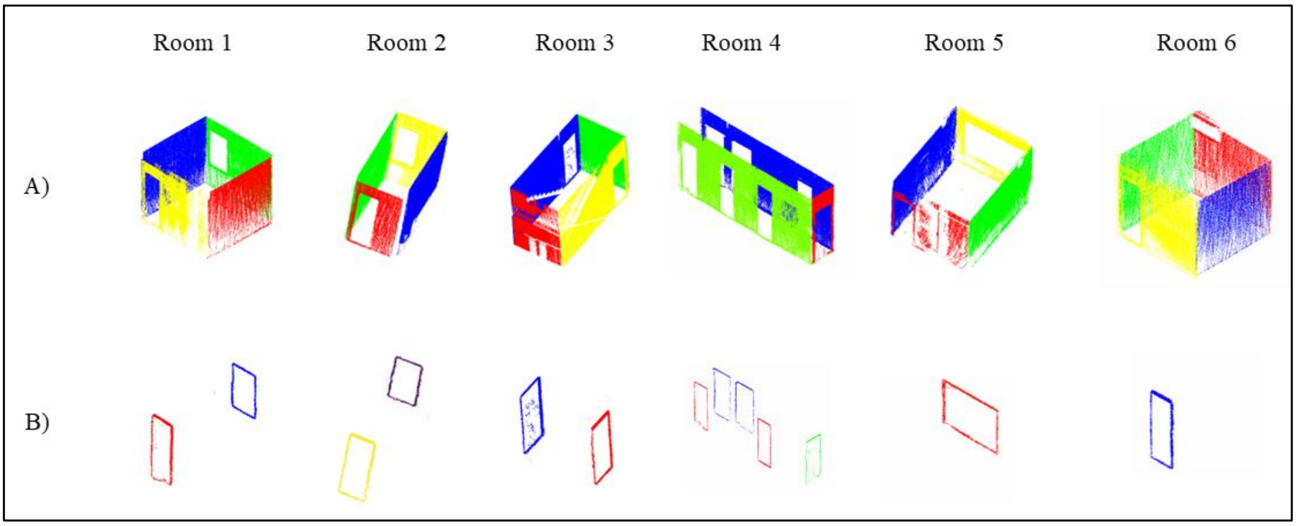

Figure 8. Door and window segmentation results for six selected rooms of CS1 and CS2, A) planes continuing walls and B) segmented doors and windows.

The door segmentation results had significant precision for both case studies, highlighting that the method was effective in detecting doors (Table 3 and Figure 9 A). All doors in CS1 were correctly detected, while only six doors in CS2 were not detected. The most likely explanation for this malfunctioning is that four of these undetected doors were closed during the data capturing, one was embedded with multiple doors in a single wall, and the sixth door was only partially captured. As a result, this contributed to a lower recall for CS2.

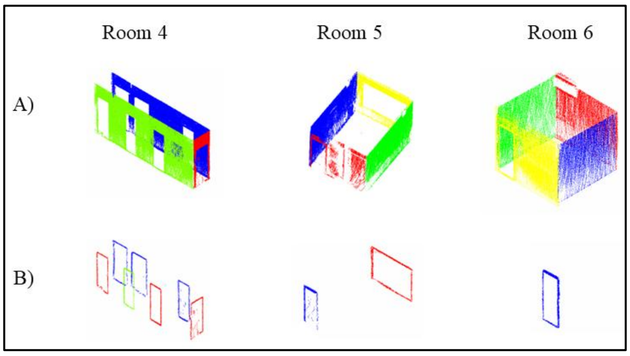

Figure 8. Reprocessing results of doors and windows segmentation, A) planes continuing walls and B) segmented doors and windows.

Overall, the method correctly detected and classified 52 doors across both case studies and mistakenly classified only two segments as doors in CS2. These segments have similar dimensions and locations to doors, which resulted in the method to classify them as doors. On the other hand, the proposed method detected 24 windows across the case studies (Figure 9B). Five windows were not detected because four of them were closed during data capture and the fifth was partially captured. Two segments were incorrectly classified as windows because of their shape and location similar to actual windows. In general, the method can effectively perform open door and window segmentation; closed elements remain undetected, as one could reasonably expect.

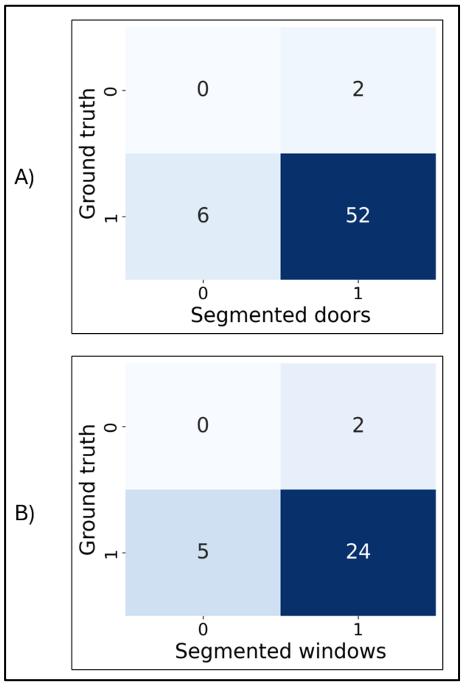

Figure 9. Confusion matrix of segmentation results for all doors and windows in the two case studies, A) for doors and B) for windows.

5. CONCLUSION

This paper explores the potential use of SAM for indoor point cloud segmentation, demonstrating its ability to effectively segment rooms, doors, and windows. The proposed SAM-based approach achieves significant performance across different architectural case studies, highlighting SAM’s adaptability to varying indoor environments. A key advantage of SAM is its ability to generate accurate masks without requiring any pretraining process, making it highly flexible and efficient. However, challenges persist in handling complex environments and detecting closed doors and windows. The study underscores that pre-trained vision foundation models can contribute to advancing automated indoor segmentation and 3D reconstruction. In future experiments, SAM will be explored to detect and segment furniture and other building elements.

ACKNOWLEDGMENTS

This work was partially supported by human resources grant RYC2020-029193-I funded by MCIN/AEI/10.13039/501100011033 and FSE ‘El FSE invierte en tu futuro’, by grant ED431F 2022/08 funded by Xunta de Galicia, Spain-GAIN, and by the projects PID2021-123475OA-I00 funded by MCIN/AEI/10.13039/501100011033/ FEDER, UE and PCI2022-132943 and CNS2022-135730, funded by MCIN/AEI/10.13039/501100011033 and by European Union NextGenerationEU/PRTR. The statements made herein are solely the responsibility of the authors.

REFERENCES

Bi, X., Chen, D., Chen, G., Chen, S., Dai, D., Deng, C., ... & Zou, Y. (2024). Deepseek llm: Scaling open-source language models with longtermism. *arXiv preprint arXiv:2401.02954*.

Chai, C., Gao, Y., Li, H., & Zhu, X. (2024). Corrosion SAM: adapting segment anything model with parameter-efficient fine-tuning for structural corrosion inspection.

Díaz-Vilariño, L., Khoshelham, K., Martínez-Sánchez, J., & Arias, P. (2015). 3D modeling of building indoor spaces and closed doors from imagery and point clouds. *Sensors*, *15*(2), 3491-3512.

Díaz-Vilariño, L., Verbree, E., Zlatanova, S., & Diakité, A. (2017). Indoor modelling from SLAM-based laser scanner: Door detection to envelope reconstruction. In *ISPRS Geospatial Week 2017* (pp. 345-352). ISPRS.

Frías, E., Balado, J., Díaz-Vilariño, L., & Lorenzo, H. (2020). Point cloud room segmentation based on indoor spaces and 3D mathematical morphology. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *44*, 49-55.

Gourguechon, C., Macher, H., & Landes, T. (2023). Room Point Clouds Segmentation: a New Approach Based on Occupancy and Density Images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 93-100.

Hazzaa, F., Udoidiong, I., Qashou, A., & Yousef, S. (2024). Segment Anything: A Review. *Mesopotamian Journal of Computer Science*, *2024*, 150-161.

Khoshelham, K., Vilariño, L. D., Peter, M., Kang, Z., & Acharya, D. (2017). The ISPRS benchmark on indoor modelling. In *ISPRS Geospatial Week 2017* (pp. 367-372). ISPRS.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015-4026).

Macher, H., Landes, T., & Grussenmeyer, P. (2017). From point clouds to building information models: 3D semi-automatic reconstruction of indoors of existing buildings. *Applied Sciences*, *7*(10), 1030.

Mehranfar, M., Vega-Torres, M. A., Braun, A., & Borrmann, A. (2024). Automated data-driven method for creating digital building models from dense point clouds and images through semantic segmentation and parametric model fitting. *Advanced Engineering Informatics*, *62*, 102643.

Montero, R. S., & Bribiesca, E. (2009). State of the art of compactness and circularity measures. In *International mathematical forum* (Vol. 4, No. 27, pp. 1305-1335).

Ochmann, S., Vock, R., Wessel, R., & Klein, R. (2016). Automatic reconstruction of parametric building models from indoor point clouds. *Computers & Graphics*, *54*, 94-103.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models Are Unsupervised Multitask Learners. *OpenAI Blog*, *1*, 9.

Staats, B. R., Diakité, A. A., Voûte, R. L., & Zlatanova, S. (2017). Automatic generation of indoor navigable space using a point cloud and its scanner trajectory. In *ISPRS Geospatial Week 2017* (pp. 393-400). ISPRS.

Sun, Y., Zhang, X., & Miao, Y. (2024). A review of point cloud segmentation for understanding 3D indoor scenes. *Visual Intelligence*, *2*(1), 14.

Tang, S., Huang, J., Cai, B., Du, H., Zhou, B., Zhao, Z., ... & Guo, R. (2024). Back to geometry: Efficient indoor space segmentation from point clouds by 2D–3D geometry constrains. *International Journal of Applied Earth Observation and Geoinformation*, *135*, 104265.

Túñez-Alcalde, R. M., Albadri, M. S., González-Cabaleiro, P., Fernández, A., & Díaz-Vilariño, L. (2024). A Top-Down Hierarchical Approach for Automatic Indoor Segmentation and Connectivity Detection. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *10*, 289-296.

Wang, B., Chen, Z., Li, M., Wang, Q., Yin, C., & Cheng, J. C. (2024). Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes. *Automation in Construction*, *162*, 105384.

Yang, Y., Wu, X., He, T., Zhao, H., & Liu, X. (2023). Sam3d: Segment anything in 3d scenes. *arXiv preprint arXiv:2306.03908*.

Ye, Z., Lin, W., Faramarzi, A., Xie, X., & Ninić, J. (2025). SAM4Tun: No-training model for tunnel lining point cloud component segmentation. *Tunnelling and Underground Space Technology*, *158*, 106401.

Ye, Z., Lovell, L., Faramarzi, A., & Ninić, J. (2024). Sam-based instance segmentation models for the automation of structural damage detection. *Advanced Engineering Informatics*, *62*, 102826.