2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED DIMENSION-AWARE 2D-TO-BIM RECONSTRUCTION THROUGH CROSS-MODAL TEXT-GEOMETRY ALIGNMENT

Xing Liang[1], Tomohiro Fukuda[1], and Nobuyoshi Yabuki[2,3]
[1]Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, The University of Osaka, Osaka, 565-0871, Japan
[2]Advanced Research Laboratories, Tokyo City University, Setagaya-ku, Tokyo, Japan
[3]The University of Osaka, Suita, Osaka, Japan (Professor Emeritus)

Abstract

Digitising existing buildings supports energy-efficient retrofitting and lifecycle analysis, yet many are documented only by legacy drawings with dimensions embedded as textual annotations. Existing methods often rely on heuristic global scaling over metric fidelity. This paper presents a vision-based framework integrating optical character recognition (OCR), OCR-guided denoising, semantic segmentation, and grid-based alignment. The primary contribution is a cross-modal text–geometry alignment strategy that reconstructs a metric reference grid from annotations and links it to building geometry. A second contribution is an Industry Foundation Classes-based evaluation framework. Results demonstrate reliable recovery of absolute dimensions for dimension-aware BIM reconstruction.

Introduction

The building sector accounts for nearly 40% of global energy consumption (El Hafdaoui et al., 2023), highlighting the need to rapidly digitise existing building stock for energy-efficient retrofits, structural analysis, and lifecycle applications. Many advanced architecture, engineering, and construction (AEC) workflows rely on semantically rich and geometrically precise building information models (BIMs) (Schönfelder et al., 2023). However, most existing buildings predate BIM adoption and are documented primarily by legacy two-dimensional (2D) floorplan drawings (Lu and Lee, 2017). These drawings contain critical geometric, semantic, and dimensional information. Consequently, automated 2D-to-BIM reconstruction has become an important research topic (Lim et al., 2018).

Typical 2D-to-BIM pipelines involve floorplan analysis followed by BIM instantiation. While BIM generation can often be handled deterministically once reliable inputs are available, the floorplan analysis stage remains challenging. Many existing methods prioritise topological correctness over engineering-level geometric precision, often relying on heuristic global scaling or normalised coordinates that are insufficient for millimetre-sensitive BIM reconstruction (Wang and Pajarola, 2025).

Achieving engineering-grade precision requires not only detecting walls and openings, but also correctly interpreting dimensional annotations and associating them with the corresponding building geometry. This task is difficult because engineering drawings contain diverse drafting conventions, rotated or occluded texts, and annotation-induced noise (Pizarro et al., 2022).

Existing studies often assume that relative geometry is sufficient or use annotations only for coarse scale estimation, without establishing explicit component-level correspondences between dimension texts and building elements (Lee et al., 2025). Moreover, reconstruction quality is usually assessed with computer-vision-oriented metrics such as intersection over union (IoU), which are insensitive to millimetre-level geometric deviations. As a result, a model may appear visually correct while remaining inadequate for engineering use.

To address these limitations, this paper proposes an automated framework for dimension-aware 2D-to-BIM reconstruction from annotated raster floorplans. Its primary methodological contribution is a cross-modal text–geometry alignment strategy that reconstructs a metric reference grid from dimensional annotations and links it to segmentation-derived architectural geometry at the component level. Its second main contribution is an IFC-based evaluation framework for engineering-level assessment of reconstruction quality at the object and parameter level.

Related work

Research on 2D-to-BIM reconstruction for existing buildings generally follows a two-stage paradigm: (1) automated floorplan analysis, which extracts geometric and semantic information from legacy drawings, and (2) BIM instantiation, which converts these extracted entities into structured, parameterised building components. BIM instantiation is relatively mature and can often be handled deterministically once reliable inputs are available, using commercial platforms, programmatic APIs, or open-standard libraries (Lu et al., 2020). As a result, the main technical divergence among existing 2D-to-BIM approaches lies in floorplan analysis, where drawing format, quality, and interpretation strategy strongly affect reconstruction accuracy and robustness.

The performance of automated floorplan analysis is influenced by drawing format. Vector-based CAD drawings preserve editable information such as layers, blocks, and dimension objects, enabling rule-based or API-driven methods to achieve high-fidelity geometric reconstruction and component classification (Zhao et al., 2021). However, in practical engineering contexts, access to native CAD files is often restricted by software dependencies, version incompatibilities, or confidentiality constraints (Dixit et al., 2019). As a result, floorplans are more commonly exchanged as raster images, such as scanned PDFs, which irreversibly lose embedded semantics and introduce noise, resolution variability, and drafting inconsistencies. These limitations have driven increasing interest in learning-based methods designed specifically for raster floorplans.

Existing research on raster-based floorplan analysis largely falls into three parallel streams: (1) geometric analysis, which reconstructs architectural structures such as walls and room boundaries; (2) symbolic object detection and recognition, which identifies standardised elements such as doors, windows, and fixtures; and (3) text detection and recognition, which extracts annotations including room names and dimensional values.

In geometric analysis, wall detection is typically treated as the core task and formulated as pixel-level semantic or instance segmentation to generate wall masks (Sun et al., 2025). More recent approaches attempt to directly predict vectorised primitives, such as corners and edges, using end-to-end models that produce machine-readable spatial layouts (Hu et al., 2024). Nevertheless, most geometric methods operate in normalised image coordinates or relative spatial representations, deriving absolute dimensions only through assumed global scale factors or omitting metric information altogether (Su et al., 2023). Although such representations are sufficient for layout understanding or navigation-oriented applications, they fall short of the millimetre-level geometric accuracy required for engineering-grade BIM reconstruction.

In parallel, symbolic object detection leverages general-purpose frameworks, such as YOLO variants, to localise and classify architectural symbols including doors, windows, and sanitary fixtures (Xu et al., 2024; Jamieson et al., 2025). These models typically output class-labelled bounding boxes or segmentation masks, enabling structured interpretation of floorplan content. However, they are often evaluated on simplified or schematic drawings with sparse or intentionally removed textual annotations. As a result, they are not designed to handle the dense and overlapping annotations commonly found in real-world engineering drawings, which limits their applicability to practical 2D-to-BIM scenarios.

In contrast, text-centric methods follow the standard scene text understanding pipeline, detecting text regions followed by character recognition, and have demonstrated success in handling arbitrarily oriented and stylised annotations (Schönfelder et al., 2024). Yet, their integration into geometry-aware reasoning remains limited. Textual annotations are frequently overlaid on or adjacent to walls, and auxiliary lines, such as dimension leaders, can substantially interfere with structural segmentation and geometric inference.

Consequently, most existing pipelines either ignore textual information entirely or use it only for coarse global scaling, without establishing explicit, component-level associations between dimensional annotations and building elements (Pokuciński and Mrozek, 2024). This reveals a critical gap: although geometric, symbolic, and textual analyses have each advanced substantially, they remain largely decoupled in practice. Existing workflows still rely heavily on heuristic assumptions, without explicitly linking dimensional annotations to building geometry or evaluating reconstruction quality at the engineering level. This limitation reduces metric fidelity and constrains the practical applicability of 2D-to-BIM reconstruction for engineering use.

Methodology

Framework overview

The proposed methodology introduces a multi-stage, vision-based framework for the automated analysis of annotated floorplans to achieve geometry-accurate 2D-to-BIM reconstruction. As illustrated in Figure 1, the framework integrates OCR, OCR-guided denoising, semantic segmentation, and geometric alignment to recover component-level dimensional information from legacy raster floorplans.

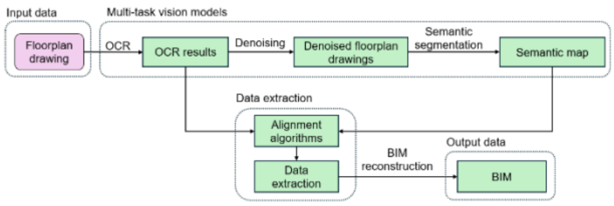


Figure 1: Overall workflow of proposed method

The methodology emphasises the cross-modal integration of heterogeneous visual cues, including textual annotations, auxiliary reference lines, and architectural geometry. This design addresses a critical bottleneck in AEC workflows, namely the lack of robust mechanisms for translating disparate recognition outputs into geometric representations.

The workflow takes raster floorplans as input and does not assume pre-cleaned or simplified documents, as shown in Figure 2(a). Textual and geometric information are processed in parallel and subsequently aligned through a dedicated data extraction and association stage, enabling staged BIM automation with dimensionally grounded component reconstruction.

OCR training and configuration

OCR is applied directly to the input floorplan to detect and recognise textual elements, including dimensional annotations. A two-stage OCR pipeline is employed, consisting of a CRAFT text detector (Baek et al., 2019) and a TPS-STN + CRNN recogniser (Shi et al., 2016) trained with CTC loss.

The OCR model is fine-tuned using 1,000 annotated floorplan images sampled from CubiCasa5k (Kalervo et al., 2019) and CVC-FP (de las Heras et al., 2015), with an 8:1:1 train/validation/test split; ResBIM (Liang et al., 2026) is used exclusively for external validation. Overlapping tiling and print-scan-style augmentations are adopted to handle large drawing canvases and small text

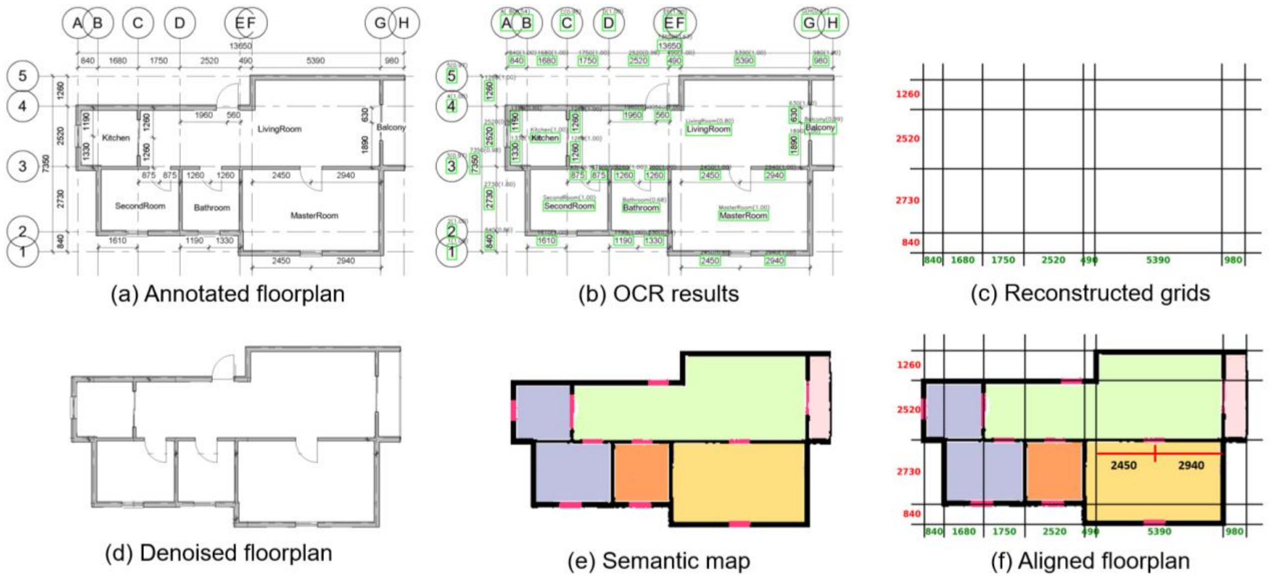


Figure 2: Overall process of multi-stage vision-based framework. (a) Input annotated floorplan; (b) OCR detection and recognition results; (c) reconstructed metric reference grid from dimensional annotations; (d) OCR-guided denoised floorplan; (e) semantic segmentation map of the denoised floorplan; and (f) final aligned floorplan obtained by aligning (c) and (e).

instances. Recognised texts are normalised and converted into structured representations, including numeric values, orientations, bounding-box coordinates, and confidence scores. These structured OCR outputs are treated as intermediate dimensional cues and are subsequently used in later cross-modal alignment and geometric reasoning stages.

OCR-guided denoising

Prior to semantic segmentation, an OCR-guided denoising step is applied to suppress annotation-induced artifacts that may interfere with geometric extraction. OCR outputs provide spatial and directional cues that localise regions associated with dimensional annotations. Within these regions, subtraction-based operations are applied to remove annotation lines and leaders that are geometrically aligned with detected text.

To further reduce non-structural artifacts, a thin-line removal procedure is applied to suppress auxiliary reference lines, including grids and leaders, which are typically rendered with thinner strokes than architectural components. Morphological and geometric filtering is used conservatively to preserve wall and opening continuity. The resulting denoised image retains the structural geometry required for segmentation while significantly reducing false positives caused by annotation-related noise, as shown in Figure 2(d).

Semantic segmentation

Semantic segmentation of architectural components is formulated as a multi-task learning problem that jointly models architectural boundaries and interior spatial regions. The adopted U-net encoder-decoder architecture (Ronneberger et al., 2015) employs a shared convolutional encoder and two task-specific decoder branches, enabling efficient feature sharing across related segmentation tasks.

The first decoder branch focuses on boundary segmentation, producing pixel-level masks for structural elements such as walls and openings. The second branch performs region-level semantic segmentation of interior spaces. To enhance structural consistency, boundary features are incorporated into the region prediction process through a boundary-guided attention mechanism at multiple feature scales. This design helps preserve thin geometric elements and enforces spatial alignment between room regions and their enclosing structures. Direction-aware feature aggregation is applied to better capture dominant architectural orientations.

The network is fine-tuned using the R2V (Liu et al., 2017) and R3D (Liu et al., 2015) datasets, which provide pixel-level annotations for both boundaries and regions. Training employs a multi-task, class-balanced loss to mitigate severe data imbalance between large interior regions and thin structural components. Although both tasks are retained during training to leverage shared structural priors, only the boundary-related outputs, specifically wall and opening masks, are propagated to downstream geometric alignment and BIM reconstruction.

In the proposed pipeline, semantic segmentation serves as a geometric sensing module rather than an end objective. Its role is to provide topology-consistent and geometrically reliable structural primitives that can be aligned with OCR-derived dimensional annotations, thereby supporting engineering-grade geometric reconstruction.

Alignment algorithms

The alignment stage aims to establish explicit and metrically consistent correspondences between dimensional annotations extracted by OCR and the architectural geometry obtained from semantic segmentation. Instead of assuming a global scale factor, the proposed method reconstructs a metric reference grid directly from dimensional annotations embedded in the floorplan drawings. This grid serves as an intermediate geometric representation that bridges raster-level geometry and real-world distances, enabling component-level metric reconstruction.

As illustrated in Algorithm 1, the process begins by analysing OCR outputs, as shown in Figure 2(b), where each detected text instance is represented by its recognised string, bounding box, orientation, and confidence score. High-confidence numeric annotations with consistent orientation are first selected as candidates potentially associated with grid spacing. Drawing-convention-aware spatial heuristics are then applied to identify annotations that likely describe distances between parallel grid lines. Specifically, vertical grid spacings are inferred from horizontally oriented dimension texts whose bounding boxes lie at similar vertical positions, reflecting the common drafting practice in which distances between vertical elements are annotated using horizontal dimension strings placed between the corresponding grid lines.

Candidate annotations are subsequently grouped based on the proximity of their bounding-box coordinates and sorted according to spatial order. Adjacent annotations within each group are interpreted as describing consecutive grid intervals, yielding an ordered set of grid spacings associated with both pixel-domain distances and OCR-derived metric values.

Because OCR outputs may contain recognition errors, a local verification and correction mechanism is introduced prior to global consistency checks. The algorithm estimates a pixel-to-metric conversion ratio by computing the median ratio between OCR-reported distances and their corresponding pixel gaps measured from bounding-box positions or dimension leader extents. Each grid spacing is then re-evaluated by comparing its OCR-derived value with the geometry-inferred distance. When deviations exceed a predefined tolerance, OCR values are locally corrected using geometry-consistent estimates, thereby enforcing metric consistency at the grid-interval level.

After individual grid spacings are verified and corrected, the algorithm detects total grid-length annotations by comparing detected dimension values with cumulative grid lengths. Grid lines are then placed cumulatively according to the corrected distances, forming a globally consistent metric grid in real-world units, as illustrated in Figure 2(c). The semantic map is subsequently scaled along the horizontal and vertical axes to match the grid extent and aligned to the grid coordinate system, as demonstrated in Figures 2(e) and 2(f). Semantic labels for doors, windows, and room regions are transferred to the aligned floorplan, enabling component-level metric interpretation.

The current alignment strategy incorporates several drafting-convention-aware assumptions. It assumes that dimension texts associated with a given grid interval exhibit consistent orientation, preserve relatively regular spatial ordering, and remain sufficiently close to the referenced structural location for local grouping and verification to be meaningful. These assumptions are generally consistent with common engineering drafting practice. However, robustness may decrease for non-standard layouts, such as offset dimension texts, long auxiliary leader lines, overlapping dimension chains, or highly irregular annotation structures.

Algorithm 1: Grid alignment with OCR-guided verification

Input: OCR detections $\mathcal{O} = \{o_i\}_{i=1}^{N}$, $o_i = (s_i, b_i, \theta_i, c_i)$ where s_i is recognized text, $b_i = (x_i, y_i, w_i, h_i)$ is the bounding box, θ_i is text orientation, c_i is confidence; semantic segmentation map $\mathcal{S}$

Output: Metric vertical grid $\mathcal{G}^v$ and grid-aligned geometry

Step 1: Select candidate vertical grid spacings;
Initialize candidate set $\mathcal{C}^v \leftarrow \emptyset$;
foreach $o_i \in \mathcal{O}$ **do**
  if $c_i \geq \tau_c$ *and* $\theta_i = 0$ **then**
    $\mathcal{C}^v \leftarrow \mathcal{C}^v \cup \{o_i\}$; // horizontal text, high confidence

Step 2: Group candidates by vertical alignment;
Partition $\mathcal{C}^v$ into groups $\{\mathcal{C}_k^v\}$ such that $|y_i - y_j| < \tau_y$ for $o_i, o_j \in \mathcal{C}_k^v$; // similar height

Step 3: Extract vertical grid spacings;
Initialize spacing set $\Delta^v \leftarrow \emptyset$;
foreach *group* $\mathcal{C}_k^v$ **do**
  Sort $\mathcal{C}_k^v$ by x coordinate;
  for $j = 1$ **to** $|\mathcal{C}_k^v| - 1$ **do**
    $\delta x_j = |x_{j+1} - x_j|$; // pixel distance between adjacent grid leaders
    $d_j^{ocr} = \mathrm{val}(s_j)$; // OCR-derived spacing value (metric)
    $\Delta^v \leftarrow \Delta^v \cup \{(\delta x_j, d_j^{ocr})\}$; // store pixel gap and OCR distance

Step 4: Verify and correct OCR-derived distances;
Estimate pixel-to-metric ratio $\rho = \mathrm{median}\left(\frac{d_j^{ocr}}{\delta x_j}\right)_{(\delta x_j, d_j^{ocr}) \in \Delta^v}$; // distance per pixel estimated from grid spacings
foreach $(\delta x_j, d_j^{ocr}) \in \Delta^v$ **do**
  $d_j^{px} = \rho \cdot \delta x_j$; // distance inferred from pixel gap
  if $|d_j^{px} - d_j^{ocr}| > \epsilon$ **then**
    $d_j^{ocr} \leftarrow d_j^{px}$; // correct OCR error using geometric consistency

Step 5: Detect total vertical grid length (if available);
Compute cumulative length $L^v = \sum_{(\delta x_j, d_j^{ocr}) \in \Delta^v} d_j^{ocr}$; // sum of verified vertical grid spacings
foreach $o_k \in \mathcal{O}$ **do**
  if $|val(s_k) - L^v| < \epsilon$ **then**
    Mark o_k as total vertical grid length annotation;

Step 6: Construct metric vertical grid and align geometry;
Construct vertical grid $\mathcal{G}^v = \mathrm{cumsum}(\{d_j^{ocr}\}_{(\delta x_j, d_j^{ocr}) \in \Delta^v})$; // metric grid positions
Align semantic segmentation map $\mathcal{S}$ to $\mathcal{G}^v$; // assign coordinates to segmented map

Finally, the aligned geometry and associated semantic information are used to instantiate building elements as IFC entities. The resulting BIM preserves the topological relationships derived from semantic segmentation while incorporating explicit, grid-consistent dimensional information. By using the reconstructed grid as an intermediary between textual annotations and raster geometry, the proposed alignment strategy reliably recovers absolute measurements without relying on heuristic global scaling, thereby providing a robust metric foundation for accurate 2D-to-BIM reconstruction and downstream BIM-based analysis.

Results and discussion

Quantitative results

The quantitative evaluation addresses three complementary aspects: OCR recognition performance, semantic segmentation accuracy, and engineering-level BIM reconstruction quality.

For OCR evaluation, commonly used metrics in computer vision and document analysis include the character error rate, CER, and string similarity measures such as the FuzzyWuzzy ratio. These metrics assess character-level recognition accuracy and are sensitive to minor transcription errors, making them suitable for algorithm tuning and benchmarking. However, they are not well suited to engineering-oriented reconstruction tasks. In the context of dimensional annotations, even a single-digit error may cause substantial geometric distortion in the reconstructed BIM, making partial string similarity inadequate for downstream evaluation.

Therefore, a strict OCR accuracy metric is adopted to reflect whether recognised dimension values are directly usable for geometric reconstruction. OCR accuracy is defined as:

$$OCR\ accuracy = \frac{N_{correct}}{N_{GT}} \quad (1)$$

where $N_{correct}$ is the number of correctly recognised dimension annotations and N_{GT} is the total number of ground-truth dimension annotations. A prediction is considered correct only if the numeric value is fully matched. Under this strict criterion, the OCR module achieves an accuracy of 90.2% on the test set and 85.7% on the ResBIM dataset, indicating stable cross-dataset generalisation.

Semantic segmentation performance is evaluated using the IoU metric:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

where A and B represent the predicted and ground-truth masks, respectively. The segmentation model achieves an IoU of 92.1% for wall and opening classes, demonstrating reliable extraction of structural geometry from denoised raster floorplans.

IFC-based evaluation

Although OCR accuracy and IoU are standard evaluation metrics in computer vision, they do not directly reflect reconstruction effectiveness at the engineering level, where the geometric correctness of BIM elements is critical. To address this gap, IFC-based evaluation metrics are introduced to assess reconstruction quality at the object and parameter levels.

The ResBIM dataset provides paired annotated floorplans and corresponding reference BIM models, enabling one-to-one comparison between reconstructed and ground-truth elements. Both reconstructed and reference BIMs are converted into IFC format, which provides a platform-independent representation in which each architectural element is encoded as an object with placement and geometric parameters. This object-centric representation enables direct and interpretable element-level comparison. Although this dataset supports rigorous IFC-level comparison, it does not cover the full range of drafting conventions encountered in practice. Therefore, the present evaluation should be interpreted as an initial engineering-level validation under the currently available benchmark setting.

Because the input data consist of 2D floorplans, the evaluation focuses on planar reconstruction accuracy. For walls, the origin location corresponds to the start point of the wall baseline, as shown in Figure 3(a), and the length is defined by the baseline length encoded in the IFC geometry. For doors and windows, the origin is defined by the opening base point, as shown in Figure 3(b), and dimensional parameters are retrieved from the corresponding IFC geometric definitions.

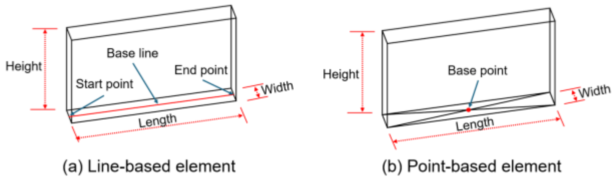


Figure 3: Two methods for representing element locations in IFC-based BIM models in the ResBIM dataset

To comprehensively evaluate geometry-accurate 2D-to-BIM reconstruction at the engineering level, three complementary IFC-based metrics are proposed to assess numerical completeness, dimensional correctness, and strict geometric accuracy.

The count error Err_{cnt} measures numerical consistency between reconstructed and ground-truth models.

$$Err_{cnt} = \frac{|N_{pred} - N_{gt}|}{N_{gt}} \quad (3)$$

where N_{pred} and N_{gt} denote the numbers of reconstructed and ground-truth IFC elements, respectively.

The dimension-level precision metric $IFC_{\mathrm{dim}\_precision}$ evaluates whether element dimensions are correctly reconstructed, independent of absolute location:

$$IFC_{dim\_precision} = \frac{TP_{dim}}{TP_{dim} + FP_{dim}} \quad (4)$$

where TP_{dim} refers to elements whose dimensional parameters are reconstructed within a predefined tolerance, and FP_{dim} corresponds to elements whose dimensions exceed this threshold. This metric evaluates dimensional correctness independently of absolute spatial alignment and is particularly sensitive to OCR and dimension-geometry association errors.

Finally, the strict geometric precision metric $IFC_{precision}$ assesses whether both dimensional parameters and origin locations are simultaneously correct:

$$IFC_{precision} = \frac{TP}{TP + FP} \quad (5)$$

where TP denotes elements whose dimensions and origin locations are simultaneously reconstructed within tolerance, and FP denotes elements failing either criterion. This metric reflects whether reconstructed elements are directly usable for downstream BIM applications without manual correction.

A tolerance of 1 mm is applied to wall dimensions and locations only to absorb minor numerical inconsistencies introduced by IFC export, unit normalisation, and floating-point conversion, rather than to relax geometric evaluation. For openings, a larger tolerance of 20 mm is adopted as a benchmark-specific evaluation threshold. In ResBIM, openings are annotated by centre-point distances rather than explicit physical widths, so opening dimensions and positions must be inferred indirectly from raster geometry and host-wall context before being converted into IFC opening parameters. This introduces higher annotation and representation uncertainty than for walls, whose baselines and lengths are more directly defined. Because openings are also smaller elements, they are more sensitive to minor segmentation or alignment shifts. The 20 mm threshold therefore accounts for annotation granularity and benchmark-specific uncertainty, rather than implying a looser engineering tolerance.

Experimental results on a randomly selected subset of 200 floorplans from ResBIM are presented in Table 1. The count error Err_{cnt} is 2.45%, indicating high numerical consistency between reconstructed and ground-truth models. At the IFC entity level, the dimension-level precision $IFC_{\mathrm{dim}\_precision}$ is 82.4%, and the strict geometric precision $IFC_{precision}$, which requires both correct dimensions and locations, is 71.1%. Notably, in 62 of the 200 floorplan reconstruction tasks, all building components are correctly predicted in both location and dimensional attributes. The average processing time per floorplan is approximately 6.5 minutes.

Table 1: Quantitative evaluation results of the proposed dimension-aware 2D-to-BIM reconstruction framework

Metric	Value
Err_{cnt}	2.5%
$IFC_{\mathrm{dim}\_precision}$	82.4%
$IFC_{precision}$	71.1%
Exact dimension match (count)	62/200
Total processing time	21.5 hours

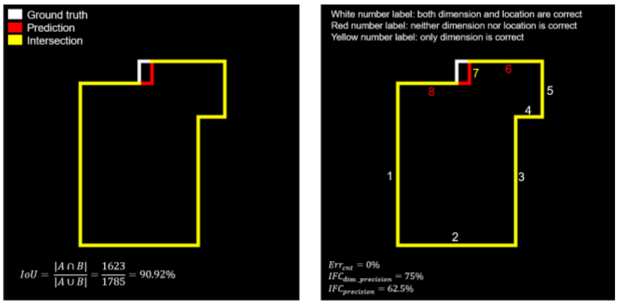
Error analysis and discussion

This study evaluates the proposed vision-based 2D-to-BIM reconstruction framework using both conventional computer vision metrics and the proposed IFC-based evaluation metrics. The OCR module achieves 90.2% accuracy on the in-domain test split and 85.7% on the ResBIM dataset for external validation, while the semantic segmentation model achieves an IoU of 92.1%. These results indicate strong performance when each visual task is evaluated independently using standard computer vision criteria.

However, when the same outputs are assessed using IFC-based reconstruction metrics, a noticeable performance drop is observed. The dimension-level accuracy and strict geometric precision reach only 82.4% and 71.1%, respectively, even after applying the pixel-level correction strategy described in Algorithm 1. This discrepancy reflects a fundamental difference between computer-vision-oriented evaluation and engineering-oriented BIM reconstruction requirements.

Conventional metrics such as OCR accuracy and IoU evaluate predictions in isolation, assessing each text string or pixel region independently. In contrast, BIM reconstruction is inherently relational and connectivity-driven. Walls act as hosts for openings, adjacent walls are expected to connect seamlessly, and local geometric inconsistencies can propagate through topological relationships. As a result, dimensional values extracted by OCR are not independent in the reconstruction context but mutually constraining.

Figure 4 provides a representative qualitative example of this discrepancy. In this case, the predicted wall mask achieves an IoU of 90.92% against the ground truth, and the reconstructed model contains the correct number of wall entities, 8 versus 8. However, only 6 of the 8 wall entities are dimensionally correct, yielding a dimension-level accuracy of 75%, and only 5 of the 8 are correct in both dimension and location, yielding a strict geometric precision of 62.5%. This example illustrates why strong image-level overlap does not necessarily translate into equally high IFC-level correctness.



(a) Pixel-based exterior-wall prediction result and IoU (b) Vectorised exterior-wall and IFC-based metric calculation

Figure 4: Representative exterior-wall example illustrating the difference between conventional computer vision metrics and engineering-oriented IFC metrics.

IFC-based evaluation is stricter because it requires each reconstructed entity to satisfy dimensional correctness, spatial placement, and geometric consistency simultaneously. Therefore, small OCR assignment errors or minor boundary shifts that have little influence on OCR accuracy or IoU may still cause reconstructed IFC entities to be judged incorrect once their baselines, origins, or dimensions exceed the evaluation tolerance. These effects are further amplified during BIM instantiation because local errors do not remain isolated at the image level but propagate through dependent geometric relationships. This explains both the drop from the higher OCR and IoU scores to lower IFC-based metrics and the additional gap between dimension-level precision, 82.4%, and strict geometric precision, 71.1%. In other words, conventional computer vision metrics tend to overestimate reconstruction quality in dimension-sensitive BIM workflows because they do not capture entity-level placement errors and topology-dependent error propagation.

Limitations

Despite its advantages, the proposed pipeline has several limitations. First, the alignment module relies on drafting-convention-aware spatial assumptions, including locally interpretable annotation placement, consistent text orientation, and regular interval ordering. Its robustness may therefore decrease for drawings with highly irregular annotation layouts, overlapping dimension chains, or long auxiliary leader lines that displace dimension texts far from the referenced geometry. In addition, the alignment algorithm degrades noticeably when OCR accuracy falls below approximately 50%, as may occur in highly cluttered or severely occluded drawings. To maintain computational efficiency, the semantic segmentation model also operates on downsampled inputs, which may amplify geometric distortions in the semantic map and reduce alignment quality.

Second, IFC-based evaluation is constrained by the limited availability of datasets that provide paired floorplans and reference BIM models (Pizarro et al., 2022; Liang et al., 2026). ResBIM enables direct object-level comparison, but relying primarily on a single benchmark limits the diversity of drafting conventions and annotation styles represented in the evaluation. Accordingly, the reported results should be interpreted as benchmark-based evidence rather than a complete demonstration of robustness across all real-world floorplan types.

Third, the current study is limited to planar reconstruction from floorplans and does not recover elevation-related dimensions or richer semantic attributes from multi-view drawings. In addition, although the pipeline automates several reconstruction stages, the present work does not directly quantify labour or time savings through a user study or a comparison against manual or semi-automatic workflows.

Conclusions

This paper presents a multi-stage, vision-based framework for the automated analysis of annotated architectural floorplans to enable dimension-aware 2D-to-BIM reconstruction. The main methodological contribution is the proposed cross-modal text–geometry alignment strategy, which reconstructs a metric reference grid from dimensional annotations and supports component-level metric interpretation. A second key contribution is the IFC-based evaluation framework, which enables direct object-level assessment of reconstruction quality from an engineering perspective.

The results show that conventional computer vision metrics, such as OCR accuracy and segmentation IoU, are insufficient for characterising engineering-level BIM reconstruction quality. Although the proposed workflow supports automated, dimension-aware reconstruction from annotated raster floorplans, the present study remains limited to planar, benchmark-based validation and does not directly quantify labour or time savings through user studies or comparisons with manual or semi-automatic workflows.

Future directions

Several limitations remain and motivate future research. First, the current framework focuses on planar reconstruction from floorplans and does not recover elevation-related dimensions or richer semantic attributes from multi-view drawings. Future work should therefore incorporate additional drawing types, such as elevations and sectional views, to support more complete BIM reconstruction.

Second, the present alignment strategy remains partly rule-based and relies on drafting-convention-aware assumptions. Learning-based alignment methods trained on annotated geometry–text correspondences may improve robustness to OCR errors, irregular annotation layouts, and heterogeneous drawing styles.

Third, the current IFC-level evaluation is constrained by the limited availability of paired floorplan–BIM datasets. Extending the evaluation to more diverse real-world drawings and additional benchmarks will be important for testing generalisability across different drafting conventions and building standards.

Together, these directions support more comprehensive and reliable automation of existing-building digitisation in real-world AEC workflows.

Acknowledgements

This research was partially supported by the Telecommunications Advancement Foundation's overseas travel expense assistance.

References

Baek, Y., Lee, B., Han, D., Yun, S. & Lee, H. (2019) Character Region Awareness for Text Detection. arXiv [Preprint].

Dixit, M.K., Venkatraj, V., Ostadalimakhmalbaf, M., Pariafsai, F. & Lavy, S. (2019) Integration of facility management and building information modeling (BIM). Facilities, 37(7/8), pp. 455–483.

El Hafdaoui, H., Khallaayoun, A. & Ouazzani, K. (2023) Activity and efficiency of the building sector in Morocco: A review of status and measures in Ifrane. AIMS Energy, 11(3), pp. 454–485.

Hu, S., Wu, W., Su, R., Hou, W., Zheng, L. & Xu, B. (2024) Raster-to-Graph: Floorplan Recognition via Autoregressive Graph Prediction with an Attention Transformer. Computer Graphics Forum, 43(2).

Jamieson, L., Moreno-Garcia, C.F. & Elyan, E. (2025) Towards fully automated processing and analysis of construction diagrams: AI-powered symbol detection. International Journal on Document Analysis and Recognition, 28(1), pp. 71–84.

Kalervo, A., Ylioinas, J., Häikiö, M., Karhu, A. & Kannala, J. (2019) Cubicasa5k: A dataset and an improved multi-task model for floorplan image analysis. In Image Analysis: 21st Scandinavian Conference, SCIA 2019, Norrköping, Sweden, June 11–13, 2019, Proceedings 21. Springer, Cham, pp. 28–40.

de las Heras, L.-P., Terrades, O.R., Robles, S. & Sánchez, G. (2015) CVC-FP and SGT: a new database for structural floor plan analysis and its groundtruthing tool. International Journal on Document Analysis and Recognition (IJDAR), 18, pp. 15–30.

Lee, H., Jang, S., Lee, J., David Jeong, H. & Lee, G. (2025) Automated BIM generation from inconsistent multi-view raster architectural drawings with missing dimensions. SSRN Preprint [Preprint].

Liang, X., Yabuki, N. & Fukuda, T. (2026) Fully automated synthetic BIM dataset generation using a deep learning-based framework. Automation in Construction, 181, p. 106584.

Lim, J., Janssen, P. & Stouffs, R. (2018) Automated generation of BIM models from 2D CAD drawings. In CAADRIA 2018 - 23rd International Conference on Computer-Aided Architectural Design Research in Asia: Learning, Prototyping and Adapting. The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), pp. 61–70.

Liu, C., Schwing, A.G., Kaustav Kundu, ⇤, Urtasun, R. & Fidler, S. (2015) Rent3D: Floor-Plan Priors for Monocular Layout Estimation. In In Computer Vision and Pattern Recognition (CVPR), Boston, 2015.

Liu, C., Wu, J., Kohli, P. & Furukawa, Y. (2017) Raster-to-vector: Revisiting floorplan transformation. In Proceedings of the IEEE International Conference on Computer Vision, pp. 2195–2203.

Lu, Q., Chen, L., Li, S. & Pitt, M. (2020) Semi-automatic geometric digital twinning for existing buildings based on images and CAD drawings. Automation in Construction, 115, p. 103183.

Lu, Q. & Lee, S. (2017) Image-Based Technologies for Constructing As-Is Building Information Models for Existing Buildings. Journal of Computing in Civil Engineering, 31, p. 4017005.

Pizarro, P.N., Hitschfeld, N., Sipiran, I. & Saavedra, J.M. (2022) Automatic floor plan analysis and recognition. Automation in Construction, 140, p. 104348.

Pokuciński, S. & Mrozek, D. (2024) Methods and Applications of Space Understanding in Indoor Environment—A Decade Survey. Applied Sciences (Switzerland). Multidisciplinary Digital Publishing Institute (MDPI).

Ronneberger, O., Fischer, P. & Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241.

Schönfelder, P., Aziz, A., Faltin, B. & König, M. (2023) Automating the retrospective generation of As-is BIM models using machine learning. Automation in Construction, 152, p. 104937.

Schönfelder, P., Stebel, F., Andreou, N. & König, M. (2024) Deep learning-based text detection and recognition on architectural floor plans. Automation in Construction, 157, p. 105156.

Shi, B., Wang, X., Lyu, P., Yao, C. & Bai, X. (2016) Robust Scene Text Recognition with Automatic Rectification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4168–4176.

Su, J.-W., Tung, K.-Y., Peng, C.-H., Wonka, P. & Chu, H.-K. (2023) SLIBO-Net: Floorplan Reconstruction via Slicing Box Representation with Local Geometry Regularization. In Advances in Neural Information Processing Systems, 2023.

Sun, Q., Fang, C., Liu, S., Sun, Y., Shang, Y. & He, Y. (2025) PolyGraph: A Graph-Based Method for Floorplan Reconstruction From 3D Scans. IEEE Transactions on Visualization and Computer Graphics, 31(10), pp. 7350–7362.

Wang, S. & Pajarola, R. (2025) Eliminating Rasterization: Direct Vector Floor Plan Generation with DiffPlanner. IEEE Transactions on Visualization and Computer Graphics, 31.

Xu, Z., Jha, N., Mehadi, S. & Mandal, M. (2024) Multiscale object detection on complex architectural floor plans. Automation in Construction, 165.

Zhao, Y., Deng, X. & Lai, H. (2021) Reconstructing BIM from 2D structural drawings for existing buildings. Automation in Construction, 128, p. 103750.