

RESEARCH

Open Access



Intelligent assessment system of material deterioration in masonry tower based on improved image segmentation model

Jianshen Zou¹ and Yi Deng^{1*}

Abstract

Accurate and timely data collection of material deterioration on the surfaces of architectural heritage is crucial for effective conservation and restoration. Traditional methods rely heavily on extensive field surveys and manual feature identification, which are significantly affected by objective conditions and subjective factors. While machine vision-based methods can help address these issues, the accuracy, intelligence, and systematic nature of material deterioration assessment for large-scale masonry towers with complex geometries still require significant improvement. This research focuses on the architectural heritage of masonry towers and proposes an intelligent assessment system that integrates an improved YOLOv8-seg machine vision image segmentation model with refined 3D reconstruction technology. By optimizing the YOLOv8-seg model, the system enhances the extraction capabilities of both detailed and global features of material deterioration in masonry towers. Furthermore, by complementing it with image processing methods for the global visualization of large-scale objects, this research constructs a comprehensive intelligent assessment process that includes "deterioration feature extraction—global visualization—quantitative and qualitative comprehensive assessment." Experimental results demonstrate that the intelligent assessment system significantly improves the performance of target feature extraction for material deterioration in masonry towers compared to existing methods. The improved model shows improvements of 3.39% and 4.55% in the key performance metrics of mAP50 and mAP50-95, respectively, over the baseline model. Additionally, the efficiency of global feature extraction and visualization of material deterioration increased by 66.36%, with an average recognition accuracy of 95.78%. Consequently, this system effectively overcomes the limitations and subjective influences of field surveys, enhancing the objectivity and efficiency of identifying and analyzing material deterioration in masonry towers, and providing invaluable data support for the subsequent preservation and restoration efforts.

Keywords Masonry tower, Intelligent assessment system, Image segmentation, Material deterioration

Introduction

Architectural heritage stands as a testament to the rich history and remarkable civilization of a nation and its people. It encapsulates the profound cultural essence of the nation, serving as a pinnacle of human intellect and

perseverance, and possesses immense scientific, cultural, and artistic value. Consequently, the preservation and transmission of architectural heritage have acquired a paramount importance. However, the passage of time subjects the surface materials of architectural heritage to the ravages of natural calamities such as wind erosion, earthquakes, and floods, as well as human-induced damages such as graffiti and other external environmental factors. This, in turn, leads to material deterioration including damages, cracks, and corrosion. These material deteriorations are often the result of physical changes,

*Correspondence:

Yi Deng

dengyi@gzhu.edu.cn

¹ School of Architecture and Urban Planning, Guangzhou University, Guangzhou 510006, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

chemical reactions, and environmental factors such as temperature, humidity, and climate. Failure to promptly detect and address these material deteriorations not only impairs the performance and aesthetics of architectural heritage materials, but also eventually leads to damage to building components and jeopardizes the overall structural integrity and stability of the architectural heritage [1]. This, in turn, increases the risk of unforeseen damage and collapse [2]. Significantly, massive and towering masonry towers of architectural heritage are particularly vulnerable to external influences such as wind loads and earthquakes. Consequently, conducting timely and systematic deterioration surveys and statistical analysis becomes crucial. However, the current statistical analysis and assessment of material deterioration in masonry tower architectural heritage still relies on manual visual inspections and experiential judgments [3]. This antiquated approach demands extensive time and effort, posing challenges to the swift and accurate acquisition of deterioration data. Hence, an urgent and automated method is imperative to identify material deterioration in masonry towers.

The continuous advancement of machine vision technology has revolutionized the preservation, archaeological exploration, and restoration of architectural heritage. Increasingly, projects are utilizing machine vision for image recognition tasks, including unsupervised classification (Fuzzy K-Means algorithms) [4], supervised classification (Maximum Likelihood classification algorithms) [5], and multispectral image analysis [6] and so on, to identify and detect material-related deterioration in architectural heritage. The application of these digital techniques based on machine vision effectively reduces the likelihood of unnecessary damage to buildings or cultural relics [7], thus playing a pivotal role in the protection, digitization, and sustainable development of architectural heritage. Recent innovations in convolutional neural networks, from the development of models such as AlexNet [7], VGG [8] to ResNet [9], have driven substantial progress in machine vision technology. These advances, particularly in object detection and image segmentation, provide novel approaches for accessing the current conditions and issues affecting architectural heritage. Furthermore, they facilitate the completion of the complex and repetitive tasks of deterioration statistics with greater efficiency, while simultaneously reducing the risk of secondary damage during the investigation process. Table 1 presents a compilation of recent relevant studies from the fields of architecture and engineering that employ object detection and image segmentation techniques to facilitate condition surveys. These studies encompass a variety of tasks, including the identification of structural crack damage, the detection of surface

defects and decay in architectural heritage, and so on. In the field of object detection, Kwon et al. [10], Mansuri et al. [11], and Pathak et al. [12] employed a region-based convolutional neural network algorithm R-CNN and Fast R-CNN to detect damages in built heritage. Yan et al. [13], Rout et al. [14], Zhang et al. [15], Mishra et al. [16] and others employed enhanced versions of the YOLO series models for detection tasks, achieving highly accurate detection results. In terms of image segmentation, Bruno et al. [17], Xu et al. [18], Kim et al. [19], Wang et al. [20], Hou et al. [21], Hatir et al. [22], Altaweel et al. [23], and others employed the Mask R-CNN model to locate and segment the deterioration conditions of architectural heritage, archaeological site, cultural relic, traffic roads, and concrete structures. These studies conducted quantitative assessments of deterioration based on the predicted mask morphology. Liu et al. [24], Banasiak et al. [25] and Stoean et al. [26] employed the U-Net model to tackle concrete crack, archaeological monuments features and metal heritage assets material degradation segmentation tasks, yielding a commendable performance. In the aforementioned studies, object detection and image segmentation methodologies primarily relied on deep learning-based convolutional neural networks. Among these techniques, those focused on object detection aim to identify targets on object surfaces. However, object detection still employed rectangular bounding boxes that included redundant background information, which hindered accurate target quantification [27]. In terms of image segmentation, although segmentation masks can precisely delineate the boundaries and spatial information of the affected areas, existing methods are often limited to segmenting local images and lack intuitive visualization from a global perspective.

In recent years, significant advances in digital surveying techniques such as 3D laser scanning and photogrammetry have provided more convenient approaches for acquiring spatial information and conducting condition surveys of architectural heritage. These techniques offer greater convenience and efficiency compared to traditional site surveys. Notably, refined 3D reconstruction models, enriched with high-resolution texture information through the fusion of multi-source data, can provide a comprehensive view of architectural heritage in certain cases. However, the extraction of feature data from these refined models still predominantly relies on labor-intensive manual processes, lacking efficient automated methods. Therefore, by leveraging the global perspective afforded by 3D reconstruction models and the precise localization features of image segmentation techniques, researchers have made further strides in analyzing relevant features and localizing damage targets in architectural heritage from a global perspective. For

Table 1 Recent research on object detection and image segmentation in the field of heritage and engineering

Task	References	Model	Application scenarios
Detect	[13]	YOLOv4	Damage detection of shedthn tiles in traditional garden architecture
	[10]	R-CNN	Damage detection of stone cultural heritage property
	[11]	Fast R-CNN	Defect detection of built heritage
	[12]	Fast R-CNN	Damage detection of heritage structure surfaces
	[14]	Improved YOLOv7	Pothole detection of traffic road
	[15]	Improved YOLOX	Defect detection of artworks
	[16]	YOLOv5	Defect detection of cultural heritage structure
	[30]	MLNN	Non-invasive damage detection of monument heritage
	[31]	VGG-16	Defect and deterioration detection of buildings' condition
	[32]	Multiple Models	Spatial distribution and size detection of cultural heritage
Segment	[33]	Multiple Models	Cracks segmentation of earthen heritage site surface
	[17]	Mask R-CNN	Decay segmentation of built heritage
	[18]	Mask R-CNN	Cracks segmentation of road pavement
	[19]	Mask R-CNN	Cracks segmentation of concrete structure
	[20]	Mask R-CNN	Architectural form segmentation of traditional village
	[21]	Mask R-CNN	Gold foil shedding segmentation of stone cultural heritage
	[22]	Mask R-CNN	Deterioration segmentation of stone culture heritage
	[23]	Mask R-CNN	Segmentation of archaeological remains pits
	[34]	Mask R-CNN	Defect segmentation of marquetry heritage
	[24]	U-net	Cracks segmentation of concrete
	[25]	U-Net	Segmentation of archaeological features
	[26]	U-Net	Material degradation state segmentation of metal heritage assets
	[28]	Mask R-CNN	Crack segmentation of engineering infrastructure
	[29]	Unknown CNN	Segmentation of large-scale surveys of historic masonry
	[35]	Improved YOLOv5	Limestone peeling segmentation of cultural heritage buildings
	[36]	DeeplabV3+	Plant growth status segmentation of stone masonry heritage
	[37]	Rank-R-FNN	Material defects segmentation of cultural heritage monuments
	[38]	cGAN	Corrupted glyphs segmentation of archaeological artefacts

instance, Kalfarisi et al. [28] successfully localized and segmented cracks in engineering infrastructure by combining the Mask R-CNN model with 3D reality models. Louis et al. [29] integrated high-precision laser scanning and unmanned aerial vehicle (UAV) photogrammetry data to generate high-resolution facade images of a specific architectural heritage, and integrated them with image segmentation models to extract and analyze surface brick textures. Idjaton et al. [35] generated high-resolution image data of wall spalling deterioration using high-precision 3D models and localized wall deterioration areas using the YOLOv5 model incorporating the Transformer architecture, achieving favorable recognition results. Liu et al. [36] employed a combination of DeeplabV3+ and photogrammetry to monitor plant growth status at a stone masonry heritage site. Although these integrated methods combining 3D reconstruction models and image segmentation have advanced global feature analysis and qualitative labelling of architectural heritage, they still lack comprehensive quantitative and

qualitative assessments, which limits their intuitive application in the preservation and restoration of architectural heritage.

In conclusion, while current research has made certain progress in the intelligent recognition of material deterioration in architectural heritage using machine vision technology, there are still limitations in terms of general applicability and practical implementation. Specifically, the complex geometrics and large size of masonry tower architectural heritage present hurdles in intelligently extracting material deterioration features. These towers exhibit multi-categorical and multi-scaled surface deterioration features, with larger-scale deterioration targets tending to dominate the model-learning process, which impairs the model's capacity to accurately recognize the deterioration types of smaller-scale objects. Consequently, optimizing the modules, algorithm architectures, and loss functions of existing image segmentation models tailored to the specific characteristics of material deterioration in masonry towers is crucial. This optimization

aims to develop a model capable of correctly classifying and identifying these deteriorations. Furthermore, while current methods integrating refined 3D reconstruction technology and image segmentation model achieve the global localization and labelling of material deterioration in architectural heritage, they still lack comprehensive global visualization and both quantitative and qualitative assessment methods. This deficiency reduces their effectiveness in addressing the demanding workloads and low efficiency inherent in current material deterioration inspection tasks. Based on the preceding literature review and summary, this study proposes an intelligent assessment system comprising three key components: "deterioration feature extraction—global visualization—quantitative and qualitative comprehensive assessment." This system improves existing image segmentation algorithm frameworks and integrates refined 3D reconstruction technology to intelligently identify various types and scales of material deterioration in masonry towers in a global view. In the meanwhile, by incorporating global visualization methods suitable for large-scale objects and comprehensive quantitative and qualitative assessments, this system improves the efficiency of current condition surveys and data analysis of material deterioration in masonry tower architectural heritage, providing reliable data support for developing subsequent preservation strategies and estimating restoration project costs.

Method

This study establishes an intelligent assessment system that integrates an improved YOLOv8-seg machine vision image segmentation model with refined 3D reconstruction technology for architectural heritage. By optimizing the YOLOv8-seg base model, the system enhances its ability to capture both fine-scale features and global characteristics of material deterioration in masonry towers. This is further complemented by image processing methods for global visualization of large-scale objects, constructing a comprehensive intelligent assessment process encompassing "deterioration feature extraction—visualization—quantitative and qualitative comprehensive assessment". This approach improves the objectivity and efficiency of identifying and analyzing material deterioration in masonry towers, overcoming the limitations of previous research that relies solely on local image data for feature extraction and lacks further in-depth analysis, thus providing robust scientific data support for the protection and restoration of masonry tower architectural heritage. The overall framework of the proposed intelligent assessment system is illustrated in Fig. 1, which includes three key stages: deterioration feature extraction, global visualization, and comprehensive assessment. Firstly, image datasets relevant to this study are collected

and annotated from open-source platforms based on the customized classification of material deterioration in masonry towers. Secondly, 3D laser scanning and photogrammetry technologies are employed to acquire point cloud and image data of the research objects. These data are then used to generate orthophotos from a vertical projection perspective via a refined 3D reconstruction method based on multi-source data fusion. Concurrently, targeted improvements are applied to existing image segmentation models to better meet the needs of deterioration feature extraction. Following multiple rounds of model accuracy validation, the system is capable of conducting batch prediction of material deterioration, global visualization, and comprehensive quantitative and qualitative assessments on the research cases.

This study employs an enhanced YOLOv8-seg model to facilitate intelligent recognition of material deterioration in masonry towers. Based on the Ultralytics algorithmic framework, the YOLO model is widely acknowledged for its real-time detection and segmentation capabilities, offering exceptional speed and accuracy across various scenarios. Furthermore, the YOLOv8 version has been integrated into the Ultralytics repository, forming an algorithmic framework with scalability, flexibility, and a wide range of potential applications, making the YOLOv8-seg model highly suitable for this study. The improvements of the segmentation model will consider the unique characteristics of material deterioration in masonry towers, such as cracks and defects with elongated morphology and small scale, as well as mildew and wall detachment with significant area coverage. These enhancements include the introduction of the RFACnv (Receptive Field Attention Convolution) [39], MHSA (Multi-Head Self-Attention) mechanism [40], and the SlideLoss category-balancing function [41], which enable the model to adapt to the segmentation tasks associated with material deterioration in masonry towers and achieve intelligent recognition.

The enhanced YOLOv8-seg model exhibits several key characteristics. Firstly, it enhances detailed scale features extraction by introducing RFACnv, which allows the model to better capture boundary and detail information from images, thus improving its ability to extract subtle targets. Secondly, it improves global feature extraction by incorporating the MHSA mechanism, which enables the model to better learn contextual information within images and enhances its ability to extract global features. Lastly, by fusing the aforementioned modules, along with the incorporation of a category-balancing loss function, the segmentation performance and generalization ability of the model can be better improved, allowing for more accurate identification and segmentation of material

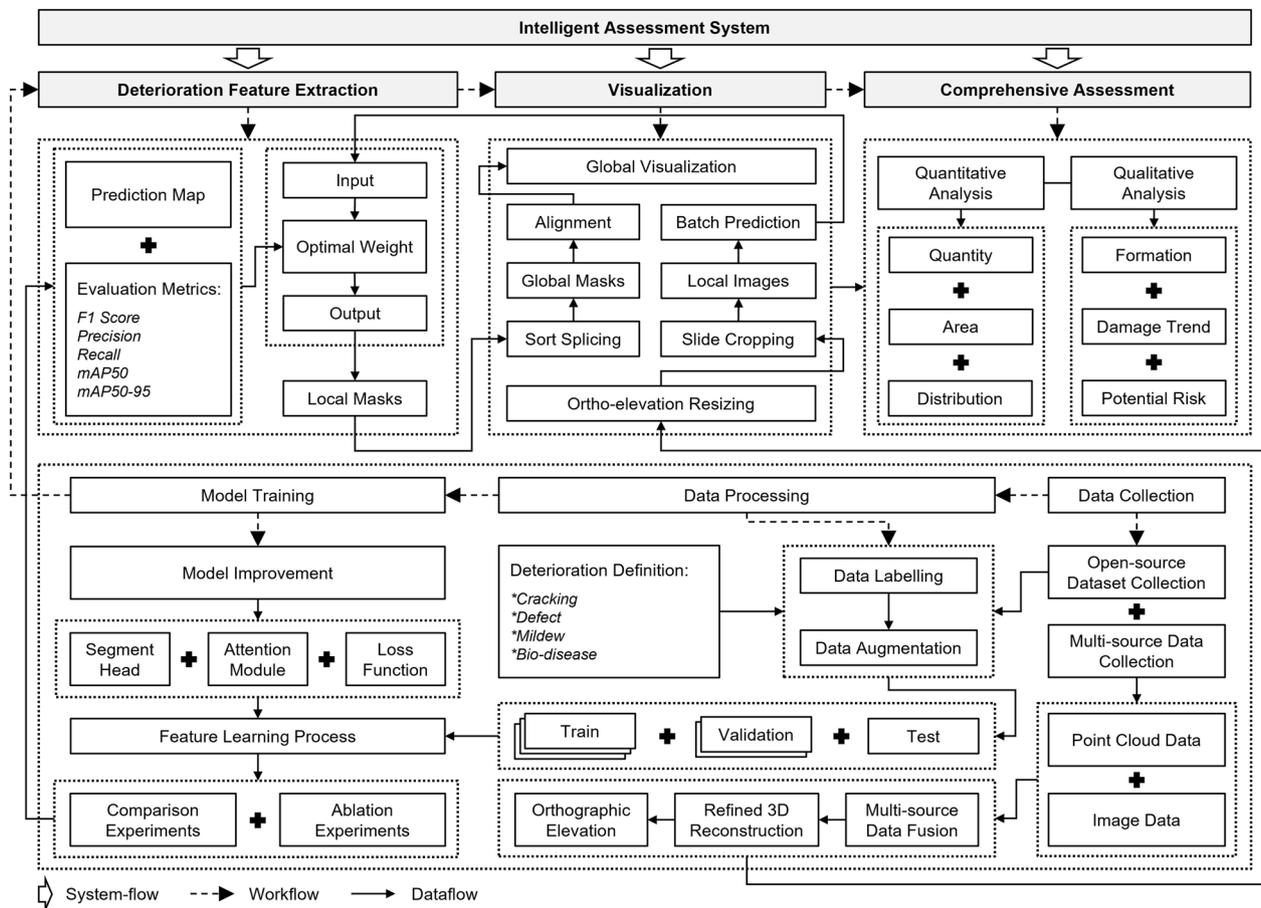


Fig. 1 Research framework

deterioration features in masonry towers. The specific model framework is illustrated in Fig. 2.

RFACConv

RFACConv is a lightweight plug-and-play module proposed by Zhang et al. [39] in 2023. In standard convolutional neural networks, the receptive field refers to the local region of input data that a convolutional layer receives. During the model-training process, standard convolution utilizes parameter sharing and a sliding window of receptive fields to extract feature information from images, addressing the issues of parameter and computational complexity in fully connected neural networks. However, the parameter-sharing mechanism also constrains the network’s ability to extract features from disparate positions within the image, thereby impeding further enhancement of the model’s performance. RFACConv combines the spatial attention mechanism with

the receptive field’ features, allowing the convolutional neural network to adaptively adjust the processing of the receptive field based on the characteristics of each region in the image. By emphasizing spatial features within the receptive field, RFACConv addresses the problem of parameter sharing in standard convolutional kernels. This enables the network to efficiently and accurately capture and process local features in the image, enhancing its ability to extract features from subtle targets and improving overall model performance. Additionally, RFACConv utilizes average pooling to aggregate all feature information within each receptive field, reducing the model’s parameter and computational complexity. Based on these advantages, this study integrates RFACConv into the segmentation head of the YOLOv8-seg algorithm framework to accommodate the complexity and diversity of material deterioration segmentation tasks in masonry towers. The computation process can be represented by Formula (1) [39]:

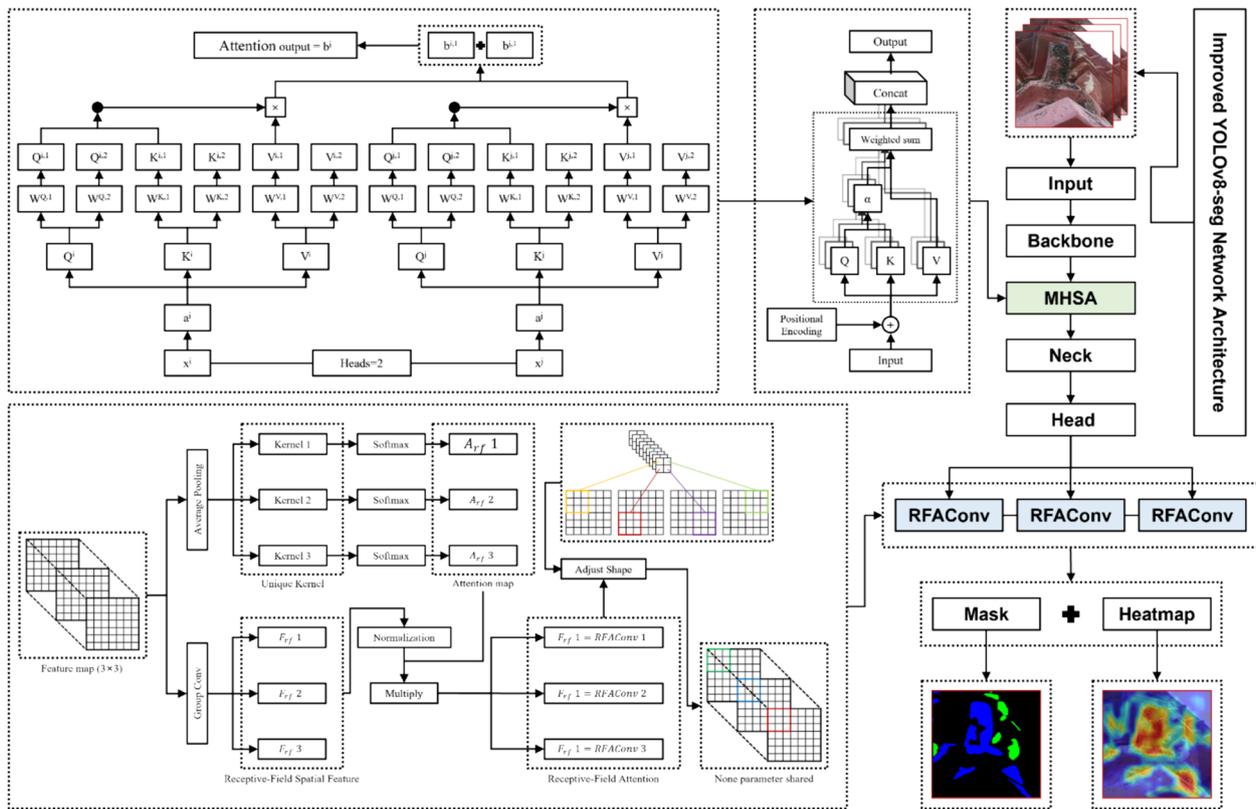


Fig. 2 Improved YOLOv8-seg model network architecture

$$\begin{aligned}
 F &= \text{Softmax}\left(g^{i \times i}(\text{AvgPool}(X))\right) \\
 &\quad \times \text{ReLU}\left(\text{Norm}\left(g^{k \times k}(X)\right)\right) \\
 &= A_{rf} \times F_{rf}
 \end{aligned}
 \tag{1}$$

where $g^{i \times i}$ represents grouped convolution with a size of $i \times i$, k represents the size of the convolution kernel, Norm represents normalization, X represents the input feature map, and F represents the result obtained by multiplying the attention map A_{rf} with the transformed receptive field spatial feature F_{rf} .

Multi-head self attention

Although the aforementioned methods can enhance the model’s ability to focus on subtle target information, YOLOv8-seg, as a convolutional neural network model, continues to encounter challenges in the extraction of global features. Therefore, the MHSA (Multi-Head Self-Attention) mechanism is integrated to enhance the model’s ability to extract global features. MHSA is a deep learning module based on the attention mechanism, initially proposed by Vaswani et al. [40] in 2017. It builds upon the self-attention mechanism by stacking

self-attention blocks in parallel to enhance the effectiveness of its layers. Each self-attention block consists of three different feature spaces: Q (query), K (key) and V (value). Each head utilizes different learned weights W^{Q_i} , W^{K_i} and W^{V_i} to generate different sets of Q^i , K^i , and V^i . This design permits disparate heads to concurrently focus on different information, including global and local information, thereby enabling the model to learn the interdependencies between image features from multiple perspectives. This module not only helps the model better understand the relationships between features but also improves the model’s segmentation ability for targets of different scales and positions. Furthermore, the Multi-Head Self Attention mechanism enables the model to establish long-range dependencies between different positions, thereby capturing the global contextual information of the targets. This, in turn, deepens the model’s understanding of relationships between different objects in the image, thus improving its ability to extract global features from images. Based on these advantages, in this study integrates the MHSA module into the backbone of the YOLOv8-seg algorithm framework to enhance the model’s ability to extract global features from images.

The computation process can be expressed in Eqs. (2) to (4) [40]:

$$MultiHead(Q, K, V) = Concat(head1, \dots, head2)W^o \tag{2}$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$ (3)

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where Q, K , and V represent query, value, and key respectively, $\frac{1}{\sqrt{d_k}}$ is the scaling factor, and d_k is the dimension of the key matrix.

Slide loss function

In most cases, models are more readily able to identify material deterioration of masonry towers, such as extensive surface peeling and mildew, which are characterized by a high quantity, clear boundaries, and prominent features. However, recognizing less frequent and smaller cracks and defects is more challenging. This discrepancy in recognition ability is reflected during the training process, as different loss values correspond to different categories. Furthermore, the uneven distribution of different material deterioration categories in the dataset results in class imbalance in segmentation. Category imbalance can cause the model to lean towards learning the simpler categories, leading to inefficient training [42] and affects the performance and generalization ability of the model. To address this issue, this study builds upon the integration of the aforementioned modules and employs the SlideLoss category-balancing function [41] to compensate for this deficiency. Presented by Yu et al. [41], SlideLoss is an adaptive category-balancing function that adjusts parameters based on the IoU values of all predicted results. It sets a threshold μ , designating categories with IoU values less than μ as difficult categories, and those greater than μ as simple categories. The algorithm

assigns greater weights to categories that are challenging to predict, thereby increasing their influence on the loss calculation. This enables the model to focus more on these challenging categories during the training process. The introduction of the SlideLoss category-balancing function allows the model to balance its attention across different categories, improving its generalization ability and overall performance. Its computation process can be expressed in Eq. (5) [41]:

$$f(x) = \begin{cases} 1 & x \leq \mu - 0.1 \\ e^{1-\mu} & \mu < x < \mu - 0.1 \\ e^{1-x} & x \geq \mu \end{cases} \tag{5}$$

where μ represents the threshold value, $x \leq \mu - 0.1$ indicates the difficult category, $\mu < x < \mu - 0.1$ represents the difficult category with assigned weighted values, and $x \geq \mu$ represents the simple category.

Experiment

Refined 3D reconstruction

The objective of the refined 3D reconstruction of architectural heritage is to combine 3D laser scanning and photogrammetry technology to accurately restore the geometric shape, details, and surface textures of architectural heritage. This process can provide valuable and authentic digital resources for the research, conservation, and exhibition of architectural heritage. The overall refined 3D reconstruction process is illustrated in Fig. 3. This study focuses on the refined 3D reconstruction of the Huiguang Tower, a national cultural heritage site located at No. 18 Huiguang Road, Lianzhou City, Guangdong Province. Built during the Song Dynasty, the tower is a nine-story hexagonal masonry structure with a total height of 49.866 m, including a spire height of 7.766 m. The main body of the tower is coated in white plaster, while the columns, brackets, beams, and doorways are covered with red plaster. Additionally, as the number of floors increases, the area of each floor gradually decreases, forming a shape of successive setback

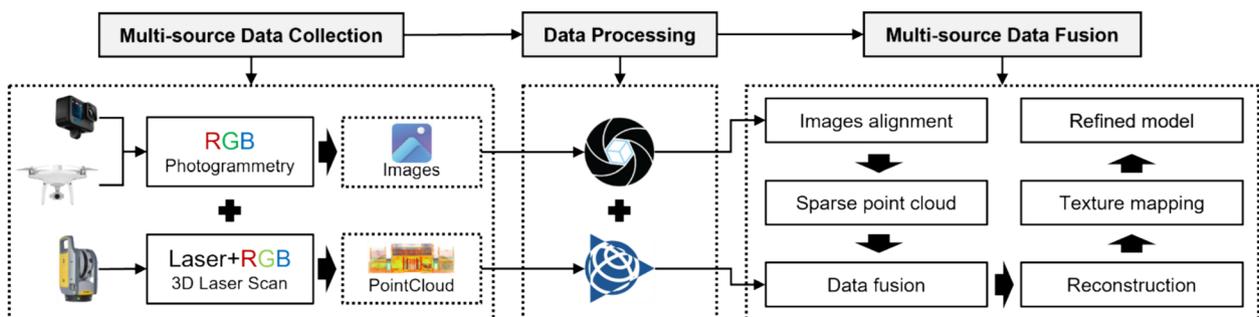


Fig. 3 Refined 3D reconstruction process based on multi-source data fusion

design. The foundation of the Huiguang Tower is consists of artificially processed red clay mixed with pebbles. The underlying rock surface exhibits significant undulations, and the shallow bedrock depth, along with geological conditions such as caves and heavily weathered limestone, have led to uneven settlement. Consequently, the tower has tilted more than 2% towards the southwest direction. The general aerial view of this tower is shown in Fig. 4c.

The collection of multi-source data involves acquiring of 3D laser point-cloud data and photogrammetric data. Known as a reality capture technology, 3D laser scanning technology captures the true external shape and internal structure of the scanned objects. In this digital survey, a ground-based station laser scanner (Trimble X7) was employed to collect 3D laser point-cloud data from the Huiguang Tower, with a total of 210 scanning stations. The resulting 3D laser point-cloud is illustrated in Fig. 4a. Photogrammetry technology holds significant application value across fields, including surveying, engineering, geographic information systems (GIS), and remote sensing, etc. It captures image data of the target and applies photogrammetry

principles for geometric derivation and calculation to obtain 3D spatial position and shape information of the target. During the process of photogrammetry-based 3D reconstruction, the algorithm calculates the distances between various points in 3D among multi-view images, enabling the reconstruction of the target's spatial morphology in the form of a 3D point-cloud. To ensure the algorithm can identify common feature points quickly and accurately, the collected images must achieve an overlap of approximately 70%. Furthermore, the increasing prevalence of digital cameras and the ongoing evolution of photogrammetry software have led to a gradual decline in the cost of photogrammetry, further promoting its widespread application in various fields. In this research, the outdoor image data of the experimental subject was captured using UAV (DJI PHANTOM 4, 5472×3648 px) oblique photography, while the indoor image data was collected by close-range photography (GoPro 11, 5568×4872 px). The resolution of the captured images is of great importance for the accuracy of 3D reconstruction, as it is directly proportional to the reconstruction's texture details and precision. The preliminary alignment of the

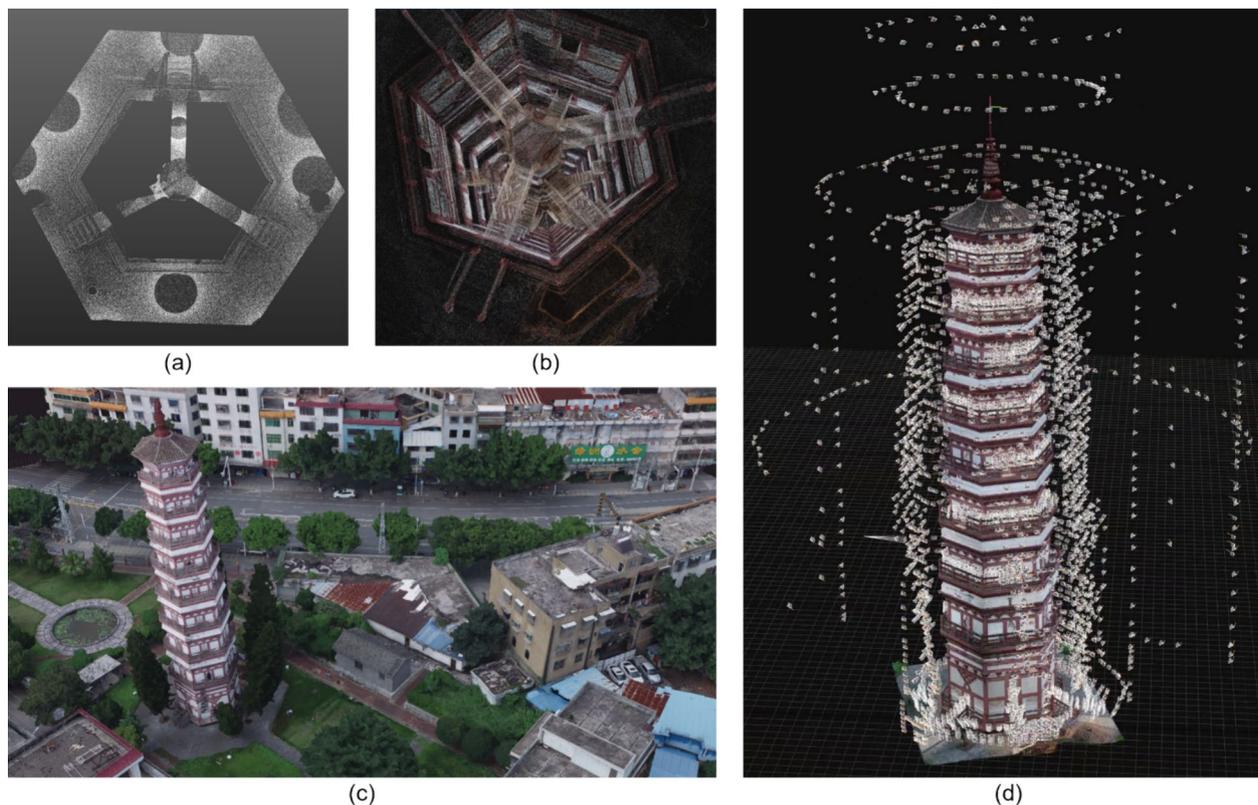


Fig. 4 Overview and multi-source data collection process of Huiguang Tower (**a** first floor slice of 3D laser scanning model of Huiguang Tower; **b** photogrammetric preliminary aligned point-cloud; **c** aerial view of photogrammetric 3D reconstruction modelling; **d** refined 3D reconstruction model based on the fusion of multi-source data)

point-cloud derived from photogrammetry is shown in Fig. 4b.

Following the collection of multi-source data for the Huiguang Tower, the 3D laser point cloud data is used as the reference control benchmark for the overall contour. High-resolution photogrammetry texture data was then integrated with it to generate the final refined 3D reconstruction model of the masonry tower, as shown in Fig. 4d. The reconstruction of this refined 3D model overcomes the challenges posed by the large scale of the tower and the limitations of on-site conditions. Additionally, the texture generated from the refined 3D reconstruction model consists of 59 high-resolution maps at 16k resolution, with each texture pixel corresponding to a spatial size of 0.45 mm. This high-precision texture ensures the accuracy and reliability of the subsequent material deterioration feature extraction process for the masonry tower.

Model training

Material deterioration definition

Architectural heritage material pathology refers to phenomena that cause abnormalities or damage to the structural safety and aesthetic value of architectural heritage due to natural and artificial forces. The classification and definition of material deterioration directly influence the effectiveness of architectural heritage restoration work. To ensure a scientific nature classification method, this study defines the material deterioration of masonry towers based on three principles: objectivity, structural rationality, and unique classification. In consideration of the research object of this study, the deterioration types are defined according to the kinds of deteriorations affecting brick and stone materials. Based on the definitions of stone deterioration from the ICOMOS glossary [43], as well as the morphological and color characteristics of masonry towers under natural conditions and the properties of their construction materials, the deterioration is categorized into four typical types: cracking, defect, mildew, and bio-disease. However, it should be noted that material deterioration can occur concurrently with multiple categories, such as the overlap of wall defects and mildew or bio-disease. Therefore, the annotation process of the dataset must be based on observed phenomena as precisely as possible through descriptive observations [44]. This is to ensure that the model is able to learn and extract deterioration characteristics with the requisite accuracy.

Dataset annotation and preprocessing

Once the annotation types of the dataset had been defined, the corresponding image data were collected from an open-source online platform, resulting in a total

collection and annotation of 1468 images. The dataset was then divided into three important subsets: training set, validation set, and test set. The training set accounted for 70% of the total data, while the remaining 30% is split into a validation set and a test set, with 20% and 10% respectively. This division helps to prevent overfitting of the model and ensures accurate predictions during the subsequent model testing. The performance of image segmentation models is significantly dependent on the quality and scale of the data. To enhance the model's generalization ability and prevent overfitting, various data augmentation techniques such as flipping, blurring, contrast adjustment, brightness adjustment, and noise addition are applied to the training set, which collectively enhance the quality and diversity of the dataset.

Model assessment metrics

In this study, four metrics were employed to quantitatively assess the performance of the model in identifying material deterioration in masonry tower, including F1 score, precision, recall, and mAP. The F1 score is the harmonic mean of precision and recall, which can be used to measure the average segmentation performance of the model for masonry tower material deterioration. Precision is the ratio of the number of true positive samples correctly classified to the number of false positive samples correctly classified. Recall is the proportion of the actual regions of material deterioration correctly identified by the model, reflecting the segmentation capability of the model. mAP is the mean of the average precision for multiple categories, serving as an important indicator to evaluate the overall performance of the model. The specific calculation methods for these assessment metrics are as follows:

$$F1score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$mAP = \left(\frac{1}{K} \right) \sum_{i=0}^k (AP)_i \quad (9)$$

where TP , FP , and FN represent true positives, false positives, and false negatives. k represents the number of segmentation categories, i represents the average precision calculated for category i out of k categories, and AP stands for the mean average precision for the category. By calculating these assessment metrics, the performance of the model in the task of segmenting material

Table 2 Comparative experimental results of different models

Model	F1 score	Precision	Recall	mAP50	mAP50-95
Gold-YOLO [45]	0.583	0.645	0.532	0.570	0.350
YOLOv5x-seg [47]	0.611	0.683	<u>0.553</u>	0.606	0.374
ASF-YOLO [46]	0.619	<u>0.732</u>	0.537	0.605	<u>0.382</u>
YOLOv8x-seg [48]	0.620	0.728	0.539	0.598	0.378
YOLOv9c-seg [49]	<u>0.627</u>	0.731	0.549	<u>0.607</u>	0.379
Ours	0.638	0.753	0.553	0.619	0.396

The highest value is highlighted in bold, and the second best value is underlined

deterioration of masonry towers can be evaluated from different points of view.

Comparison verification

To evaluate the performance of the enhanced YOLOv8-seg model in the context of material deterioration segmentation in masonry towers, five image segmentation models based on the YOLO algorithm framework were selected for comparison, including Gold-YOLO [45], ASF-YOLO [46], YOLOv5x-seg [47], YOLOv8x-seg [48], and YOLOv9c-seg [49]. All models were evaluated under identical configuration environments and using the same image dataset, with the maximum number of parameters utilized for training. The training environment is configured uniformly with CUDA 11.7 and Pytorch 1.13.1. The Stochastic Gradient Descent (SGD) optimizer is employed with momentum set to 0.9, weight decay

at 0.0005, and a batch size of 8. The input image size is standardized to 640×640 pixels. Furthermore, all experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU workstation equipped with 24 GB of memory.

The training results are presented in Table 2. The comparative analysis was conducted based on the highest F1 score, with the highest value highlighted in bold and the second-best value underlined. The experimental results demonstrate that the enhanced model achieved superior performance across all metrics. Specifically, compared to the baseline model, the improved model demonstrates an increase of 2.82% in the F1 score and 3.32% in precision. Additionally, it exhibits improvements of 3.39% and 4.55% in the mAP50 and mAP50-95 metrics, respectively.

In the subsequent visualization evaluation, four images containing deterioration information at different scales were selected for segmentation mask prediction, as shown in Fig. 5. The improved model, integrating the RFACnv, MHSA and SlideLoss category-balancing function, exhibits enhanced precision in delineating details near the object boundaries when predicting masks for large-scale defects, mold growth, and small-scale cracks, closely aligning with annotated masks. To provide further intuitive performance analysis, the Grad-CAM algorithm is employed to visualize the trained model. Grad-CAM uses gradient information to evaluate the importance of various spatial locations in convolutional layers [50], revealing the model's attention on different prediction categories. As illustrated in Fig. 6, the improved model not only accurately identifies and

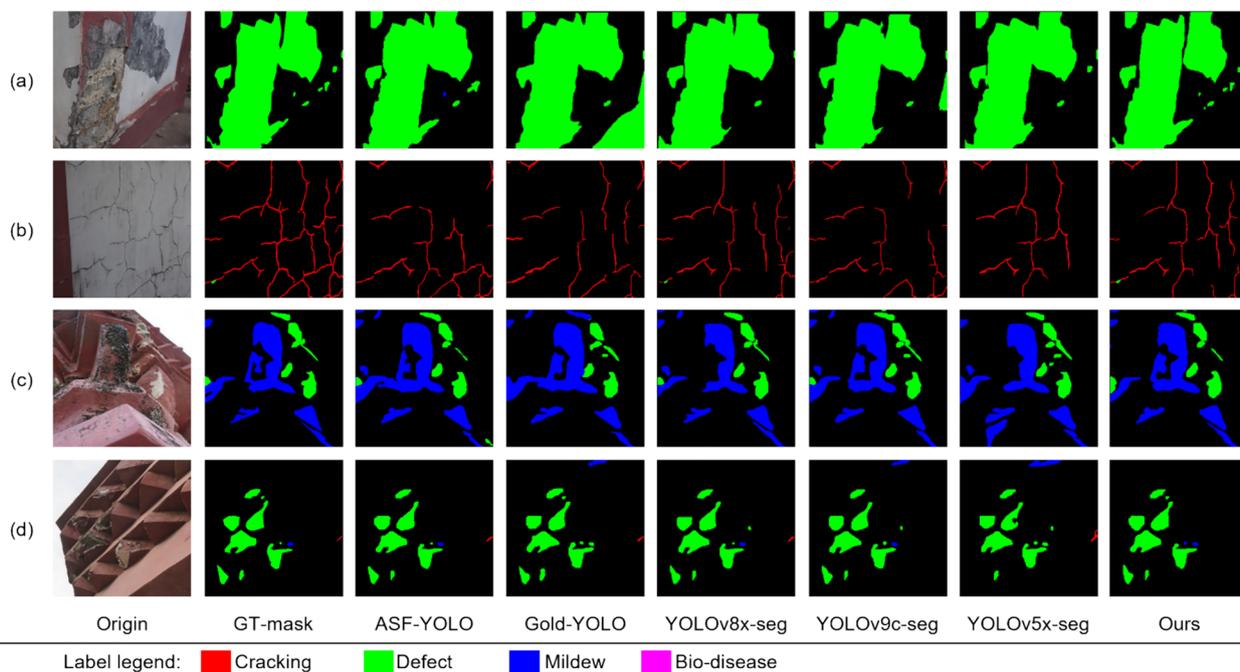


Fig. 5 Segmentation mask visualization results

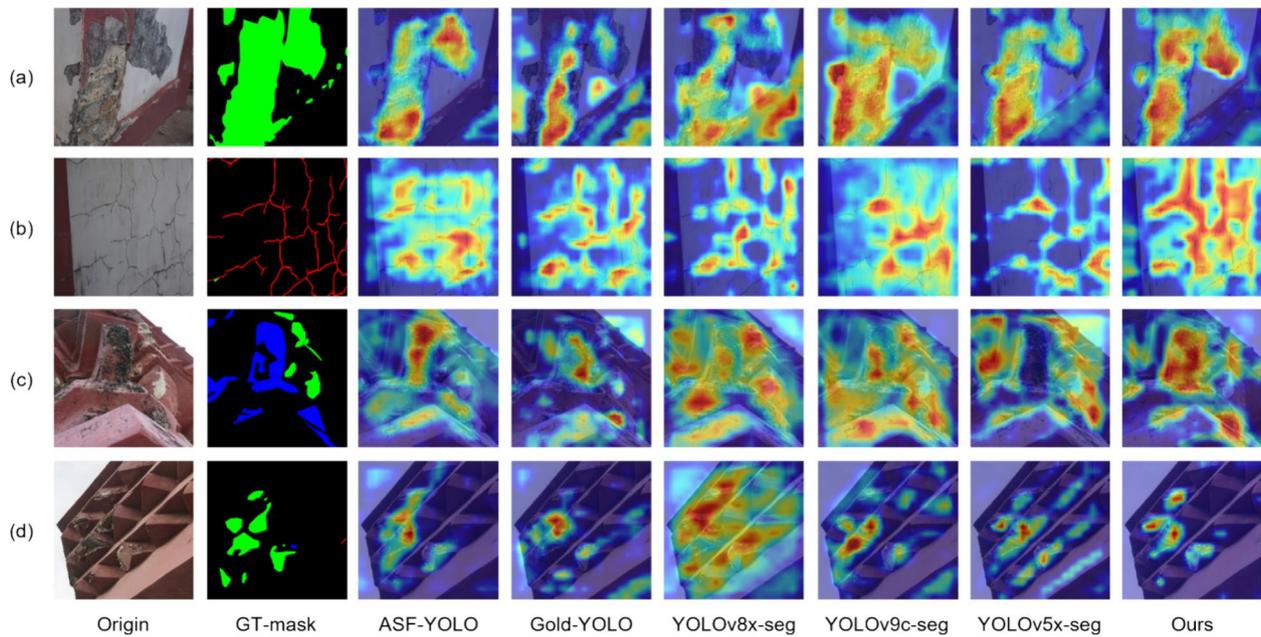


Fig. 6 Grad-CAM visualization results

highlights the key areas of large-scale predicted objects but also captures linear aggregation features that consistent with the morphology of long and small-scale cracks, and exhibits reduced noise artifacts in output predictions. A series of comparative experiments demonstrate that the improved model performs better in the segmentation task of material deterioration in masonry towers, enabling more accurate identification of such deterioration.

Ablation experiment

To validate the effectiveness of the incorporated modules in this study, an ablation experiment was performed, as shown in Table 3, where "Base" represents the original YOLOv8-seg model framework. The experimental results demonstrate that the RFACnv and MHSA modules improve the F1 score metric by 0.48% and 0.32%, respectively, and the precision metric by 0.14% and 3.32%, respectively. This indicates that both modules can enhance the base model's ability to extract detailed

features and global characteristics. Moreover, integrating these two modules into the base model framework demonstrates a collective enhancement in segmentation performance, highlighting their complementary benefits.

Global feature extraction and visualization

To further analyze the characteristics of material deterioration in Huiguang Tower, this study performed intelligent extraction and visualization of global features from the output results. Given the considerable size of the Huiguang Tower, with an aspect ratio of approximately 4.5:1, the orthographic elevation generated by the refined 3D reconstruction model require uniform slide-cropping before image batch prediction. This adjustment is necessary to match the initial input size of 640×640 pixels set during the model training. Subsequently, these cropped images are batch-predicted using the optimal weights validated by multiple rounds of accuracy checks. The predicted masks were then integrated and stitched together

Table 3 Results of ablation experiments

Base	Modules		Metrics				
	RFACnv	MHSA	F1 Score	Precision	Recall	mAP50	mAP50-95
√			0.620	0.728	0.539	0.598	0.378
√	√		<u>0.623</u>	0.729	<u>0.544</u>	<u>0.607</u>	0.378
√		√	0.622	<u>0.753</u>	0.531	0.601	<u>0.382</u>
√	√	√	0.638	0.753	0.553	0.619	0.396

The highest value is highlighted in bold, and the second best value is underlined

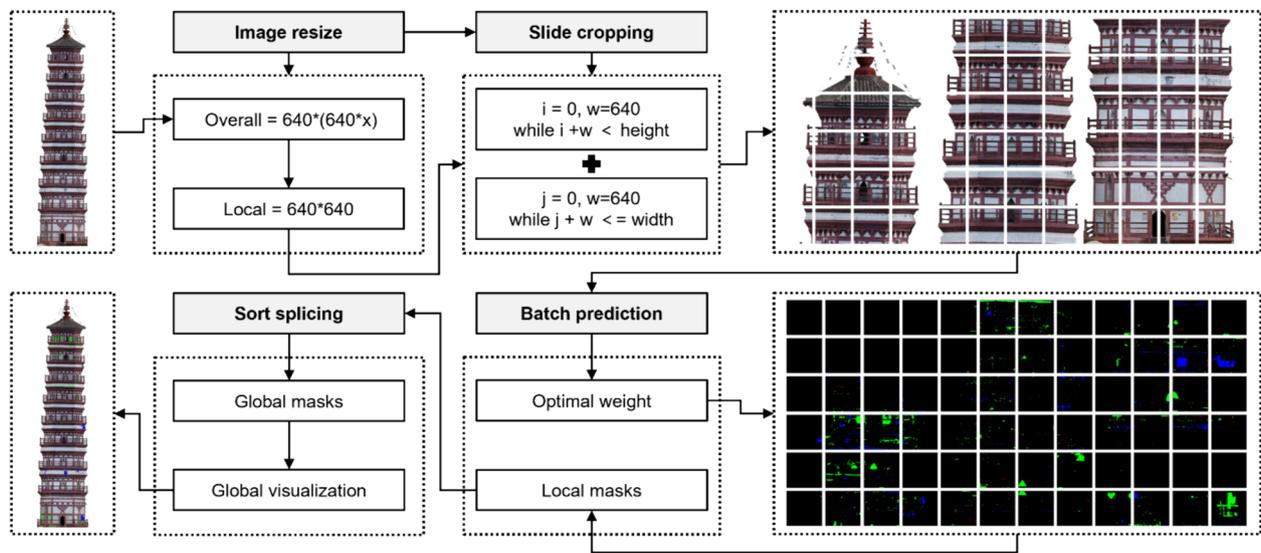


Fig. 7 Global visualization process for material deterioration in Masonry Towers

based on sorted prediction results to precisely align with the original orthographic elevation. This process facilitates the global visualization of material deterioration in Huiguang Tower. The overall global visualization process is illustrated in Fig. 7.

Comprehensive assessment

Upon completion of the global visualization of material deterioration in the Huiguang Tower, it is necessary to perform comprehensive quantitative and qualitative assessments of the output results. The quantitative analysis involves statistical data analysis of the number, area, and distribution of material deterioration, while the qualitative analysis includes preliminary judgments on the formation, damage trends, and potential risks associated with these deteriorations. These comprehensive assessments facilitate the identification and comprehension of the prevailing trends of material deterioration in the Huiguang Tower, as well as providing crucial guidance and a foundation for future decision-making regarding prevention and restoration.

Figure 9 presents the results of global mask visualization of deteriorations on the six elevations of the Huiguang Tower. The color legend facilitates intuitive observation of the spatial distribution, coverage, and development trends of material deterioration on the tower’s surface. According to the general plan in Fig. 8, elevations three, four, and five are south-facing. It is clearly observed in Fig. 9 that these three south-facing elevations, which are more susceptible to sunlight and high temperatures, exhibit a greater prevalence of defects and mildew issues on the walls. Furthermore,

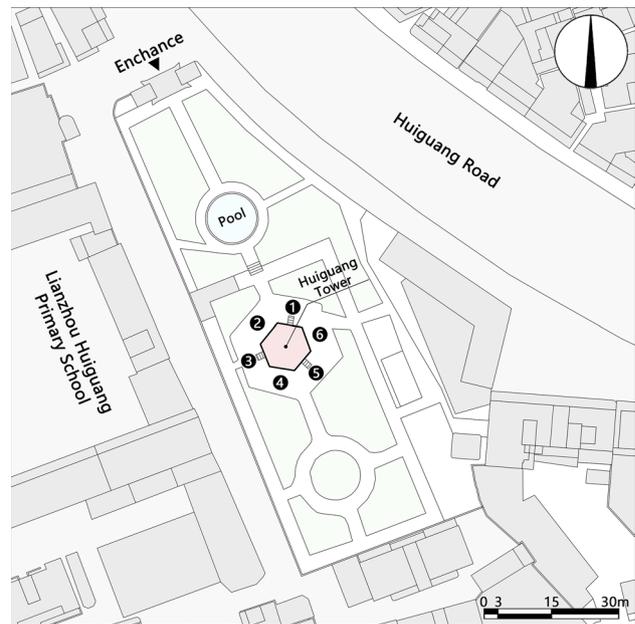


Fig. 8 General plan of Huiguang Tower

the extensive growth of mildew at the bottom of these three south-facing elevations results in constant dampness, facilitating moisture penetration into the wall materials. This, in turn, causes the formation of minor cracks and defects on the surrounding walls. Over-time, this process causes the façade materials to gradually lose their adhesive properties, eventually resulting in plaster detachment and increase mildew formation. The sequential material deterioration can weaken the

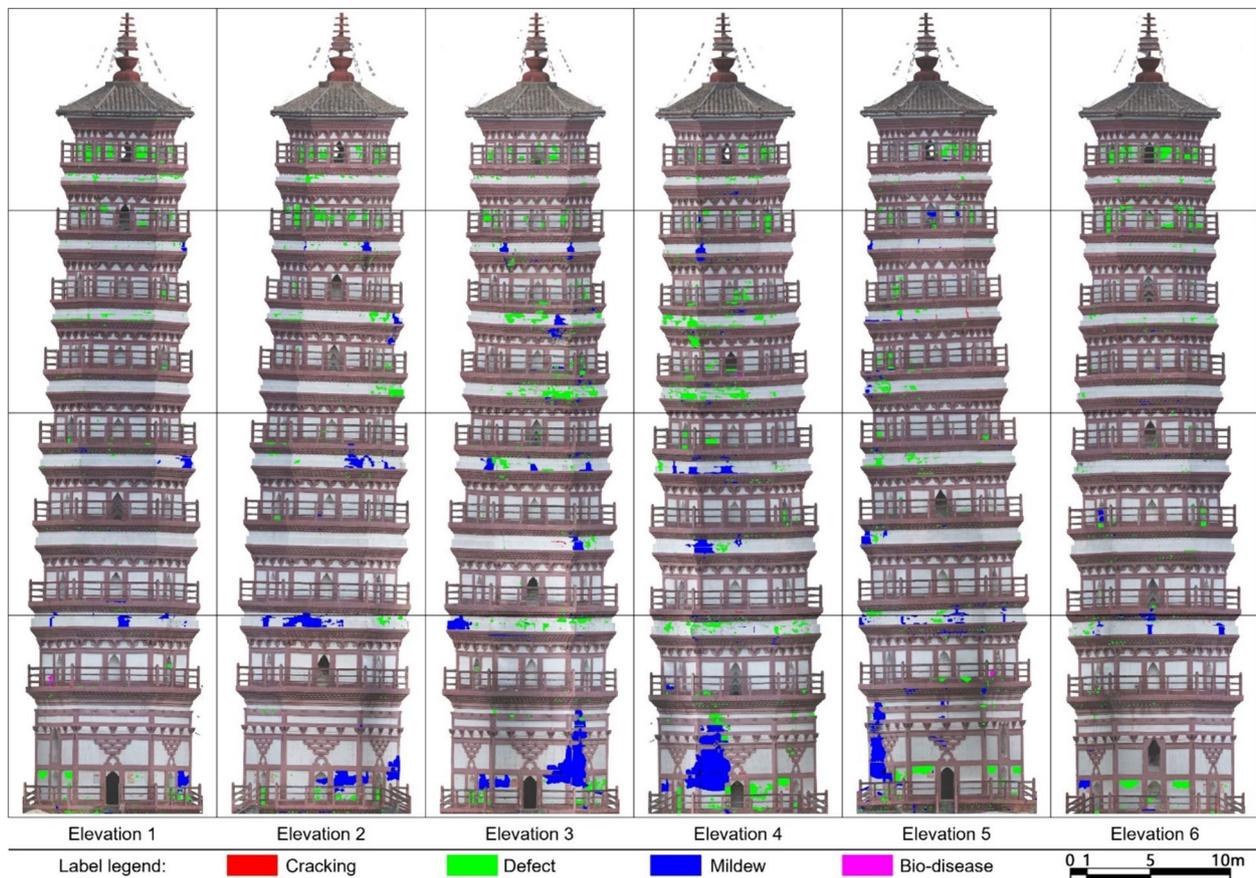


Fig. 9 Global visualization results of material deterioration on six elevations of the Hui-guang Tower

Table 4 Quantitative statistics on material deterioration of Hui-guang Tower

	Cracking		Defect		Mildew		Bio-disease		Total	
	Number	Area (m ²)	Number	Area (m ²)	Number	Area (m ²)	Number	Area (m ²)	Number	Area (m ²)
Elevation 1	27	0.04	664	6.83	73	3.15	2	0.17	766	10.19
Elevation 4	47	0.13	1110	14.49	247	11.97	5	0.05	1413	26.63
Total	74	0.17	1774	21.32	320	15.12	7	0.21	2175	36.82

structural integrity of the walls, altering their load distribution and increasing overall instability. Regarding significant defects on the upper levels of the walls, their prevalence is likely due to their higher exposure to wind and other external forces, which intensify the erosion of wall materials and exacerbate defects on the upper walls.

Tables 4, 5, 6 and Fig. 10 present the quantitative results of the global features of material deterioration in the Hui-guang Tower, using elevation 1 and 4 as examples. To provide a detailed presentation of the quantitative results, this research subdivides the statistical data on material

deterioration for each floor of the masonry tower. In terms of quantity statistics, the number of material deterioration instances on each elevation can be counted by batch predicting the slide-cropped orthographic images. For area statistics, the multi-category masks generated from the batch prediction are used to calculate the percentage of area ratio affected by material deterioration on each elevation, thereby quantifying the extent of deterioration spread. Since the research object is a hexagonal masonry tower, the non-vertical projection areas on both sides of each elevation can be calculated as half of the front projection area. The actual area conversion

Table 5 Material deterioration number distribution of each layer in Huiguang Tower (number)

	Elevation 1				Elevation 4				Total
	Cracking	Defect	Mildew	Bio-disease	Cracking	Defect	Mildew	Bio-disease	
F1	7	133	14	0	3	127	39	2	325
F2	20	41	0	2	9	78	17	1	168
F3	0	66	5	0	14	187	30	0	302
F4	0	73	14	0	13	159	48	0	307
F5	0	100	38	0	6	176	42	0	362
F6	0	64	1	0	1	76	16	2	160
F7	0	40	1	0	0	82	18	0	141
F8	0	41	0	0	1	142	37	0	221
F9	0	106	0	0	0	83	0	0	189
Total	27	664	73	2	47	1110	247	5	2175

Table 6 Material deterioration area distribution of each layer in Huiguang Tower (m²)

	Elevation 1				Elevation 4				Total
	Cracking	Defect	Mildew	Bio-disease	Cracking	Defect	Mildew	Bio-disease	
F1	0.02	1.61	1.20	0.00	0.04	2.06	7.82	0.01	12.76
F2	0.02	0.17	0.70	0.17	0.03	2.44	0.46	0.01	3.99
F3	0.00	0.16	0.01	0.00	0.02	0.40	1.24	0.00	1.83
F4	0.00	0.27	1.01	0.00	0.01	0.77	1.55	0.00	3.61
F5	0.00	0.27	0.01	0.00	0.01	2.87	0.11	0.00	3.27
F6	0.00	0.78	0.00	0.00	0.01	3.02	0.15	0.03	3.98
F7	0.00	0.39	0.23	0.00	0.00	0.87	0.52	0.00	2.02
F8	0.00	1.48	0.00	0.00	0.01	1.27	0.12	0.00	2.89
F9	0.00	1.69	0.00	0.00	0.00	0.79	0.00	0.00	2.48
Total	0.04	6.83	3.15	0.17	0.13	14.49	11.97	0.05	36.82

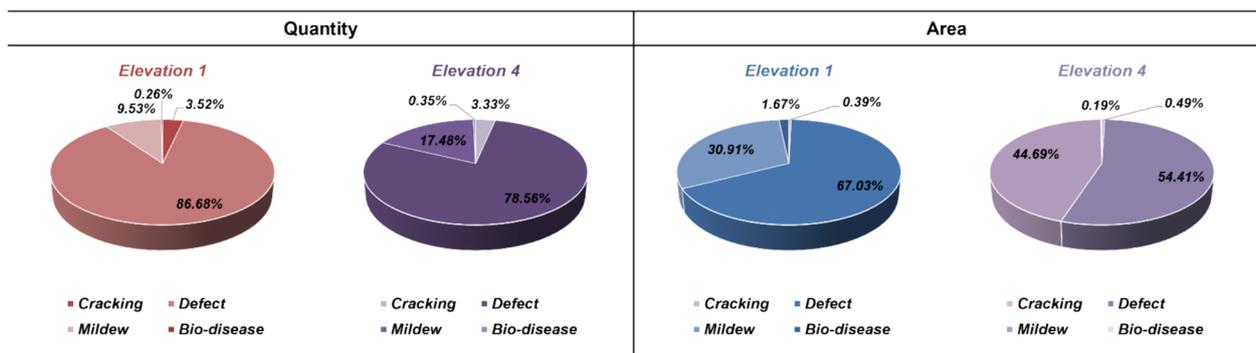


Fig. 10 Number and area ratio of material deterioration in Huiguang Tower

can be obtained by comparing the real area measured on-site with the pixel count of the corresponding vertical projection from the 3D reconstruction. A comprehensive quantitative analysis reveals that defect issues are the most prevalent in the Huiguang Tower, followed by

mildew, while cracking and bio-disease are relatively rare. A comparative analysis between the first and the fourth elevations reveals that the number and area of material deterioration on the fourth elevation are approximately twice and 2.5 times those of the first elevation.

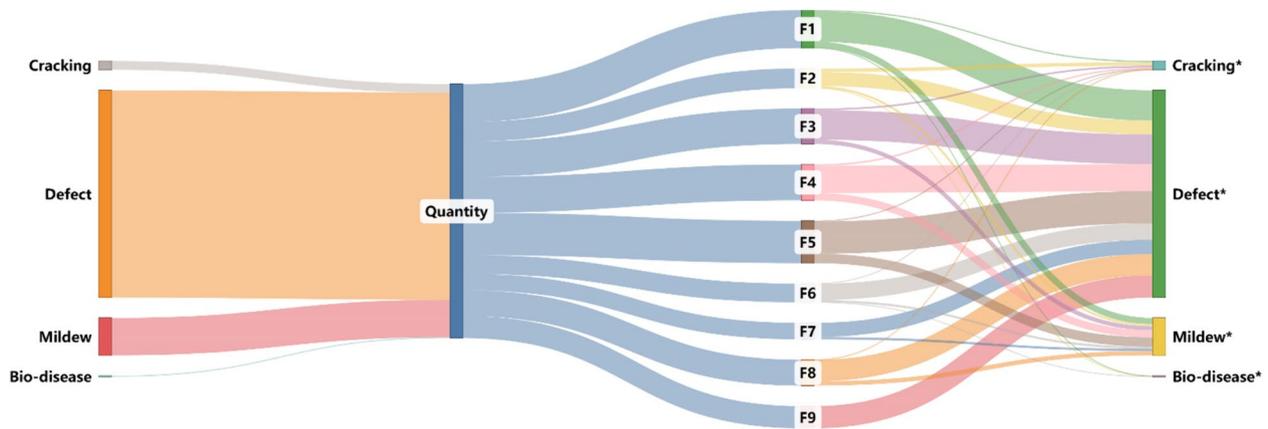


Fig. 11 Sankey chart of Huiguang Tower’s material deterioration number distribution

To visually present the global features of the material deterioration predicted by the model, this study employs Sankey diagrams for data visualization. As illustrated in Figs. 11 and 12, Sankey diagrams, with their distinctive data flow display method, clearly depict the distribution and flow of material deterioration data across each floor of the masonry tower. They effectively present the allocation and interaction relationships between different types of data within a complex network, aiding in the deep exploration of critical information. On the left side of the diagrams, the total quantity and area of material deterioration on each floor are displayed, while the right side indicates the distribution of various types of material deterioration data across the floors, forming an intuitive four-level data flow approach. The extended branches in the diagram indicate the direction of data flow, with the width of the branches indicating the volume of the corresponding information flow. The total width of the branches at both ends is equal. The general distribution of data indicates that the quantity of material

deterioration is relatively consistent across each floor of the masonry tower, while the area of material deterioration on the bottom floor is nearly three to six times that of other floors. This suggests that the bottom floor is more susceptible to the combined effects of multiple factors, including ground dampness, water accumulation, heavy loads, and foundation settlement, which may lead to extensive material deterioration. A quantitative analysis of the relationships between data types reveals that although the quantity of defects on the surface exceeds the number of mildew issues, both are similar in terms of area. Therefore, defects are characterized by a large quantity but small area distribution, while mildew shows a lower quantity but larger area. From the perspective of individual data flow, the distribution quantity of defects and mildew is similar on each floor. However, large-scale defects are primarily distributed in the lower and middle-upper floors, while large-scale mildew is primarily found on the bottom and mid-level floors. In terms of bio-disease, the corners of the tower cloister tend to accumulate

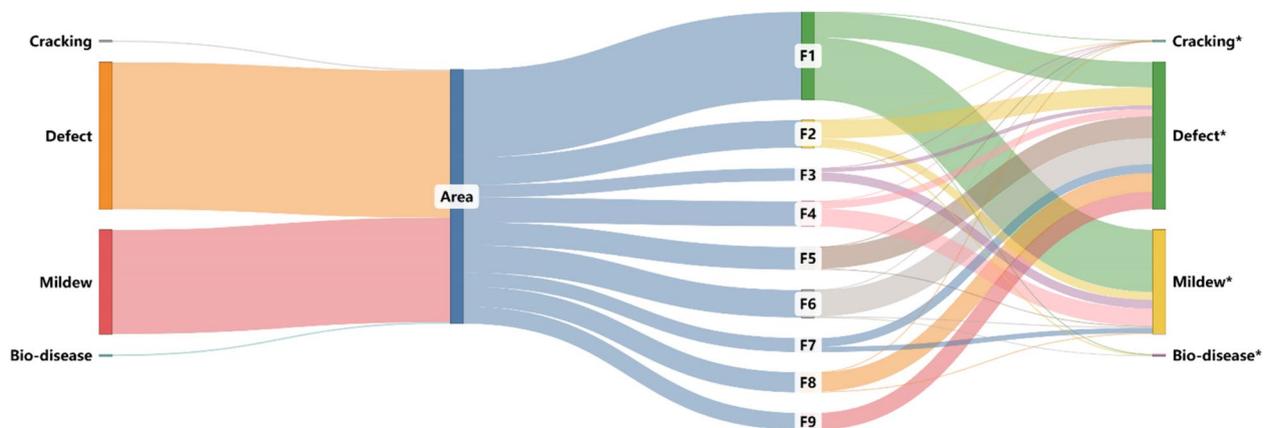


Fig. 12 Sankey chart of Huiguang Tower’s material deterioration area distribution

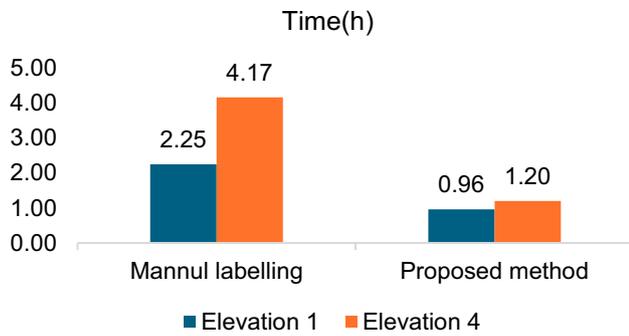


Fig. 13 Time consumption of different labeling methods for Huiguang Tower material deterioration

moisture, creating a relatively moist environment conducive to plant growth. Additionally, the bottom floor is more susceptible to ground dampness, resulting in the majority of bio-disease being concentrated on the bottom floor of the masonry tower. The prolonged invasion of bio-disease can lead to the formation of cracks and holes on the wall surface, potentially infiltrating the wall's

Table 7 Time consuming of global feature extraction and visualization method proposed in this research (s)

	Elevation 1	Elevation 4	Total
Batch cropping (s)	0.251	0.457	0.708
Batch prediction (s)	58.206	68.131	126.157
Batch splicing (s)	1.876	1.321	3.197
Visual post-processing (s)	3399.84	4264.92	7664.76

interior structure. This leads to soil and moisture accumulation on the wall surface, significantly increasing the weight and load on both the walls and the overall structure, thereby affecting the overall stability and structural integrity of the masonry tower. Furthermore, the majority of cracks are also relatively concentrated on the bottom floor, which is related to the higher loads borne by the bottom floor and foundation settlement. Instabilities or uneven foundation settlement can cause uneven loads on the walls, making the walls of the lower floor more prone to cracks. These analysis results not only provide a macro view of the distribution of material deterioration in the Huiguang Tower, enabling the identification of high-incidence areas and the severity of material deterioration, but also offer crucial data support and a decision-making basis for further preservation and restoration efforts.

Discussion

The validation and analysis results of the intelligent assessment and visualization method proposed in this study are presented in Figs. 13, 14, and Table 7. Figure 13 compares the time consumption differences between the intelligent extraction and visualization method for the global features of material deterioration in the Huiguang Tower and the traditional manual annotation method. As illustrated in Table 7, the proposed method requires a mere 0.708 s for batch sliding cropping, 126.157 s for batch prediction, and 3.197 s for batch stitching of the entire image from the vertical projection images generated by the refined 3D reconstruction model of the

	Elevation 1			Elevation 4		
GT-mask (pixels)	16651	155857	123182	149204	74914	16675
Prediction map						
Prediction mask (pixels)	16771	166311	115021	144256	72364	15367
Overlap ratio	99.28%	93.71%	93.37%	96.68%	96.60%	92.16%
GT-mask (pixels)	264410	58129	24519	137517	52016	47419
Prediction map						
Prediction mask (pixels)	264803	56397	24915	132632	50621	49674
Overlap ratio	99.85%	97.02%	98.41%	96.45%	97.32%	95.46%
Average ratio	96.94%			95.78%		
Label legend:	Cracking	Defect	Mildew	Bio-disease	GT-mask	

Fig. 14 Validation of the accuracy of the proposed method

Huiguang Tower. Including post-processing and adjustments for visualization, the total time for the entire process is only 2.16 h, which is 66.36% more efficient than the 6.42 h required by the manual annotation method. This improvement in efficiency effectively enhances preliminary investigation of material deterioration in masonry towers, making it a more viable and practical approach for large-scale and detailed assessments.

Figure 14 presents the results of the accuracy validation for material deterioration feature extraction in masonry towers based on the intelligent assessment system proposed in this study. During the validation process, local vertical projection images obtained by batch cropping from Elevation 1 and Elevation 4 are selected, covering various scales and types of material deterioration. The results indicate that the proposed assessment system excels in locating and segmenting areas of material deterioration in masonry towers. The overlap between the predicted masks generated by the system and the original annotated masks generally exceeds 92%, with an average overlap of 95.78%, demonstrating the high accuracy of the method in extracting material deterioration features in masonry towers. Despite limitations in data quality, which may result in some instances of missed deterioration detection, classification errors or false positives in the predicted masks generated by the improved image segmentation model during validation, the approach based on refined 3D reconstruction technology and machine vision demonstrates higher accuracy in most cases compared to manual identification under field conditions. This indicates its practical value. Furthermore, these prediction errors can be further controlled through in-depth research. The inconsistent quality of datasets collected from open-source platforms may partially affect the performance of the model. To improve the accuracy and generalizability of the model, future research will focus on improving the quality of the datasets. This includes incorporating high-resolution 3D reconstruction images with refined feature details to enrich the training data, thereby reducing the occurrence of classification errors and false positives.

In conclusion, the intelligent assessment system proposed in this paper effectively addresses the issues of efficiency and accuracy in the investigation of material deterioration in masonry towers. By employing an intelligent methodology for the global feature extraction and visualization, along with comprehensive quantitative and qualitative assessments, it offers clear and intuitive data support for a deeper understanding of the root causes, distribution patterns, damage trends, and severity of material deteriorations in masonry towers. This approach not only improves the efficiency of investigating and documenting the preservation status of masonry towers but

also provides valuable guidance for subsequent protection and restoration efforts. An in-depth analysis of the characteristics of material deterioration allows us to infer the degradation mechanisms of local structures within the masonry tower, thereby providing a scientific basis for formulating restoration interventions and sustainable protection plans. This method can significantly mitigate the occurrence and progression of material deterioration, reduce long-term repair costs, and extend the service life of the masonry towers. Consequently, it enables them to continue playing their important roles in public use, social education, and cultural heritage conservation, maintaining their historical value. Therefore, the methodologies and findings of this study have significant practical and applied value for the protection and management of masonry tower architectural heritage, offering an innovative solution for the digital preservation of architectural heritage.

Conclusion

To address the current issues of time-consuming and labor-intensive inspections of material deterioration on the surfaces of masonry towers, this study constructs an intelligent assessment system that integrates an improved YOLOv8-seg machine vision image segmentation model with refined 3D reconstruction technology of architectural heritage. By optimizing the YOLOv8-seg base model, this system improves its ability to capture both fine-scale features and global characteristics of material deterioration in masonry towers. Furthermore, by complementing it with image processing methods for global visualization of large-scale objects, this research establishes a comprehensive intelligent assessment process that includes “deterioration feature extraction—global visualization—quantitative and qualitative comprehensive assessment”. This system improves the objectivity and efficiency of identifying and analyzing material deterioration in masonry towers, addressing the limitations of existing research that relies solely on local image data for feature extraction and lacks in-depth analysis. The utilization of automated and intelligent methodologies in current condition surveys serves to minimize the potential for errors associated with manual visual inspection, while simultaneously reducing the associated workload. This approach provides a robust scientific data foundation for the protection and restoration of masonry tower architectural heritage. Experimental results demonstrate that the incorporation of the RFACnv module, the MHSA module, and the SlideLoss category-balancing function into the YOLOv8-seg algorithm framework enhances its ability to locate and segment fine-scale objects and global features, making it suitable for the intelligent identification of multiple categories and

scales of material deterioration in masonry towers. The improved model demonstrates enhanced capabilities for extracting material deterioration features in masonry towers, with increases of 3.39% and 4.45% in the mAP50 and mAP50-95 metrics, respectively, compared to the baseline model. Additionally, the efficiency of global feature extraction and visualization of material deterioration using the proposed method increased by 66.36% compared to existing manual annotation methods, with an average recognition accuracy of 95.78%, showing a more efficient and accurate solution for surveying material deterioration in masonry towers. Furthermore, the subsequent comprehensive quantitative and qualitative assessment of material deterioration data provides new insights and methods to understand the sources, distribution, damage trends, and severity of material deterioration on masonry towers. The research also validates the practicality and feasibility of machine vision technology in the preservation of architectural heritage, offering digital technology support for the future protection and restoration of masonry towers.

However, the dataset used in this study is relatively limited in size, with only four primary categories of deterioration defined for segmentation, lacking detailed subcategories and differentiation of various deterioration severity levels. Additionally, the quality of datasets collected from open-source platforms is inconsistent. To address these limitations, future work will focus on expanding the scale and diversity of the datasets, particularly by integrating high-resolution 3D reconstruction images with refined feature details. The objective of this approach is to enhance the effectiveness of the model during training, improve its generalization capabilities and increase the accuracy of identification. Despite the improved model demonstrating satisfactory performance in the intelligent identification of material deterioration in masonry towers, it has a considerable number of parameters that require significant computational resources. Consequently, future work will focus on compressing and pruning the model to reduce the size of the parameters and the computational overhead while further enhancing the performance of the model to make it suitable for a broader range of application scenarios. Finally, it is necessary to further explore methods for global risk diagnosis and continuous monitoring of masonry towers based on the existing intelligent assessment system for material deterioration, providing more comprehensive and suggestive data analysis for subsequent protection, restoration and long-term maintenance of masonry towers.

Acknowledgements

Not applicable.

Author contributions

YD mainly contributed to the development of research ideas, experimental methods and suggestions; JZ contributed to the overall experiments, data curation and manuscript writing. All authors approved the final manuscript.

Funding

No funding was received for this work.

Availability of data and materials

The data used and analyzed during the study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 31 May 2024 Accepted: 10 July 2024

Published online: 23 July 2024

References

- Mishra M, Lourenço PB. Artificial intelligence-assisted visual inspection for cultural heritage: state-of-the-art review. *J Cult Herit.* 2024;66:536–50. <https://doi.org/10.1016/j.culher.2024.01.005>.
- Nugraheni DMK, Nugroho AK, Dewi DIK, Noranita B. Deca Convolutional Layer Neural Network (DCL-NN) method for categorizing concrete cracks in heritage building. *Int J Adv Comput Sci.* 2023;14(1):722–30. <https://doi.org/10.14569/UACSA.2023.0140180>.
- Li RX, Geng GH, Wang XZ, Qin YL, Liu YY, Zhou PB, et al. LBCapsNet: a lightweight balanced capsule framework for image classification of porcelain fragments. *Herit Sci.* 2024;12(1):133. <https://doi.org/10.1186/s40494-024-01250-0>.
- Armesto-González J, Riveiro-Rodríguez B, González-Aguilera D, Rivas-Brea MT. Terrestrial laser scanning intensity data applied to damage detection for historical buildings. *J Archaeol Sci.* 2010;37(12):3037–47. <https://doi.org/10.1016/j.jas.2010.06.031>.
- Del Pozo S, Herrero-Pascual J, Felipe-García B, Hernández-López D, Rodríguez-González P, González-Aguilera D. Multispectral radiometric analysis of façades to detect pathologies from active and passive remote sensing. *Remote Sens-Basel.* 2016;8(1):80. <https://doi.org/10.3390/rs8010080>.
- Valença J, Gonçalves L, Júlio E. Damage assessment on concrete surfaces using multi-spectral image analysis. *Constr Build Mater.* 2013;40:971–81. <https://doi.org/10.1016/j.conbuildmat.2012.11.061>.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* 2017;60(6):84–90. <https://doi.org/10.1145/3065386>.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint.* 2014. <https://doi.org/10.48550/arXiv.1409.1556>.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv preprint.* 2016. <https://doi.org/10.48550/arXiv.1512.03385>.
- Kwon D, Yu J. Automatic damage detection of stone cultural property based on deep learning algorithm. *Int Arch Photogramm.* 2019;42–2(W15):639–43. <https://doi.org/10.5194/isprs-archives-XLII-2-W15-639-2019>.
- Mansuri LE, Patel D. Artificial intelligence-based automatic visual inspection system for built heritage. *Smart Sustain Built Environ.* 2022;11(3):622–46. <https://doi.org/10.1108/SASBE-09-2020-0139>.
- Pathak R, Saini A, Wadhwa A, Sharma H, Sangwan D. An object detection approach for detecting damages in heritage sites using 3-D point clouds

- and 2-D visual data. *J Cult Herit.* 2021;48:74–82. <https://doi.org/10.1016/j.culher.2021.01.002>.
13. Yan LA, Chen YL, Zheng L, Zhang Y. Application of computer vision technology in surface damage detection and analysis of shedthn tiles in China: a case study of the classical gardens of Suzhou. *Herit Sci.* 2024;12(1):72. <https://doi.org/10.1186/s40494-024-01185-6>.
 14. Rout NK, Dutta G, Sinha V, Dey A, Mukherjee S, Gupta G. Improved Pot-hole Detection Using YOLOv7 and ESRGAN. arXiv preprint. 2023. <https://doi.org/10.48550/arXiv.2401.08588>.
 15. Zhang ZY, Zhang H, Hu J, Sfarra S, Mostacci M, Wang Y, et al. Defect detection: an improved YOLOX network applied to a replica of "The Birth of Venus" by Botticelli. *J Cult Herit.* 2023;62:404–11. <https://doi.org/10.1016/j.culher.2023.06.018>.
 16. Mishra M, Barman T, Ramana G. Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage. *J Civ Struct Heal Monit.* 2024;14(1):103–20. <https://doi.org/10.1007/s13349-022-00643-8>.
 17. Bruno S, Galantucci RA, Musicco A. Decay detection in historic buildings through image-based deep learning. *Vitruvio.* 2023;8:6–17. <https://doi.org/10.4995/vitruvio-jats.2023.18662>.
 18. Xu XY, Zhao M, Shi PX, Ren RQ, He XH, Wei XJ, et al. Crack detection and comparison study based on faster R-CNN and mask R-CNN. *Sensors.* 2022;22(3):1215. <https://doi.org/10.3390/s22031215>.
 19. Kim B, Cho S. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Struct Control Hlth.* 2019;26(8): e2381. <https://doi.org/10.1002/stc.2381>.
 20. Wang WK, Shi Y, Zhang J, Hu LJ, Li S, He D, et al. Traditional village building extraction based on improved mask R-CNN: a case study of Beijing, China. *Remote Sens.* 2023;15(10):2616. <https://doi.org/10.3390/rs15102616>.
 21. Hou ML, Huo DX, Yang Y, Yang S, Chen HW. Using mask R-CNN to rapidly detect the gold foil shedding of stone cultural heritage in images. *Herit Sci.* 2024;12(1):46. <https://doi.org/10.1186/s40494-024-01158-9>.
 22. Hatir E, Korkanç M, Schachner A, Ince I. The deep learning method applied to the detection and mapping of stone deterioration in open-air sanctuaries of the Hittite period in Anatolia. *J Cult Herit.* 2021;51:37–49. <https://doi.org/10.1016/j.culher.2021.07.004>.
 23. Altaweel M, Khelifi A, Shana'ah MM. Monitoring looting at cultural heritage sites: applying deep learning on optical unmanned aerial vehicles data as a solution. *Soc Sci Comput Rev.* 2024;42(2):480–95. <https://doi.org/10.1177/08944393231188471>.
 24. Liu ZQ, Cao YW, Wang YZ, Wang W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Automat Constr.* 2019;104:129–39. <https://doi.org/10.1016/j.autcon.2019.04.005>.
 25. Banasiak PZ, Berezowski PL, Zapłata R, Mielcarek M, Duraj K, Stereńczak K. Semantic segmentation (U-Net) of archaeological features in airborne laser scanning—example of the Białowieża forest. *Remote Sens.* 2022;14(4):995. <https://doi.org/10.3390/rs14040995>.
 26. Stoean R, Bacanin N, Stoean C, Ionescu L, Atencia M, Joya G. Computational framework for the evaluation of the composition and degradation state of metal heritage assets by deep learning. *J Cult Herit.* 2023;64:198–206. <https://doi.org/10.1016/j.culher.2023.10.007>.
 27. Zhu RX, Hao FQ, Ma DX. Research on polygon pest-infected leaf region detection based on YOLOv8. *Agriculture.* 2023;13(12):2253. <https://doi.org/10.3390/agriculture13122253>.
 28. Kalfarisi R, Wu ZY, Soh K. Crack detection and segmentation using deep learning with 3D reality mesh model for quantitative assessment and integrated visualization. *J Comput Civil Eng.* 2020;34(3):04020010. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000890](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000890).
 29. Vandenabeele L, Loverdos D, Pfister M, Sarhosis V. Deep learning for the segmentation of large-scale surveys of historic masonry: a new tool for building archaeology applied at the Basilica of St Anthony in Padua. *Int J Archit Herit.* 2023;1:1–13. <https://doi.org/10.1080/15583058.2023.2260771>.
 30. Perumal R, Venkatchalam SB. Non invasive decay analysis of monument using deep learning techniques. *Trait Signal.* 2023;40(2):639–46. <https://doi.org/10.18280/ts.400222>.
 31. Perez H, Tah JHM, Mosavi A. Deep learning for detecting building defects using convolutional neural networks. *Sensors.* 2019;19(16):3556. <https://doi.org/10.3390/s19163556>.
 32. Monna F, Rolland T, Denaire A, Navarro N, Granjon L, Barbé R, et al. Deep learning to detect built cultural heritage from satellite imagery. Spatial distribution and size of vernacular houses in Sumba, Indonesia. *J Cult Herit.* 2021;52:171–83. <https://doi.org/10.1016/j.culher.2021.10.004>.
 33. Zhang Y, Zhang ZY, Zhao W, Li Q. Crack segmentation on earthen heritage site surfaces. *Appl Sci-Basel.* 2022;12(24):12830. <https://doi.org/10.3390/app122412830>.
 34. Garrido I, Erazo-Aux J, Lagüela S, Sfarra S, Ibarra-Castanedo C, Pivarčiová E, et al. Introduction of deep learning in thermographic monitoring of cultural heritage and improvement by automatic thermogram pre-processing algorithms. *Sensors.* 2021;21(3):750. <https://doi.org/10.3390/s21030750>.
 35. Idjaton K, Janvier R, Balawi M, Desquesnes X, Brunetaud X, Treuillet S. Detection of limestone spalling in 3D survey images using deep learning. *Automat Constr.* 2023;152: 104919. <https://doi.org/10.1016/j.autcon.2023.104919>.
 36. Liu Z, Brigham R, Long ER, Wilson L, Frost A, Orr SA, et al. Semantic segmentation and photogrammetry of crowdsourced images to monitor historic facades. *Herit Sci.* 2022;10(1):27. <https://doi.org/10.1186/s40494-022-00664-y>.
 37. Tzortzis IN, Rallis I, Makantasis K, Doulamis A, Doulamis N, Voulodimos A. Automatic inspection of cultural monuments using deep and tensor-based learning on hyperspectral imagery. 2022 IEEE International Conference on Image Processing (ICIP): IEEE; 2022. p. 3136–40.
 38. Melnik G, Yekutieli Y, Sharf A. Deep segmentation of corrupted glyphs. *ACM J Comput Cult Herit.* 2022;15(1):1–24. <https://doi.org/10.1145/3465629>.
 39. Zhang X, Liu C, Yang D, Song T, Ye Y, Li K, et al. Rfaconv: innovating spatial attention and standard convolutional operation. arXiv preprint. 2023. <https://doi.org/10.48550/arXiv.2304.03198>.
 40. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30:5998–6008.
 41. Yu Z, Huang H, Chen W, Su Y, Liu Y, Wang X. Yolo-facev2: a scale and occlusion aware face detector. arXiv preprint. 2022. <https://doi.org/10.48550/arXiv.2208.02019>.
 42. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision. 2017. p. 2980–8.
 43. Cartwright TA, Bourguignon E, Bromblet P, Cassar J, Charola AE, De Witte E, et al. ICOMOS-ISCS: illustrated glossary on stone deterioration patterns. International Council of Monuments and Sites; 2008.
 44. Li H. Classification of deterioration states of historical stone relics and its application. *Sci Conserv Archaeol.* 2011;22(1):1–6. <https://doi.org/10.16334/j.cnki.cn31-1652/k.2011.01.002>.
 45. Wang C, He W, Nie Y, Guo J, Liu C, Wang Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism. *Adv Neural Inf Process Syst.* 2024;36:51094–112.
 46. Kang M, Ting C-M, Ting FF, Phan RC-W. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image Vision Comput.* 2024;147: 105057. <https://doi.org/10.1016/j.imavis.2024.105057>.
 47. Jocher G. YOLOv5 by Ultralytics. 2020. <https://github.com/ultralytics/yolov5>.
 48. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO. 2023. <https://github.com/ultralytics/ultralytics>.
 49. Wang C, Yeh I, Liao H. YOLOv9: learning what you want to learn using programmable gradient information. arXiv preprint. 2024. <https://doi.org/10.48550/arXiv.2402.13616>.
 50. Woo S, Park J, Lee J-Y, Kweon IS. Cbam: convolutional block attention module. Proceedings of the European conference on computer vision (ECCV). 2018. p. 3–19.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.