Voxel-Based YOLO Framework for Classifying Korean Traditional Wooden Architectural Components

Minseo Kim[1], Nasrulloh Yahyozoda[1], Yongsik Lee[2], Seungjae Lee[2], Seungjoon Yang[1,*]
[1]UNIST, [2]ETRI
{mskim620, nasrulloh, *syang}@unist.ac.kr, {ys.lee, seungjlee}@etri.re.kr

한국 전통 목조 건축 부재 분류를 위한 복셀 기반 YOLO 프레임워크

김민서 [1], Nasrulloh Yahyozoda[1], 이용식 [2], 이승재 [2], 양승준 [1,*]
[1] 울산과학기술원, [2] 한국전자통신연구원

Abstract

Korean traditional wooden architecture, valued for its unique cultural significance, faces the risk of loss due to aging and limited restoration resources. This study proposes a YOLO-based 3D classification framework to categorize structural components from gongpo system using voxelized point clouds. By extending YOLO to 3D space, the proposed method enables accurate and automated classification of architectural elements, providing a scalable approach for preserving and revitalizing Korea's wooden heritage.

Ⅰ. Introduction

Korean traditional wooden architecture, exemplified by temples, palaces, and hanok houses, reflects Korea's unique cultural identity but faces challenges of degradation and insufficient documentation. Traditional survey methods capture limited detail, while 3D scanning technologies generate rich point clouds that remain difficult to interpret. To bridge this gap, we adapt YOLO, a real-time object detection model, for 3D classification of architectural components. Applied to voxelized data, the model identifies structural elements crucial for analysis and reconstruction. This approach contributes to cultural preservation by enabling efficient digital archiving and supporting future restoration of Korea's wooden heritage.

Ⅱ. Proposed Method

This study is based on a dataset composed of components from the gongpo system—such as gwihandae, dori, bo, salmi, and soro—which plays a critical role in supporting and distributing roof loads in Korean traditional wooden architecture. While previous research has explored AI-based object detection of gongpo components from 2D architectural drawings [1], this study extends into 3D space. Specifically, the dataset consists of CAD models of gongpo components, from which point clouds were generated through random sampling. To address the irregularity of point cloud, we adopted voxelization, converting each point cloud into a fixed-size occupancy voxel grid representation, regardless of color.

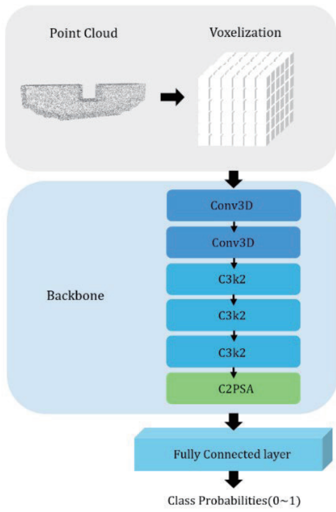

Figure 1. Overall pipeline of the proposed framework

Voxelization was performed with a grid resolution of 128^3, which provided a balance between preserving fine geometric details and maintaining computational feasibility. Each voxel grid was normalized to fit within a unit cube to ensure scale invariance across architectural components. Data augmentation strategies, including random rotations, scaling, and translation, were applied to increase robustness against viewpoint variations.

Although YOLO (You Only Look Once) [2] is primarily developed for 2D object detection, its efficiency and real-time capability make it attractive for 3D architectural preservation tasks. In our approach, we adapted a YOLO-based classification framework to voxelized point cloud data, Figure 1.

Voxel grids were treated as 3D occupancy tensors of shape [B,1,128,128,128], where B denotes batch size. Binary occupancy values (0/1) were used to reduce computational overhead and to emphasize structural geometry.

To adapt YOLO for voxelized point clouds, we employed a YOLOv11-inspired backbone [3] where all convolutions were extended to 3D. The backbone begins with two downsampling convolutional layers (64 and 128 channels), followed by stacked C3k2 residual blocks (256–1024 channels) that capture increasingly abstract geometric patterns. At the deepest stage, a C2PSA (Cross-Stage Partial Self-Attention) block refines high-level features by emphasizing discriminative structural cues.

The detection head was replaced with a lightweight classification head, consisting of 3D global average pooling, a fully connected layer, and a softmax activation to produce categorical probabilities across 9 architectural classes. This design allows efficient feature reuse, multi-scale abstraction, and robust classification of both coarse structural elements and fine details. The classification process can be formulated as Eq. (1):

$$p(y \mid X) = Softmax(W_2 \cdot \sigma(W_1 \cdot f(X) + b_1) + b_2) \quad (1)$$

Where f(X) denotes the pooled YOLO backbone features, W_1, W_2 are the classification weights, and σ represents the ReLU activation.

The normalized confusion matrix in Figure 2 shows that the voxel-based YOLO framework performs well across most Korean traditional wooden architectural components. Classes such as kidoong, bo, salmi, and changbang achieve perfect accuracy, while gwihandae (0.83) and cheomja (0.82) show minor misclassifications. Most confusion occurs between geometrically similar components—such as gwihandae vs. dori, soro vs. joodu, and soro vs. salmi—reflecting their highly similar voxelized shapes. The background is perfectly separated. Overall, the model effectively distinguishes most structural elements, though subtle geometric overlaps suggest that higher voxel resolution, class-specific augmentations, or additional feature descriptors could further improve performance.

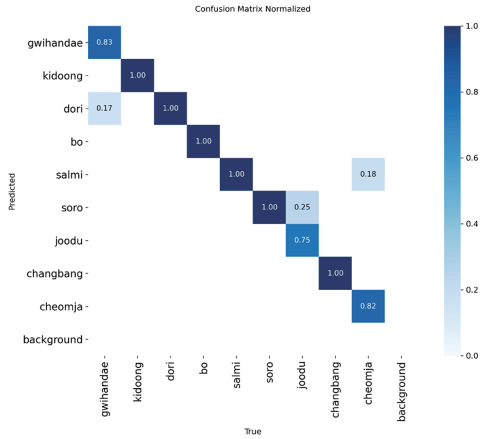

Figure 2. Confusion matrix of the proposed framework on traditional wooden architecture components

Ⅲ. Conclusion

In this study, we proposed a YOLO-based classification framework for voxelized point clouds of Korean traditional wooden architecture. By extending YOLO from 2D object detection into the 3D domain, our approach achieved high accuracy in classifying major structural components such as kidoong, bo, salmi, and changbang. However, it also revealed challenges in distinguishing geometrically similar elements like gwihandae and cheomja. This work demonstrates the potential of deep learning to make complex 3D scan data more interpretable, thereby supporting digital documentation and analysis of traditional architecture. In the broader context of cultural heritage, the proposed framework can contribute to efficient digital archiving and provide a foundation for reconstruction, and educational applications of Korea's wooden heritage.

ACKNOWLEDGMENT

This work was supported by the 2025 Cultural Heritage Smart Preservation & Utilization R&D Program of Korea Heritage Service, National Research Institute of Cultural Heritage (Project Name:Development of AI-based analysis and design technology of traditional wooden architecture, Project Number: RS-2024-00396158, contribution rate 70%), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00359358,contribution rate 30%).

REFERENCES

[1] J. Lee *et al.*, " A Study on The Relationship between the Drawing Representation and the Gongpo Recognition Process in AI Object Detection of Jusimpo Gongpo", Proceedings of the AIK conference Volume 42, No. 2, 2022.

[2] J. Redmon *et al.*, "You Only Look Once: Unified, Real-Time Object Detection", CVPR 2016.

[3] R. Khanam *et al.*, "YOLOv11: An Overview of the Key Architectural Enhancements", 2024. (https://arxiv.org/abs/2410.17725).