

Article

A Deep Learning-Based Decision-Support Framework for Assessing the Conservation Condition and Protection Priority of Huizhou Historic Buildings

Jing Sun ^{1,2,3,*} , Zhongxu Xie ^{1,2} and Yuanjie Li ^{1,2}

¹ School of Architecture and Spatial Planning, Anhui Jianzhu University, Hefei 230009, China; xiezhongxu@stu.ahjzu.edu.cn (Z.X.); lyj12575@stu.ahjzu.edu.cn (Y.L.)

² Provincial Key Laboratory of Huizhou-Style Architecture, Anhui Jianzhu University, Hefei 230009, China

³ China-Portugal Joint Laboratory of Cultural Heritage Conservation Science, Soochow University, Suzhou 215123, China

* Correspondence: sunjing@ahjzu.edu.cn; Tel.: +86-15056116375

Abstract

Efficient conservation prioritization of Huizhou historic buildings requires joint consideration of visible deterioration and value-bearing architectural elements. This study developed a deep learning-based decision-support framework using an original field dataset of 500 facade images and 2197 audited annotations from five towns in Shexian County, China. Separate YOLOv11n and YOLOv11n-seg models detected six decorative or typological element classes and segmented cracks, spalling, and stains, respectively. Model outputs were converted into a pathology severity index (*PSI*) based on mask-pixel-area ratios and a decorative value index (*DVI*) based on weighted element counts; the two indices were combined into a conservation priority index (*CPI*), whose weighting was examined through sensitivity analysis. On the validation set, decorative-element detection yielded a *mAP@50* of 0.655, and pathology segmentation yielded a mask *mAP@50* of 0.586. In a preliminary application to four held-out buildings, model-derived priority categories matched the blind ratings of three conservation experts in three cases. The framework offers interpretable evidence for preliminary screening and resource-allocation discussions, but it does not replace field diagnosis. Larger balanced datasets, external building-level validation, and metric calibration are required before regional deployment.

Keywords: Huizhou historic buildings; deep learning; YOLOv11; YOLOv11-seg; instance segmentation; architectural pathology; decorative elements; conservation priority; decision support



Received: 26 June 2026

Revised: 21 July 2026

Accepted: 22 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Huizhou historic buildings constitute a distinctive category of Chinese vernacular architectural heritage. Their whitewashed masonry walls, black-tiled roofs, horse-head gables, and carved brick, stone, and timber components record regional construction practices, clan organization, ritual order, and merchant culture. In particular, doors and windows concentrate the brick, stone, and timber carving traditions through which artistic craftsmanship and social meaning are expressed [1]. The conservation significance of these buildings therefore resides in both their surviving material fabric and their value-bearing decorative and typological elements.

Conservation priority cannot consequently be inferred from deterioration extent alone. It requires a combined reading of physical condition, typological integrity, architectural significance, and the likely loss of heritage value. Aydin [2] recently formalized this principle by integrating structural vulnerability, typological degradation, and heritage value in a conservation priority index. In China, Wang et al. [3] used a matter–element model and the analytic hierarchy process to rank historic buildings exposed to subway construction. These studies demonstrate the need for multidimensional prioritization while also showing that the reliability of the result depends on how condition and value evidence are acquired and weighted.

A broader international literature has developed multi-criteria decision-making approaches for heritage conservation. Nadkarni and Puthuvayi [4] showed that such methods have been used for condition assessment, significance evaluation, renovation ranking, budget allocation, and adaptive reuse. Barbati et al. [5] combined multicriteria ranking with portfolio optimization to prioritize interventions on historic buildings under resource constraints. Maselli et al. [6] developed an AHP-based method that balances architectural conservation with social, cultural, and economic values in adaptive-reuse decisions. Abdul Huq and Puthuvayi [7] quantified priority weights for the performance assessment of urban heritage conservation projects, whereas Salerno [8] demonstrated that sensitivity analysis is necessary to test whether multicriteria rankings remain stable when weights change. These contributions provide transparent decision structures, but they generally depend on manually compiled inventories, expert scores, or project records. Decorative value is commonly represented as an aggregate criterion rather than as image-derived evidence about specific facade elements.

In parallel, computer vision has expanded the capacity for rapid and repeatable heritage inspection. Mishra and Lourenço [9] reviewed artificial-intelligence-assisted visual inspection across cultural heritage and identified object detection and semantic or instance segmentation as increasingly important tools. Ararat and Karadayi Yenice [10] similarly found growing use of artificial intelligence in historic-building damage assessment, while noting limitations related to small datasets, contextual transferability, and practical validation. Earlier task-specific studies recognized weathering in stone monuments [11], detected damage in historic masonry using mobile deep learning [12], and implemented smartphone-based recognition of cracks, stains, mould, and paint deterioration [13].

Recent studies have become more closely aligned with architectural-heritage survey conditions. Karimi et al. [14] detected multiple defect types in patterned Portuguese tiles, demonstrating both the potential and the visual ambiguity of decorative surfaces. Chen et al. [15] identified several pathologies in historic masonry buildings in Suzhou. Singh et al. [16] used YOLOv8 for both damage detection and segmentation at Darbhanga Fort, while Zheng et al. [17] combined YOLOv8-based recognition with damage quantification for traditional gray-brick buildings in Macau. Li et al. [18] addressed limited masonry-heritage data and mobile deployment through augmentation and model compression, and Seo et al. [19] combined infrared imagery with multi-label classification for masonry heritage. Lin et al. [20] improved YOLOv8-based surface-defect detection for contemporary buildings. Collectively, these studies advance automated recognition, but their outputs are usually reported as classes, boxes, masks, or accuracy metrics and are seldom connected to explicit heritage-value indicators and building-level conservation priorities.

Three more specific gaps follow from the two bodies of literature. First, a representation gap remains for Huizhou facades: existing datasets and models do not jointly encode their irregular pathologies and culturally significant decorative or typological elements. Second, an evidence-integration gap separates the two research streams: heritage-priority models rely mainly on manual condition and value scores, whereas computer-vision studies

generally stop at recognition outputs. A reproducible procedure for converting pathology masks and detected value-bearing components into complementary building-level indices remains insufficiently developed. Third, a validation and uncertainty gap persists because expert elicitation, normalization references, classification thresholds, and sensitivity to alternative weights are not always reported together. These gaps limit both the conservation interpretability of automated inspection and the reproducibility of resulting priority rankings.

Accordingly, this study addresses three research questions: (1) What level of performance can the YOLOv11n and YOLOv11n-seg models achieve in identifying visible facade deterioration and decorative or typological elements in Huizhou historic buildings? (2) How can segmentation masks and object-detection outputs be converted into interpretable indices of deterioration severity and decorative value? (3) To what extent are the resulting conservation-priority categories robust to alternative index weights and consistent with a preliminary blind assessment by conservation experts?

Rather than proposing a new deep-learning architecture, this study integrates established detection and segmentation techniques into a task-specific conservation decision-support workflow. A YOLOv11n object-detection model identifies decorative and typological elements, while a YOLOv11n-seg model delineates visible deterioration masks. The study makes three principal contributions: (1) a dual-task annotation and analysis scheme adapted to Huizhou facade characteristics; (2) an explicitly documented translation of model outputs into deterioration severity and decorative value indices using mask-area ratios, component frequencies, and independently elicited expert weights; and (3) an examination of the resulting priorities through weight sensitivity analysis and a preliminary blind expert assessment involving four pilot buildings and three experts. The framework is intended for rapid screening and preliminary resource-allocation discussions; definitive diagnosis and intervention design continue to require on-site professional investigation.

2. Materials and Methods

2.1. Study Area and Data Sources

A dedicated facade-image dataset was developed for Huizhou historic buildings. Images were collected between June 2023 and September 2024 in five towns in She County, Anhui Province, China. These towns contain historic buildings representing diverse settlement patterns, topographic conditions, construction periods, and states of preservation, thereby increasing the environmental and architectural diversity of the dataset.

Image acquisition combined ground-based and unmanned aerial vehicle (UAV) photography. Ground-based images were acquired using a mobile phone (iPhone XR, Apple Inc., Cupertino, CA, USA), which documented facades, decorative components, and visible deterioration in detail, whereas UAV images were acquired using a drone (SZ DJI Technology Co., Ltd., Shenzhen, China), which provided complementary views of roof forms, overall building morphology, and the surrounding historic environment. The field campaign produced 500 quality-screened facade images suitable for model development. Each image was linked to a unique building identifier and the corresponding survey record. The reported dataset size therefore refers to the original curated image collection.

The dataset included images acquired under different weather and illumination conditions. Such variation is particularly relevant in Huizhou settlements, where narrow lanes, projecting eaves, and layered facades produce pronounced shadows and uneven illumination. Images representing different materials, deterioration conditions, and repair states were retained to improve model robustness under realistic conservation-survey conditions. The overall research and decision-support workflow is summarized in Figure 1.

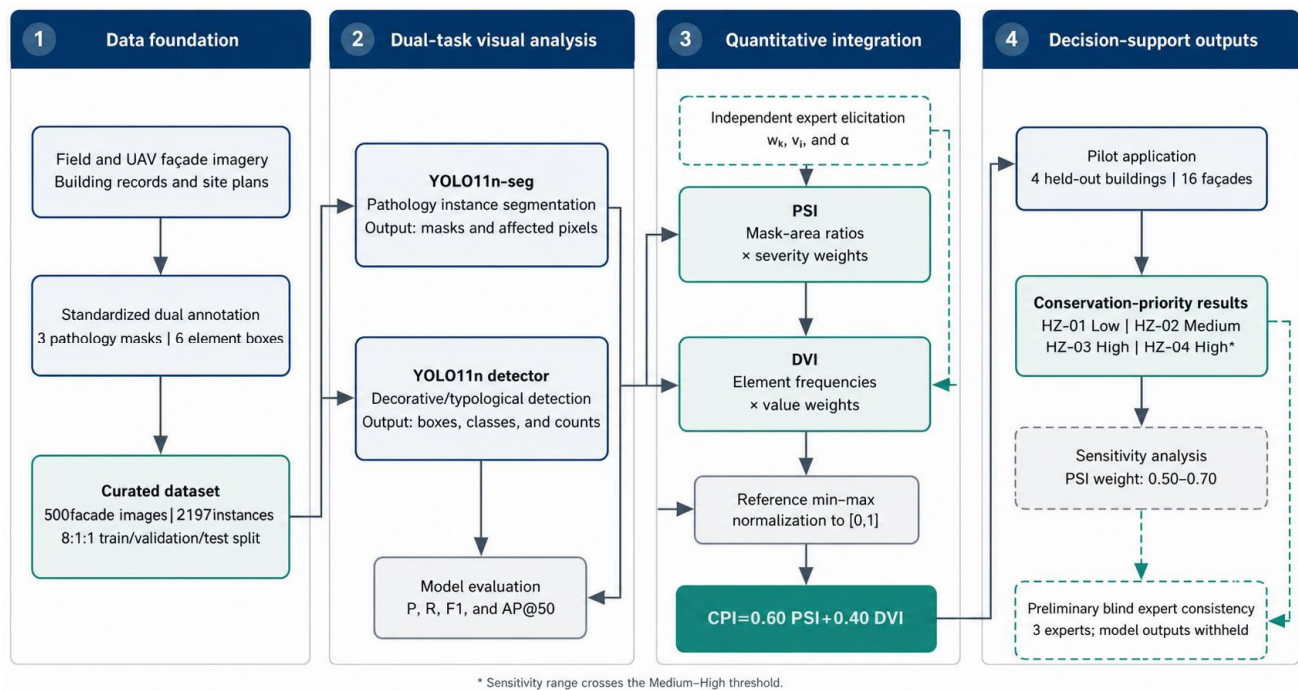


Figure 1. Research methodology and decision-support workflow.

2.2. Image Preprocessing and Annotation Scheme

Preprocessing involved quality screening and image standardization. Images affected by severe blur, overexposure, underexposure, substantial occlusion, or insufficient interpretability were excluded. The retained images were standardized in JPG format, and model inputs were resized to 1280×1280 pixels. During training, random rotation, horizontal flipping, mosaic augmentation, and color perturbation were applied to improve robustness to variations in viewpoint, illumination, and partial occlusion. These operations are consistent with established image-augmentation strategies for deep-learning training [21].

The pathology categories were defined according to the visible deterioration represented in the dataset and recent experimental descriptions of gray- and blue-brick deterioration [22–24]. Three classes (crack, spalling, and stain) were selected for instance segmentation. Six decorative or typological classes (door, window, door canopy, window head, polychrome painting, and horse-head gable) were selected for object detection. Pathologies were annotated as pixel-level instance masks using X-AnyLabeling (version 4.0.0-beta.7, CPU build; CVHub, 2023; open-source software), whereas decorative and typological elements were annotated as bounding boxes using LabelImg (version 1.8.6; Tzutalin, 2015; open-source software). The resulting classification system is summarized in Table 1.

Table 1. Classification system of pathologies and decorative or typological elements.

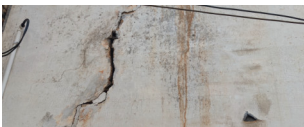
Category	Attributes	Images	Feature
Crack (CR)	Pathology		Linear cracking in the plaster layer, masonry, or structural surfaces.
Spalling (SP)	Pathology		Plaster layer detachment, brick exposure, surface scaling, or material loss.

Table 1. *Cont.*

Category	Attributes	Images	Feature
Stain (ST)	Pathology		Moisture marks, rainwater stains, wall deposits, and dark stained areas.
Door (D)	Decoration		Door frames built with polished blue bricks, heavy wooden doors, often solid wood.
Window (W)	Decoration		Mostly square or rectangular, window frames made of polished blue bricks.
Door Canopy (DC)	Decoration		Imitation wood-structure archway above the door.
Window Head (WH)	Decoration		Brick corbeling structures and decorative bands above windows.
Polychrome Painting (PP)	Decoration		Painted colorful patterns on the wall surface.
Horse-head Gable (HG)	Decoration		Stepped gable walls rising above the roof.

2.3. Annotation Protocol and Dataset Split

A consistent annotation protocol was established to reduce class ambiguity and inter-annotator variation. Decorative and typological elements were delineated according to their visible physical boundaries. Door annotations included the entrance opening and its visible frame, while window annotations included the opening and visible frame. Door canopies and window heads encompassed the decorative or structural elements above the corresponding openings, and horse-head gables were annotated along their visible stepped outlines.

For pathology segmentation, cracks were represented by narrow polygons following their visible widths rather than by center lines. Spalling annotations followed the boundaries of detached material, plaster loss, or exposed substrate. Stains were delineated according to visible discoloration, textural differences, water traces, or surface deposits. Ambiguous cases were reviewed collectively to improve annotation consistency.

After annotation, three experts experienced in the conservation of Huizhou historic buildings reviewed class assignments and boundary consistency. The audited annotations comprised 2197 instances across the 500 facade images: 136 crack, 632 spalling, and 373 stain masks, together with 187 doors, 473 windows, 62 door canopies, 102 window heads, 120 polychrome paintings, and 112 horse-head gables. Class-specific image counts indicate images containing at least one instance of the relevant class and are non-additive because one image may contain several classes. The dataset was divided into training, validation,

and test subsets at a ratio of 8:1:1. Table 1 summarizes the classification scheme, and Figure 2 presents representative annotation examples used to verify the consistent application of these conventions.



Figure 2. Annotation examples of pathologies and decorative or typological elements.

2.4. Model Development and Training

Two complementary computer-vision tasks were implemented. YOLOv11n was used to detect decorative and typological elements and generate class labels, bounding boxes, and confidence scores. YOLOv11n-seg was used for pathology instance segmentation and generated class labels, confidence scores, bounding boxes, and pixel-level masks. The task design draws on established instance-segmentation principles [25], the one-stage YOLO detection paradigm [26], subsequent real-time detector development [27], and current object-detection synthesis [28]. This separation reflects the different analytical roles of the outputs: element detection supports heritage-value assessment, whereas pathology segmentation enables quantitative estimation of projected affected areas.

All experiments were conducted on a 64-bit Windows 11 workstation equipped with an AMD Ryzen 7 5800H CPU and an NVIDIA GeForce RTX 3060 Laptop GPU. The software

environment comprised Ultralytics 8.4.48, Python 3.10.20, PyTorch 2.0.0 + cu118, and CUDA-enabled acceleration. The same image resolution and GPU configuration were used during validation to maintain comparability.

Pathology segmentation was performed using the lightweight YOLOv11n-seg architecture initialized from the COCO-pretrained yolo11n-seg.pt checkpoint [29]. Decorative and typological elements were detected using the corresponding YOLOv11n architecture initialized from yolo11n.pt. Both models were fine-tuned for 300 epochs using an input resolution of 1280×1280 pixels, a batch size of 4, and a single GPU. Except for task-specific heads and loss functions, the same nominal training settings were applied to both tasks. Ultralytics automatic optimizer selection was enabled; the optimizer selected by the framework and its effective initial learning rate were recorded in the training logs. Both models were fine-tuned from pretrained weights rather than trained from random initialization.

2.5. Performance Metrics

Precision (P), recall (R), $F1$ -score, average precision (AP), and mean average precision (mAP) were used to evaluate model performance. Precision measures the proportion of correct predictions among all positive predictions, while recall measures the proportion of ground-truth targets correctly detected. The $F1$ -score is the harmonic mean of precision and recall. AP is calculated as the area under the precision-recall curve for a single class, and mAP is the mean AP over all classes, following established object-detection evaluation practice [29,30]. For the segmentation task, both box and mask metrics were calculated, but mask $mAP@50$ was regarded as the more relevant indicator for pathology area quantification.

$$P = \frac{TP}{FP + TP'} \quad (1)$$

$$R = \frac{TP}{FN + TP'} \quad (2)$$

$$\begin{cases} AP = \int_0^1 P(r)dr \\ mAP = \frac{1}{C} \sum_{i=1}^C AP_i \end{cases} \quad (3)$$

$$F1 = \frac{2PR}{P + R} \quad (4)$$

2.6. Construction of the Conservation Priority Index

To operationalize the outputs of the two models, four held-out Huizhou historic buildings were used as a pilot case application. For each building, pathology masks were converted into a Pathology severity index (PSI), while detected decorative and typological elements were converted into a decorative value index (DVI). The two normalized indices were then integrated into a conservation priority index (CPI). This application was intended to examine the internal operation of the decision-support procedure rather than to constitute large-scale external validation.

For each building, the raw pathology score was calculated as the weighted sum of the predicted pixel-area ratios of all pathology classes, whereas the raw decorative-value score was calculated as the weighted sum of the detected instances of all decorative and typological elements. Both scores were normalized to the range $[0, 1]$ using a reference min–max transformation. Zero represented the theoretical absence of pathology or value-bearing elements, while the maximum raw score observed among the four pilot buildings served as the pilot-specific upper reference. This procedure retained nonzero index values for buildings with detected features; however, the upper references should be regarded

as pilot calibration anchors rather than universally applicable thresholds. The normalized indices were calculated as follows:

$$PSI = \frac{\sum_{k=1}^N \left(\frac{A_k}{A_f} \times w_k \right) - \min(S)}{\max(S) - \min(S)}, \quad (5)$$

$$DVI = \frac{\sum_{j=1}^M (N_j \times v_j) - \min(V)}{\max(V) - \min(V)}, \quad (6)$$

$$CPI = 0.60 \times PSI + 0.40 \times DVI. \quad (7)$$

where A_k is the union area of the predicted mask pixels for pathology class k ; A_f is the total number of pixels within the evaluated facade image or manually defined facade region of interest; w_k is the normalized pathology-severity weight; N_j is the number of non-duplicate detected elements of class j ; and v_j is the corresponding normalized heritage-value weight. Both A_k and A_f were obtained directly from the Python mask-pixel-counting code applied to the model-generated masks. Therefore, A_k/A_f is a dimensionless projected pixel-area ratio, not a calibrated area in square meters. Overlapping masks of the same class were merged before pixel counting to prevent double counting.

Pilot-specific normalization bounds were defined using zero as the theoretical minimum and the highest observed raw score as the upper reference. The resulting upper bounds were 7.7025 for PSI , derived from HZ-04, and 9.80 for DVI , derived from HZ-03. These values are calibration anchors for the four-building pilot sample rather than universal standards for Huizhou historic buildings; larger regional samples are required to establish transferable reference bounds.

The baseline CPI assigned weights of 0.60 to PSI and 0.40 to DVI . Using equal-width thresholds, CPI values were classified as Low (<0.33), Medium ($0.33\text{--}0.67$), or High (≥ 0.67). Sensitivity analysis varied the PSI weight from 0.50 to 0.70 at intervals of 0.05, with the DVI weight adjusted complementarily.

3. Results

3.1. Model Validation Performance

Table 2 reports class-wise validation results for pathology segmentation and decorative-element detection. Among the pathology classes, cracks produced the highest mask-level values (precision 81.2%, recall 72.0%, and $mAP@50$ 72.2%), while stains yielded a mask $mAP@50$ of 62.5%. Spalling was the principal limitation: its mask precision was 73.2%, but recall was only 32.8%, resulting in a mask $mAP@50$ of 41.2%. Thus, predicted spalling masks were often credible when present, although many ground-truth instances were missed. The class-dependent results support preliminary screening for the better-performing pathology categories but do not justify a uniform performance claim.

For decorative-element detection, doors achieved the highest $mAP@50$ of 89.8%, followed by windows at 79.7% and horse-head gables at 71.5%. Door canopies, window heads, and polychrome paintings achieved $mAP@50$ values of 59.2%, 50.1%, and 42.7%, respectively. Although horse-head gables attained relatively high precision and $mAP@50$, their recall was only 47.0%, indicating that the model produced comparatively conservative predictions and missed a substantial proportion of instances.

Overall, model performance varied considerably among classes. The regular geometry and visually distinctive boundaries of doors and windows may have facilitated their detection. In contrast, the irregular boundaries and heterogeneous surface appearance of spalling probably contributed to its low recall. The visual complexity and compara-

tively limited representation of polychrome paintings may also have constrained their detection performance.

Table 2. Class-wise validation metrics for pathology segmentation and decorative-element detection.

Class	Box			Mask		
	Precision	Recall	mAP@50	Precision	Recall	mAP@50
Crack	0.812	0.720	0.716	0.812	0.720	0.722
Spalling	0.715	0.321	0.410	0.732	0.328	0.412
Stain	0.877	0.557	0.639	0.877	0.557	0.625
Door	0.886	0.973	0.898	/	/	/
Window	0.769	0.616	0.797	/	/	/
Door Canopy	0.602	0.600	0.592	/	/	/
Window Head	0.548	0.500	0.501	/	/	/
Polychrome Painting	0.543	0.400	0.427	/	/	/
Horse-head Gable	0.850	0.470	0.715	/	/	/

Table 3 compares YOLOv11n-seg and YOLOv8n-seg under identical evaluation conditions. Their overall mask *mAP@50* values were nearly identical (0.586 and 0.585, respectively), so the comparison does not establish general superiority. YOLOv11n-seg nevertheless achieved higher recall for cracks, spalling, and stains (0.720, 0.328, and 0.557) than YOLOv8n-seg (0.667, 0.284, and 0.521), and its crack mask *mAP@50* was higher (0.722 versus 0.685). Conversely, YOLOv8n-seg produced higher overall mask precision and marginally higher *mAP@50* for spalling and stains. YOLOv11n-seg was retained because preliminary conservation screening in this study prioritized reducing missed pathology instances; this is a task-specific choice, not a claim of universal architectural superiority.

Table 3. Comparative validation performance of YOLOv11n-seg and YOLOv8n-seg.

Model	Class	Images	Instances	Box			Mask		
				P	R	mAP@50	P	R	mAP@50
YOLOv11n-seg	All	40	261	0.801	0.533	0.588	0.807	0.535	0.586
	Crack	6	12	0.812	0.720	0.716	0.812	0.720	0.722
	Spalling	23	134	0.715	0.321	0.410	0.732	0.328	0.412
	Stain	31	115	0.877	0.557	0.639	0.877	0.557	0.625
YOLOv8n-seg	All	40	261	0.900	0.460	0.561	0.948	0.490	0.585
	Crack	6	12	0.834	0.583	0.623	0.954	0.667	0.685
	Spalling	23	134	0.914	0.276	0.410	0.939	0.284	0.424
	Stain	31	115	0.952	0.521	0.651	0.952	0.521	0.646

3.2. Preliminary Blind Expert Consistency Assessment

The trained models were applied to four Huizhou historic buildings, each represented by four facade images, resulting in 16 images in total. None of these images were used for model training. The pathology masks were converted into class-specific pixel-area ratios, while decorative-element detections were converted into weighted frequency scores. Given the limited number of buildings and images, this analysis is presented as a pilot application rather than a general validation of large-scale survey performance. The site information and model outputs for the pilot buildings are presented in Figures 3 and 4, respectively.



Figure 3. Basic information for the four pilot buildings: site plans and front elevations.



Figure 4. Detection and segmentation outputs for the four pilot buildings.

A structured two-stage expert-elicitation procedure was used to derive the assessment weights. The panel comprised three specialists with relevant experience in historic-building conservation, heritage-value assessment, and traditional-building repair. Before receiving any case-building images or model outputs, each expert independently completed two scoring matrices: one addressing the conservation consequences of the pathology classes and the other evaluating the heritage significance of the decorative and typological elements. Each item was accompanied by a definition and a five-point scale, on which 1 represented a low consequence or limited heritage value, 3 represented a moderate level, and 5 represented a very high consequence or value. The resulting normalized pathology-severity and heritage-value weights are reported in Tables 4 and 5, respectively.

Table 4. Expert-derived severity weights for pathology types.

Pathology Type	Mean Expert Score	Normalized Weight	Conservation Interpretation
Crack	4.80	1.00	Structural deformation/cracking; highest potential risk.
Spalling	4.10	0.85	Surface detachment affects authenticity and durability.
Stain	2.90	0.60	Visual deterioration related to dampness or poor drainage.

Table 5. Expert-derived value weights for decorative and typological elements.

Element Type	Mean Expert Score	Normalized Weight	Value Interpretation
Polychrome painting	4.90	1.00	High iconographic and artistic information.
Door canopy	4.60	0.94	Important entrance ornament and ritual expression.
Window head	4.20	0.86	Representative facade ornament.
Horse-head gable	4.00	0.82	Strong regional typological identity.
Door	3.70	0.76	Functional and spatial-order element.
Window	3.30	0.67	Frequent components with moderate individual value.

The independent scores were recorded before aggregation. For each item, the arithmetic mean of the three scores was divided by the highest mean score in the corresponding matrix to obtain the normalized class weight. The anonymized score distributions and item definitions were subsequently discussed to confirm a consistent interpretation, but the original individual scores were not revised. At the criterion level, the experts assigned weights of 0.60 to physical condition (*PSI*) and 0.40 to decorative value (*DVI*). No detection boxes, segmentation masks, building-level indices, or model-derived priorities were available to the experts during this weighting procedure. Because the 60:40 allocation represented elicited expert judgment rather than a statistically estimated parameter, its influence was examined through sensitivity analysis.

The recalculated building-level indices and model-derived priorities are reported in Table 6. A separate blind procedure was used to assess preliminary consistency with expert judgment. The experts received only the site plans, original unannotated facade photographs, and standardized case identifiers. They did not have access to the detection boxes, segmentation masks, *PSI*, *DVI*, *CPI* values, model-derived priorities, or the ratings of the other experts. Each expert independently assigned a conservation-priority score from 1 to 5. Mean scores below 2.5 were classified as Low, scores from 2.5 to below 4.0 as Medium, and scores of 4.0 or above as High. All expert ratings were locked before the model results were disclosed and compared. Given the sample of only four buildings and three experts, this procedure represents a preliminary blind consistency assessment rather than definitive external validation. The resulting expert classifications are summarized in Table 7.

Table 6. Quantitative assessment and conservation priority of the selected historic buildings.

ID	CR (%)	SP (%)	ST (%)	PSI	DVI	CPI	Model Priority
HZ-01	0.000	0.000	0.000	0.000	0.242	0.097	Low
HZ-02	0.000	1.110	8.810	0.809	0.687	0.760	High
HZ-03	0.000	0.100	12.300	0.969	1.000	0.981	High
HZ-04	1.730	5.770	1.780	1.000	0.078	0.631	Medium

Table 7. Expert-based assessment of building conservation priorities.

ID	Expert 1	Expert 2	Expert 3	Mean Expert Score	Expert Priority	Consistency
HZ-01	1	2	1	1.33	Low	Consistent
HZ-02	3	3	4	3.33	Medium	One-category difference
HZ-03	5	4	5	4.67	High	Consistent
HZ-04	3	4	3	3.33	Medium	Consistent

Under the reference normalization defined in Section 2.6, HZ-03 achieved the highest baseline *CPI* (0.981), combining a high *PSI* (0.969) with the maximum *DVI* (1.000). HZ-02 ranked second and was classified as High priority (*CPI* = 0.760), with a *PSI* of 0.809 and a *DVI* of 0.687. HZ-04 was classified as Medium priority at the baseline weighting (*CPI* = 0.631). Its raw pathology score, derived from crack, spalling, and stain area ratios of 1.73%, 5.77%, and 1.78%, respectively, defined the pilot-specific *PSI* upper reference and therefore produced a *PSI* of 1.000. However, only one door was detected, resulting in a comparatively low *DVI* of 0.078. Because its sensitivity range (0.539–0.723) crossed the Medium–High threshold, the priority of HZ-04 should be regarded as weight-sensitive rather than as a robust categorical assignment. HZ-01 remained low priority (*CPI* = 0.097) because no pathology masks were predicted, although its *DVI* of 0.242 reflected the presence of detected decorative elements. The resulting ranking (HZ-03 > HZ-02 > HZ-04 > HZ-01) therefore reflected the combined influence of predicted deterioration and detected value-bearing elements rather than pathology extent alone.

Sensitivity analysis varied the *PSI* weight from 0.50 to 0.70, with the *DVI* weight adjusted complementarily from 0.50 to 0.30. HZ-01, HZ-02, and HZ-03 consistently retained their Low, High, and High priority classifications, respectively. Their *CPI* ranges were 0.073–0.121, 0.748–0.772, and 0.978–0.985. In contrast, the *CPI* of HZ-04 increased from 0.539 to 0.723 and crossed the Medium–High threshold when the *PSI* weight reached 0.65, indicating sensitivity to the relative weighting of physical condition and decorative value. Nevertheless, the overall ranking remained stable across all five scenarios: HZ-03 > HZ-02 > HZ-04 > HZ-01. The resulting *CPI* ranges are presented in Figure 5.

After the blind expert ratings had been finalized, the expert- and model-derived priority categories agreed for three of the four buildings (75%). HZ-02 represented the only discrepancy: the experts assigned it Medium priority, whereas the baseline model classified it as High priority. This difference persisted across the tested *PSI–DVI* weighting scenarios, suggesting that it may reflect factors considered by the experts but not fully represented by the two quantitative indices. HZ-04 was classified as Medium priority by both approaches under the baseline weighting, although its model-derived category changed to High when greater weight was assigned to *PSI*. Given the small sample, the observed agreement should be interpreted only as preliminary evidence of consistency rather than validation of the framework’s general applicability.

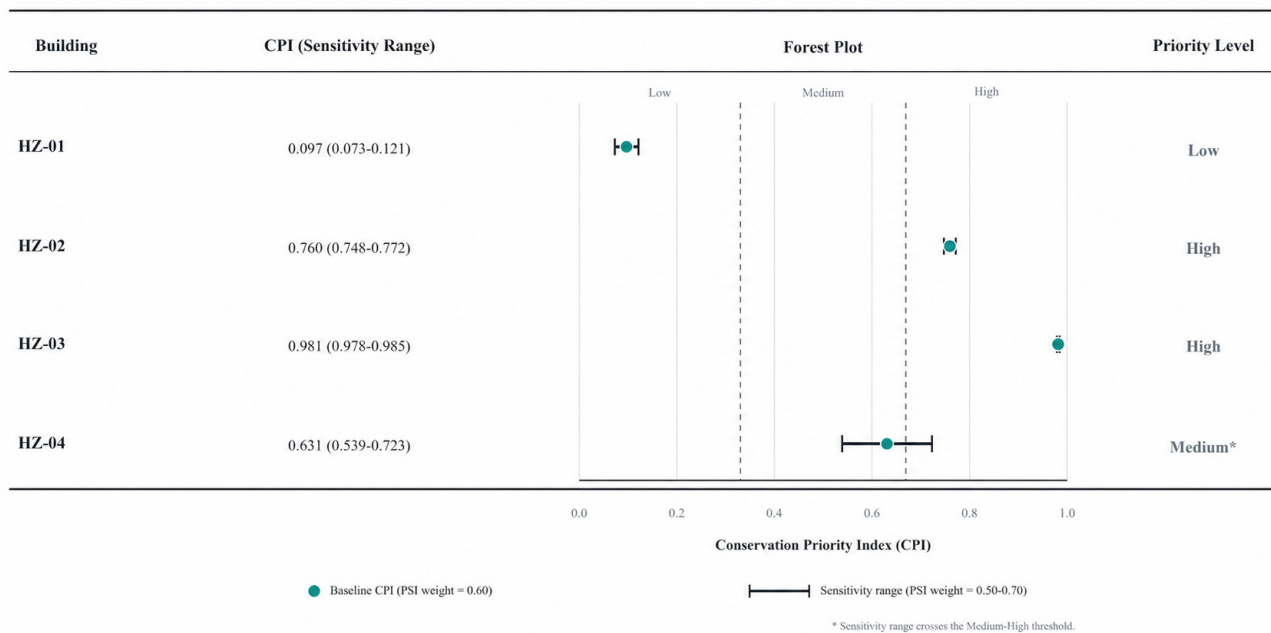


Figure 5. Sensitivity analysis of the Conservation Priority Index for the four pilot buildings under varying PSI weights.

4. Discussion

4.1. From Visual Recognition to Conservation Decision Support

A central contribution of the proposed framework is the translation of computer-vision outputs into quantitative information relevant to conservation management. Previous image-based studies have often focused primarily on whether deterioration can be detected and how accurately it can be classified. Conservation planning, however, also requires information on the extent of visible deterioration, the heritage value of the building and its components, and the relative urgency of intervention. The priority models of Barbati et al. [5] and Maselli et al. [6] show how heterogeneous criteria can be organized for decisions, but they do not automatically derive facade-condition evidence from images. Pixel-level mask quantification and weighted building-level indices provide a transparent link between visual recognition and preliminary conservation prioritization.

This translation is particularly relevant to Huizhou historic buildings, whose heritage value is often embodied in entrance canopies, window heads, polychrome paintings, and horse-head gables. Prioritization based solely on total deterioration area may overemphasize extensive but relatively superficial staining while underrepresenting buildings that contain significant decorative elements. By integrating *PSI* and *DVI*, the *CPI* incorporates both physical condition and detected heritage-value elements into building-level screening. Nevertheless, the current *CPI* does not determine whether an individual pathology directly overlaps with or affects a specific decorative component. Such spatial relationships require image registration, component-level overlap or proximity analysis, and professional field verification.

4.2. Role of Instance Segmentation in Pathology Quantification

Bounding boxes are appropriate for locating and counting relatively regular elements such as doors and windows, but they provide unreliable estimates of irregular pathology areas. A bounding box inevitably includes undamaged background pixels and may therefore substantially overestimate the affected area. Instance segmentation delineates the visible contours of cracks, spalling, and stains more precisely and is consequently more suitable for calculating projected affected-area ratios and severity indices. This choice is

supported by the mask-based localization principle established by He et al. [25] and by recent heritage applications that segment irregular tile and masonry damage [14,16].

This advantage is accompanied by greater annotation costs and stricter image-quality requirements. Fine cracks may occupy only a few pixels in distant facade images, while spalling and stains can be visually confounded by uneven illumination, wall texture, surface deposits, and previous repairs. Future improvements should therefore include larger and more balanced class samples, high-resolution crop-based training, standardized annotation protocols, and orthorectified facade imagery where feasible.

4.3. Qualitative Spatial Patterns of Pathology Distribution

Recurring spatial patterns were qualitatively noted during annotation and expert review. Because facade zones were not systematically coded, and no zone-specific statistical analysis was conducted, these observations should not be interpreted as evidence of statistically non-random distributions. Instead, they represent qualitative patterns supported by established knowledge of masonry deterioration and provide hypotheses for subsequent quantitative investigation.

- (a) **Spalling.** Spalling was repeatedly observed around openings, wall corners, eaves, material interfaces, and the margins of previous repairs. Huizhou facades commonly combine gray-brick masonry, lime-based mortar, or render, stone plinths, and adjacent timber elements, which differ in stiffness, porosity, and moisture response. Eaves, window heads, sills, and door canopies may intercept and redirect wind-driven rain, causing concentrated runoff and repeated wetting–drying cycles at edges and joints. Water penetration, binder leaching, salt crystallization, and differential hygrothermal movement may progressively weaken adhesion between surface layers and the masonry substrate. Dense or incompatible repair mortars may further displace the evaporation front toward repair margins and promote edge detachment [22–24].
- (b) **Staining.** Stains were frequently noted at wall bases, below openings, along visible runoff paths, and within shaded or recessed areas. At wall bases, runoff, splashback, inadequate surface drainage, and capillary moisture transport through porous brick and mortar may maintain elevated moisture levels. Below windows and projecting ornaments, discontinuities in the water-shedding surface can concentrate runoff into vertical bands. Reduced drying in sheltered areas may also encourage particle deposition, salt migration, and biological growth. Staining should therefore be interpreted as a potential visual indicator of moisture transport or drainage deficiencies rather than solely as chromatic deterioration. Its underlying cause, however, cannot be determined from RGB images alone [22–24].
- (c) **Cracking.** Cracks were qualitatively observed near opening corners, material transitions, repair boundaries, and other discontinuities in the wall fabric. Openings reduce wall continuity and can generate localized stress concentrations. Differential settlement, interaction between timber components and masonry infill, moisture- and temperature-related movement, salt crystallization, and incompatible rigid repairs may also generate or widen discontinuities at material interfaces [22–24]. These mechanisms are plausible interpretations based on construction characteristics and deterioration theory, but they were not directly measured in the present study.

Future dataset development should assign each pathology mask to predefined facade zones, such as the wall base, wall field, opening perimeter, corner, eave, or parapet, material interface, and repair patch. Zone-specific frequencies and affected-area ratios should be calculated relative to the available area of each zone rather than from raw counts alone. Reporting zone denominators, confidence intervals, and building-level clustering would permit chi-square screening or mixed-effects regression to test whether particular pathology

classes are overrepresented in specific locations. The facade-zone definitions and coded spatial data should also be released with the dataset to support reproducibility.

4.4. Limitations and Future Work

The proposed framework provides transparent quantitative evidence for preliminary screening but should not be regarded as a substitute for professional conservation judgment. Conservation priority also depends on authenticity, integrity, material compatibility, reversibility, historical significance, current use, safety, and available resources, none of which can be fully inferred from facade images. This limitation is consistent with multidimensional heritage-priority research, which combines condition with typological, cultural, social, and economic criteria [2,4,6]. The present *CPI* integrates building-level physical condition and decorative value but does not directly quantify the spatial coincidence between pathology and individual value-bearing components. Future research should incorporate mask-component overlap and proximity measures, followed by on-site verification of the affected fabric.

The current *PSI* is based on projected mask-pixel ratios. Although the pilot images were acquired using a broadly standardized protocol, perspective distortion, occlusion, lens geometry, facade framing, and uncertainty in the evaluation region can affect the denominator and reduce cross-building comparability. Future surveys should record camera-calibration parameters and control points, generate orthorectified facade images or photogrammetric models, and compare pixel-based estimates with manually measured or metrically calibrated reference areas. Agreement may be evaluated using mean absolute error, root mean square error, intraclass correlation coefficients, or Bland–Altman analysis, as appropriate.

The normalization bounds and *CPI* classification thresholds were calibrated using only the four pilot buildings and therefore represent pilot-specific reference values rather than regional standards. A larger stratified sample covering different settlements, construction periods, materials, facade orientations, environmental exposures, and conservation conditions is required to establish stable regional reference distributions. Expert-derived pathology and heritage-value weights should also be re-estimated using a documented Delphi method or analytic hierarchy process. Uncertainty should be propagated through bootstrap or Monte Carlo simulation, while future sensitivity analyses should jointly vary pathology weights, decorative-element weights, normalization bounds, *CPI* weights, and classification thresholds.

Finally, although the dataset contains 500 facade images, its scale and class composition remain limited for deep-learning applications, particularly for visually small or heterogeneous classes. The pilot application was further restricted to four buildings, 16 facade images, and three experts; expert-model consistency must therefore be interpreted as preliminary. Future studies should conduct building-level external validation across multiple settlements, cameras, seasons, and environmental conditions; preserve source-level separation among dataset subsets; increase minority-class representation; and involve a larger independent expert panel. Inter-rater agreement should be quantified using an appropriate statistic, such as Fleiss' kappa or an intraclass correlation coefficient.

5. Conclusions

This study developed an interpretable decision-support framework that integrates pathology segmentation, decorative-element detection, and expert-informed assessment for the preliminary conservation prioritization of Huizhou historic buildings. The main conclusions are as follows:

1. YOLOv11n-seg generated pixel-level pathology masks for calculating the Pathology severity index (*PSI*), while YOLOv11n detected decorative and typological elements for calculating the decorative value index (*DVI*). Their integration into the conservation priority index (*CPI*) translated computer-vision outputs into transparent building-level evidence.
2. Model performance was class-dependent. The mask *mAP@50* values for cracks, stains, and spalling were 0.722, 0.625, and 0.412, respectively, while doors and windows achieved detection *mAP@50* values of 0.898 and 0.797. YOLOv11n-seg and YOLOv8n-seg achieved similar overall mask *mAP@50* values, but YOLOv11n-seg provided higher class-wise recall and was therefore better aligned with preliminary screening requirements.
3. In the four-building pilot application, HZ-03 and HZ-02 were classified as High priority, HZ-04 as Medium priority, and HZ-01 as Low priority. The overall ranking remained stable under alternative *PSI–DVI* weights, although HZ-04 crossed the Medium–High threshold. Blind expert assessment agreed with the baseline model for three buildings (75%), providing preliminary evidence of consistency.

The framework can support rapid condition screening, inspection documentation, and preliminary resource allocation, but it is not intended to replace professional diagnosis. Future research should expand the dataset and conduct external building-level evaluations, improve the metric calibration of pathology areas, and establish transferable regional normalization standards. Model outputs should ultimately be interpreted alongside field investigations and professional conservation judgment.

Author Contributions: Conceptualization, J.S. and Z.X.; methodology, J.S. and Z.X.; formal analysis, J.S. and Z.X.; investigation, J.S. and Z.X.; data curation, J.S. and Z.X.; writing—original draft preparation, J.S. and Z.X.; writing—review and editing, Z.X. and Y.L.; visualization, Z.X. and Y.L.; funding acquisition, J.S. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2024 Key Project of the Humanities Research Program for Higher Education Institutions of the Anhui Provincial Department of Education (grant no. 2024AH052356), the Anhui Jianzhu University 2025 New Era Education Quality Project for Postgraduate Education (grant no. 2025ghjc016), and the Anhui Provincial University Humanities and Social Sciences Research Key Project (grant no. 2025AHGXSK30481).

Data Availability Statement: The 500 facade images contain identifiable privately owned heritage properties and site-location information and are therefore not deposited in an unrestricted public repository. De-identified images, YOLO-format bounding-box and segmentation annotations, dataset-split lists, and the trained model weights are available from the corresponding author for non-commercial academic research upon reasonable request, subject to owner/privacy review and a data-use agreement. Training settings and aggregate evaluation results are reported in the article to support methodological reproducibility.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cao, Z.; Mustafa, M.B. A Study of Ornamental Craftsmanship in Doors and Windows of Hui-Style Architecture: The Huizhou Three Carvings (Brick, Stone, and Wood Carvings). *Buildings* **2023**, *13*, 351. [\[CrossRef\]](#)
2. Aydin, M.F. An Integrated Structural-Typological-Value Sensitivity Model (STVSM) for Vulnerability Assessment and Conservation Prioritization of Historic Buildings. *Int. J. Archit. Herit.* **2026**, 1–22. [\[CrossRef\]](#)
3. Wang, Q.; Yang, C.; Lu, J.; Wu, F.; Xu, R. Analysis of Preservation Priority of Historic Buildings along the Subway Based on Matter-Element Model. *J. Cult. Herit.* **2020**, *45*, 291–302. [\[CrossRef\]](#)
4. Nadkarni, R.R.; Puthuvayi, B. A Comprehensive Literature Review of Multi-Criteria Decision Making Methods in Heritage Buildings. *J. Build. Eng.* **2020**, *32*, 101814. [\[CrossRef\]](#)

5. Barbati, M.; Figueira, J.R.; Greco, S.; Ishizaka, A.; Panaro, S. A Multiple Criteria Methodology for Priority Based Portfolio Selection. *Socio-Econ. Plan. Sci.* **2023**, *88*, 101595. [\[CrossRef\]](#)
6. Maselli, G.; Cucco, P.; Nesticò, A.; Ribera, F. Historical Heritage-MultiCriteria Decision Method (H-MCDM) to Prioritize Intervention Strategies for the Adaptive Reuse of Valuable Architectural Assets. *MethodsX* **2024**, *12*, 102487. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Abdul Huq, S.; Puthuvayi, B. Assessing the Performance of Urban Heritage Conservation Projects-Influencing Factors, Aspects and Priority Weights. *Built Herit.* **2024**, *8*, 4. [\[CrossRef\]](#)
8. Salerno, E. Identifying Value-Increasing Actions for Cultural Heritage Assets through Sensitivity Analysis of Multicriteria Evaluation Results. *Sustainability* **2020**, *12*, 9238. [\[CrossRef\]](#)
9. Mishra, M.; Lourenço, P.B. Artificial Intelligence-Assisted Visual Inspection for Cultural Heritage: State-of-the-Art Review. *J. Cult. Herit.* **2024**, *66*, 536–550. [\[CrossRef\]](#)
10. Ararat, M.B.; Karadayi Yenice, T. The Use of Artificial Intelligence in Damage Assessment of Historical Buildings. *Front. Archit. Res.* **2026**, *15*, 554–565. [\[CrossRef\]](#)
11. Hatir, M.E.; Barstugan, M.; Ince, I. Deep Learning-Based Weathering Type Recognition in Historical Stone Monuments. *J. Cult. Herit.* **2020**, *45*, 193–203. [\[CrossRef\]](#)
12. Wang, N.; Zhao, X.; Zhao, P.; Zhang, Y.; Zou, Z.; Ou, J. Automatic Damage Detection of Historic Masonry Buildings Based on Mobile Deep Learning. *Autom. Constr.* **2019**, *103*, 53–66. [\[CrossRef\]](#)
13. Perez, H.; Tah, J.H.M. Deep Learning Smartphone Application for Real-Time Detection of Defects in Buildings. *Struct. Control Health Monit.* **2021**, *28*, e2751. [\[CrossRef\]](#)
14. Karimi, N.; Mishra, M.; Lourenço, P.B. Deep Learning-Based Automated Tile Defect Detection System for Portuguese Cultural Heritage Buildings. *J. Cult. Herit.* **2024**, *68*, 86–98. [\[CrossRef\]](#)
15. Chen, X.; He, J.; Wang, S. Deep Learning-Driven Pathology Detection and Analysis in Historic Masonry Buildings of Suzhou. *npj Herit. Sci.* **2025**, *13*, 197. [\[CrossRef\]](#)
16. Singh, S.K.; Maity, D.; Kumawat, P.K. Deep Learning-Based Damage Detection and Segmentation in the Battledore of Darbhanga Fort. *J. Cult. Herit.* **2025**, *73*, 510–523. [\[CrossRef\]](#)
17. Zheng, L.; Zheng, J.; Chen, Y.; Zheng, Y.; Lao, W.; Chen, S. Gray Brick Wall Surface Damage Detection of Traditional Chinese Buildings in Macau: Damage Quantification and Thermodynamic Analysis Method via YOLOv8 Technology. *Appl. Sci.* **2025**, *15*, 6665. [\[CrossRef\]](#)
18. Li, X.; Huang, Y.; Zhou, F.; Yang, Z. Intelligent Detection and Model Lightweight of Masonry Historical Building Damages Based on Deep Learning. *Structures* **2025**, *81*, 110363. [\[CrossRef\]](#)
19. Seo, H.; Raut, A.D.; Chen, C.; Zhang, C. Multi-Label Classification and Automatic Damage Detection of Masonry Heritage Building through CNN Analysis of Infrared Thermal Imaging. *Remote Sens.* **2023**, *15*, 2517. [\[CrossRef\]](#)
20. Lin, X.; Meng, Y.; Sun, L.; Yang, X.; Leng, C.; Li, Y.; Niu, Z.; Gong, W.; Xiao, X. Building Surface Defect Detection Based on Improved YOLOv8. *Buildings* **2025**, *15*, 1865. [\[CrossRef\]](#)
21. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [\[CrossRef\]](#)
22. Yue, J.; Li, Y.; Luo, Z.; Huang, X.; Kong, Q.; Wang, Z. Study on Deterioration Law and Mechanism of Gray Brick Due to Salt Crystallization. *Materials* **2022**, *15*, 2936. [\[CrossRef\]](#) [\[PubMed\]](#)
23. An, L.; Qiao, Z.; Wang, J.; Wang, F. Experimental Study on Weathering Mechanism of Ancient Bricks in Jiayuguan Wei-Jin Tombs, Gansu, China. *Herit. Sci.* **2024**, *12*, 114. [\[CrossRef\]](#)
24. Zhang, H.; Zhang, J.; Zhang, L.; Ying, Y.; Wang, N.; Li, D.; Xu, H. Ancient Architectural Pathology of Blue Bricks and Brick Carvings in Northwest China: Example from the White Temple Tower. *Case Stud. Constr. Mater.* **2024**, *20*, e03357. [\[CrossRef\]](#)
25. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969. [\[CrossRef\]](#)
26. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
27. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. [\[CrossRef\]](#)
28. Zaidi, S.S.A.; Ansari, M.S.; Aslam, A.; Kanwal, N.; Asghar, M.; Lee, B. A Survey of Modern Deep Learning Based Object Detection Models. *Digit. Signal Process.* **2022**, *126*, 103514. [\[CrossRef\]](#)

29. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision-ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 740–755. [[CrossRef](#)]
30. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.