2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

TRAINING DATASET COMPOSITION STRATEGIES FOR IMPROVING FLOOR PLAN AND ELEVATION RECOGNITIONN

Hyunjun Lee[1], Suhyung Jang[1], and Ghang Lee[1,2]
[1]Department of Architecture and Architectural Engineering, Yonsei University, Republic of Korea
[2]Institute for Advanced Studies, Technical University of Munich, Germany

Introduction

Architectural drawings play a key role in representing the spatial organization and structural features of buildings (Liu et al., 2017). Floor plans illustrate the layout of architectural components from a top-down view, while elevations present the vertical context of buildings without perspective distortion. These differences reflect the distinct visual features emphasized by each type: floor plans focus on object-level precision, whereas elevations balance object-level detail with overall building context. Legacy data for many existing buildings remain as rasterized 2D architectural drawings rather than BIM models (Zhao et al., 2021), so a deep learning–based automated recognition process is essential for digitizing them.

In large-scale drawings, dense clusters of small elements challenge deep learning models to capture both fine-grained object details and overall context simultaneously. To address this challenge, image tiling has become a common strategy, dividing large images into smaller patches to improve recognition of localized features. While image tiling strategy improves general image recognition, its effect on architectural drawing recognition remains underexplored (Jeong and Kim, 2022; Unel et al., 2019).

The study investigates how different dataset composition strategies influence model performance in architectural drawing recognition. Specifically, we evaluate three dataset types: original, tiled, and combined. We train an instance segmentation model and compare its performance on floor plans and elevations. This analysis identifies optimal dataset strategies for each drawing type by showing how local and global feature balance affects recognition accuracy, thereby improving automated recognition of large-scale drawings.

Methodology

We used architectural drawing image dataset from 'The Open AI Dataset Project (AI-Hub, South Korea)', with a resolution of 4,963×3,509 pixels. From this dataset, floor plan and elevation images were extracted and annotated for three classes: wall, door, and window. To investigate the effect of dataset design, three dataset types were constructed: original (full images), tiled (fixed-size patches), and combined (original + tiled). Floor plans were tiled into 1,500×1,500 patches and elevations into 1,024×1,024 patches, with 10% overlap to capture objects at tile boundaries. Without overlap, context loss could lead to missed or fragmented detections. Examples of the dataset are shown in Figure 1, and the dataset configurations are provided in Table 1.

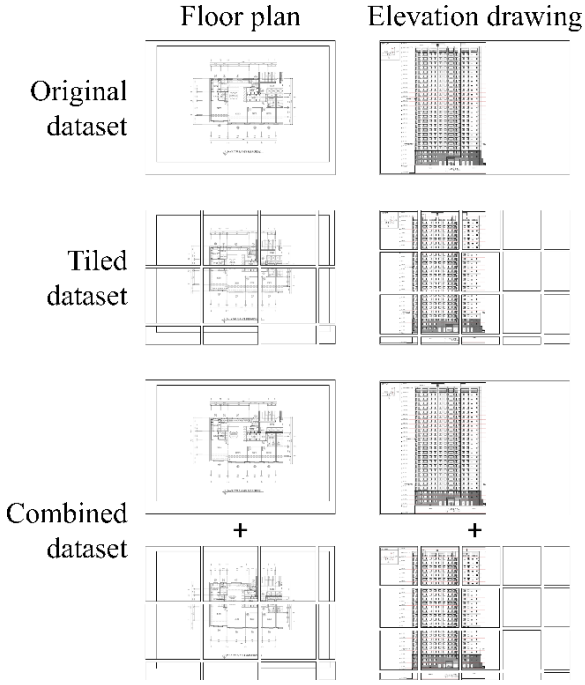

Figure 1: Examples of architectural drawing datasets

YOLO11s-seg, the latest You Only Look Once (YOLO) instance segmentation model (Redmon et al., 2016), was trained separately for each dataset. Focal Loss was used to address class imbalance, and the AdamW optimizer (Loshchilov, I., 2017) was employed with a learning rate of 0.01 and batch size of 16 for 100 epochs. The model is evaluated using metrics in object recognition: Precision, Recall, mAP@0.5, and mAP@0.5:0.95. These metrics assess the model's ability to detect and classify individual architectural components accurately across varying dataset configurations.

Table 1: Training dataset configuration

Drawing type	Dataset	Training	Validation
Floor plans	Original	7,442	1,860
	Tiled	50,792	12,698
	Combined	58,234	14,558
Elevations	Original	1,147	288
	Tiled	12,445	3,112
	Combined	13,592	3,400

Research findings

The evaluation results revealed clear performance differences among the three dataset types: original, tiled, and combined. For floor plan recognition, the model trained on the tiled dataset showed the highest performance, achieving a precision of 0.899, recall of 0.857, mAP@0.5 of 0.908, and mAP@0.5:0.95 of 0.6 (Table 3). This indicates that tiling enhances local feature extraction by enlarging the relative size of small architectural objects within each tile, allowing the model to detect small components more effectively than when trained on full images. In contrast, the original dataset, which preserves global context but lacks focus on small objects, performed worse. The combined dataset slightly improved over the original but underperformed compared to tiling alone, possibly due to inconsistent feature representations.

Table 2: Floor plan recognition results

Dataset type	Precision	Recall	$mAP_{0.5}$	$mAP_{0.5:0.95}$
original	0.632	0.519	0.507	0.218
tiled	**0.899**	**0.857**	**0.908**	**0.6**
combined	0.625	0.524	0.512	0.222

For elevation recognition, the best performance was achieved with the combined dataset (Table 4). The model trained on this dataset achieved a precision of 0.906, recall of 0.764, mAP@0.5 of 0.819, and mAP@0.5:0.95 of 0.606. Elevations typically depict both the overall building façade and small components like windows. By training on both original and tiled images, the model effectively learns global and local features. The original dataset underperformed due to insufficient detail resolution, while the tiled dataset, although emphasizing local features, lacked vertical context of building, making the combined approach most effective.

Table 3: Elevation recognition results

Dataset type	Precision	Recall	$mAP_{0.5}$	$mAP_{0.5:0.95}$
original	0.577	0.344	0.355	0.165
Tiled	0.847	0.710	0.779	0.508
Combined	**0.906**	**0.764**	**0.819**	**0.606**

These findings emphasize that dataset composition strategies must be tailored to drawing types. Floor plans benefit from local detail emphasis through tiling, while elevations require a balanced mix of global and local features. Thus, aligning dataset composition with each drawing's visual characteristics maximizes model effectiveness, improves 2D-to-BIM conversion quality by bridging rasterized archives and digital twins, and enables practitioners to tailor dataset design for specific project goals.

However, the study has limitations. The tiling was done with a fixed size, assuming uniform image dimensions. Additionally, the original dataset had significantly fewer samples, which have biased results since larger datasets typically yield better performance. Lastly, the study used only one recognition model, limiting generalizability.

To address these limitations, future work will explore adaptive, content-aware tiling that adjusts patch size to image characteristics, assemble larger and more balanced datasets, and benchmark multiple deep learning models.

References

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).

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

Jeong, D., & Kim, J. (2022, December). Road damage detection using yolo with image tiling about multi-source images. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 6401-6406). IEEE.

Ozge Unel, F., Ozkalayci, B. O., & Cigla, C. (2019). The power of tiling for small object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 0-0).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Loshchilov, I. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.