

RESEARCH

Open Access



Application of computer vision technology in surface damage detection and analysis of shedthin tiles in China: a case study of the classical gardens of Suzhou

Lina Yan^{1,2}, Yile Chen¹, Liang Zheng^{1*} and Yi Zhang³

Abstract

In computer artificial intelligence, there is great potential in research on the protection of Suzhou's traditional gardens, a world cultural heritage site. As a special material in Suzhou's traditional garden architecture, shedthin tile is widely used in roof base laying and is one of the important materials for building roofs. However, professionals need to reach the roof and spend much time and effort assessing the damage before repairing it. Therefore, the main goals of this study are to investigate a machine learning-based method for finding targets and determining the type of surface damage on a shedthin tile using the YOLOv4 model trained in this study. Using 500 shedthin tile on-site photos as training samples, the model was trained for 750 epochs. The main results of this study are as follows: (1) An object detection method based on machine learning can efficiently and accurately identify damage content, overcoming the manpower and time–cost limitations of traditional assessment methods. (2) The detection model in this study has an accuracy of 85.89% for water stain recognition of shedthin tiles, 93.29% for surface scaling, 87.37% for color aberration, and 96.15% for too wide a gap. The comprehensive accuracy is 90.20%, which meets the basic testing requirements. (3) The model demonstrated its robustness and reliability in complex environments in application tests in actual scenarios, providing a methodological reference for computer vision and target detection technology in cultural heritage protection.

Introduction

Research background

The Classical Gardens of Suzhou refer to a collection of meticulously designed gardens located in the city of Suzhou, which is situated in the Jiangsu Province of China. These gardens have been recognized and

inscribed on the prestigious UNESCO World Heritage List [1]. Encompassing a duration of nearly a millennium, ranging from the Northern Song to the late Qing dynasties (eleventh–nineteenth centuries), the aforementioned gardens, predominantly erected by erudite individuals, established a set of fundamental characteristics in classical Chinese garden architecture [2]. During the mid-Ming to early-Qing dynasties, a significant number of landscape gardens thrived, leading to the establishment of approximately 200 private gardens. Currently, Suzhou boasts a total of 69 meticulously conserved gardens (Fig. 1), each of which has been officially recognized and classified as a "national heritage site" under the protection of relevant authorities [3]. In 1997 and 2000, the UNESCO organization designated

*Correspondence:

Liang Zheng

2009853GAT30002@student.must.edu.mo

¹ Faculty of Humanities and Arts, Macau University of Science and Technology, Avenida Wai Long, Taipa 999078, Macau, China

² Shanghai Total Architectural Design & Urban Planning Co., Ltd., Shanghai, China

³ Architectural Landscape Design Department, Shanghai GOODLINKS International Design Group, Shanghai, 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.

results and technical achievements have been achieved in these fields.

Computer vision technology in architectural heritage protection

The study of computer vision technology in architectural or cultural heritage areas is a rapidly developing field. This technology can help people better understand, protect, and manage these precious cultural heritage sites through the use of image processing and analysis algorithms. For example, image recognition algorithms can be used to identify specific building elements, such as columns, walls, and roofs [26]. Classifying and documenting images of architectural heritage objects are performed with the help of deep learning technology [27, 28]. Other scholars use this technology to detect and identify the distribution and size of vernacular houses in Sumba, Indonesia [29]. This information can be used for building restoration and conservation. Additionally, issues such as damage and cracks in buildings can be detected using object detection algorithms, helping people better manage and maintain buildings [30–32]. An increasing number of archeological projects, architectural heritage survey projects, and building restoration projects use computer vision technology to identify and analyze images to reduce unnecessary damage to buildings or cultural relics and achieve nondestructive testing.

Problem statement and objectives

There are currently many research results and technological applications in architectural heritage field. However, from the standpoint of building materials, many nonrenewable, distinctive, and local building materials have been created by ancient craftsmen. Past analysis models in civil engineering, such as asphalt, concrete, and pavement image recognition, do not match the shedthin tiles of Suzhou Classical Gardens. If image analysis via computer vision technology can automatically reveal the type of damage suffered by shedthin tile surfaces, this approach can also reduce the required labor costs to a certain extent. Therefore, the efficiency of the daily maintenance of architectural heritage should improve. Therefore, this article uses Suzhou Classical Gardens as an example to construct a YOLOv4 machine learning model to verify machine learning accuracy and automatically detect damage experienced by shedthin tiles in Suzhou Classical Gardens. In this article, the following five questions were explored:

- (1) According to on-site investigations and photography, how many damage type categories apply to shedthin tiles in Suzhou Classical Gardens?
- (2) How can machine learning facilitate the development of core technologies for detecting each type of damage?
- (3) What are the results of photoidentification and analysis of damage types on shedthin tiles in Suzhou Classical Gardens?
- (4) How effective is the trained machine learning model?
- (5) How accurate is automatic detection compared to manual identification?

Shedthin tiles in Suzhou classical gardens

Analysis of the characteristics of the Chinese shedthin tiles

The shedthin tile is a kind of thin brick used for roof bases in southern Jiangsu. It is located between the rafter and roof. It plays an important role in water and dustproofing the roof base. It also makes the roof surface smooth and aesthetically pleasing (Fig. 2). The earliest use of the shedthin tiles in architecture can be traced back to the Song Dynasty (960 AD). By the Ming and Qing Dynasties (1368–1840 AD), it was only used on small buildings. Since laying shedthin tiles requires considerable workmanship, it is usually used in higher-grade buildings. Shedthin tiles come in long strips and square gray bricks; the specifications are not uniform. Shedthin tiles are used in different specifications according to the size and grade of different buildings. Among them, the long type is mostly used on the roofs of corridors and pavilions, with a general specification of 220 mm*110 mm*20 mm. The square shedthin tiles are less common, with general specifications of 220 mm*220 mm*25 mm or 300 mm*300 mm*30 mm, and are used in high-level halls or ordinary halls. Modern shedthin tile specifications are smaller, with three specifications: 210 mm*100 mm*13 mm, 220 mm*100 mm*16 mm, and 220 mm*90 mm*12 mm.

Due to the thinness of the shedthin tiles, they easily deform during firing. Therefore, before firing, the blank is cut into two pieces with thin iron wire. The two pieces are stacked together during firing and then separated with a tile knife after firing. This firing method gives shedthin tiles both positive and negative sides. When lying, the front side (the flat and smooth side) faces downward as the viewing surface. The fair-faced shedthin tiles are commonly used in the Suzhou area, and the method is more specific. Fair-faced shedthin tiles requires a second coat of gray and grout after firing to unify the color of the tiles. When laying, a white line was added on the long side of the ornamental surface to make the joints between bricks neater and more aesthetically pleasing. Today, the “fair-faced shedthin tiles” method can still be seen in most historically protected ancient buildings in

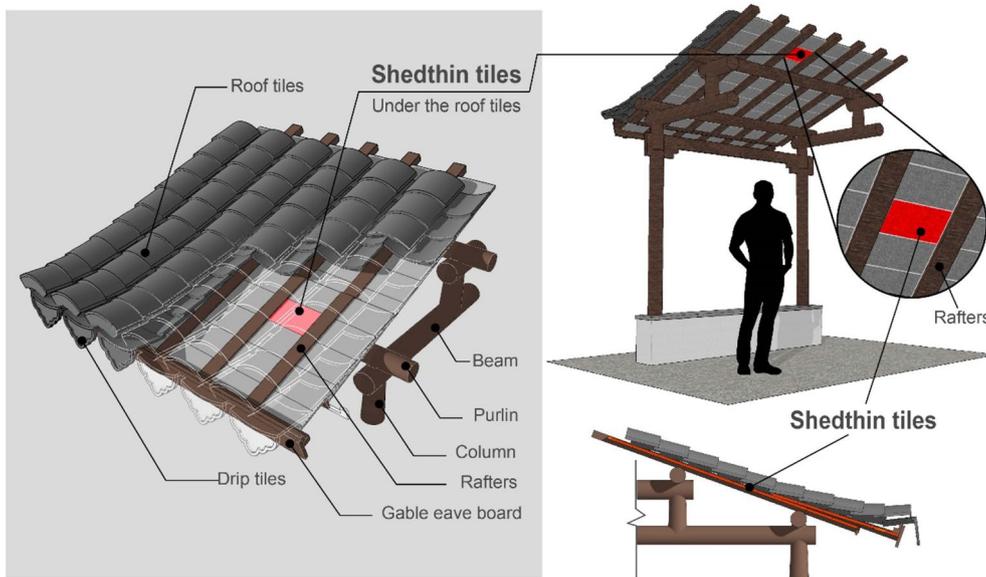


Fig. 2 Shedthin tile location on the roof. (Drawn by the author)



(1) Shedthin tile on the roof of Truth Ting in Humble Administrator's Garden



(2) Shedthin tile on the roof of Small Feihong, Humble Administrator's Garden

Fig. 3 Shedthin tiles photographed during a field trip. (Image source: photographed by the author)

Suzhou and the Jiangnan area, such as the tops of pavilions and corridors in The Humble Administrator's Garden (Fig. 3).

Analysis of damage types and factors in shedthin tiles

Most of the pavilions and corridors in Suzhou Classical Gardens use shedthin tiles as the roof base laying material. Today, these preserved, historically protected buildings have experienced hundreds or even nearly a thousand years of history. Although it has undergone many repairs, due to the loss of firing techniques and the gradual reduction in the number of craftsmen, the number of old shedthin tiles of good traditional quality has gradually disappeared. The protection and repair of shedthin tiles are urgent (Fig. 4).

There are two main reasons for damage to shedthin tiles: climate and construction process problems. The first is climate issues related to the service life of the shedthin tiles. Rainwater and moisture will keep shedthin tiles in a wet state for a long time, and the bricks will freeze in the winter. When the water saturation exceeds 80%, the tile will be damaged after experiencing multiple freeze–thaw cycles. Moisture can also cause mold to grow and accelerate brick damage. Workmanship issues are related to the improper construction of shedthins tiles. As mentioned before, since shedthin tiles need to be separated with a tile knife after firing, the thickness will vary during the separation process. This requires craftsmen to polish the tiles to uniform specifications. During the polishing process,



(1) Repair site of shedthhin tiles on a roof in The Humble Administrator's Garden

(2) Craftsmen are repairing the roof and replacing damaged tiles and shedthhin tiles

Fig. 4 On-site image of the shedthhin tiles restoration site. (Image source: photographed by the author)

surface damage occurs, exposing the internal bricks. The internal brick body is more fragile than the surface, which reduces the service life of the brick and increases the chance of breakage.

Based on the above two main reasons, the researchers summarized four common damage types through analysis (Fig. 5): water staining, surface scaling, color aberration, and excessive gap.

(1) Water staining: the shedthhin tiles have certain moisture-proof and breathable properties; however, long-term moisture and freeze–thaw cycles cause water to stay inside the brick, forming water stains, which accelerate brick aging. There are two types of water stains on shedthhin tiles: surface water stains and penetrating water stains. Surface water stains are strips of water stains left behind when rainwa-

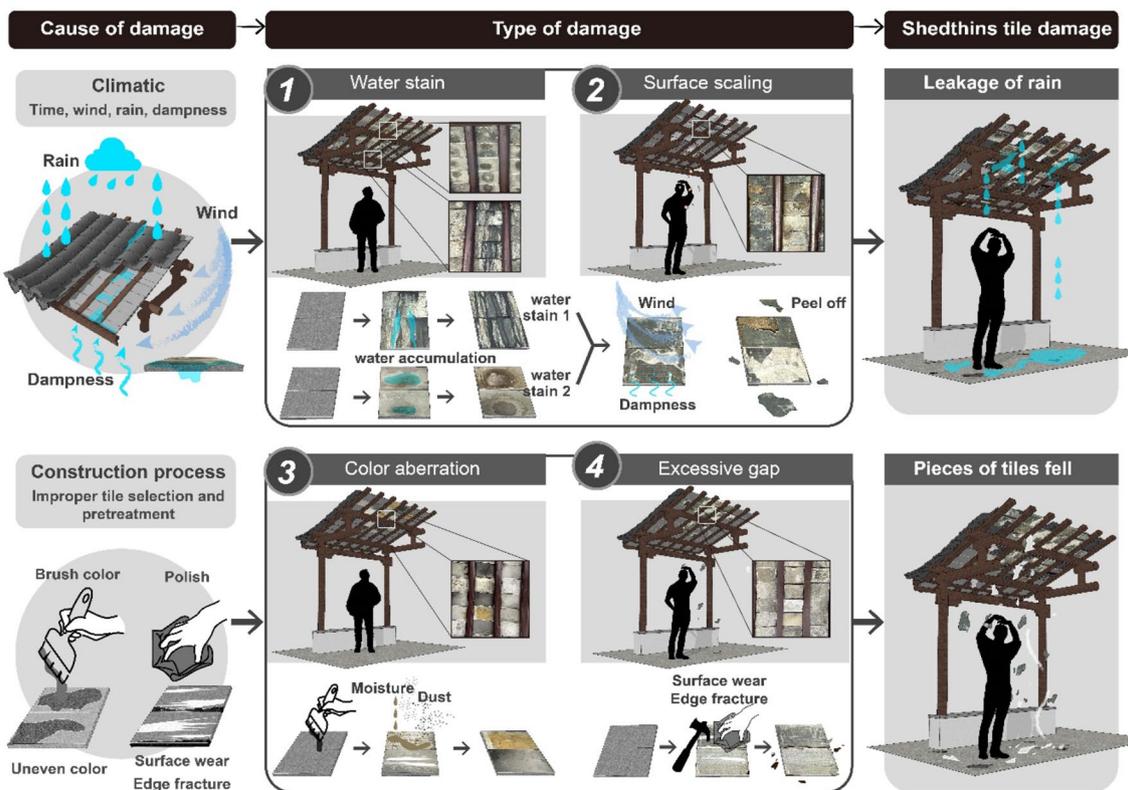


Fig. 5 Causes and processes of damage to shedthhin tiles. (Drawn by the author)

ter leaks from the roof and flows onto the surface of the shedthin tile, damaging the mortar on its surface. Penetrating water stains occur when rainwater flows onto the shedthin tiles in the horizontal laying direction of the roof. After a long period of retention, the rainwater penetrates into the interior of the tiles to form circular water stains. The occurrence of these two water stains on the shedthin tiles indicates a roof leak.

- (2) Surface scaling: under the influence of long-term rain and moisture, the mortar on the surface of the shedthin tiles gradually degrades, and the deeper areas penetrated by water form a bulging air layer as a result. These bulging air layers accelerate aging due to moisture-laden winds, which cause the surface to peel off. Since the shedthin tiles is located on the inside of the roof, the bulging mortar on the surface will fall downward with gravity and affect the area where people walk.
- (3) Color aberration: the surface of the fair-faced shedthin tiles is usually painted with gray water and grout to unify the color of the bricks. Bricks of different thicknesses are also polished. Color differences are caused by improper repair processes. There are two types of color differences. First, the color of the gray water and grouting mixture is not uniform because the gray color will differ between batches, producing different shades. Second, wear marks will be produced after polishing. These wear marks can easily accumulate dust and oil from the air, resulting in yellowing and deterioration of the brick surface.
- (4) Excessive gap: shedthin tiles are thin and not particularly hard. In Suzhou World Heritage buildings, the practice of retaining as many old bricks as possible is usually adopted. The service life of most shedthin tiles has been exceeded. The specifications of old bricks continue to shrink due to wear and tear, and when they are reused in their original positions, they cannot be aligned with other bricks, resulting in leaks and excessively large seams between spliced bricks. Excessively large brick joints not only cause serious rain leakage and dust fall but may also cause rafters on both sides to fall and injure pedestrians.

According to the samples collected during the on-site inspection, the shedthin tiles in Suzhou's world heritage buildings are generally seriously damaged. Historical buildings must retain their original traces during repairs to maximize their authenticity while also ensuring the continued functional integrity of the building. This contradiction is becoming increasingly prominent, making

protection and repair more difficult. If the roof is not protected, severe damage or even roof collapse will occur. To improve the intelligence and sustainability of world heritage buildings, this research used a machine learning method that can be used to quickly detect real-time damage to shedthin tiles.

Materials and research process

Photo image collection source

The samples for this study were collected from the buildings and corridors in The Humble Administrator's Garden, The Lingering Garden, and The Master-of-Nets Garden. In 1997, these three gardens were recognized as UNESCO World Heritage Sites. The shedthins tiles in this architecture is highly representative. During the field investigation, a large number of shedthin stile samples were collected—a total of 670—and 500 valid samples were obtained after screening. Among them, there were 230 valid samples of shedthin tiles from The Humble Administrator's Garden, 137 valid samples of shedthin tiles from The Lingering Garden, and 133 valid samples of shedthin tiles from The Master-of-Nets Garden (Table 1).

The shedthin stile sample collection locations were selected on the inner roofs of buildings and corridors with obvious damage features and good lighting and were shot with high-definition cameras to ensure sample clarity. After the samples were collected, the damage characteristics were manually classified, and the four damage types summarized above were obtained (Fig. 6).

By finely dividing these four visually distinct damage types, the YOLOv4 model can learn and distinguish the differences between them more accurately. This approach enables accurate damage detection and analysis of shedthins tiles in practical applications, providing technical support for protecting and restoring ancient buildings.

Research process

This study adopted an experimentally verified and effective research process and systematically explored the application possibility and practical value of computer vision technology in shedthin tile damage detection in Jiangnan Classical Gardens through a series of steps from data collection to final application [33]. In this process, an automatic detection method for shedthin tile damage was proposed and verified based on the YOLOv4 model (Fig. 7). By integrating computer vision technology with the actual needs of ancient building protection, this research aims to construct a model that can accurately and quickly detect damage to shedthin tiles. Its ability to provide solid scientific and technological support for

Table 1 Shedthin stile image collection locations

UNESCO ID	Name & Location	Coordinates	Acquisition time	Collection position	Quantity
813-001	The Humble Administrator's Garden 	N31 19 27.721 E120 37 44.62	12/07/2023	Gallery of the Wuzhu Youju Pavilion (Wu Bamboo Cabin) Gallery of the Jianshan Building(See Hill House) Gallery of the Yi Two Pavilion Complex Gallery Xiaofei hong (Small Rainbow) Bridge Gallery of the Dezhen(Truth) Pavilion	230
813-002	The Lingering Garden 	N31 18 56.2 E120 35 33.515	09/07/2023	Gallery of the Hanbi Cottage Gallery of the Banye Thatched Cottage Gallery of the Wenmu Xixiang Xuan Gallery of the Qingfeng Chiguan Gallery of the Guanyun Viewing Platform Gallery of the Wufengxian Hall	137
813-003	The Master-of-Nets Garden 	N31 17 52.66 E120 38 2.418	11/07/2023	Gallery of the Yuedao Fenglai Pavilion(Moon Comes with the Breeze Pavilion) Gallery of the Zhuwai Yizhixuan Pavilion(Bamboo Branch Pavilion) Gallery of the Zhuwai Yizhixuan Pavilion(Bamboo Branch Pavilion) Gallery of the Xiaoshan Conggui Pavilion(Small Hill and Osmanthus Fragrance Pavilion) Gallery of the Zhuoying Shuige(Cear Water Pavilion)	133

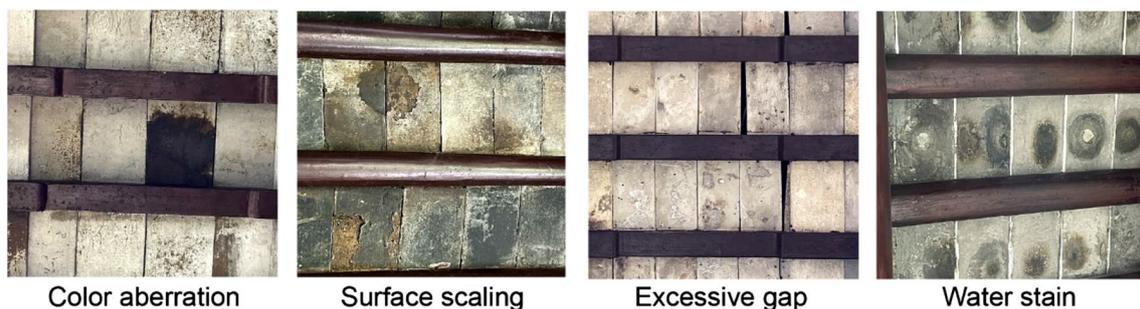


Fig. 6 On-site collection of shedthin tiles damage types. (photographed by the author)

the long-term preservation of Jiangnan Classical Gardens will be utilized in actual cultural relic protection work.

- (1) Data collection: the diversity and representativeness of the data are directly related to the stability and generalizability of the model. Therefore, dur-

ing the data collection stage, this study places special emphasis on obtaining fully diverse and highly representative shedthin tiles image data. In several typical locations in Suzhou gardens (Fig. 1), 500 high-definition shedthin tile images containing rich damage types were collected. (In terms of dam-

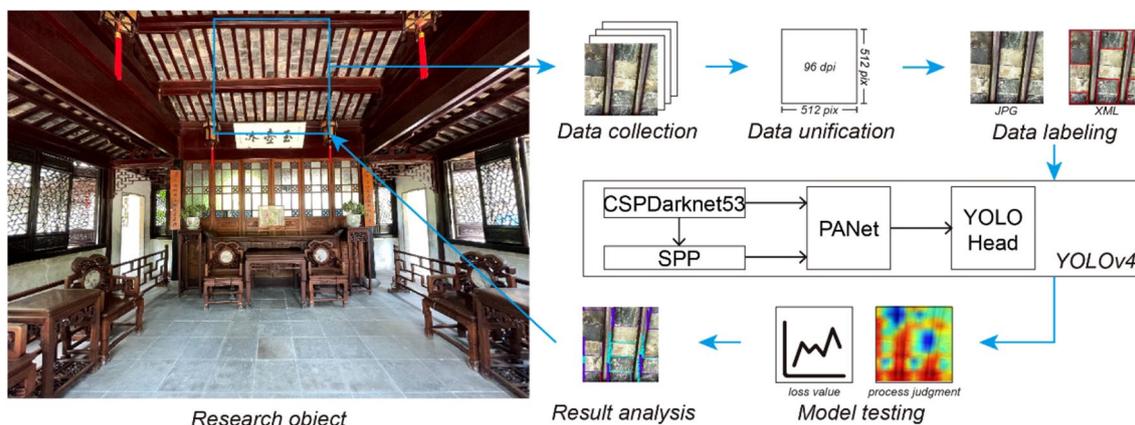


Fig. 7 Research process. (Drawn by the author)

age types, the dataset includes 163 images of water stains, 164 images of surface scaling, 95 images of color aberration, and 78 images of excessive gaps.) To cover as many actual scene changes as possible, image collection was carried out under different weather conditions (sunny, cloudy, and rainy days) and light conditions (sunlight, shadow, and artificial light sources). The data collection process lasted for one month to capture the different manifestations of damage in various environments. In addition, to ensure that the data collected can reflect the characteristics of various damage types, including but not limited to water staining, surface scaling, color aberration, and too wide a gap, the diversity of the data comes not only from differences in the type of damage but also from the multiple dimensions of severity, size, shape, and color of the damage.

- (2) **Data processing:** a series of image preprocessing strategies are adopted in the data processing stage to optimize the quality of the images and ensure the consistency of the data input, thereby improving the efficiency and effectiveness of model training. These include histogram equalization processing and noise filtering technology. The purpose of this approach is to eliminate the influence of environmental factors such as light and shadow, reduce random noise, and enhance the clarity of damaged features in the image. In addition, this study performed image size standardization to ensure that all the images input to the YOLOv4 model had the same resolution and size. Furthermore, image size standardization was performed by adjusting all the images to a uniform resolution of 512×512 pixels. This was achieved through a combination of resizing and mosaicking, with an overlapping strategy applied during the mosaicking process to ensure

that complete features of the shedthin tiles were present in at least one image. This step ensures the stability of model training and reduces the computational complexity caused by inconsistent image sizes and proportions. Finally, in the training stage of the model, code for data enhancement operations, such as rotating and flipping images, is also added to expand the diversity of the dataset and enhance the generalizability and robustness of the model.

- (3) **Data annotation:** the data annotation stage is key for ensuring that the YOLOv4 model learns accurate features; thus, the research team will focus on providing accurate and consistent annotation information for each image. In this article, the professional image annotation tool LabelImg is used to draw accurate bounding boxes for the damage phenomena in each image [34]. Each damage type is assigned a unique code based on its name to build an exact correspondence. During the annotation process, this study focused on the consistency and accuracy of the annotations and implemented two rounds of annotation review and calibration to ensure the accuracy of each annotation box and category label. In addition, to further improve the generalization ability of the model, the different stages and degrees of various types of damage are covered during labeling, ensuring that the model can make accurate predictions when facing damage of different severities.
- (4) **Model training:** in the model training stage, YOLOv4 was selected as the training model. The decision to utilize YOLOv4, instead of more competitive algorithms such as Faster-RCNN, was based on a balance of several factors. First, YOLOv4's single-stage detection architecture ena-

bles simultaneous prediction of object categories and locations in a single network pass, offering a more streamlined process than the two-stage architecture of Faster R-CNN. This results in faster processing speeds, which are essential for real-time or near-real-time object detection. Second, the architecture of YOLOv4 includes advanced feature extraction technologies such as DarkNet-53, spatial pyramid pooling (SPP), and path aggregation network (PANet), which enhance detection accuracy by capturing detailed image contexts. Third, although Faster-RCNN may provide more detailed detection in some scenarios, it tends to have higher computational complexity. YOLOv4 strikes a better balance between performance and complexity, fitting the objectives and resource constraints of this study. Although YOLOv4 is not the latest version of the you only look once (YOLO) series, preliminary testing of the newer YOLOv8 model revealed only minor improvements in performance and efficiency compared to YOLOv4, with no significant differences. Given the research team's expertise in YOLOv4, YOLOv4 was used as the machine learning model for this study. Specifically, this study first conducts pretraining on the VOC2007 open source image dataset to enable the model to learn general image features and subsequently fine-tunes it on the shedthin stile image dataset annotated in this study to ensure that the model can adapt and recognize shedthin stile damage-specific characteristics. During the model training process, the cross-entropy loss function is used to quantify the difference between the model predictions and the real labels, and the Adam optimizer is used to iteratively update the weight of the model to minimize the loss during the training process. Model training at this stage is divided into two key stages, the freezing and unfreezing stages, to ensure that the model can learn specific features for damage detection in shedthin tiles while making full use of pretraining knowledge. In the freezing stage, the weights of the model backbone (feature extraction network) remain unchanged; that is, the weights are "frozen" to utilize the common features in the pre-trained model. A total of 10 epochs of training are performed using a batch size of 2, and the learning rate is set to 0.001. These settings allow the model to quickly adapt to the characteristic representation of brick damage. During the unfreezing phase, the model backbone is "unfrozen", allowing the weights to be updated so that the model further learns the specific characteristics of shedthin tile damage. This phase involves a longer period of training, with a

total of 750 epochs. While keeping the batch size at 2, the learning rate is adjusted to 0.0001 to fine-tune the model and optimize its performance on the shedthin tile damage detection task. Additionally, extensive data augmentation techniques are employed to enhance the model's ability to generalize and improve its robustness. These include image transformations such as random rotations (ranging from -30 to 30 degrees), horizontal and vertical flips, scaling, and translations. Variations in brightness and contrast simulate different lighting conditions. Furthermore, random cropping is used to encourage the model to recognize damage features in varying spatial contexts. These data augmentation steps are critical in preparing the model to handle diverse real-world scenarios and contributed significantly to its improved performance in detecting and analyzing shedthin tile damage.

- (5) Model testing: the main goal of the model testing phase is to assess the model's performance on a different dataset from the training model to predict the impact of the model during actual deployment. This study carefully prepares a diverse test set, including a total of 40 images of four types and varying degrees of damage to shedthins tiles, to comprehensively examine the generalization ability of the model. In model testing, this study employs multiple performance metrics. In terms of algorithm indicators, the average precision (AP) and miss rate (MR) are used. These two indicators can reflect the accuracy and MR of the model. The algorithmic indicators usually do not fully reflect the actual detection capabilities of the model. To reflect the quality of the model in practical applications, the final model detection results are manually judged and counted one by one to obtain the final model accuracy.
- (6) Results analysis: the results analysis phase of model testing aims to gain an in-depth understanding of model performance and potential room for improvement. At the macro level, this study summarizes the comprehensive performance of the model on the overall test set and analyzes whether the model can stably and accurately identify and locate different damage types. At the microlevel, this study focuses on the detection effect of the model on each damage type and explores the advantages and disadvantages of the model in identifying different damage types, such as whether there is a higher detection rate for certain damage types or specific damage states, sensitivity or bias. Through an in-depth analysis of the model's performance under various scenarios and conditions, this

study clarifies the challenges that the model may face in practical applications and how to optimize the model further to address these challenges. In addition, this study analyzes possible failure cases of the model and explores the possible underlying factors, such as data imbalance and insignificant features, to provide inspiration for subsequent model improvement and optimization.

Through in-depth research and exploration of these six steps, this study developed a computer vision model that can accurately identify and locate damage to shedthins tiles.

Model settings

The network framework of the YOLOv4 model is shown in Fig. 8. The overall architecture consists of the DarkNet53 backbone network, the PANet feature extraction layer integrated with SPP structure, a step in the 'convolutional neural network that turns the picture into a feature vector, and the final detection head. DarkNet53 serves as the backbone network for extracting basic features from images. In the feature extraction layer, PANet enhances the feature expression ability of the model at different levels through top-down and bottom-up information flow guidance to balance the gradient propagation of features and the propagation of activation values, improving the model's learning ability. By introducing the SPP structure, multiscale feature fusion is achieved,

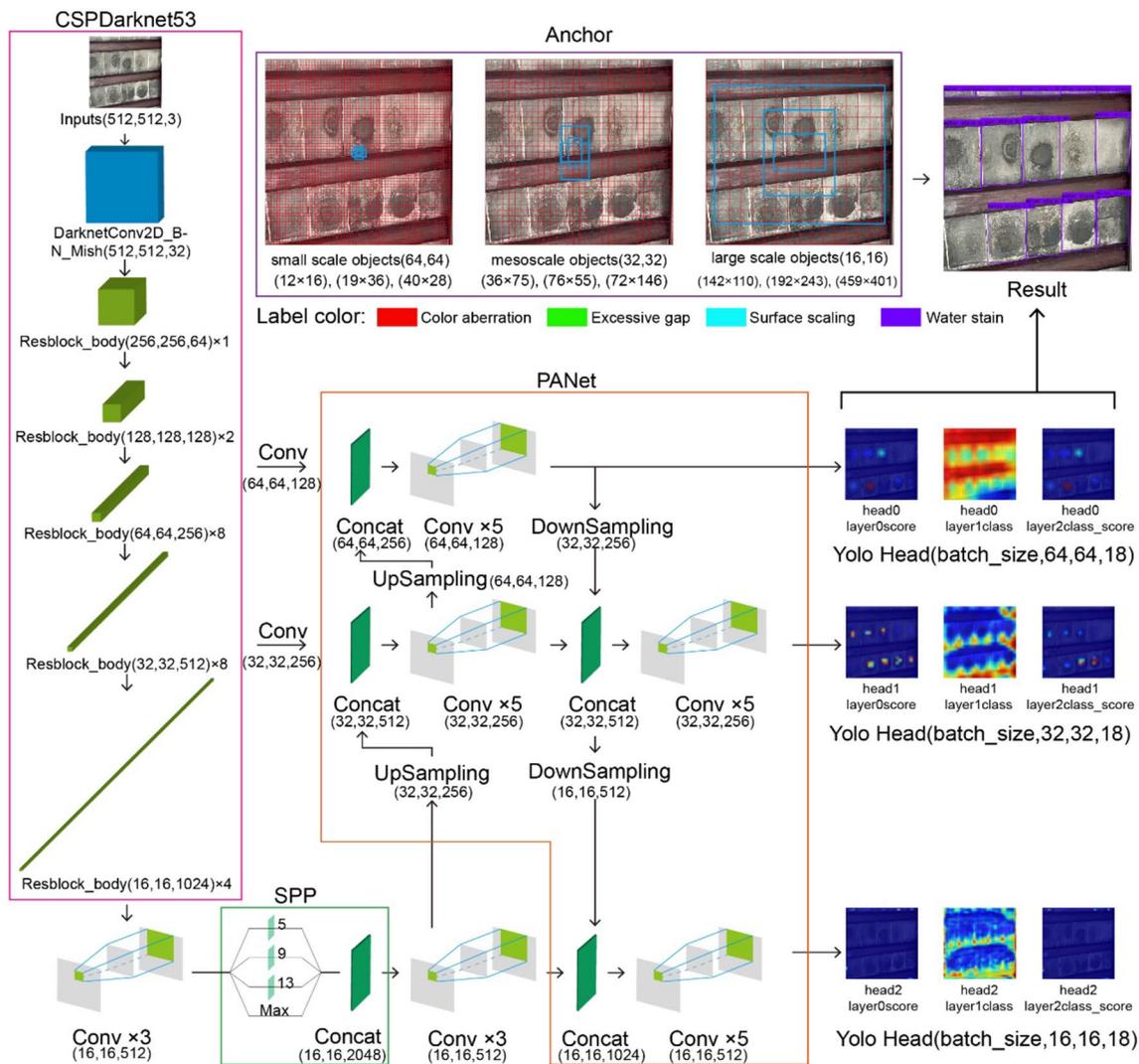


Fig. 8 YOLOv4 network framework. (Drawn by the author)

allowing the model to adapt to and accurately identify targets of different sizes. The detection head predicts the bounding box coordinates and categories of the target at multiple scales. The generator is responsible for the final target detection and classification, using the generated feature maps to predict bounding boxes, target scores, and category scores. The detection head makes predictions at multiple scales and can detect targets of different sizes. Strategies such as the Mish activation function, CIoU loss function, and learning rate cosine annealing scheduling are also used during the model training process to ensure that the model achieves stable and accurate detection results in complex scenes and diverse target scales [34].

This study also made a series of refinements to YOLOv4 to improve its efficiency in shedthn tile damage detection. This study precisely customized the output layer of the model and established four category nodes corresponding to the four brick damage types determined in the study. After in-depth analysis of the training dataset and multiple experiments, the sizes and proportions of the predefined anchor boxes in YOLOv4 were adjusted to more accurately capture damage characteristics of various sizes and shapes. In the data preprocessing process, two different data enhancement techniques, namely, color dithering and size transformation, are used to simulate a variety of lighting, environmental, and viewing angle conditions. To fully broaden the generalizability of the model during the training process, it must still maintain robust detection performance in diverse practical application scenarios.

In the model training process, this study adopted a staged strategy to more accurately fine-tune the YOLOv4 model to adapt to the specific task of identifying tile damage. In the first step, the model backbone weights were frozen to ensure the stability of the feature extraction network. The shedthins tile damage dataset was subsequently used for preliminary fine-tuning. This stage was trained for 10 epochs, using a batch size of 2 and a learning rate set to 0.001. The purpose was to allow the model to learn the basic feature expression of shedthn tile damage based on fully borrowing pretraining knowledge. In the second stage, the weights of the model backbone are unfrozen, allowing the weights of the feature extraction network to be updated and optimized, adapting to the characteristics of shedthn and tile damage at a deeper level. At this stage, the training cycle of the model reaches 750 epochs, the batch size is maintained at 2, and the learning rate is moderately reduced to 0.0001 to ensure that the model can converge stably during the process of deep fine-tuning. Notably, after 750 generations of training, the loss value of the model did not further decrease. This may mean that the model reached the optimization

limit on the current dataset, so this study chose to terminate training at this stage. In subsequent experiments, among the hundreds of weight files generated in these 750 epochs, the model with the best performance was selected based on the performance of the validation set as the basis for further experiments and analysis.

Discussion: automatic recognition result analysis

Model test

The model testing step is one of the key steps in this study and is evaluated in detail and systematically in three parts. A detailed description of these three parts is provided below.

LOSS value

Figure 9 shows that the model learns and converges in important ways during the training process. This is done by examining how the training loss and validation loss change over time. The loss value in the initial stage experienced violent fluctuations, especially for large outliers that appeared in the training loss value. This may be due to the rapid adaptation of the model to the data and the rapid optimization of parameters in the early stages. As the number of epochs increases, the training loss value steadily decreases and finally approaches a relatively stable level, indicating that the model is continuing to learn and gradually converges. Although the verification loss value remains relatively stable throughout the training process, its slight fluctuations and differences from the training loss may indicate a certain degree of overfitting in the model. On this basis, in this study, several key models were selected for subsequent in-depth analysis and comparison. The 90th epoch model exhibited the lowest validation loss (loss = 3.82), which indicated that it performed well on the validation set and exhibited good

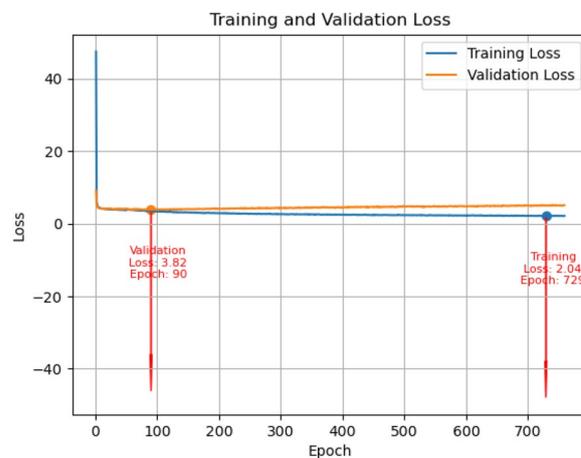


Fig. 9 LOSS values of the model training logs. (Drawn by the author)

generalizability. The 729th epoch model showed the lowest training loss (loss=2.04) on the training set, indicating a strong fitting ability. The 750th epoch model is the model when the training cycle is completed, and its performance reflects the learning results of the model under the full amount of training data. The 100th epoch model serves as a randomly selected reference point, providing this study with a performance snapshot in the early training stage.

AP (average precision) and MR (miss rate)

Before in-depth analysis of the model test results at different stages, it is necessary to clarify several core evaluation indicators: AP (average accuracy) reflects the model's ability to detect various shedthin tile damage types (such as surface scaling and color aberration) with accuracy. The MR is closely related to the damaged areas that may be ignored when the model is applied in the field, and these areas may be critical repair and protection areas. The ground truth represents the real and conclusive

annotation information in the dataset, including the category and location information of the object, and is the basis for model training and evaluation. On this basis, by observing and analyzing the performance of the model in each training stage, some phenomena and patterns of the model can be discovered (Fig. 10).

In general, the longer the training period is, the better the model should perform. However, in this study, the min validation loss model at 90 epochs achieved an mAP of 43.57% (mean average precision), which was much greater than the 30.88% mAP of the max-epoch model at 750 epochs. This reveals that the model locked in the key information of shedthin tile damage features in a relatively short training period and that later training may have paid too much attention to noise or secondary features in the training data, resulting in no further improvement in performance. In detecting various damage types, a EG (excessive gap) generally has a high AP value, which may be related to its visually prominent feature—a wider gap can make an obvious difference in the

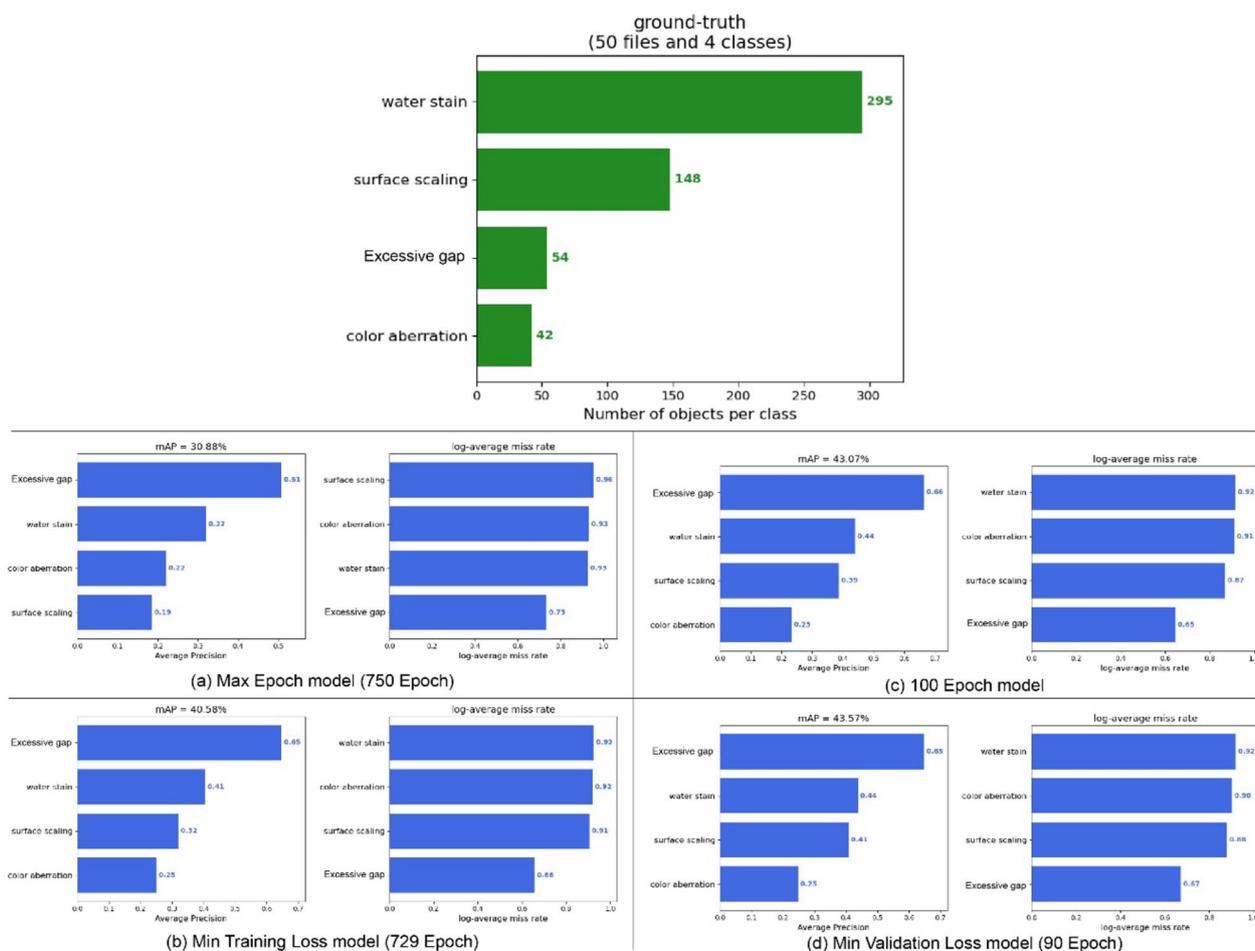


Fig. 10 Comparison of AP and MR indices between different models. (Drawn by the author)

image. The performances of SS (surface scaling) and CA (color aberration) are relatively weak, especially in the max-epoch model, where the AP value of SS is only 0.19 and the MR value is as high as 0.96. This may reflect the complexity of the visual characteristics of these two types of impairment and the potentially less noticeable forms of impairment. Considering the sample distribution of each category in the ground truth (WS (water stain): 295, SS: 148, EG: 54, CA: 42), the difference in sample size between different categories may have a certain impact on model learning. Further research may find optimization directions for balancing sample distributions or improving model structures.

Comparison and analysis

In the model measurement and analysis process, six representative test photos were selected as experimental samples to further verify the detection performance of each model through actual application scenarios. All the models use unified parameter settings during the testing phase; that is, the confidence threshold (confidence) is 0.1, and the intersection-over-union (IoU) threshold (nms_iou) of nonmaximum suppression is set to 0.3 to ensure the comparability of the outputs of each model. The experimental results are shown in Fig. 11. There are common phenomena of missed detection and false detection in the detection of each model, but the degrees of detection are different. The model trained for 750 epochs exhibited more severe missed and false detections. For example, in Figs. 12A2, C2 and 13B2, and, the number of shedthin tile damage incidents recognized by the model is significantly less. In Fig. 12E2, the model incorrectly identifies "surface scaling" (SS) as CA. Relatively speaking, the minimum validation loss model performs more prominently, with relatively fewer missed detections and false detections. It can identify shedthin tile damage more accurately in the six test images. However, the model still misses situations when the damaged areas are dense or widely distributed. As shown in Fig. 12D5, although the damage type of the entire photo is SS, the model can only detect part of the damage type and mis-detect it as WS in some areas. This reveals, to a certain extent, the similarity in certain visual features between SS and WS, making it more challenging for the model to distinguish between them. Taken together, these results are generally consistent with the conclusions of the previous indicator analysis stage.

In summary, the min validation loss (90 epochs) yields better detection results, and this study selects this loss as a model for further application. Moreover, this study exported and evaluated the output of each detection head in the YOLOv4 model to obtain a better understanding of how the model works on the

inside during the damage detection process. These output results are presented in the form of heatmaps, with particular attention given to their ability to identify shedthin tile damage at different scales. In YOLO's architecture, different detection heads are usually responsible for identifying targets of different scales. By analyzing their output, we can further understand the model's adaptability and recognition strategies for object scales. As shown in Fig. 12, due to the relatively short shooting distance, head0 and head1 show more significant feedback, mainly focusing on smaller-scale target detection, while head2's response to the image is relatively weak, which also confirms that it is usually responsible for detecting the characteristics of larger-scale targets. In the multilayer output of the model, the "score" layer functions as a quick location target, highlighting the approximate location of the damaged area in the image. The "class" layer is responsible for determining the category of the located area, that is, distinguishing different types of shedthin tile damage. The "class_score" layer combines the information from the first two layers and outputs the final heatmap, which reveals the model's judgment and confidence in different damage types in each area. Finally, the model locks and outputs the final detection result through the pre-defined anchor box strategy.

Model application

In the practical application stage of the model, this research adopted a real-scene experiment method, specifically taking an original picture from the interior of the building in the Jiangnan Classical Garden. Its size is 6000×4000 , which is much larger than the 512×512 image used in the model training and testing stages to simulate and evaluate the detection effect of the model in actual field applications. Figure 13 presents the results of this application experiment. Given the larger size of the captured images and the longer shooting distance, the model was able to capture more visual brick damage phenomena. In the generated heatmap, the "score" layer of head2 successfully locks the exact location of the damage, and the "class" layer also achieves relatively accurate classification of damage types. The overall output results mainly focus on two damage types, "surface scaling" (SS) and WS, which are consistent with the actual situation on site. In addition, the performance of the model in terms of error correction capabilities is worthy of attention. For example, in the "score" layer of head0, there are large areas of false positives (red areas). However, these locations are not strongly reflected in the corresponding "class" layer. This shows that the model suppresses these erroneous detection responses during the type judgment

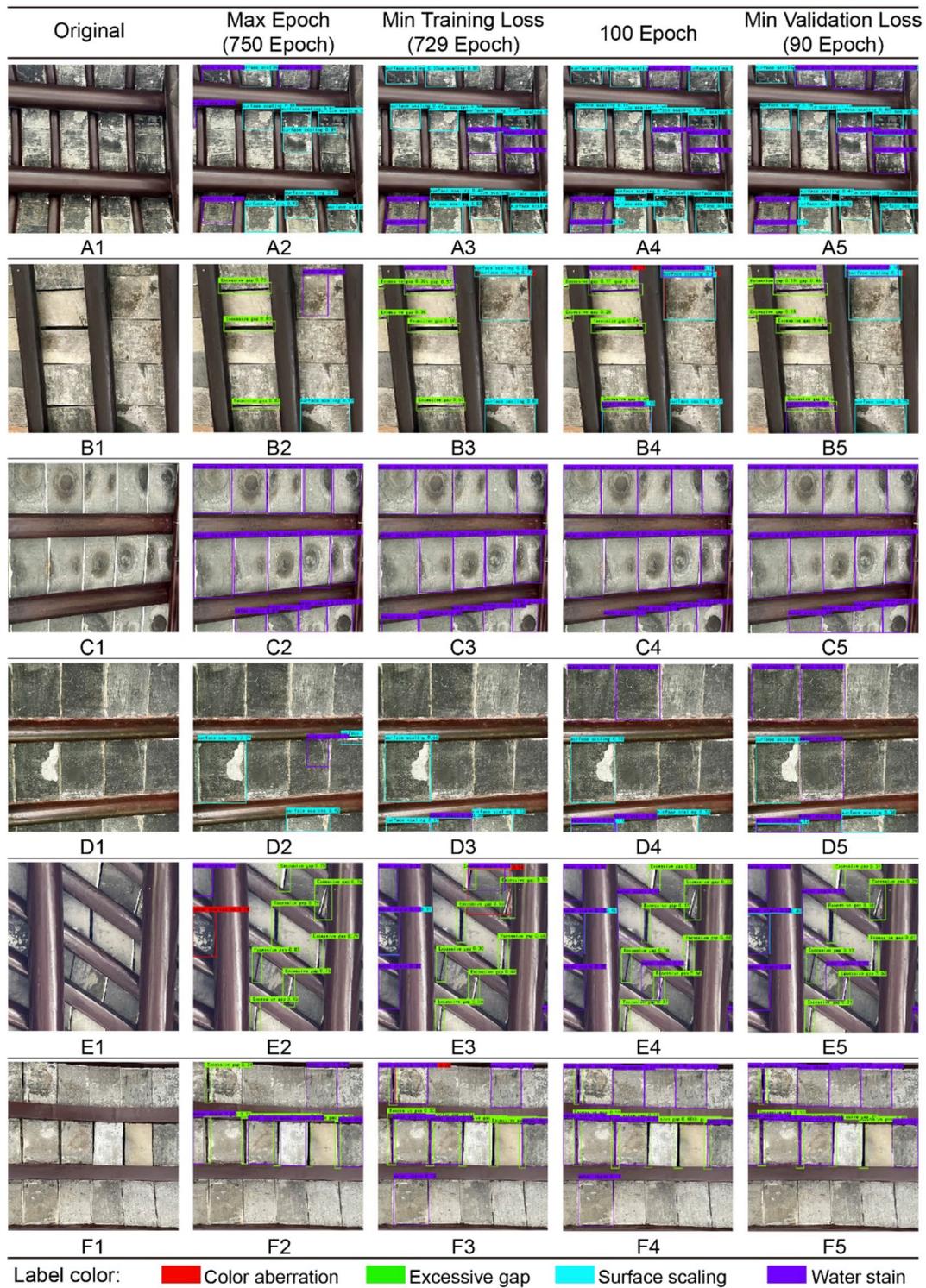


Fig. 11 Model test results for different epochs. (Drawn by the author)

stage, so these false positives are not reflected in the final detection results. These findings highlight the robustness and reliability of the model in practical applications.

To further analyze the effectiveness of the model in practical application, the following section will focus on the detection accuracy of the model for different types

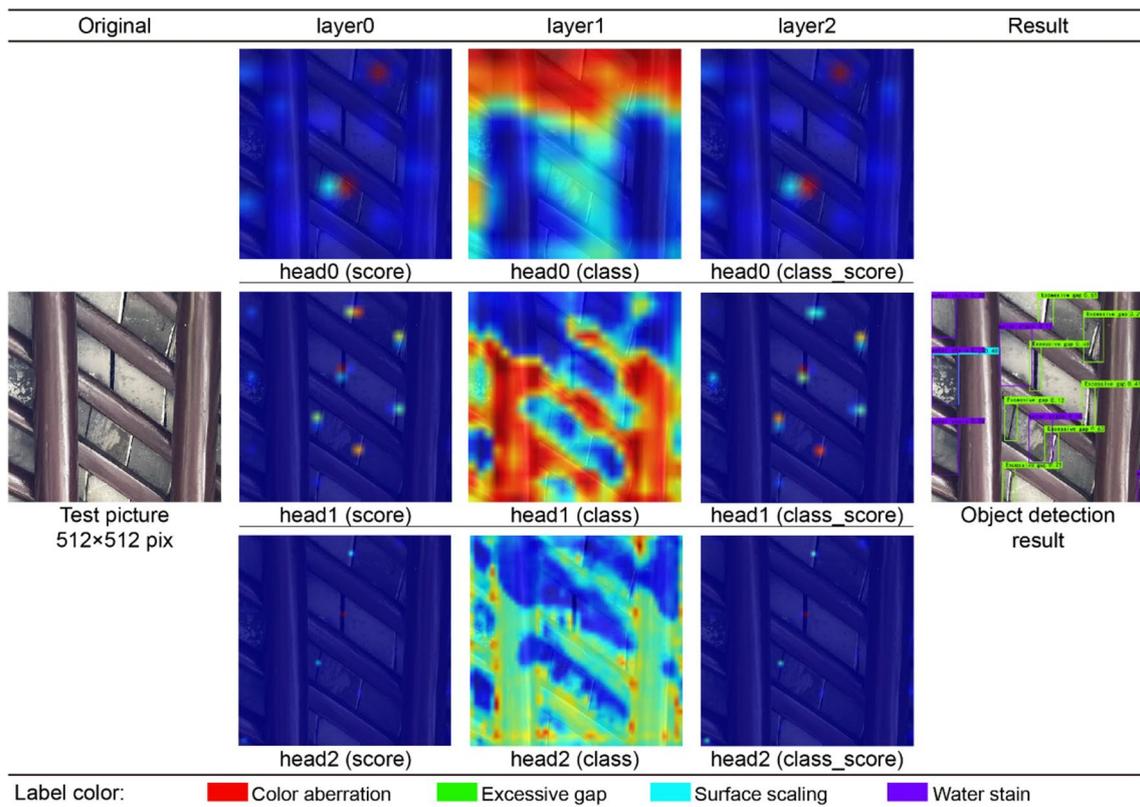


Fig. 12 Heatmap analysis of test images. (Drawn by the author)

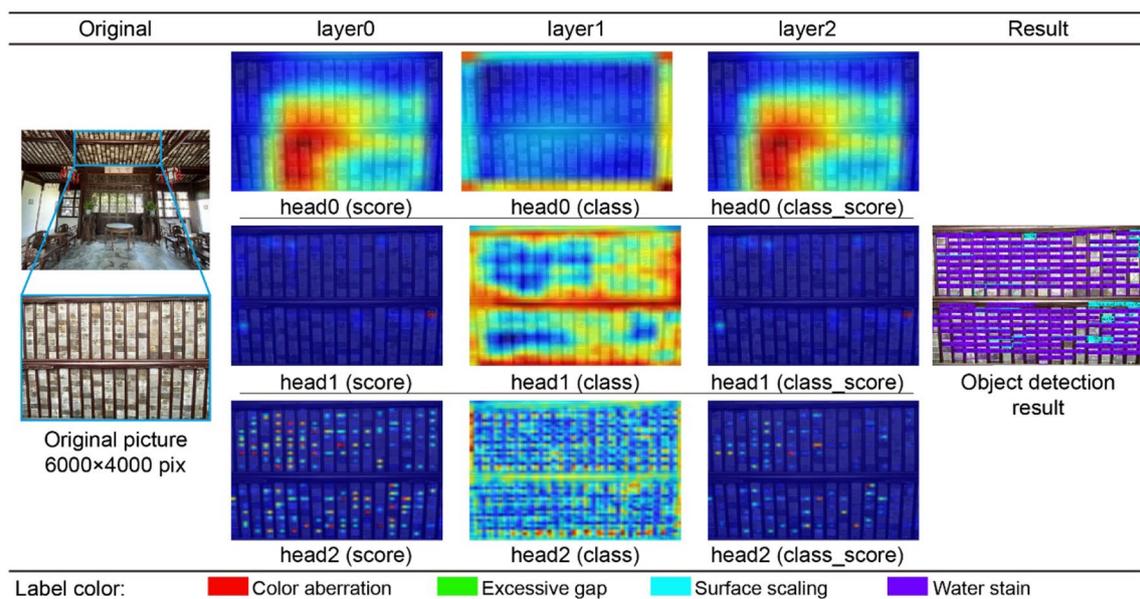


Fig. 13 Heatmap analysis via on-site photo inspection. (Drawn by the author)

of brick damage. Thus, four sets of real-life images with different damage types as the main subject were collected in an outdoor corridor environment in Jiangnan Classical Gardens for model testing. Since the corridor is located outdoors, its environmental factors are complex and changeable. A variety of factors, such as rainfall, moisture, and air pollutants, may affect damage formation and model detection, increasing the diversity and complexity of damage types and manifestations.

Figure 14 shows the detection results of four groups of experiments. In the image dominated by CA in Project 1, the model successfully captured all CA damage and identified some "surface scaling" (SS) damage. In the images dominated by the EG in Project 2, the model accurately located each EG and could also detect other types of damage, but some SS were misjudged as WS. In the images dominated by SS in Project 3, the model generally performs well, but there are still a few cases where SS is misdetected as WS. In the image dominated by WS in Project 4, the model successfully completed the detection task, every WS was accurately detected, and there were no false detections or missed detections.

The above experimental results indicate that, in most cases, the model can ideally complete various damage detection tasks. However, in some cases, especially in SS and WS detection, the model exhibits a certain degree of confusion, misdetecting SS as WS. However, overall, the detection performance of the model is excellent, laying a solid foundation for its deployment in practical application scenarios.

Manual validation of the models

Finally, the researchers evaluated the accuracy of the model (Fig. 15) by performing batch detection on 500 images. In this process, 20 images can be processed per second, demonstrating the efficiency of the model detection process. The researchers manually checked and evaluated all the test results, and the main conclusions are as follows: (1) the shedthin tsTile WS recognition accuracy is 85.89%, the surface scaling recognition accuracy is 93.29%, the CA recognition accuracy is 87.37%, the too-wide gap accuracy is 96.15%, and the comprehensive accuracy is 90.20%. (2) Among the various damage types, the shedthins tile, which is too wide, has the highest recognition accuracy. Water staining has a lower recognition rate because it is similar to surface scaling damage. (3) Considering both efficiency and accuracy, this model has advantages in damage detection in shedthin tiles (Fig. 15).

Analysis of potential in cultural heritage protection

In model testing and application, this study provides profound experimental insights and multidimensional evaluation perspectives. (1) Through comprehensive evaluation of models at different training stages, researchers have observed that, in some cases, an increase in the number of epochs does not always lead to an improvement in mAP. For example, the mAP of the 750-epoch model is lower than that of the 90-epoch model. However, in most cases, the model shows relatively consistent detection trends. (2) In terms of damage type detection accuracy, the EG usually obtains higher AP values. The "surface scaling" (SS) rate is relatively low,

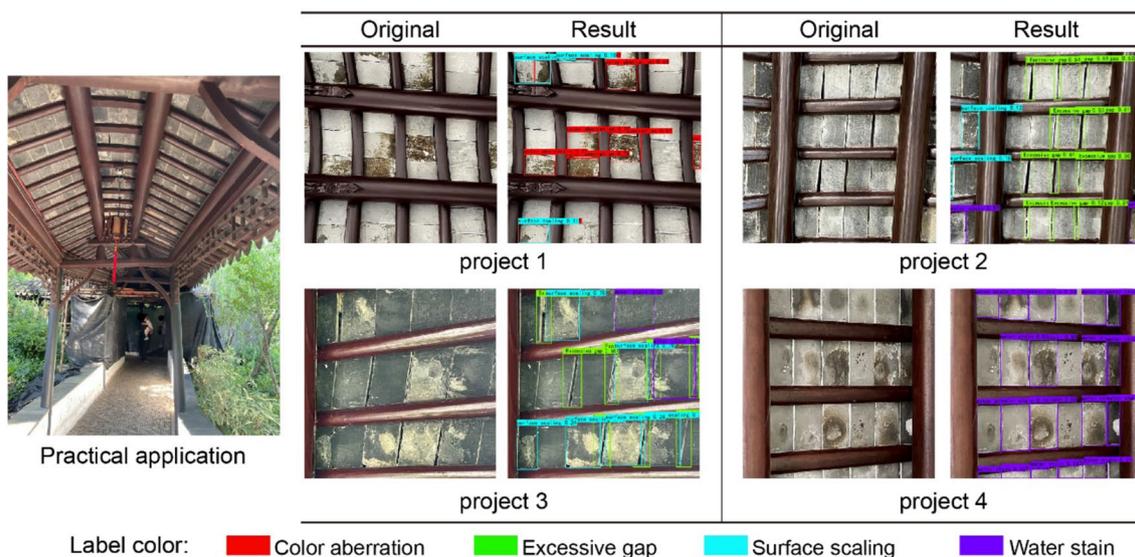
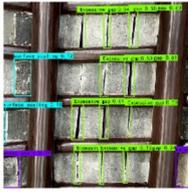
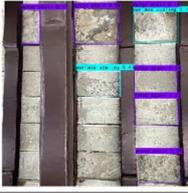
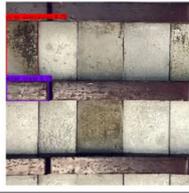


Fig. 14 Detection results for different types of shedthin tile damage. (Drawn by the author)

	Water stain	Surface scaling	Color aberration	Excessive gap	Total
Label quantity	163	164	95	78	500
Accuracy quantity					451
	140	153	83	75	
Accuracy rate	85.89%	93.29%	87.37%	96.15%	90.20%
Quantity of errors					49
	23	11	12	3	
Error rate	14.11%	6.71%	12.63%	3.85%	9.80%

Label color: ■ Water stain ■ Surface scaling ■ Color aberration ■ Excessive gap

Fig. 15 Model accuracy statistics. (Drawn by the author)

which may be related to the difference in visual characteristics between these two types of damage. (3) The model has demonstrated its robustness and reliability in complex environments in application tests in actual scenarios. For example, when detecting outdoor corridor environments, the model can still locate and identify various types of shedthin tile damage relatively accurately, although environmental factors (such as rainfall, moisture, and air pollutants) are complex and changeable. However, the model exhibits a degree of confusion in the identification of certain damage types, especially SS and WS. This implies that in future model optimization, we may need to strengthen the model's ability to distinguish damage types with similar characteristics further. (4) In the heatmap analysis of the model on different detection heads, this study provides an in-depth understanding of the working mechanism of the model in damage detection and classification. For example, the model can accurately capture the location of damage in the "score" layer, perform type judgment in the "class" layer, and output the final detection result through the "class_score" layer. This analysis not only deepens this study's understanding of the working principle of the model but also provides a direction for further optimization of the model in the future.

In summary, although the current model has shown considerable capabilities and potential in the application

of brick damage detection, further improvements in accuracy, robustness, and identification of specific damage types are still needed. Future work will focus more on how to further improve the detection performance of the model through algorithm optimization, data enhancement, and other strategies to play a greater role in the practical application of cultural relic protection and restoration.

Conclusion

In the complex field of cultural heritage protection, the shedthin tiles of Jiangnan Classical Gardens has become a research object that cannot be ignored because of its unique cultural and historical value. Automatic detection and analysis of damage to shedthins tiles, especially when faced with complex and changeable on-site environments, are both technical challenges and practical needs for cultural relic protection. In this context, this article highlights the advantages of computer vision and target detection technology. In particular, the YOLOv4 model automatically detects four damage types: WS, surface scaling, CA, and a gap in shedthin tiles. By collecting and sorting 500 diverse shedthin tile images and performing meticulous data preprocessing and annotation, the researchers trained and optimized the model to adapt to the needs of this specific task.

This study has three main strengths. (1) Improve the efficiency of cultural relic protection: This research can significantly improve the efficiency and accuracy of cultural relic protection by realizing automated and intelligent detection of damage to shedthins tiles and reducing errors and omissions caused by manual inspections. This approach also reduces the workload of manual inspection. (2) The application of computer vision in cultural relic protection should be broadened. Through specific experiments and analysis, this article verifies the feasibility and practicability of computer vision technology in cultural relic protection and provides a reference and inspiration for future research in similar fields. (3) Rich experimental analysis: Through multiangle and multidimensional experimental analysis, this article explores in depth the working mechanism and performance of the model in shedthin tile damage detection. This study provides valuable insights and data support for further optimization and adjustment of the model.

Considering model optimization, algorithm innovation, and the expansion of practical applications, this research can be expanded in the following directions: (1) Model optimization and algorithm innovation: Although existing models have demonstrated certain detection capabilities, there is still room for improvement in accuracy and robustness. Future research can further explore model structures and algorithms that are more suitable for evaluating the damage characteristics of shedthins tiles. (2) Attempts at multimodal learning: Considering the complexity of cultural relic protection, we can attempt to introduce multimodal learning in the future, such as combining multiangle images captured by drones or three-dimensional information captured by laser scanning to improve the detection capabilities of the model. (3) Practical application and system deployment: Future research can also focus more on the deployment and optimization of the model in actual application scenarios. For example, the performance of models under different environments and lighting conditions could be explored, or the models could be embedded into mobile terminals or drone devices to provide real-time technical support for on-site cultural relic protection.

At the intersection of science and culture, applying computer vision technology to cultural relic protection is not only a manifestation of technological innovation but also a modern practice of cultural inheritance. In the future, we will continue to explore new research topics in this field, such as the detection and quantification of damage severity. It is anticipated that the research in this article can provide some reference for related fields and be further promoted and applied in future research and practice.

Appendix

Machine learning environment

Machine learning environment: The operating system is Windows 11 (X64), the CUDA version is 11.5, the deep learning framework is PyTorch (1.13.0), and the graphics card and processor are a GeForce GTX 3070 (16 G) and an AMD Ryzen 9 5900HX (3.30 GHz), respectively.

Author contributions

Conceptualization, LY and LZ; methodology, LZ; software, LY and LZ; validation, LY and LZ; formal analysis, LY; investigation, YZ; resources, YZ and XL; data curation, YZ; writing—original draft preparation, YC and LY; writing—review and editing, LY and LZ; visualization, LY, YC, LZ and YZ. All authors have read and agreed to the published version of the manuscript.

Funding

This research received funding by Guangdong Provincial Department of Education's key scientific research platforms and projects for general universities in 2023: Guangdong, Hong Kong, and Macau Cultural Heritage Protection and Innovation Design Team (Funding Project Number: 2023WCXTD042).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable for studies not involving humans or animals.

Competing interests

The authors declare no competing interests.

Received: 23 December 2023 Accepted: 17 February 2024

Published online: 29 February 2024

References

1. UNESCO. Classical Gardens of Suzhou. 2023. <https://whc.unesco.org/en/list/813/>. Accessed 13 Sept 2023.
2. Henderson R. The gardens of Suzhou. Philadelphia: University of Pennsylvania Press; 2012.
3. Xu Y. The Chinese city in space and time: the development of urban form in Suzhou. Honolulu: University of Hawaii Press; 2000.
4. Yu L, Kim C, Kim H. Exploring visitors' subjectivities toward authenticity of the Suzhou UNESCO heritage site using Q-method. *J Tour Hospit*. 2021;10:460.
5. Jiang J, Zhou T, Han Y, Ikebe K. Urban heritage conservation and modern urban development from the perspective of the historic urban landscape approach: a case study of Suzhou. *Land*. 2022;11(8):1251. <https://doi.org/10.3390/land11081251>.
6. Pannu A. Artificial intelligence and its application in different areas. *Artif Intell*. 2015;4(10):79–84.
7. Das S, Dey A, Pal A, Roy N. Applications of artificial intelligence in machine learning: review and prospect. *Int J Comput Appl*. 2015;115(9):31.
8. Kumar K, Thakur GSM. Advanced applications of neural networks and artificial intelligence: a review. *Int J Inform Technol Comput Sci*. 2012;4(6):57.
9. Li Q, Zheng L, Chen Y, Yan L, Li Y, Zhao J. Nondestructive testing research on the surface damage faced by the Shanhaiguan Great Wall based on

- machine learning. *Front Earth Sci.* 2023;11:1225585. <https://doi.org/10.3389/feart.2023.1225585>.
10. Chen CH, editor. *Handbook of pattern recognition and computer vision*. Singapore: World scientific; 2015.
 11. Wright J, Ma Y, Mairal J, Sapiro G, Huang TS, Yan S. Sparse representation for computer vision and pattern recognition. *Proc IEEE.* 2010;98(6):1031–44. <https://doi.org/10.1109/JPROC.2010.2044470>.
 12. Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: a survey. *Proc IEEE.* 2023. <https://doi.org/10.1109/JPROC.2023.3238524>.
 13. Zou X. A review of object detection techniques. In 2019 International conference on smart grid and electrical automation (ICSGEA), IEEE. 2019, August; 251–254. <https://doi.org/10.1109/ICSGEA.2019.00065>.
 14. Gupta AK, Seal A, Prasad M, Khanna P. Salient object detection techniques in computer vision—a survey. *Entropy.* 2020;22(10):1174. <https://doi.org/10.3390/e22101174>.
 15. Pathak AR, Pandey M, Rautaray S. Application of deep learning for object detection. *Proc Comput Sci.* 2018;132:1706–17. <https://doi.org/10.1016/j.procs.2018.05.144>.
 16. Cazzato D, Cimarelli C, Sanchez-Lopez JL, Voos H, Leo M. A survey of computer vision methods for 2d object detection from unmanned aerial vehicles. *J Imaging.* 2020;6(8):78. <https://doi.org/10.3390/jimaging6080078>.
 17. Prasad DK, Prasath CK, Rajan D, Rachmawati L, Rajabaly E, Quek C. Challenges in video based object detection in maritime scenario using computer vision. arXiv preprint. 2016. [arXiv:1608.01079](https://arxiv.org/abs/1608.01079). <https://doi.org/10.48550/arXiv.1608.01079>
 18. Parker JR. *Algorithms for image processing and computer vision*. Hoboken: John Wiley & Sons; 2010.
 19. Latif A, Rasheed A, Sajid U, Ahmed J, Ali N, Ratyal NI, Khalil T. Content-based image retrieval and feature extraction: a comprehensive review. *Math Probl Eng.* 2019. <https://doi.org/10.1155/2019/9658350>.
 20. Kodituwakku SR, Selvarajah S. Comparison of color features for image retrieval. *Indian J Comput Sci Eng.* 2004;1(3):207–11.
 21. Smeulders AW, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. *IEEE Trans Pattern Anal Mach Intell.* 2000;22(12):1349–80. <https://doi.org/10.1109/34.895972>.
 22. Tieu K, Viola P. Boosting image retrieval. *Int J Comput Vision.* 2004;56:17–36. <https://doi.org/10.1023/B:VISI.0000004830.93820.78>.
 23. Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *Int J Med Informatics.* 2004;73(1):1–23. <https://doi.org/10.1016/j.ijmedinf.2003.11.024>.
 24. Rui Y, Huang TS, Chang SF. Image retrieval: past, present, and future. *J Vis Commun Image Represent.* 1999;10(1):1–23.
 25. Qiao L, Li Y, Chen D, Serikawa S, Guizani M, Lv Z. A survey on 5G/6G, AI, and robotics. *Comput Electr Eng.* 2021;95:107372. <https://doi.org/10.1016/j.compeleceng.2021.107372>.
 26. Lafarge F, Keriven R, Brédif M. Insertion of 3-D-primitives in mesh-based representations: towards compact models preserving the details. *IEEE Trans Image Process.* 2010;19(7):1683–94. <https://doi.org/10.1109/TIP.2010.2045695>.
 27. Oses N, Dornaika F, Moujahid A. Image-based delineation and classification of built heritage masonry. *Remote Sensing.* 2014;6(3):1863–89. <https://doi.org/10.3390/rs6031863>.
 28. Llamas J, Leronés PM, Medina R, Zalama E, Gómez-García-Bermejo J. Classification of architectural heritage images using deep learning techniques. *Appl Sci.* 2017;7(10):992. <https://doi.org/10.3390/app7100992>.
 29. Monna F, Rolland T, Denaire A, Navarro N, Granjon L, Barbé R, Chateau-Smith C. Deep learning to detect built cultural heritage from satellite imagery: Spatial distribution and size of vernacular houses in Sumba, Indonesia. *J Cult Heritage.* 2021;52:171–83. <https://doi.org/10.1016/j.culher.2021.10.004>.
 30. Akinosho TD, Oyedele LO, Bilal M, Ajayi AO, Delgado MD, Akinade OO, Ahmed AA. Deep learning in the construction industry: a review of present status and future innovations. *J Build Eng.* 2020;32:101827. <https://doi.org/10.1016/j.jobe.2020.101827>.
 31. Pan Y, Zhang L. Roles of artificial intelligence in construction engineering and management: a critical review and future trends. *Autom Constr.* 2021;122:103517. <https://doi.org/10.1016/j.autcon.2020.103517>.
 32. Flah M, Nunez I, Ben Chaabene W, Nehdi ML. Machine learning algorithms in civil structural health monitoring: a systematic review. *Arch Comput Methods Eng.* 2021;28:2621–43. <https://doi.org/10.1007/s11831-020-09471-9>.
 33. Zheng L, Chen Y, Yan L, Zhang Y. Automatic detection and recognition method of Chinese clay tiles based on YOLOv4: a case study in Macau. *Int J Archit Heritage.* 2023. <https://doi.org/10.1080/15583058.2023.2246029>.
 34. Yakovlev A, Lisovychenko O. An approach for image annotation automatization for artificial intelligence models learning (Підхід до автоматизації анування зображень для навчання моделей штучного інтелекту). *Адаптивні системи автоматичного управління.* 2020; 1(36): 32–40. <https://doi.org/10.20535/1560-8956.36.2020.209755>

Publisher's Note

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