Jordan Journal of Civil Engineering, Volume **20**, No. **4**, **2026**
DOI: https://doi.org/10.14525/JJCE.v20i4.04

Jordan Journal of Civil Engineering

Journal homepage: https://jjce.just.edu.jo

Digital Protection of Rural Architectural Heritage by Integrating Computer Vision and Big Data Analysis

Chen Chen [1)*]***, Zhangying Chen*** [1)]

[1)] *School of Business Management, Hangzhou Polytechnic, Hangzhou, 311400, China.*
** Corresponding Author. Email: cclouis11@163.com*

ARTICLE INFO

Article History:
Received: 12/8/2025
Accepted: 6/6/2026

ABSTRACT

With the rapid development of the social economy, the survival of rural architectural heritage is under threat, and digital protection is particularly important. The aim of this study is to enhance the detection accuracy of building inspection under the digital protection framework. The novelty of this study lies in the domain-specific adaptation and synergistic configuration of existing modules into the improved You Only Look Once, version 8 (YOLOv8) algorithm and the construction of the digital protection framework integrating Principal Component Analysis (PCA) with the enhanced YOLOv8. The average accuracy of the improved YOLOv8 algorithm reaches 98.69%, which is superior to other algorithms. The recognition accuracy rate of this framework for four types of building components is as high as 98.2%, which is superior to similar frameworks. The innovation of this study lies in the framework-level integration of PCA preprocessing with the enhanced YOLOv8 algorithm within a complete digital protection pipeline encompassing multi-source data acquisition, intelligent detection, and three-dimensional modeling. This integration enhances feature extraction, reduces the dimension, and makes the detection system more efficient and accurate. The contributions of this research are in two aspects. On the one hand, it demonstrates through systematic ablation that the synergistic configuration of existing modules achieves superior performance in rural architectural heritage detection, providing a high-performance algorithm that can be applied to similar digital protection frameworks within comparable architectural heritage contexts. On the other hand, it provides a comprehensive digital protection framework, significantly enhancing the accuracy of building inspections, which is crucial for the conservation of rural architectural heritage. However, it is important to note that the study's data collection primarily focused on wooden structures, limiting the generalizability of the results to other building types, such as brick and stone structures. Furthermore, the reported high performance values are achieved under relatively controlled conditions with high-quality data, and the framework's robustness under adverse scenarios, including occlusion, lighting changes, structural degradation, background clutter, noisy scans, and out-of-distribution samples, remains to be systematically evaluated. Additionally, the enhanced computational complexity of the improved YOLOv8 algorithm may hinder real-time application due to increased inference time. These findings provide a solid theoretical foundation and practical approaches for the research on the protection of architectural heritage.

Keywords: Improved YOLOv8, Principal component analysis, Rural area, Architectural heritage, Digital protection.

INTRODUCTION

As an important carrier of human history and culture, the protection of architectural heritage is particularly important. However, with the rapid advancement of urbanization, many Rural Architectural Heritages (RAHs) are gradually disappearing in the tide of the times, and a large number of regional cultures are also facing the risk of gradual loss. Although digital technology has been broadly employed in the realm of RAH protection due to its advantages of integrity, permanence, and efficiency, the current protection methods have the problem of low detection accuracy in building inspection (Muangasame, 2023; Hassanvand et al., 2023). Although many scholars have conducted relevant research, the results are still not satisfactory. In response to the problem of the disappearance of the rich architectural heritage in Michoacán due to social changes and the replacement of traditional skills by modern systems, Sanchez-Calvillo et al. used digital methods, such as audiovisual technology and drone surveys to record and disseminate Michoacán's architectural heritage, but the results were not satisfactory (Sanchez-Calvillo et al., 2023). Kabadayi and Erdoğan proposed a digital technology that combines laser scanning and 360° imaging technology to address the issue of poor preservation of architectural heritage, but it is only at the theoretical level (Kabadayi & Erdoğan, 2022). Of course, there have been some successful studies, such as Qi et al.'s proposal of a digital reconstruction method for architectural heritage protection and restoration based on generative adversarial neural networks to address the issue of heritage preservation. Experimental outcomes revealed that this method was effective (Qi et al., 2024). Guo et al. proposed a laser radar and Unmanned Aerial Vehicle (UAV) mapping technology to enhance the validity and precision of digital protection of tomb sites. Experimental findings revealed that this technology could achieve digitalization of tomb sites (Guo et al., 2022). Regarding the issue of how to use digital technology to protect architectural heritage, Cline et al. proposed a methodological approach. Through comparative analysis, the results showed that by combining different digital representation methods, more comprehensive support could be provided for architectural heritage analysis (Cline et al., 2024).

Finding an effective and accurate method to identify and detect the collected RAH data can help improve the efficiency of digital protection of RAH. You Only Look Once, version 8 (YOLOv8) is a computer vision algorithm that has advantages, such as high efficiency and accuracy, and is widely used in fields, such as construction and manufacturing (Talaat & ZainEldin, 2023). Principal Component Analysis (PCA) is a big data analysis technique that has advantages, such as reducing data dimensionality and improving computational efficiency. It is widely used in fields, such as data analysis and feature extraction (Marukatat, 2023). Many scholars have conducted related research. Chen et al. designed a model that integrates multi-head attention and YOLOv8 to improve the detection accuracy of the characteristics of Chinese Tang Dynasty wooden structures. The experiment outcomes revealed that the accuracy and recall of the model were 95.6% and 85.6% (Chen et al., 2025). In response to the problem of poor real-time crack detection in residential historical buildings, Zhang et al. introduced an improved YOLOv8-n model for real-time monitoring of cracks in wooden structural components of residential historical buildings. The model was compared and analyzed with the original YOLOv8-n model, and the results showed significant improvement (Zhang et al., 2025). Yang et al. proposed an automatic damage detection model that combined YOLOv8 and feature fusion to improve the low accuracy of wall crack detection in RAH. Experimental findings revealed that the model had high detection accuracy (Yang et al., 2025). Turbay et al. introduced a PCA-based evaluation method to address the risks faced by architectural heritage due to human or natural threats. The method was applied to church analysis and the results showed its effectiveness (Turbay et al., 2024). Pang et al. introduced an approach that combined PCA and learning networks to tackle the issue of low efficiency caused by excessive human intervention in architectural heritage protection. Experimental results showed that this method had high robustness (Pang et al., 2025).

Despite these research efforts, their application in RAH protection remains inadequate, particularly in improving the precision of target detection and recognition. Existing approaches still struggle to achieve high-precision identification of specific, complex building components. Moreover, current studies often focus on singular defect types, lacking a comprehensive framework that integrates efficient data processing with a versatile, high-precision algorithm tailored for the diverse elements of RAH. Therefore, the field still lacks a unified system capable of accurately classifying a wide range of architectural components with the required precision. Therefore, to improve the detection accuracy, the study first constructs an improved YOLOv8 algorithm. Subsequently, PCA algorithm is used to perform data dimensionality reduction on the collected multi-source data of RAH. Finally, a digital protection framework for RAH based on improved YOLOv8 and PCA is constructed, aiming to improve the efficiency of RAH protection. The novelty of this study consists in the strategic configuration and synergistic integration of partial convolution, Efficient Multi-scale Attention Module (EMA), Wise-inner-MPDIOU loss function, Down down sampling module, and quadruple detection layer from FasterNetblock to improve YOLOv8 for RAH detection, which is combined with PCA algorithm to offer a theoretical foundation to research in the realm of architectural heritage protection. It should be clarified that Efficient Multi-scale Attention Module, partial convolution, Down downsampling module, quadruple detection layer from FasterNet block, and Wise-inner-MPDIOU loss function are existing techniques adapted from prior studies; the novelty resides in their domain-specific configuration and synergistic integration validated by ablation experiments, not in the individual modules themselves.

MATERIALS AND METHODS

Improved YOLOv8 Algorithm Construction

With the development of science and technology, digital technology is widely used in the protection of architectural heritage. By using computer vision technology to detect and identify RAH data, maintenance support can be provided to relevant personnel, thereby improving protection efficiency. However, the current detection and recognition methods have the problem of low accuracy. Therefore, the study introduces the YOLOv8 algorithm and improves it to enhance its detection and recognition accuracy. Meanwhile, digital technologies, such as drones and 3D laser scanning, are used to collect RAH data, and PCA algorithm is used to reduce dimensionality and extract features from the data. Finally, the dimensionality reduced data is combined with the improved YOLOv8 algorithm to construct a digital protection framework for RAH based on the improved YOLOv8 and PCA. Before constructing a digital protection framework for RAH, it is necessary to improve the YOLOv8 algorithm. YOLOv8 is a computer vision algorithm broadly employed in fields, such as video surveillance and image analysis (Shen et al., 2023). The basic framework of YOLOv8 algorithm is presented in Figure 1 (Sapkorta & Karkee, 2024).

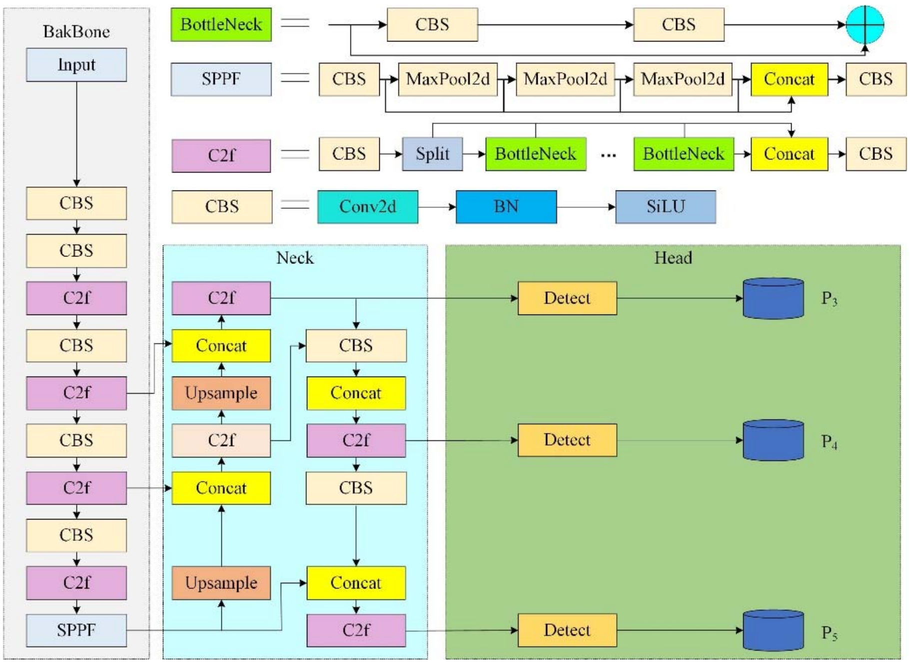


Figure 1. Basic structure of YOLOv8

Note: C2f = Cross Stage Partial Feature; BN = Batch Normalization; SPPF = Spatial Pyramid Pooling Fast; BN = Batch Normalization = CBS = Convolution Batch Normalization Activation.

In Figure 1, Batch Normalization (BN) is used to accelerate training and improve model stability. P3, P4, and P5 are detection layers used to detect targets of different sizes; Two-dimensional convolution (Conv2d) is a convolution operation used to process image data. From Figure 1, the YOLOv8 algorithm first utilizes multiple Convolution Batch Normalization Activation (CBS) and CSP Bottleneck with 2 Convolutions (C2f) layers in the backbone network, as well as the Spatial Pyramid Pooling Fast (SPPF) layer, to extract features from the input data. Then, the features extracted by the backbone network are fused and multi-scale processed through multiple C2f, upsampling layers, and concatenation layers in the neck. Finally, it utilizes multiple detection layers on the head to output the final detection results. However, the YOLOv8 algorithm has a problem of poor accuracy in detecting small targets. To address this issue, the study first utilizes partial convolution and EMA in FasterNetBlock to improve C2f. Partial convolution is a convolution operation used to handle missing data in images. It has the ability to deal with image incompleteness without sacrificing image quality and is widely applied in fields, such as image restoration (Wang et al., 2025). EMA is a technique used to enhance the feature extraction ability of models. It improves the performance of models by focusing on information of different scales in images. EMA has the advantage of capturing the features of images at different scales, which can significantly improve the accuracy of tasks, such as object detection (Luo et al., 2024). The improved C2f (IC2f) is presented in Figure 2.

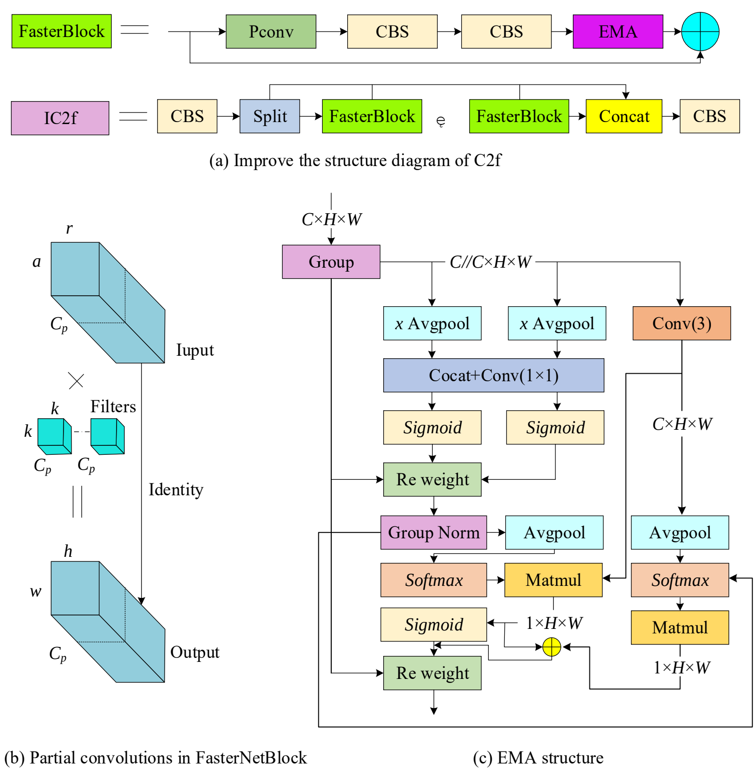


Figure 2. IC2f structure diagram: Feature extraction is optimized by introducing partial convolution and multi-scale attention mechanisms in the YOLOv8 model

In Figure 2, Pconv represents the partial convolution in FasterNetBlock, IC2f is the improved C2f, H and W are the vertical and horizontal dimensions of the image or FM, C represents the number of channels, C_p is the quantity of channels in the partial convolution operation, Conv is the convolution operation, $GroupNorm$ represents group normalization, $matmul$ represents matrix multiplication, and $Re\ wegiht$ represents weight calculation. k is the convolution kernel. From Figure 2 (a), the processing flow of IC2f first uses the CBS module to perform preliminary processing on the input FM, and then divides the FM and performs in-depth processing through multiple parallel Faster Blocks. Each FasterBlock includes partial convolution, CBS module, 1×1 convolution, and EMA to enhance feature extraction capabilities. The processed FMs are concatenated and further fused through another CBS module to ultimately output FMs for object detection. Among them, Pconv sets the proportion of input channels participating in convolution calculation to 1/4 to improve computing speed and performance, and the remaining channels are directly passed through Identity to output without participating in the operation. This process is equivalent to 1/16 of the quantity of ordinary convolution floating-point operations, and its calculation expression is presented in Equation (1).

$$H \times W \times k^2 \times C_p^2 \quad (1)$$

Meanwhile, the memory access is 1/4 of that of a regular convolution, and its calculation expression is presented in Equation (2).

$$H \times W \times 2C_p + k^2 \times C_p^2 \approx H \times W \times 2C_p \quad (2)$$

The process of EMA, as shown in Figure 2 (c), first groups the number of channels of the received FM for parallel processing. Secondly, it performs an average pooling operation on each group. Then, a 1×1 convolutional layer is used for feature fusion to generate attention weights. Before generating attention weights, it is necessary to perform a dimension change operation on the output of the 1×1 convolutional layer to match the dimension of the input FM, which can be represented by Equation (3).

$$Z_c = 1/(H \times W) \sum_i^H X_c\,(i,j) \quad (3)$$

In Equation (3), Z_c is the average value of the cth channel and X_c is the value of the FM at position (i,j). Next, the obtained weight is re-weighted onto the original FM, and the FM is further processed through group normalization and matrix multiplication to enhance the representation ability of the model. Then, the processed FMs are merged through concatenation and fused through a 3×3 convolutional layer. Finally, the results are obtained at the output. The multi-scale feature of EMA enables the model to simultaneously consider the micro and macro information of the image, which is particularly important for handling scenes with complex backgrounds and diverse target sizes. By weighting and normalizing feature maps of different scales, EMA helps suppress unimportant background noise and highlight key target features, thereby improving the discrimination and robustness of features. In addition, the dynamic adjustment capability of EMA enables the model to adaptively adjust the attention distribution according to the changes in input data, which can enhance the model's processing ability in dynamic scenes or when the target is moving rapidly. In addition, to improve the classification ability of YOLOv8 in detection and recognition, Wise-inner-MPDIOU is used as the loss function, and its calculation is presented in Equation (4).

$$L_{total} = L_{Wise-MPDIoU} + IoU - IoU^{inner} \quad (4)$$

In Equation (4), L_{total} is the Wise-inner-MPDIOU loss function, IoU is the traditional intersection-to-union ratio loss employed to measure the degree of overlap between the predicted box (PB) and the real box (RB). IoU^{inner} is the Inner IoU loss, which considers the internal regions of the PB and the RB to enhance the rationality of the regression. $L_{Wise-MPDIoU}$ is the Wise-MPDIOU loss function, which combines IoU and MPDIoU to measure the difference between PBs and RBs. The calculation expression for $L_{Wise-MPDIoU}$ is presented in Equation (5).

$$L_{Wise-MPDIoU} = IoU - d_1^2/(w^2 + h^2) - d_2^2/(w^2 + h^2) \quad (5)$$

In Equation (5), d_1 and d_2 are the lengths between the upper left and lower right corner points of the PB and the RB, w and h are the heights and widths of the maximum bounding box (BB) that merges the PB and the RB. The calculation expression of IoU^{inner} is presented in Equation (6).

$$IoU^{inner} = \frac{(min(b_p^w, b_i)) - max\,(b_p^w, b_i)) * (min(b_p^h, b_i) - max\,(b_p^h, b_i))}{(w^w * h^w) * (ratio)^2 + (w * h) * (ratio)^2 - inter} \quad (6)$$

In Equation (6), b_p^w and b_p^h are the width and height of the inner BBs of the PB and the RB, b_i is the width and height of the inner BB, $inter$ is the intersection area of the inner BBs, and $ratio$ is the scaling factor. Finally, to extract and fuse features more validly and improve the comprehensive detection capability of the model, the ADown downsampling module and quadruple detection layer P2 are introduced to improve YOLOv8. Based on the preceding information, the study will integrate the

IC2f and Wise-inner-MPDIOU loss functions into YOLOv8 to construct an improved YOLOv8 algorithm. The improved YOLOv8 algorithm structure is presented in Figure 3.

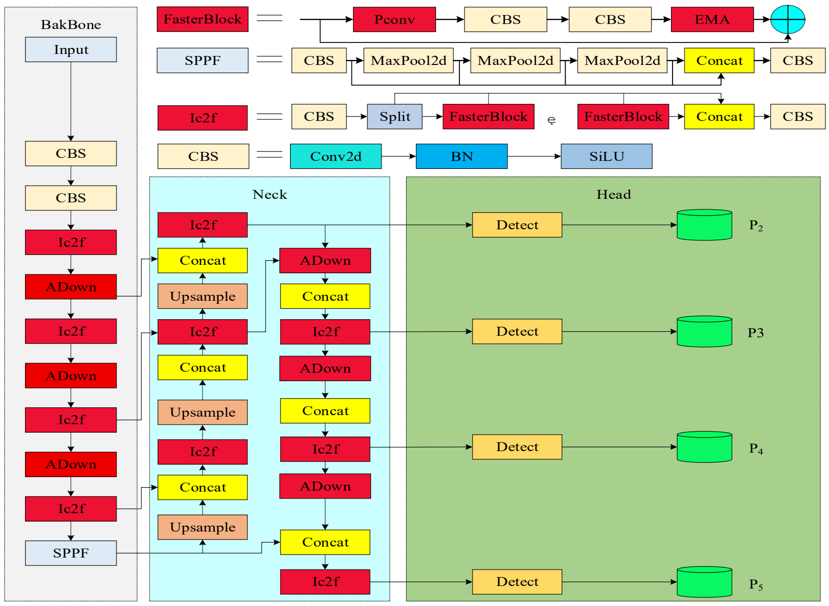


Figure 3. Structure diagram of the improved YOLOv8 algorithm

In Figure 3, the improved YOLOv8 algorithm backbone network consists of 2 CBS, 4 IC2f, 3 ADown downsampling, and 1 SPPF. Among them, CBS is used for preliminary processing of input FMs, IC2f is used to enhance feature extraction capabilities, and ADown downsampling is used to capture various feature information. The neck is composed of 5 IC2f, 6 Concat, 3 Upsample, and 3 ADown downsamples. Among them, Concat is used for FM stitching, and Upsample is used for upsampling. To improve the detection precision of small targets, the model adds a quadruple downsampling detection layer P2 on the original basis. Therefore, the FM resolution of detection layer P2 is 160×160, the FM resolution of P3 is 80×80, the FM resolution of P4 is 40×40, and the FM resolution of P5 is 20×20.

Digital Protection Framework for RAH Based on Improved YOLOv8 and PCA

Due to the complex factors, such as the structure and geometric shape of rural buildings, to ensure the accuracy of the information obtained, it is essential to collect multi-source information before using the improved YOLOv8 algorithm to detect and identify their details. Therefore, a multi-source data acquisition and processing method is designed, as shown in Figure 4.

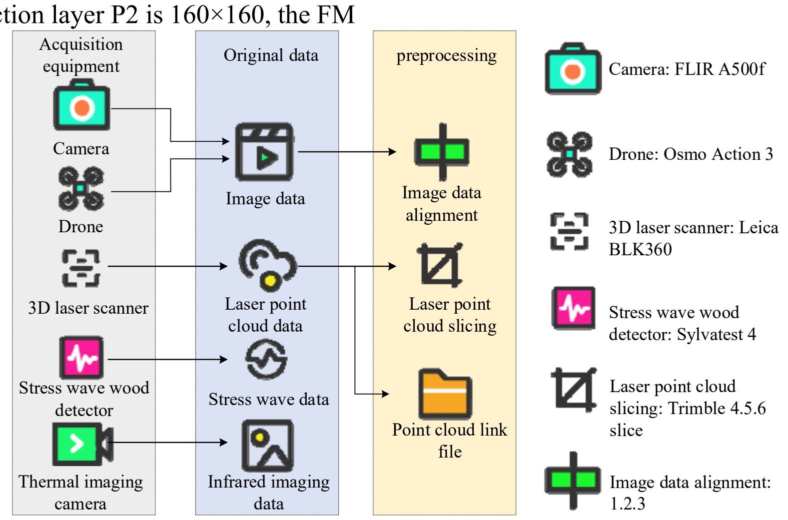


Figure 4. Multi-source data acquisition and processing methods
(icons source from: https://iconpark.oceanengine.com/official)

As shown in Figure 4, this method first uses the Leica BLK360 3D laser scanner to obtain laser point cloud data (LPCD) inside the building. The DJI Osmo Action 3 motion camera captures internal image data of buildings. DJI Mavic 3E drone collects external image data of buildings. Sylvatest 4 ultrasonic wood detector and FLIR A500f thermal imaging camera collect indoor data of buildings. Secondly, RealityCapture software is used to automatically align the image data collected by drones and motion cameras. Trimble software is used to slice the LPCD obtained by the 3D laser scanner. Meanwhile, slice data is preprocessed using methods, such as translation and rotation. Next, roboflow annotation software is used to annotate the sliced point cloud data. The annotation method is based on annotating the abbreviation of the component name as the label name, and obtaining the section and plan datasets. After the annotation is completed, roboflow annotation software is used to rotate and mirror the data for data augmentation. After augmentation, the data is exported to Txt format and packaged to obtain an LPCD file, which is used to improve the YOLOv8 algorithm for detection and recognition. The data collected by Sylvatest 4 ultrasonic wood detector and FLIR A500f thermal imaging camera is kept as a backup. Due to the large scale of building data, algorithms may experience reduced computational efficiency during computation. For this purpose, the PCA algorithm is studied for dimensionality reduction of LPCD. When choosing PCA as the dimensionality reduction method, the study considered its computational efficiency and stability when dealing with large-scale datasets. PCA extracts the principal components of data through linear transformation, making it both fast and effective when dealing with high-dimensional data. Compared with t-SNE, PCA is more efficient in computation. Although t-SNE performs well in visualizing high-dimensional data, its computational cost is relatively high and it is not suitable for large-scale datasets. Compared with LDA, PCA does not rely on category labels, and thus is more applicable in unsupervised learning scenarios. The PCA algorithm process is presented in Figure 5 (Addab & Al Rammahi, 2024).

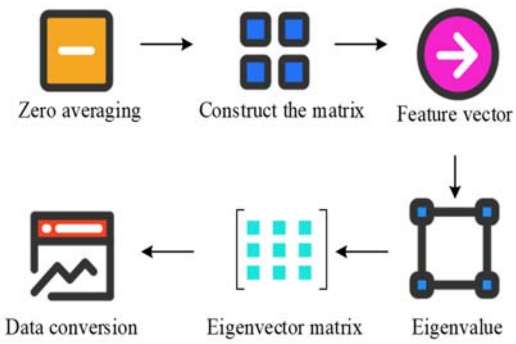


Figure 5. The process of the PCA algorithm
(icons source from: https://iconpark.oceanengine.com/official)

From Figure 5, the PCA algorithm first performs zero mean processing on the data, removing the mean of each feature in the dataset. The calculation expression is presented in Equation (7).

$$Q=\begin{bmatrix} q_{11}-\overline{u_1} & q_{12}-\overline{u_2} & \cdots & q_{1t}-\overline{u_t} \\ q_{21}-\overline{u_1} & q_{22}-\overline{u_2} & \cdots & q_{2t}-\overline{u_t} \\ \vdots & \vdots & \ddots & \vdots \\ q_{o1}-\overline{u_1} & q_{o2}-\overline{u_2} & \cdots & q_{ot}-\overline{u_t} \end{bmatrix} \tag{7}$$

In Equation (7), Q is the zero mean data matrix, o is the total quantity of samples, t is the total quantity of features, q_{ij} is the observation value of the jth feature of the ith sample, and $\overline{u_{ij}}$ is the average value of any feature in the original data matrix. The calculation expression of $\overline{u_i}$ is presented in Equation (8).

$$\overline{u_i}=1/o\sum_i u_{ij} \tag{8}$$

Secondly, the calculation expression for the covariance matrix I is presented in Equation (9).

$$I = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1t} \\ \varphi_{21} & \varphi_{21} & \cdots & \varphi_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{t1} & \varphi_{t2} & \cdots & \varphi_{tt} \end{bmatrix} \tag{9}$$

In Equation (9), $\varphi_{ij} = \varphi_{ji}$ is the covariance between feature i and feature j. The acquisition of the value of φ_{ij} can be represented by Equation (10).

$$\varphi_{ij} = \psi[(Q(:,i) - \psi(Q(:,i)))(Q(:,i) - \psi(Q(:,i))] \tag{10}$$

In Equation (10), ψ is the expectation. Subsequently, the eigenvalues of matrix φ_{ij} and their corresponding eigenvectors are to be calculated and sorted in descending order. Then, the principal components (PCs) are ranked based on the amount of feature information they contain, with the one containing the most features designated as the first PC, followed by the second PC, and so on. This approach enables the identification of all PCs of the data. Assuming that the vector matrix of the first PC is denoted as ψ, the calculation expression is presented as shown in Equation (11).

$$\begin{aligned} &max \tfrac{1}{t}\textstyle\sum_{i=1}^{t} |Q'(:,i) \cdot \psi|^2 \\ &\quad s.t. \psi^T\psi = 1 \end{aligned} \tag{11}$$

In Equation (11), ψ is the first PC. Finally, a projection matrix is constructed, and the original data is projected onto the selected PCs to obtain the dimension-reduced data. Ultimately, by integrating the improved YOLOv8 algorithm, the PCA algorithm, and the collected multi-source data, a digital conservation framework for RAH based on the improved YOLOv8 and PCA is established. This framework is illustrated in Figure 6.

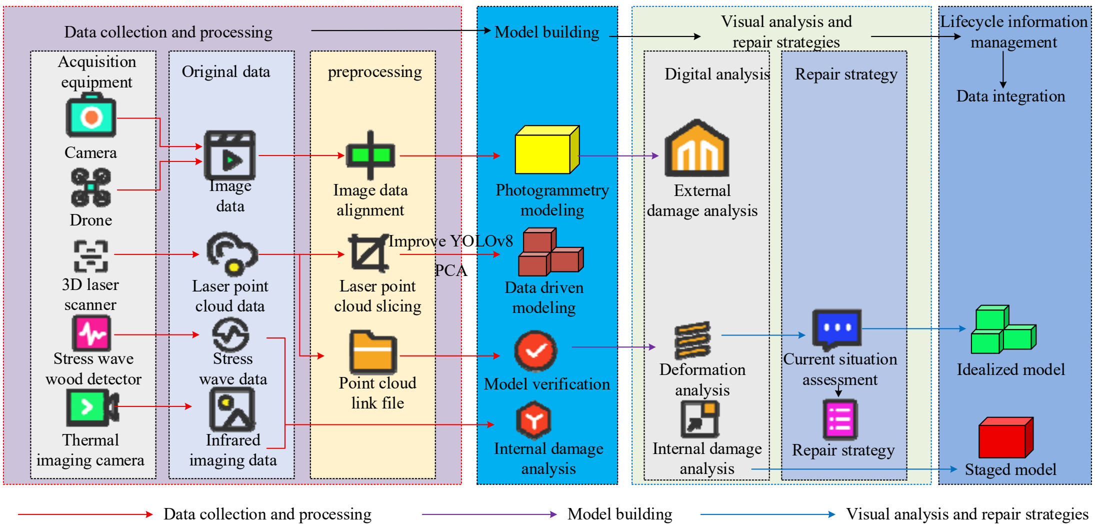


Figure 6. A digital protection framework for RAH based on improved YOLOv8 and PCA (icons source from: https://iconpark.oceanengine.com/official)

From Figure 6, within this framework, following the acquisition and processing of multi-source data, the sliced point cloud data undergoes dimensionality reduction through PCA. Subsequently, it is categorized into a test set and a training set to train and improve the YOLOv8 algorithm for building detection and recognition. Next, digital modeling is conducted, employing the Dynamo tool to establish a three-dimensional spatial topological model based on the aligned data. This model encompasses the building's roof curves, planar column grids, and fundamental component attribute units of ground-supported columns. Concurrently, Revit is utilized to verify the generated spatial model using data from the point cloud data folder. Following this, visual analysis is performed. In the context of visual analysis, external damage analysis involves conducting grid-based comparative analysis using previously measured stress wave and thermal imaging data to assess damage changes in components, such as roofs, walls, and ground bases. Additionally, CloudCompare is employed to analyze deformations in walls and columns based on the model verification results. Subsequently, upon completion of the analysis, the current conditions of column bodies and roof sides are evaluated, and repair strategies are proposed based on the evaluation outcomes. Finally, Revit software is used for information management of the conditions of various internal and external building components at different stages, including their original state, pre-repair, during repair, and post-repair. Through these steps, the digital conservation of RAH is achieved. In addition, to ensure the consistency and synchronization of multi-source data, the measures taken in the research are as follows: Firstly, time synchronization is achieved by using the same timestamp on all data collection devices to ensure the temporal consistency of data collection. Secondly, the RealityCapture software is utilized to automatically align the image data collected by drones and action cameras to ensure spatial consistency. In addition, the roboflow annotation software is used to annotate the sliced point cloud data. The abbreviations of component names are labeled as label names, and the profile and plane datasets are obtained.

Improving YOLOv8 Algorithm Implementation Details

In order to accurately reproduce the improved YOLOv8 algorithm, the research shows its parameter settings, and the specific contents are shown in Table 1.

Table 1. Improved YOLOv8 algorithm parameter settings

Parameter	Value	Remark
Enter image size	640x640	Common input sizes, suitable for most computer vision tasks
Learning rate	0.01	Initial learning rate, suitable for most deep learning tasks
Momentum	0.941	Momentum parameter for the optimizer, which helps speed up convergence
Optimizer	Stochastic Gradient Descent (SGD)	Classic optimizer, suitable for large-scale datasets
Training times	1500	Total number of training iterations, adjusted based on dataset size and model complexity
Weight decay	0.0005	L2 regularization parameter, used to prevent overfitting
Stop early	Enable	Stop training in advance based on the performance of the validation set to prevent overfitting
Batch size	16	Suitable for most hardware configurations
Data augmentation strategy	Random horizontal flipping, random cropping, color dithering	Enhance model generalization capabilities
Confidence threshold	0.5	Used to filter low confidence predictions
IoU threshold	0.5	Used for NMS to determine whether the prediction box is a positive sample
NMS settings	0.5	IoU threshold for NMS
mAP definition	mAP@0.5	Commonly used average accuracy evaluation criteria
Random seed	0	Ensure consistency of model output when running multiple times

In the above settings, mAP is the average accuracy, which is the core indicator of target detection. By calculating the average accuracy (AP) of each category and calculating the average value, a value higher than mAP indicates that the model has excellent performance in multi-category tasks and is the core standard for measuring the comprehensive capabilities of the model.

RESULTS

Performance Analysis of Improved YOLOv8 Algorithm

After constructing the improved YOLOv8 algorithm, comparative analysis experiments were conducted on it. Comparative algorithms included Particle Swarm Optimization Support Vector Machine (PSO-SVM), Convolutional Neural Network-Long

Short-Term Memory (CNN-LSTM), and Genetic Algorithm-Back Propagation Neural Network (GA-BP) algorithms. These traditional machine learning and optimization methods were selected as cross-paradigm benchmarks, because they represent the state-of-the-art approaches previously applied in architectural heritage protection and related detection tasks. Specifically, PSO-SVM combines swarm intelligence optimization with support vector classification, which has been employed for pattern recognition in heritage structure analysis. CNN-LSTM integrates spatial feature extraction with temporal sequence modeling, suitable for processing sequential inspection data. GA-BP leverages genetic algorithm optimization to enhance neural network training, representing evolutionary computation approaches in heritage detection. While these methods are not standard modern object detection baselines, their inclusion demonstrates the substantial performance advancement achieved by the proposed deep learning framework over classical paradigms. The parameter settings of the improved YOLOv8 were as follows: an input image size of 640, a learning rate of 0.01, and momentum of 0.941. The optimizer used stochastic gradient descent (SGD) with 1500 training iterations and weight decay of 0.0005. To prevent overfitting of the model, an early stopping regularization method was adopted. The dataset was sourced from ancient rural buildings in Suzhou, China. The data was collected using Leica BLK360 3D laser scanner, DJI Osmo Action 3DJI Mavic 3E drone, Sylvatest 4 ultrasonic wood detector, and FLIR A500f thermal imaging camera. Data preprocessing was automatically aligned through RealityCapture software. Trimble software was used to slice the point cloud data. Meanwhile, slice data was preprocessed using methods, such as translation and rotation. Next, roboflow annotation software was employed to annotate the sliced point cloud data. After the annotation was completed, roboflow annotation software was used to rotate and mirror the data for data augmentation. After augmentation, the data was exported to Txt format and packaged to obtain an LPCD file. There were a total of 6784 pieces of data, with 70% being the training set, 20% being the testing set, and 10% being the validation set. In addition, Matlab 2022b software was adopted for model verification in the study. The composition of the dataset is shown in Table 2.

Table 2. The composition of the dataset

Category	**Definition**	**Total sample size**	**Number of training sets**	**Number of test sets**	**The number of the validation set**
Vertical components	Including the inner columns, eave columns, gable columns, the top of the roof-raising columns, and the curved brackets at the corners, … etc.	2261	1583	452	226
Lateral component	Including four-step beams, cross girders and purlins, … etc.	2261	1583	452	226
Decorative components	Including water jet arches	2261	1583	452	226

To prevent data leakage and ensure the validity of model evaluation, the data splitting strategy adopted is as follows: Firstly, the data of all components of each building was divided into training set, test set and validation set, ensuring that the data of the same building does not appear simultaneously in the training set and the test set. Secondly, the dataset was randomly divided into 70% training set, 20% test set and 10% validation set to provide sufficient data for model training, while leaving enough data for model evaluation and tuning. Finally, the stratified sampling method was used to ensure that the proportion of each category in the training set, test set and validation set was the same as that in the entire dataset, thereby maintaining the representativeness of the data. The comparative indicators were precision, accuracy, and recall. The experiment settings are presented in Table 3.

Table 3. Specific settings of the experiment

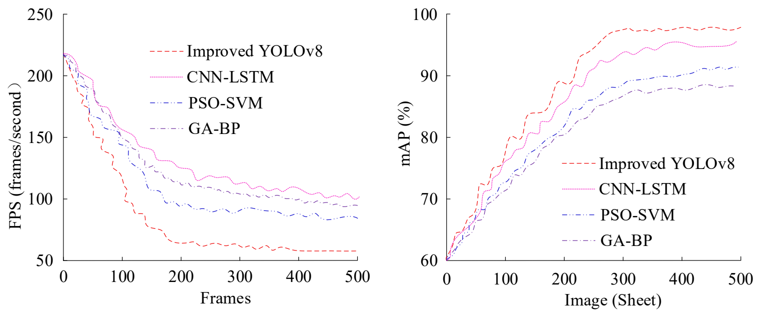
Parameter names	Parameter
Processor	Intel Core i9-9880H
Frequency	4.8GHz
Memory	64GB
Storage capacity	1TB
Framework	Pytorch 1.13.0
GPU	NVIDIA GeForce RTX 3060
Software	Matlab 2022b
Language	Python 3.7.16
Data analysis	Spss23.0

In addition, all experiments were conducted on a single NVIDIA GeForce RTX 3060 GPU with 12GB memory, CUDA 11.7 and PyTorch 1.13.0. The random seeds for weight initialization and data shuffling were fixed at 0. Training employed the SGD optimizer with initial learning rate 0.01, momentum 0.941, weight decay 0.0005, and step decay schedule (factor 0.1 at epochs 1000 and 1300) with 3 warmup epochs. Anchor boxes were generated through k-means clustering on the training set, with 3 anchors assigned to each detection layer (P2, P3, P4, P5). Data augmentation included random horizontal flipping (probability 0.5), random cropping (scale 0.5-1.0), color dithering (magnitude 0.1), and Mosaic. Inference used confidence threshold 0.5, IoU threshold 0.5, and IoU-based NMS (threshold 0.5). Single-scale inference was applied without test-time augmentation. Before conducting comparative analysis, to confirm the improvement effect of YOLOv8, the study first conducted ablation experiments in the above environment. The data was based on the self-collected data of ancient buildings in rural areas of Suzhou, China. A total of 2880 images were used, with 80% for training and 20% for testing. The ablation experiment indicators were recall rate, precision, F1 score and parameter quantity. The experiment outcomes are presented in Table 4.

Table 4. Ablation experiment outcomes

Algorithm	Recall	Precision	F1 score	Parameter quantity
YOLOv8s	85.43%	87.23%	86.31%	12.21
YOLOv8s+P2	86.27%	81.95%	84.03%	11.76
YOLOv8s+IC2f+P2	88.72%	87.46%	87.59%	10.25
YOLOv8s+ADown+IC2f+P2	90.03%	87.12%	88.57%	9.69
YOLOv8s+*Ltotal+ADown+IC2f+P2*	89.47%	90.56%	89.99%	9.03
Improved YOLOv8	98.89%	96.94%	97.88%	6.29

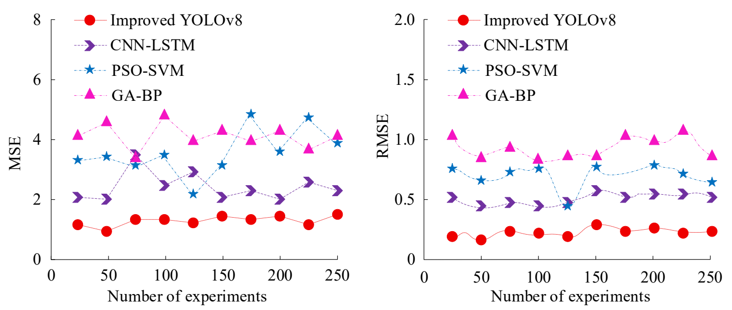
As shown in Table 4, with the increase of each module, the recall rate of the improved YOLOv8 proposed in the study increased from 85.43% before the improvement to 98.89%. The precision rate increased from 87.23% before the improvement to 96.94% after the improvement. The F1 score increased from 86.31% to 97.88% with the increase of each module. The parameter count decreased from 12.21 before the improvement to 6.29. From the above results, the improved YOLOv8 proposed in the study gradually improved performance in regard to recall, precision, and F1 score dimensions, and the improvement effect is significant. The performance evolution across module additions exhibited non-monotonic patterns that require technical explanation. First, the precision degradation observed with YOLOv8s+P2 (from 87.23% to 81.95%) was attributed to the introduction of the shallow P2 detection layer without sufficient deep feature enhancement. The P2 layer operates at 4× downsampling resolution, which increased sensitivity to small targets, but simultaneously amplified background clutter in complex rural architectural scenes. Without the complementary IC2f module for enhanced feature extraction, the model generated more false positives on textured backgrounds, degrading precision. Second, the counterintuitive parameter reduction despite module additions (from 12.21 to 6.29) was explained by the efficiency-oriented design of the incorporated components. Partial convolution processed only 1/4 of input channels, reducing floating-point operations to 1/16 of standard convolution. The EMA attention mechanism operated via channel grouping and reweighting without introducing additional trainable parameters. ADown downsampling replaced conventional strided convolution with a more compact architecture. Consequently, the cumulative effect was a parameter-efficient redesign rather than parameter expansion. Third, the substantial performance leap in the final Improved YOLOv8 (F1 from 89.99% to 97.88%) reflected synergistic module interaction rather than incremental addition. The Wise-inner-MPDIOU loss function optimized BB regression by integrating IoU, inner-IoU, and distance penalties, which became effective only when preceding modules provided sufficiently discriminative features. The P2 layer captured fine-grained localization cues, IC2f enhanced multi-scale representation, EMA suppressed background noise, and ADown preserved critical information during downsampling. This configuration enabled the loss function to converge to optimal BB predictions, producing the observed synergistic gain. The comparison results of the frames per second (FPS) and mean Average Precision (mAP) of each algorithm are shown in Figure 7.



(a) FPS comparison results (b) mAP comparison results of each algorithm

Figure 7. Comparison results of FPS and mAP for each algorithm

From Figure 7(a), the FPS values of the improved YOLOv8, PSO-SVM, CNN-LSTM, and GA-BP algorithms proposed in the study were 69.4 FPS, 98.97 FPS, 134.6 frames/second, and 123.2 FPS. Among them, the introduced improved YOLOv8 algorithm had the lowest FPS. This suggests that the added computational complexity introduced by the P2 module, IC2f module, Ltotal loss function, and ADown downsampling module may have increased inference time. As shown in Figure 7(b), the mAP value of the improved YOLOv8 algorithm proposed in the study was 98.69%, which was higher than 87.27% for PSO-SVM, 94.62% for CNN-LSTM, and 85.43% for GA-BP. The results indicated that, compared with the other algorithms, the introduced improved YOLOv8 algorithm achieved superior accuracy in terms of mAP, at the cost of reduced inference speed. These findings highlighted a trade-off between speed and accuracy. While the improved YOLOv8 algorithm offered superior accuracy, it did so at the cost of reduced speed. This trade-off was a common consideration in algorithm design and should be explicitly acknowledged. Future work could explore ways to optimize the algorithm for faster inference times without compromising accuracy, such as through model lightweighting techniques or hardware acceleration. The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) results for each algorithm are presented in Figure 8.



(a) MSE comparison results (b)RMSE comparison results of each algorithm

Figure 8. Comparison results of MSE and RMSE

Definition of MSE and RMSE. In this study, MSE and RMSE were employed to quantify the localization accuracy of predicted BBs. Specifically, these metrics were computed based on the Euclidean distance between the centers of the predicted BBs and the corresponding ground truth BBs (in pixel units). For each detected building component, the center coordinates (x, y) of the predicted box were compared with the center coordinates of the annotated ground truth box. The MSE was calculated as the average squared center distance across all test samples, while the RMSE was the square root of the MSE, representing the average center distance in original scale. Lower MSE and RMSE values indicated more accurate spatial localization of detected components. From Figure 8 (a), the introduced improved YOLOv8 had the lowest average MSE, with a value of 1.57. The average MSE of PSO-SVM was 3.62, the average MSE of CNN-LSTM was 2.47, and the average MSE of GA-BP was 2.88. As shown in Figure 8 (b), the average RMSE values of the improved YOLOv8, PSO-SVM, CNN-LSTM, and GA-BP algorithms proposed in the study were 0.52, 1.58, 0.84, and 1.46, respectively. Among them, the introduced improved YOLOv8 had the lowest average RMSE. The above outcomes revealed that compared to the comparative algorithms, the improved YOLOv8 proposed in the study had superior performance in terms of MSE and RMSE dimensions. This suggested that the improved YOLOv8 algorithm is more accurate in localizing BBs. To further verify the improved performance of the proposed YOLOv8, this study conducted a performance comparison analysis by comparing it with Faster R-CNN, RetinaNet, YOLOv5, SSD, and the unmodified YOLOv8 under the same conditions. The analysis results are shown in Table 5.

Table 5. Experimental results

Algorithm	Recall	Precision	F1 score
Improved YOLOv8	98.89%	96.94%	97.88%
Faster R-CNN	89.57%	90.46%	88.47%
RetinaNet	87.52%	89.03%	86.88%
unmodified YOLOv8	91.23%	90.46%	92.56%
YOLOv5	88.74%	86.89%	87.36%
SSD	80.07%	81.33%	79.96%

As can be seen from Table 5, the proposed improvements to the recall rate, true precision rate, and F1 score of YOLOv8 were 98.89%, 96.94%, and 97.88%, respectively, all of which were higher than those of the YOLOv5, SSD, Faster R-CNN, RetinaNet, and unmodified YOLOv8 algorithms. This result indicated that the proposed improvements to YOLOv8 had superior performance. To quantitatively evaluate the impact of PCA integration, an ablation experiment was conducted comparing the complete model with PCA preprocessing against the ablation model without PCA. The evaluation metrics included detection accuracy (mAP), inference speed (FPS), model complexity (number of parameters), and variance retention (proportion of explained variance). The experimental results are shown in Table 6.

Table 6. Results of ablation experiments with and without PCA algorithm

Indicators	Complete model (including PCA)	Ablation model (without PCA)
mAP (%)	98.69	97.88
FPS (frames/second)	69.4	72.1
Number of parameters (%)	6.29	8.15
The number of retained PCs	5	/
Proportion of explained variance (%)	98	/

The ablation experiment results in Table 6 showed that the complete model incorporating PCA outperformed the ablation model without PCA in the mAP metric, with values of 98.69% and 97.88%, respectively. This indicated that PCA had a positive impact on improving the model's detection accuracy. However, the model including PCA had a slightly lower FPS of 69.4 frames/second compared to the 72.1 frames/second of the ablation model, suggesting that PCA may slightly affect the inference speed of the model. Additionally, PCA helped reduce the number of model parameters, from 8.15% to 6.29%, which may contribute to improving the model's generalization ability. In the experiment, PCA retained 5 PCs, which explained 98% of the variance. Overall, PCA has played a role in improving model accuracy and reducing parameter numbers, but at the cost of sacrificing some speed. To quantitatively evaluate the statistical robustness of the proposed method, all experiments were repeated 5 times with different random seeds. Performance metrics are reported as mean ± standard deviation. Statistical significance was assessed using paired t-tests between the improved YOLOv8 and each baseline algorithm. The results are presented in Table 7.

Table 7. Statistical validation of performance comparisons

Algorithm	Recall	Precision	F1 score	FPS	MSE	RMSE	*p*-value
Improved YOLOv8	98.89%±0.45%	96.94%±0.38%	97.88%±0.42%	69.4±2.3	1.57±0.18	0.52±0.09	/
YOLOv5	88.74%±1.12%	86.89%±1.25%	87.36%±1.18%	/	/	/	<0.001***
Faster R-CNN	89.57%±0.98%	90.46%±0.87%	88.47%±0.92%	134.6±4.2	2.47±0.28	0.84±0.15	<0.001***
RetinaNet	87.52%±1.35%	89.03%±1.18%	86.88%±1.28%	98.97±3.1	3.62±0.35	1.58±0.22	<0.001***
unmodified YOLOv8	91.23%±0.76%	90.46%±0.82%	92.56%±0.79%	123.2±3.8	2.88±0.31	1.46±0.19	<0.001***
SSD	80.07%±1.56%	81.33%±1.42%	79.96%±1.48%	/	/	/	<0.001***

Note: ****p*<0.001 indicates statistically significant difference compared to Improved YOLOv8.

From Table 7, the improved YOLOv8 achieved the highest Recall (98.89%±0.45%), Precision (96.94%±0.38%), and F1 score (97.88%±0.42%) among all comparative algorithms, with low standard deviations indicating high reproducibility. The MSE (1.57±0.18) and RMSE (0.52±0.09) were also the lowest, demonstrating superior localization accuracy. While the FPS (69.4±2.3) was lower than Faster R-CNN and unmodified YOLOv8, this trade-off was justified by the significant accuracy gains. All baseline algorithms showed *p*-values less than 0.001, confirming that the performance improvements of the improved YOLOv8 were statistically significant rather than due to random variation.

Analysis of the Effectiveness of the Digital Protection Framework for RAH

After comparing the performance of the improved YOLOv8 algorithm, the study analyzed the effectiveness of a digital protection framework for RAH based on the improved YOLOv8 and PCA. The comparative frameworks were the PSO-SVM-based digital protection framework for RAH, the CNN-LSTM-based digital protection framework for RAH, and the GA-BP-based digital protection framework for RAH. The experiment involved selecting 26 rural buildings in a certain location in Suzhou for real-time detection and analysis. The recognition accuracy of each frame for the column head A of eaves column, four step beam B, center column C and water beam arch of wooden architecture D is presented in Figure 9.

	A	B	C	D
A	98.1	0.5	0.2	1.2
B	0.6	97.3	0	2.1
C	0	0.2	98.2	1.6
D	0.4	0.6	1.2	97.8

(a) Improved YOLOv8

	A	B	C	D
A	87.5	3.9	6.4	2.2
B	0	87.6	3.3	9.1
C	8.4	0.8	89.8	1.0
D	5.4	4.4	0	90.2

(b) PSO-SVM

	A	B	C	D
A	83.7	7.2	6.3	2.8
B	2.8	85.1	1.2	10.9
C	12.1	0	86.6	1.3
D	9.4	5.8	0	84.8

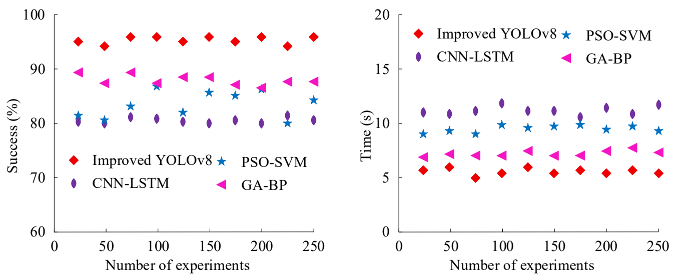
(c) CNN-LSTM

	A	B	C	D
A	84.5	11.8	0.9	2.8
B	8.4	82.8	0	8.8
C	4.3	2.7	87.1	5.9
D	3.2	2.3	3.7	90.8

(d) GA-BP

Figure 9. Recognition accuracy rates of each frame

From Figure 9 (a), the identification accuracy of the improved YOLOv8 frame for the building eaves column capitals, four step beams, central columns and water beam arch of wooden architecture was 98.1%, 97.3%, 98.2% and 97.8%, respectively, which was higher than 87.5%, 87.6%, 89.8% and 90.2% of the PSO-SVM in Figure 9 (b), and 83.7%, 85.1%, 86.6%, 84.8% of the CNN-LSTM in Figure 9 (c), and 84.5%, 82.8%, 87.1% and 90.8% of the GA-BP in Figure 9 (d). The findings mentioned above suggested that, from the perspective of recognition accuracy for the four types of building components, the proposed framework outperformed the comparative framework in regard to performance. The comparison results of the evaluation success rate and real-time feedback rate of each framework are presented in Figure 10.



(a) Success comparison results (b) Real time rate comparison results of each algorithm

Figure 10. Comparison results of evaluation success rate and feedback real-time rate

In Figure 10, the "Evaluation Success Rate" refers to the proportion of times the system correctly identifies and classifies the target building components out of a certain number of evaluation attempts. "Feedback Time" refers to the average time required for the system to receive the input data and provide the evaluation results. "Real-time Feedback" refers to the system's ability to provide feedback under real-time operation conditions. Furthermore, the study also conducted a statistical significance test on the experimental results. The *p*-value of each experiment was less than 0.001, indicating that the results were statistically significant at the 1% significance level. This suggests that our results are reliable. From Figure 10 (a), the average evaluation success rates of the improved YOLOv8 framework, PSO-SVM, CNN-LSTM, and GA-BP framework proposed in the study were 97.29%, 84.14%, 82.11%, and 89.78%, respectively. The average success rate of the digital protection framework for RAH proposed in the study was the highest. According to Figure 10 (b), the average feedback time of the digital protection framework for RAH proposed in the study was 5.17 seconds, significantly lower than the 9.67 seconds of PSO-SVM, 13.47 seconds of CNN-LSTM, and 7.89 seconds of GA-BP. The above results indicated that compared with the comparative framework, the digital protection framework for rural architecture had the best performance in regard to real-time feedback rate and evaluation success rate.

DISCUSSION

This study conducted a performance comparison analysis between the improved YOLOv8 algorithm and the digital protection framework for RAH based on PCA and improved YOLOv8. In the ablation experiment analysis, the improved YOLOv8 recall rate, precision rate, F1 score, and parameter quantity proposed by the study were 96.94%, 97.88%, and 6.29%, respectively. The precision and recall rates were significantly higher than the pre-improvement levels of 87.23% and 84.78%, respectively, while the parameter count was significantly lower than the pre-improvement level of 12.21. This result indicated that the introduction of IC2f module, Ltotal loss function, and ADown downsampling module improved the detection accuracy and precision of the algorithm. This outcome was similar to the improved YOLOv8 algorithm proposed by (Mu et al., 2025). In the comparative analysis of FPS, the FPS values of the improved YOLOv8, PSO-SVM, CNN-LSTM, and GA-BP algorithms proposed in the study were 69.4 FPS, 98.97 FPS, 134.6 FPS, and 123.2 FPS, respectively. The FPS of the improved YOLOv8 was the lowest, which suggests that the added computational complexity introduced by the P2 module, IC2f module, Ltotal loss function, and ADown downsampling module may have increased inference time. This decrease in FPS could be critically interpreted as a consequence of the added computational complexity introduced by the P2 module, IC2f module, Ltotal loss function, and ADown downsampling module. While these components improved feature extraction and detection accuracy, they also increased inference time, which may hinder real-time application. To optimize the model for real-time use, future work could explore model lightweighting techniques, such as pruning, quantization, or neural architecture search, to reduce computational load without significantly sacrificing accuracy. Additionally, hardware acceleration or algorithm-level optimizations, such as more efficient feature fusion mechanisms, may help bridge the performance gap. Sang and Peng reached consistent conclusions in their research on improving the YOLOv8 algorithm (Sang & Peng, 2025). In addition, the mAP value, average MSE value, and average RMSE of the improved YOLOv8 algorithm proposed in the study were 98.69%, 1.57%, and 0.52, respectively, all of which were superior to PSO-SVM, CNN-LSTM, and GA-BP algorithms. This result further validated the advantage of the introduced algorithm. Zhu et al. reached similar conclusions in related research (Zhu et al., 2025). In the effectiveness analysis experiment, the proposed digital framework for RAH protection achieved recognition accuracies of 98.1%, 97.3%, 98.2%, and 97.8% for four types of building components, respectively, which were significantly higher than the comparative framework. The average evaluation success rates of the proposed framework, PSO-SVM, CNN-LSTM, and GA-BP framework in the comparative analysis of evaluation success rates were 97.29%, 84.14%, 82.11%, and 89.78%, respectively. The average feedback times were 5.17s, 9.67s, 13.47s, and 7.89s, respectively. The above outcomes revealed that compared to the comparative framework, the proposed framework had the best performance. This result indicated that the integration of PCA algorithm, IC2f module, Ltotal loss function, ADown downsampling module, and P2 detection layer improved the real-time response efficiency of the framework at runtime. This conclusion was similar to the one obtained by the Prados-Peña et al. in their relevant research in 2025 (Prados-Pena et al., 2025). The drawback of this study is that most of the data collection for buildings is focused on wooden structures, while data collection for other buildings, such as brick and stone structures, has not been explored. In practical applications, the composition of RAH is diverse, and more comprehensive digital protection of architectural heritage is a further direction for research.

To explicitly distinguish between inherited, modified, and original contributions, this study categorized its methodological elements as follows. (1) Inherited components from previous literature. Efficient Multi-scale Attention Module (EMA), partial convolution, Wise-inner-MPDIOU loss function, Down downsampling module, and quadruple detection layer from FasterNetblock are existing techniques adopted directly from prior studies without structural modification. (2) Modified components. The configuration and integration strategy of these modules into YOLOv8 for RAH detection represents domain-specific adaptation. The C2f module is restructured to incorporate partial convolution and EMA, and the detection head is extended with a quadruple detection layer. (3) Original contributions of this study. The synergistic integration of these configured modules, validated by systematic ablation experiments. The framework-level integration of PCA preprocessing with the enhanced YOLOv8 algorithm within a complete digital protection pipeline encompassing multi-source data acquisition, intelligent detection, and three-dimensional modeling. The demonstration that this specific configuration achieves superior performance for RAH detection compared to generic object detection approaches.

CONCLUSION

This study effectively addressed the issue of low accuracy in building detection under current digital technology by introducing and improving the YOLOv8 algorithm. The integration of EMA, downsampling, Wise-inner-MPDIOU loss function and quadruple detection layer significantly improved the accuracy and computational efficiency of the YOLOv8 algorithm. In addition, a RAH digital protection framework, based on PCA and the improved YOLOv8, was designed. The performance comparison indicated that the enhanced YOLOv8 algorithm outperformed the PSO-SVM, CNN-LSTM and GA-BP algorithms in terms of mAP, MSE and RMSE. In addition, compared with the existing frameworks, the proposed RAH digital protection framework performed better in terms of recognition accuracy, evaluation success rate and real-time feedback rate.

However, it is important to acknowledge the limitations of this study. The experimental data was primarily collected from wooden structures in Suzhou, China, and the study did not explore data collection for other building types, such as brick and stone structures. This introduced regional data bias and limits sample diversity, as the dataset encompassed only 26 buildings with 6,784 samples, which may not adequately represent the full heterogeneity of RAH across different geographic regions and climatic conditions. The reported high performance values were achieved under relatively controlled conditions with high-quality data collection, and the dataset did not adequately represent challenging real-world scenarios, such as occluded structures, variable lighting conditions, degraded architectural elements, cluttered backgrounds, noisy point cloud scans, or out-of-distribution architectural styles. This limitation restricts the generalizability of the findings to a broader range of architectural heritage types and environmental conditions. Furthermore, the study lacks external benchmarking using publicly available datasets or third-party validation, and real deployment validation in actual heritage conservation projects remains absent. Manual annotation of building components may also introduce annotation bias, particularly for decorative elements with ambiguous boundaries. Additionally, while the improved YOLOv8 algorithm achieved higher accuracy, the added computational complexity introduced by the P2 module, IC2f module, Ltotal loss function, and ADown downsampling module resulted in a lower FPS value compared to other algorithms tested. This suggests that the current implementation may not be fully optimized for real-time applications.

The main contribution of this study is the enhancement of the YOLOv8 algorithm and the development of an effective digital protection framework for RAH. These advancements have provided an effective solution for the digital protection of rural wooden architectural heritage in the studied region, with potential for broader application subject to future validation, significantly enhancing the accuracy and precision of building inspection. This is crucial for the preservation and restoration of RAH. Future work will focus on further optimizing algorithms to adapt to real-time applications, exploring model lightweighting techniques, and expanding the framework to accommodate a wider range of building structures. Additionally, systematic robustness testing will be conducted under adverse conditions, including occlusion, lighting variation, structural degradation, background clutter, and noisy data, along with cross-regional validation to assess the framework's generalizability beyond the current dataset.

Fundings

The research is supported by Zhejiang Provincial Key Research Base of Philosophy and Social Sciences - Research Center for Cultural Development Innovation and Cultural Zhejiang Construction (No.2022JDK TZD25); and the Major Humanities and Social Sciences Research Projects in Zhejiang higher education institutions-The Identity Transition, Social Value, and Modern Development of Agricultural Culture (No.2023QN123).

REFERENCES

Addab, H.F., & Jaber Al Rammahi, A.H. (2024). Statistical assessment of water-quality parameters for different flow regimes. *Jordan Journal of Civil Engineering, 18*(2). https://doi.org/10.14525/JJCE.V 18i2.08

Chen, X., He, J., & Wang, S. (2025). Deep learning-driven pathology detection and analysis in historic masonry buildings of Suzhou. *NPJ Heritage Science, 13*(1), 1-14. https://doi.org/10.1038/s40494-025-01783-y

Cline, P., Mariotti, C., Angeloni, R., & Muñoz Cádiz, J. (2024). Architectural heritage digital representations for conservation strategies. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48*, 111-118. https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-111-2024

Guo, M., Fu, Z., Pan, D., Zhou, Y., Huang, M., & Guo, K. (2022). 3D digital protection and representation of burial ruins based on LiDAR and UAV survey. *Measurement and Control, 55*(7), 555-566. https://doi.org/10.1177/0020294 0221110949

Hasanvand, M., Nooshyar, M., Moharamkhani, E., & Selyari, A. (2023). Machine learning methodology for identifying vehicles using image processing. *Artificial Intelligence and Applications, 1*(3), 154-162. https://doi.org/10.47852/bonviewAIA3202833

Kabadayı, A., & Erdoğan, A. (2022). Application of terrestrial photogrammetry method in cultural heritage studies: A case study of Seyfeddin Karasungur. *Mersin Photogrammetry Journal, 4*(2), 62-67. https://doi.org/10.53093/mephoj.1200146

Luo, L. Q., Shao, J., Dang, W., Geng, L., Zheng, H., & Liu, C. (2024). An efficient multi-scale channel attention network for person re-identification. *The Visual Computer, 40*(5), 3515-3527. https://doi.org/10.1007/s00371-023-03049-9

Marukatat, S. (2023). Tutorial on PCA and approximate PCA and approximate kernel PCA. *Artificial Intelligence Review, 56*(6), 5445-5477. https://doi.org/10.1007/ s10462-022-10297-z

Mu, Q., Zhang, Z., Xiao, C., & Sha, R. (2025). Improved algorithm for visually impaired individuals avoid obstacles based on YOLOv8. In *Proceedings of the Fourth International Conference on Electronic Information Engineering and Data Processing (EIEDP 2025)* (Vol. 13574, pp. 1135-1143). SPIE. https://doi.org/10.1117/12. 3067095

Muangasame, K., & Tan, E. (2023). Phygital rural cultural heritage: A digitalization approach for destination recovery and resilience. *Worldwide Hospitality and Tourism Themes, 15*(1), 8-17. https://doi.org/10.1108/WHATT-08-2022-0096

Pang, B., Yang, J., Xia, T., Zhang, A., Zhang, K., Xu, Q., & Wang, F. (2025). Automated heritage building component recognition and modeling based on local features. *Journal of Cultural Heritage, 71*, 252-264. https://doi.org/10.1016/j.culher.2024.12.006

Prados-Peña, M.B., Pavlidis, G., & García-López, A. (2025). New technologies for the conservation and preservation of cultural heritage through a bibliometric analysis. *Journal of Cultural Heritage Management and Sustainable Development, 15*(3), 664-686.

Qi, Z., Kang, C., Wang, P., & Li, W. (2024). Application of digital art technology in architectural heritage preservation and digital reconstruction in the construction industry. *Computer-Aided Design and Applications, 21*(1), 14-27. https://doi.org/10.14733/cadaps.2024.S11.14-27

Sanchez-Calvillo, A., Alonso-Guzman, E.M., Solís-Sánchez, A., Martinez-Molina, W., Navarro-Ezquerra, A., & Gonzalez-Sanchez, B. (2023). Use of audiovisual methods and documentary film for the preservation and reappraisal of the vernacular architectural heritage of the State of Michoacan, Mexico. *Heritage, 6*(2), 2101-2125. https://doi.org/10.3390/heritage6020113

Sang, H., & Peng, X. (2025). Nighttime pedestrian detection using crossed wavelet fusion networks. *Journal of Electronic Imaging, 34*(2), 023067. https://doi.org/10.1117/1.JEI.34.2.023067

Sapkota, D., Ahmed, D., & Karkee, M. (2024). Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. *Artificial Intelligence in Agriculture, 13*, 84-99. https://doi.org/10.1016/j.aiia. 2024.07.001

Shen, L., Lang, B., & Song, Z. (2023). DS-YOLOv8-based object detection method for remote sensing images. *IEEE Access, 11*, 125122-125137. https://doi.org/10.1109/ACC ESS.2023.3330844

Talaat, F.M., & ZainEldin, H. (2023). An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Computing and Applications, 35*(28), 20939-20954. https://doi.org/10.1007/s00521-023-08809-1

Turbay, I., Ortiz, P., & Ortiz, R. (2024). Statistical analysis of principal components (PCA) in the study of the vulnerability of heritage churches. *Procedia Structural Integrity, 55*, 168-176. https://doi.org/10.1016/j.prostr.2024.02.022

Wang, T., Niu, Y., Zhao, W., & Research, R.P. (2025). Research on intelligent classification of limestone photomicrographs based on the improved FasterNet architecture. *Earth Science Informatics, 18*(4), 1-16. https://doi.org/10.1007/s12145-025-02028-z

Yang, S., Zheng, L., Chen, Y., Huang, Y., & Huang, Y. (2025). Artificial intelligence assists the identification and application of rural heritage wall surface damage in Zhejiang. *NPJ Heritage Science, 13*(1), 1-22. https://doi.org/10.1038/s40494-025-01725-8

Zhang, C., Zhou, L., Hu, X., & Xu, K. (2025). Real-time detection of cracks: Improved YOLOv8 network applied to wooden structure of residential historic buildings. In *Proceedings of the International Conference on Mechatronics and Intelligent Control (ICMIC 2024)* (Vol. 13447, pp. 148-157). SPIE. https://doi.org/10.1117/12. 3044980

Zhu, J., Zhou, D., Lu, R., Liu, X., & Wan, D. (2025). C2DEM-YOLO: Improved YOLOv8 for defect detection of photovoltaic cell modules in electroluminescence image. *Nondestructive Testing and Evaluation, 40*(1), 309-331. https://doi.org/10.1080/10589759.2024.2319263