

RESEARCH

Open Access



Using mask R-CNN to rapidly detect the gold foil shedding of stone cultural heritage in images

Miaole Hou^{1,3}, Dongxu Huo^{1,3}, Yue Yang^{1,3,4*}, Su Yang^{2*} and Huiwen Chen^{1,3}

Abstract

As immovable stone cultural heritage is kept in the open air, they are more susceptible to damage, and damage detection is very important for the protection and restoration of cultural heritage. This is especially true for gold-overlaid stone cultural heritage, which is usually more complicated than ordinary stone carvings. However, the detection of cultural heritage damages is mainly based on expert visual inspection, which is often subjective, time-consuming, and laborious. This paper uses the Mask R-CNN algorithm to rapidly and accurately detect the gold foil shedding of stone cultural heritage through two-dimensional images. The research data are from the high-precision images of the Dazu Thousand-Hand Bodhisattva Statue (World Heritage, UNESCO) in Chongqing, China. After cleaning and augmentation, 1900 images are input into Mask R-CNN model for training. Finally, the average precision value (AP) for detecting gold foil shedding is found to be 0.967. In order to test the performance of the model, the new images that do not participate in the training period are used, and it is found that the model can still accurately detect the gold foil shedding even if there are interference factors. This is the first attempt to detect the damages of gold-overlaid stone cultural heritage based on a deep learning algorithm, and it has achieved good results.

Keywords Dazu Thousand-Hand Bodhisattva Statue, Mask R-CNN, Damage detection, Deep learning, Gold foil shedding

Introduction

Compared with movable cultural heritage, immovable stone cultural heritage (such as ancient sites, ancient tombs, cave temples, and stone carvings) is directly exposed to the natural environment in the open air,

which is easily affected by local geological, meteorological, and hydrological conditions [1]. After long-term rainfall erosion, sudden changes in temperature and humidity, wind erosion, sand erosion, and so on, it is very easy to cause complex and widely distributed damages to stone cultural heritage such as dissolution, fracture, surface exfoliation, mechanical damage, weathering, biological colonization and so on [2–5]. In the process of investigating damages to cultural heritage, it is necessary to draw a thematic map of the damaged area to understand the spatial distribution and development trend of the damages, and then formulate a scientific and effective protection plan. The traditional method of damage detection of stone cultural heritage is that archeologists are responsible for the visual detection of various damages. Marking the disease information of cultural relics on the

*Correspondence:

Yue Yang

yoursefiam0418@hotmail.com

Su Yang

yangsu@student.cumtb.edu.cn

¹ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

² College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China

³ Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring, Beijing 100044, China

⁴ China Coal Xi'an Design Engineering Co, Ltd, Xi'an 710054, Shaanxi, 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.

image is not only a heavy workload but also requires high professional quality of on-site personnel. This is also a time-consuming and laborious task [6–8].

Considering that machine learning technology has strong feature extraction ability and can construct complex high-level features from low-level features [9–11], it provides a more efficient and accurate solution for damage detection of cultural heritage. In recent years, to find a new method to replace the traditional damage detection of cultural heritage, scholars have carried out much research on damage detection based on machine learning [20–29]. In this process, researchers have chosen different types of cultural heritage as targets for detection, such as historical buildings [23], stone monuments [29], historical glazed tiles [24, 26], wood paintings [28], and so on. To improve the accuracy of detection, they used a variety of improved machine learning algorithms, and a convolution neural network (CNN) was widely used [24–29]. Although research on the automatic detection of stone cultural heritage by computer vision has existed, especially deep learning methods, but research of gold foil shedding disease has not yet. Due to the bright color and high reflectivity of the gold foil surface, there are obvious vision disparity between the gold foil surface and the stone surface in the image. Therefore, this study focuses on detecting gold foil-related diseases in stone cultural relics covered with gold foil (specifically in the category of stone artifacts depicting Dazu Thousand-Hand Bodhisattva Statue). It demonstrates the effectiveness of the Mask-RCNN network model for such special objects, which is an important supplement to previous relevant studies.

The main contributions of this paper are as follows:

- (1) It is proved that the Mask RCNN algorithm is successful in the detection of stone relics gold disease. And the main types of detection include gold foil shedding, gold foil warping, and gold foil dust disease. This method reduces the cost of material resources and time investment when targeting large-scale detection. If the research object is captured, specific damages can be accurately detected.
- (2) There are more kinds of stone cultural relics with gold, and the visual effect is more chaotic. Mask-RCNN still has strong robustness when the damage is complex or there are interference factors.

The rest of this paper is organized as follows. Sect. “[Background and related works](#)” explains in detail the case study of this paper (i.e., Dazu Thousand-Hand Bodhisattva Statue in Chongqing, China, and the damages involved in the paper) and summarizes the related

works of damage detection of cultural heritage based on machine learning. Sect. “[Methodology](#)” introduces the workflow of this paper and explains the principle of the Mask R-CNN algorithm. Sect. “[Results and discussion](#)” presents the results of using the trained model to detect the gold foil shedding, and discusses them from the aspects of different directions, different brightness, and different distances. Sect. “[Conclusion](#)” presents the deficiencies and future works in the paper.

Background and related works

Dazu Thousand-Hand Bodhisattva Statue

The Dazu Thousand-Hand Bodhisattva Statue, which is located in the Great Mercy Storied Pavilion on the south side of Mount Baoding, Dazu District, Chongqing of China, is selected as the case study (Fig. 1). The statue is carved in the Southern Song Dynasty and has a history of more than 800 years. It is not only the largest statue carved on the cliff with stone material, color painting, and gold foil on the statue, but also known as a world cultural heritage which has high historical and artistic value [12, 13].

The Dazu Thousand-Hand Bodhisattva Statue is 7.2 m high and 12.5 m wide. It has 830 hands and 227 dharma instruments (objects with certain meanings held in the hands of Buddha statues, such as pagodas, swords and statute books). However, the statue has suffered considerable damage due to the high temperature and humidity prevalent in the Dazu district. Over time, both natural elements and human activities have significantly impacted the gold foil, color painting, and stone material of the statue, leading to a marked deterioration in its overall integrity. [14, 15].

In the protection project of the Dazu Thousand-Hand Bodhisattva Statue, in order to facilitate the restoration and research progress, the researchers divided the statue into 99 regions (i.e., the statue is divided into 9 rows horizontally and 11 columns vertically) (Fig. 2). On this basis, they also numbered all the arms and artifacts of the statue according to the area in which they were located, such as the code number “3-2-S5”, representing the fifth hands in the area of 3-2. In this paper, the hands with gold foil shedding are selected for the research [24].

The characteristics of gold foil shedding and other similar deterioration

Under the influence of geological, hydrological, atmospheric, and other factors, various damages such as weathering, flaking, and fracture may occur in the stone carvings. In order to highlight their artistic value and religious significance, some carvings are coated with gold foil on the surface. In order to protect the stone carvings, people even coat many layers of gold foil. The gold foil

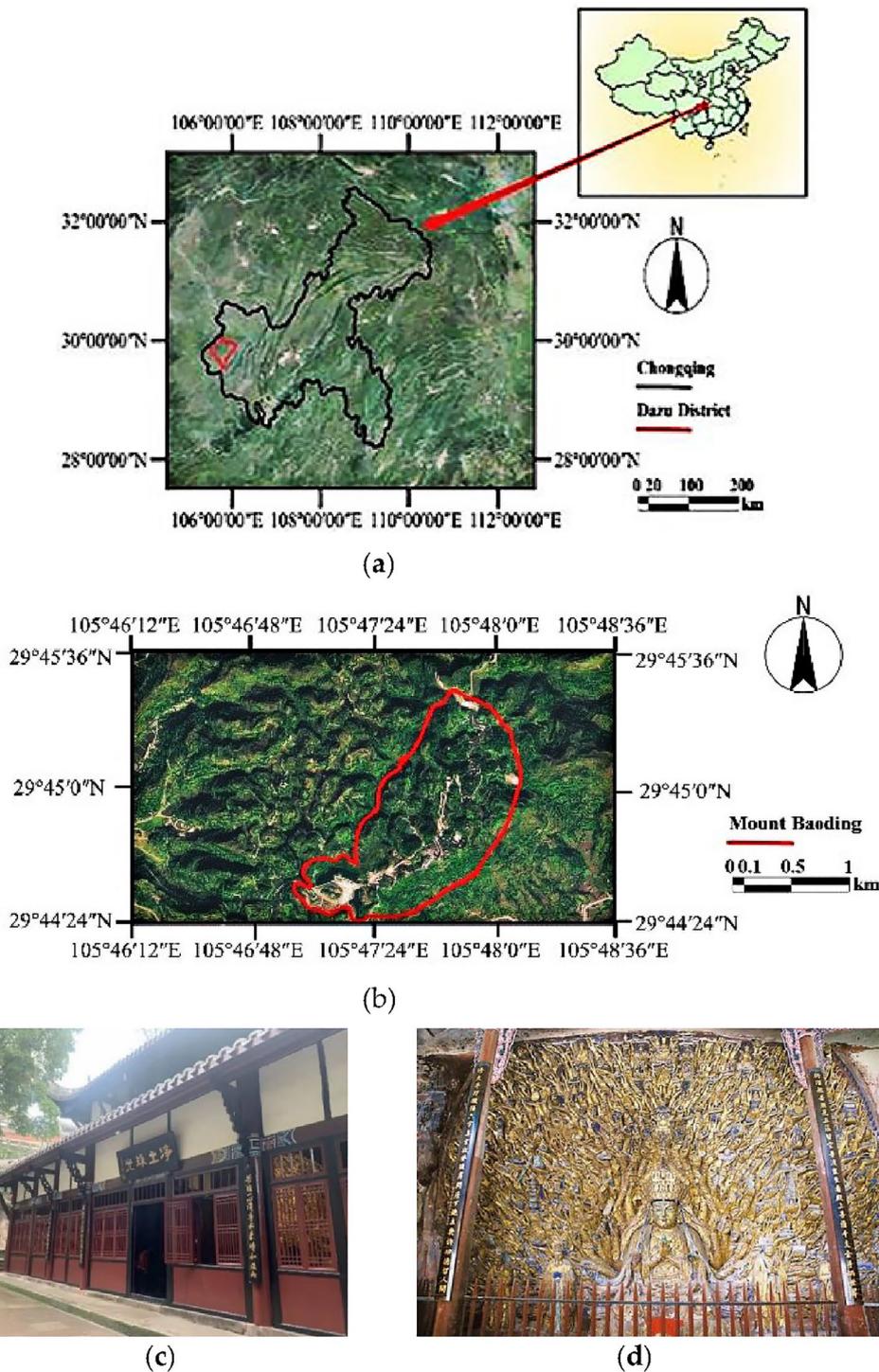
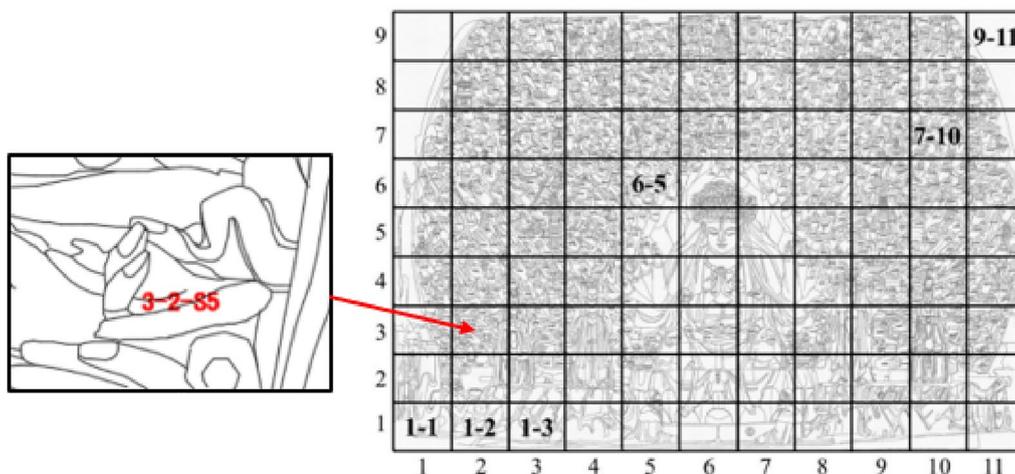


Fig. 1 **a** The location of Chongqing and Dazu District; **b** the location of the Mount Baoding; **c** the external situation of the Great Mercy Storied-Pavilion; **d** the Dazu Thousand-Hand Bodhisattva Statue before restoration

has a strong protective effect on stone material, but complex damages such as shedding, cracking, warping, and biological colonization may occur on it (Fig. 3). Among

these damages, gold foil shedding is common damage, which is widely distributed in the gold part of stone



(a)



(b)



(c)

Fig. 2 a The line drawing of the Dazhu Thousand-Hand Bodhisattva Statue; b–c hands of code number “3–2–S5”

carvings. The deterioration of gold foil not only makes the stone material unprotected but also damages the value of stone carvings seriously.

To ensure the deep learning network can precisely identify gold foil shedding without being influenced by extraneous factors, it’s crucial for the model to comprehensively understand the intricate characteristics of gold foil shedding in various conditions, as well as other damages that bear a resemblance to it. This section introduces and discusses various types of damage, including gold foil shedding, gold foil dust accumulation, gold foil warping, stone incompleteness, and the shedding of gold foil plaster. The damages cited in this paper all come from the investigation and definition of the protection project of the Dazhu Thousand-Hand Bodhisattva Statue [16]. In addition, some of the arms of the statue have Tian Yan (“The Eye of Heaven”) in the palm, while others hold dharma instruments. As the Tian Yan and dharma instruments

also have similar characteristics with the gold foil shedding, they are also introduced here.

- *Gold foil shedding* The gold foil on the surface falls off from the bedrock or plasters. The reason is that the composition of gold foil changes in the process of historical development and the gold foil falls off under the action of gravity or stress. Figure 4(a)–(c) is a schematic diagram of gold foil shedding. Among them, the characteristic of 4(a) is more obvious. After the gold foil falls off from the bedrock, the stone material is exposed, which is obviously different from the surrounding gold foil. In Fig. 4(b), the surrounding gold foil is covered with a lot of dust, which weakens its gloss and turns its color to gray, similar to the stone material exposed after the gold foil falls off, which needs to be carefully observed. In Fig. 4(c), the gold foil falls off from the plasters, and the color of

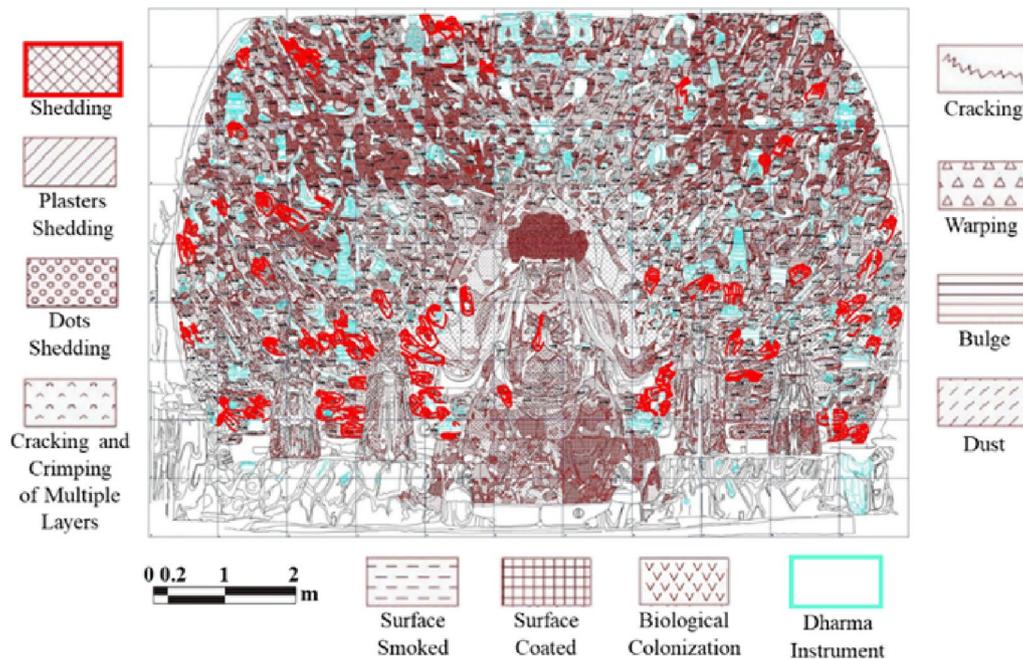


Fig. 3 Types and distribution range of gold foil damages

the surrounding gold foil changes due to rust, which is similar to the dark red plaster exposed after the gold foil falloff, which needs to be carefully observed. In the gold-overlaid stone cultural heritage, gold foil shedding is common damage, and its characteristics are variable.

- *Gold foil dust* A thick layer of dust attached to the surface of gold foil. The dust contains incense ash, paper ash, and other organic matter, and its color is dark gray or black. The existence of dust not only seriously deteriorates the value of the stone carvings, but also aggravates the weathering damage of gold foil, as shown in Fig. 4 (d).
- *Gold foil warping* In order to stick gold foil on the surface, people often brush a layer of oil before sticking it. When the surface tension of gold foil is greater than the bonding force of oil, the gold foil will crack and warp. The difference between gold foil warping and shedding is that the gold foil of the former has not completely fallen off, as shown in Fig. 4 (e).
- *Stone incompleteness* For the stone cultural heritage with a structural plane, the plane provides a channel for weathering erosion of the internal stone material, and under the action of gravity, the stone cracks, and collapses along the position of high stress or weak structure. As shown in Fig. 4(f).
- *Gold foil plasters shedding* For the gold-overlaid stone cultural heritage, the surface of the stone is covered

with multiple layers of gold foil, and there are plasters among the gold foil. As time goes on, the mineral composition of the plasters might change, and when the gold foil falls off, the plasters may also fall off under the action of their own gravity and stress, as shown in Fig. 4(g).

- *Tian Yan (“The Eye of Heaven”)* In order to show the religious significance of the Thousand-Hand Bodhisattva Statue, some arms have a “Tian Yan” in the palm of the hand, that is, an eye is engraved in the palm. Unlike the other parts of the hand, the Tian Yan is not covered with gold foil, as shown in Fig. 4(h).
- *Dharma instruments* In order to show the religious significance of the Thousand-Hand Bodhisattva Statue, some of the arms hold different shapes of dharma instruments, which are covered with color painting. If the color painting falls off the instruments, the stone material will be exposed, as shown in Fig. 4(i).

Related works

In recent years, artificial intelligence (AI) technology has been developed rapidly. Because of its high efficiency and high precision in prediction and recognition [17–19], it has been widely used in the detection of cultural heritage

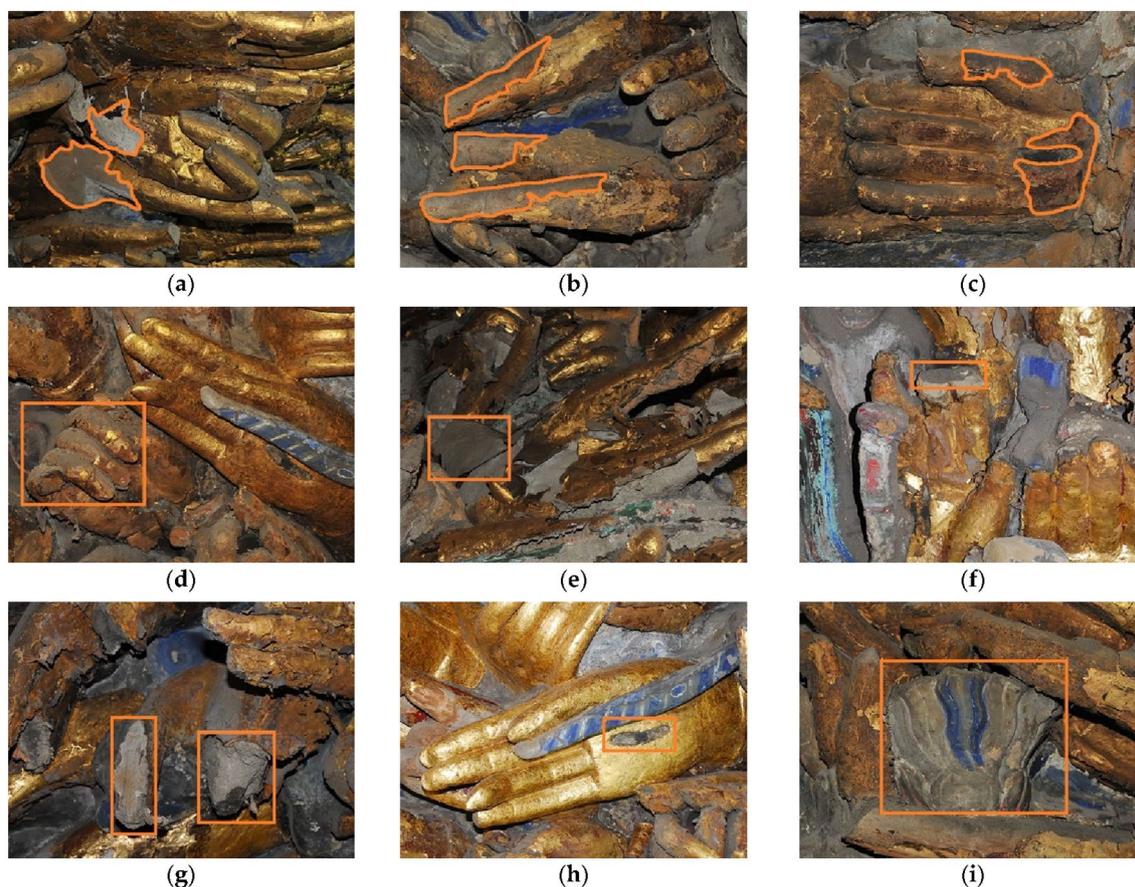


Fig. 4 The gold foil shedding and other similar deterioration. **a–c** gold foil shedding, **d** gold foil dust, **e** gold foil warping, **f** stone incompleteness, **g** gold foil plasters shedding, **h** Tian Yan, **i** dharma instruments

Table 1 Research on using deep learning algorithms to detect different cultural heritage damages

Researchers	Algorithm	Research object	Specific damages
Chaiyasarn et al. [24]	CNN, SVM	Historical masonry structures of a stupa in Wat Chai Watthanaram, Thailand	Crack
Kwon et al. [25]	Faster R-CNN	Outdoor stone cultural properties conducted by the Cultural Heritage Administration of S. Korea	Crack, loss, detachment, biological colonization
Wang et al. [26]	Faster R-CNN, Mask R-CNN	Historic glazed tiles of the Palace Museum in China	Surface damage
Wang et al. [34]	Faster R-CNN	Historic brick masonry structures of the Palace Museum	Efflorescence, spalling
Zou et al. [27]	Faster R-CNN	Components of the Forbidden City	The missing parts (Goutou, Dishui, Dingmao)
ANGHELUȚĂ et al. [36]	VGG-16 specialized convolutional network	A wood painting representing St. Constantine and Helen	Crack, blister, detachment
Hatir et al. [29]	Mask R-CNN	Yazılıkaya monuments in the Hattusa archeological site	Biological colonization, contour scaling, crack, higher plant, impact damage, microkarst, missing part
Adamopoulos et al. [23]	Supervised segmentation methods based on random decision trees, ensemble learning, and regression algorithm	A fortification in Euboea, Greece	Vegetation, moss, black crusts, lichens, missing material, dampness

damages. As shown in Table 1, most of the algorithms used in research are convolution neural networks (CNN), and the research objects are representatives of different types of cultural heritage. Various types of damages were detected. In the process of applying AI technology, some researchers used machine-learning algorithms to detect cultural heritage damages [20–23]. Nazarian et al. [20] used Support Vector Machine (SVM), Neural Network (NN), and Gaussian Naive Bayes (GNB) to train the damage assessment model and associated the stiffness of various structural components of the system with the location of the damage. It was used to evaluate the structural states of historic buildings after extreme events. Valero et al. [21] proposed an innovative strategy for the automatic detection and classification of defects in ashlar masonry walls. This classification method was based on a logistic regression multi-class classification algorithm and supplemented by the surveyor's professional knowledge to detect a single defective masonry wall. Meng et al. [22] proposed a non-destructive testing method for stone heritage based on Terahertz (THz) technology and a Support Vector Machine (SVM) machine-learning model, which was used to detect and evaluate the hollowing deterioration in the Yungang Grottoes. Adamopoulos et al. [23] took a weathered historical fortification as a case study, using a supervised segmentation method based on random decision trees, ensemble learning, and regression algorithm to automatically segment the deterioration patterns of multispectral image composites.

Other researchers used deep learning algorithms for damage detection of cultural heritage [24–29]. Chaiyasarn et al. [24] proposed an image-based crack detection system combining deep Convolution Neural Network (CNN) and Support Vector Machine (SVM) for crack detection of masonry structures in historical sites. Kwon et al. [25] used Faster R-CNN to automatically detect and classify the damages (i.e., crack, loss, detachment, and biological colonization) of outdoor stone cultural heritage. To improve the performance of damage detection, the image data set of stone heritage damage collected by the Cultural Heritage Administration of South Korea was constructed and expanded. Wang et al. [26] proposed a two-level object detection, segmentation, and measurement strategy for historical glazed tiles based on deep learning technology. The first stage of the model used Faster R-CNN to automatically detect and cut the roof images into single-glazed tile images, and the second stage used Mask R-CNN to train the cropped images, which could finally realize the damage segmentation and measurement of historical glazed tiles. Zou et al. [27] used the Faster R-CNN algorithm to detect and count the damage status of historical building components (Goutou, Dishui, and Dingmao) and then judged

the regularity of their position arrangement by marking the position of the damaged components. Angheluta et al. [28] used the VGG-16 specialized convolutional network developed at the Visual Geometry Group from Oxford University to detect cracks, blister, and detachment of wood paintings, and finally realized the detection and evaluation of damages with high visual complexity. Hatir et al. [29] used the Mask R-CNN algorithm to automatically detect nine kinds of damages of large stone heritage (biological colonization, contour scaling, crack, higher plant, impact damage, microcars, and missing part). Although there were many types of damage to be detected in the research, the recognition accuracy of the model was very high, which indicated that the object recognition algorithm could successfully detect more than one type of cultural heritage damage.

Methodology

Proposed approach

As mentioned earlier, the purpose of this paper is to use the Mask R-CNN algorithm to detect the gold foil shedding of stone cultural heritage through two-dimensional images, and the overall workflow is shown in Fig. 5. In the data preparation phase, a NIKON D300 SLR digital camera with a resolution of 12.3 million pixels was used to capture local high-definition images of stone heritage. In the process of data acquisition, due to the limited quantities of cultural heritage damages, 190 images with a resolution of 300dpi were obtained, including 131 images of 4288×2848 and 59 images of 2592×3872 . The captured images contain gold foil shedding taken from different distances and different directions, which ensures that in the process of model training, the misjudgment caused by differences in distance or direction can be avoided as much as possible.

This study uses the Mask R-CNN algorithm not only to detect the presence of gold foil shedding, but also to accurately mark the boundaries of the shedding area at the pixel level, providing more detailed spatial information. Secondly, gold foil shedding usually manifests as low contrast, texture changes, and other visual features. Mask R-CNN integrates multimodal information such as color, texture, shape, etc., which can more comprehensively capture the features of gold foil shedding and improve detection accuracy. And this detection method has better applicability and is convenient for subsequent research [29].

In order to reduce the time required for model training, all images are divided into two types, of which 1310 images of 800×1310 and 590 images of 536×800 . In addition, we augmented the sample size of the existing dataset, which can improve the quality of the dataset and enhance the generalization and robustness of the

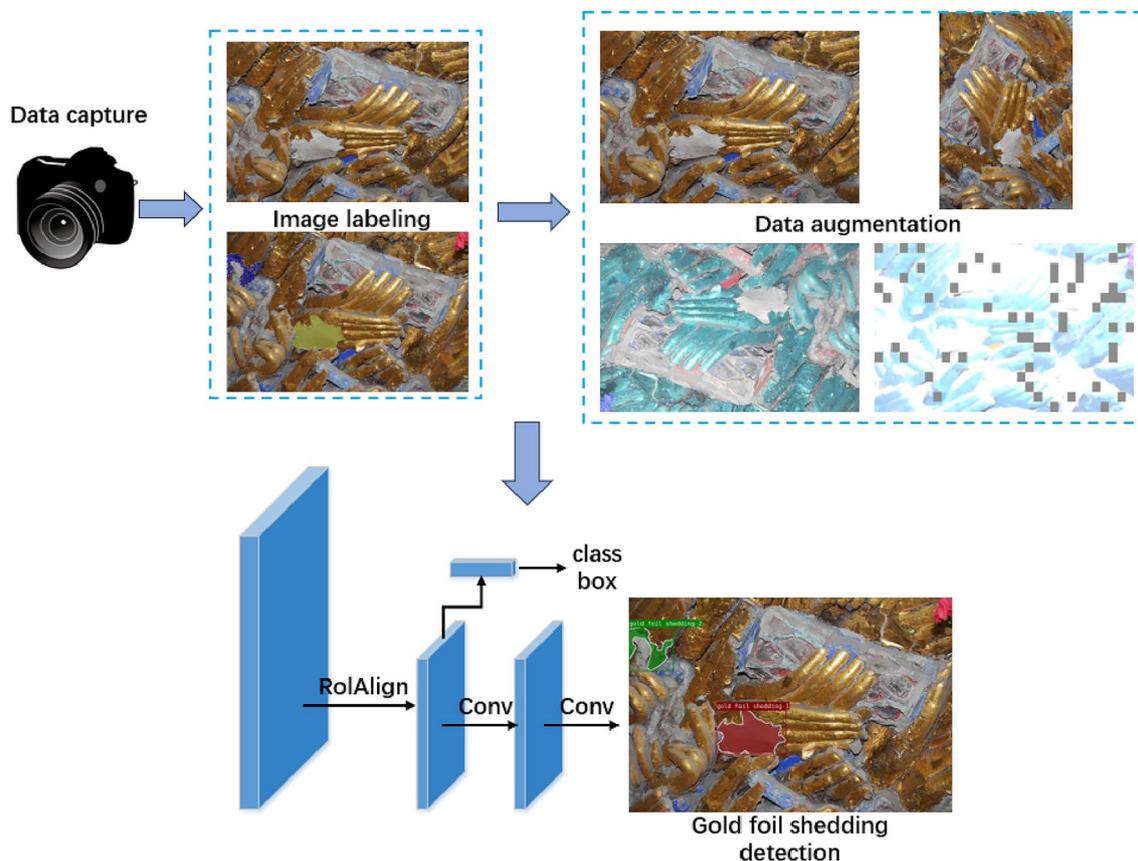


Fig. 5 The flowchart of the proposed method

model. In the image-labeling phase, we used LabelMe as open-source software to manually label all the gold foil shedding of 190 samples and generate the corresponding mask files. After the labeling was completed, all the labeled images were converted into JSON files. In the process of data augmentation, we used mirroring, rotation, random clipping, adjusting contrast, adding Gaussian noises, and other methods, and these methods can also be randomly combined.

By performing operations such as rotation and adding noise to the dataset, data augmentation was conducted, resulting in a final dataset size of 1900 images. In the dataset with 1900 images, 80% (1520) of images are used in the training and validation datasets, while the remaining 20% (380) images are used in the test set. The training process is carried out on a workstation using the GPU mode (CPU: 11th Gen Intel Core i7-11,700 @ 2.50 GHz, RAM: 16.0 GB, GPU: NVIDIA GeForce RTX 3060 Ti). The code is programmed using Python 3.7, the virtual environment is established by TensorFlow 2.5.0 and Keras 2.4.3, and the whole training process is carried out on Windows 11×64 system.

Mask R-CNN algorithm

In this paper, the Mask R-CNN algorithm is used for training. Mask R-CNN is an end-to-end, pixel-to-pixel convolution network applied to instance segmentation. As an extension of Faster R-CNN, it adds branches to evaluate the segmentation mask on each region of interest (ROI) [30]. Mask R-CNN can effectively detect objects in the image and generate a high-quality segmentation mask for each instance.

The architecture of the Mask R-CNN algorithm is shown in Fig. 6. First, the image is input into the backbone network. Backbone is a series of convolution layers used to extract the feature maps of the image. The common backbone includes VGG16, VGG19, ResNet50, ResNet101, and so on. The first layer of the backbone network is used to sense low-level features such as edges, while the later layers can sense more complex features (such as the characteristic or type of damage). In addition, neural networks with different depths can also be designed in the backbone network to achieve high-precision recognition results, but high-precision and high-speed processing cannot be obtained at the same time. In order to achieve the balance between precision and

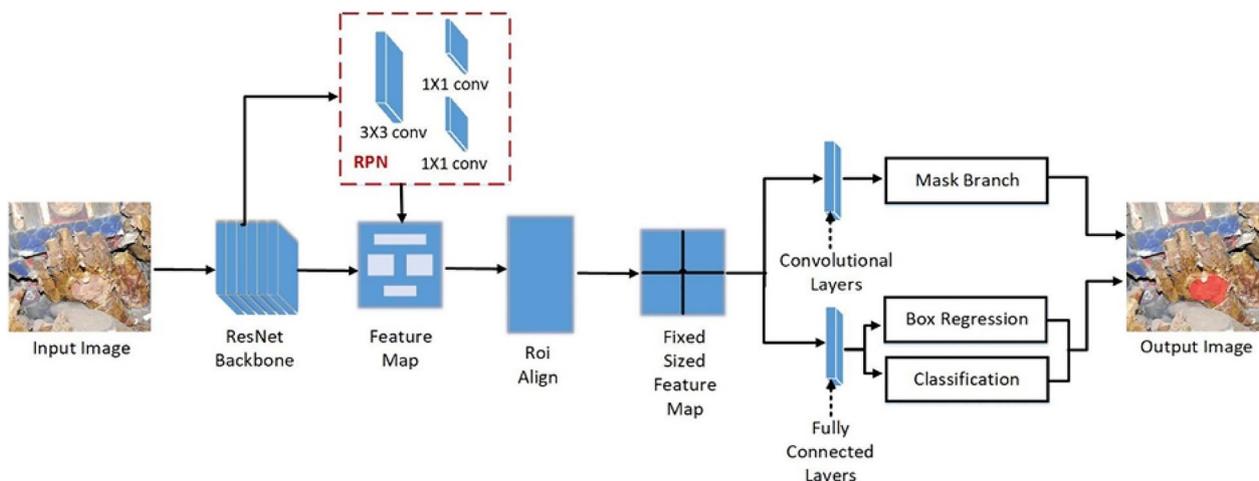


Fig. 6 The architecture of the Mask R-CNN algorithm

Table 2 The training parameters of Mask R-CNN algorithm

Pre-trained model	Batch size	Training epochs	Learning rate	Momentum	Weight decay
ResNet-101	3	40	0.001	0.9	0.0001

speed, this paper uses the ResNet101 model to achieve high-precision results in reasonable training cycles [31].

The feature maps processed by ResNet101 are then sent to the Region Proposal Network (RPN). The RPN seamlessly integrates the proposal generation process into the neural network, enabling a more efficient and accurate detection model. This integration is what truly embeds object detection into the neural network. The RPN can eliminate the ratio differences of the research object caused by different directions and distances in the image. Then the ROI align operation is

implemented on this basis. ROI alignment is proposed with a slight change of ROI Pooling in Faster R-CNN, which solves the problem of aligning the mask with the object so that all candidate targets have the same dimension. Finally, the samples are sent to the fully connected layers to determine the categories and coordinates of the detected objects, in which several boxes are created, and the one with the highest estimate is selected as the bounding box. In the last stage, the groups, which is suggested for the object in the mask branch, will be performed by mask application.

To improve the convergence speed and model performance, we considered both hardware and software conditions of the workstation in selecting the training parameters, as shown in Table 2. Among them, the mask files of the training and validation set need to be read during the training period. The process of generating the mask file is shown in Fig. 7. First, it is necessary to manually delineate the boundaries of all the gold foil

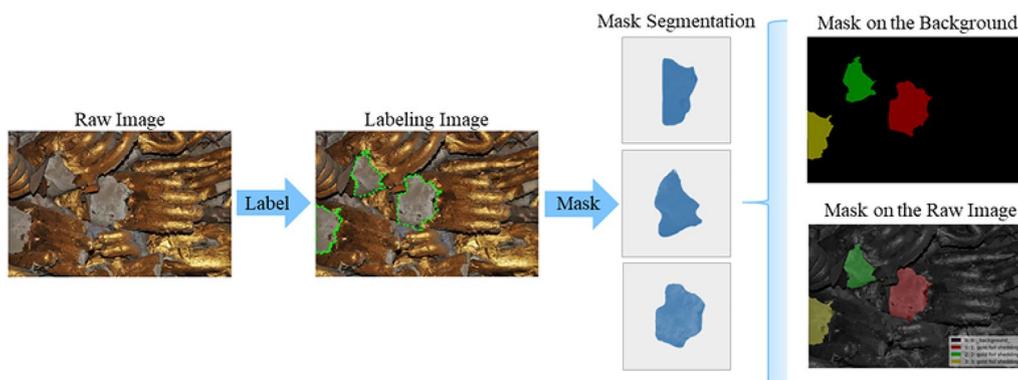


Fig. 7 The process of generating mask files

detachments in the original image. Second, all the labeled regions are segmented, and a separate mask information is generated for each region, which is conducive to the subsequent detection and segmentation of the gold foil shedding. Finally, the mask coordinates and label names corresponding to the labeled regions are converted into 8-bit mask files.

ResNet101 model

During the research, in order to use the ResNet101 pre-trained model for transfer learning, it is necessary to load the weight of ResNet101 from the base layer. The detailed training process includes two steps: (I) Add the parameter of layers="heads" to the training method. To prevent the extraction ability of the base layer from being destroyed, only the network head is trained. We freeze all the backbone layers and only train the random initialized layers. In this process, the network head is trained in 20 epochs. (II) After the randomly initialized layers are trained, the parameter of layers="all" is used to fine-tune all the layers, which can make the algorithm more adaptable to the dataset that needs training.

Results and discussion

This process takes 40 epochs for training. Figure 8 shows the loss curves for training and validation, where total loss = bbox loss + class loss + mask loss.

As shown in Fig. 8, the black curves represent the total loss, and the curves of other colors represent bbox loss, class loss, and mask loss, respectively. Figure 8(a) represents the loss curve of the training process, while Fig. 8(b) represents the loss curve of the validation process. Figure 8(a) shows that the black curve decreases

rapidly from 2.1 to stabilize at 25 epochs, and finally, the total loss converges to about 0.3. Figure 8(b) shows that the black curve decreases rapidly from 1.4 to stabilize at 25 epochs, and finally, the total loss converges to about 0.3. The curves fluctuate greatly in the process of decline and tend to stabilize at 25 epochs. By comparison, it is found that the validation loss curves in Fig. 8(b) are close to the training loss curves in Fig. 8(a), which indicates that the parameters of Mask R-CNN are not overfitting during the training process.

After the model is trained, the AP (Average Precision) is used to evaluate the detection effect of the target detector for gold foil shedding. The value of the AP is the area under the Precision/Recall curve [32, 33]. Figure 9 shows the AP value and PR curve of the gold foil shedding. The

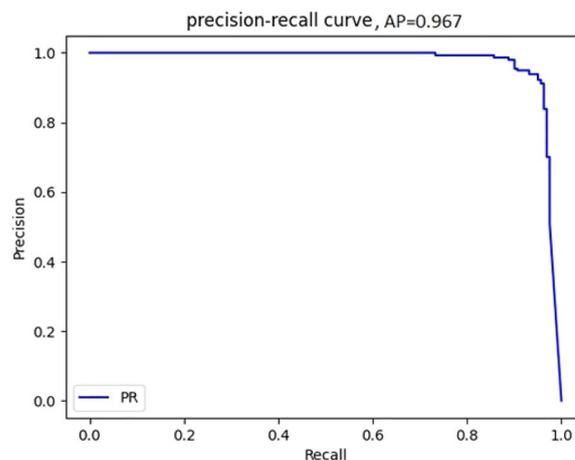


Fig. 9 Precision/Recall curve for the gold foil shedding

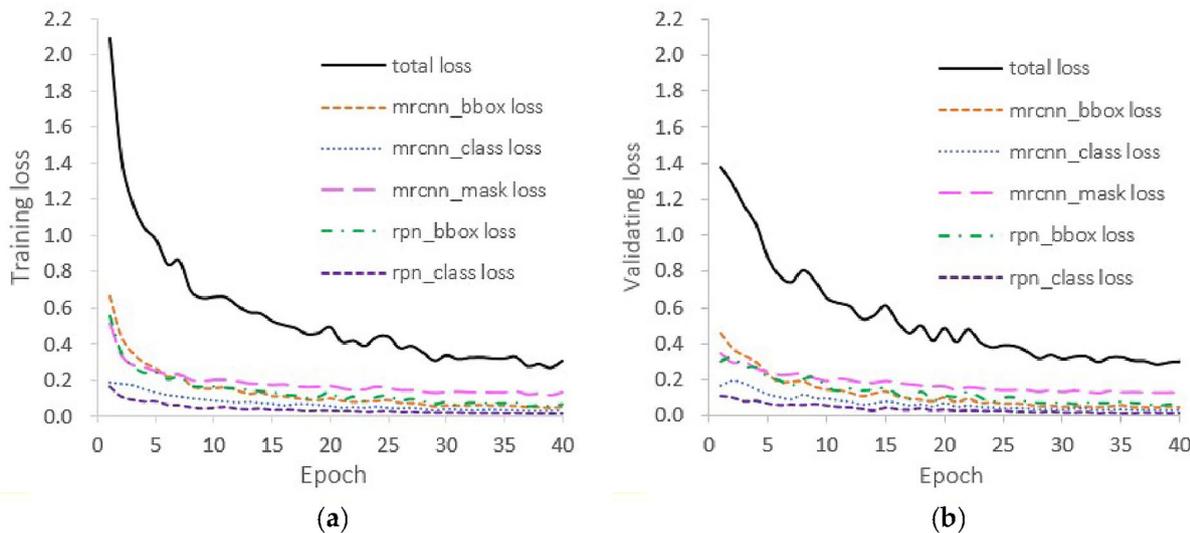


Fig. 8 **a** The training loss curve; **b** The validation loss curve

higher the AP value, the more accurate the detection result of the model. In this paper, the AP value of gold foil shedding is 0.967, indicating that the detection accuracy of the model is high.

The research data in this paper are from the protection project of the Dazu Thousand-Hand Bodhisattva Statue in China. After cleaning, labeling, and augmentation, the research data are input into the model, and then the trained model can detect gold foil shedding. To evaluate the performance of the trained model, we use 380 new images, which are not used in the training model, for testing. Figure 10(a)–(c) shows the detection results of the arm numbered 4–5-S5 taken from different directions, and the predicted damage areas match the ground truth damage very well. The gold foil shedding of 4-5-S5 is less, and the topological structure of the damage is very regular, in which the gold foil of the little finger falls off almost completely. There is a small area of gold foil shedding in the palm of the pinkie, but it has also been successfully detected.

As shown in Fig. 11, the detection results of gold foil shedding on the arm numbered 7–2-S8 show that the gold foil on the arm falls off in only one place, but the shedding area is large, most of the gold foil attached to the back of the hand has fallen off, and the topological

structure of the damage is also regular. Figure 11 (a)–(b) are 7-2-S8 taken from different directions and different brightness. Even if the external conditions change, the detection results of the model are accurate. In addition, the background of the arm also has a certain degree of interference, there are dharma instruments above and below the arm. The stone part of the instrument is very similar to the part exposed after the gold foil shedding from the arm, but the model can detect the damaged area without interference.

Figure 12 shows the detection results of the gold foil shedding on the arms numbered 5-8-S1 (upper side) and 4-8-S11 (lower side). There is a small area of gold foil shedding in the palm of 5-8-S1, while the gold foil shedding on the back of 4-8-S11 is larger, and the topological structure of the damaged area is complex. Figure 12(a)–(b) are taken from different directions. After the gold foil of these arms shed from the surface, there are some stains on the exposed stone part, but these do not affect the model’s detection of the damage area. In addition, the stone part of the dharma instruments in the image is also not detected mistakenly.

To verify the limitations of the trained model, images with complex backgrounds are selected for testing. The

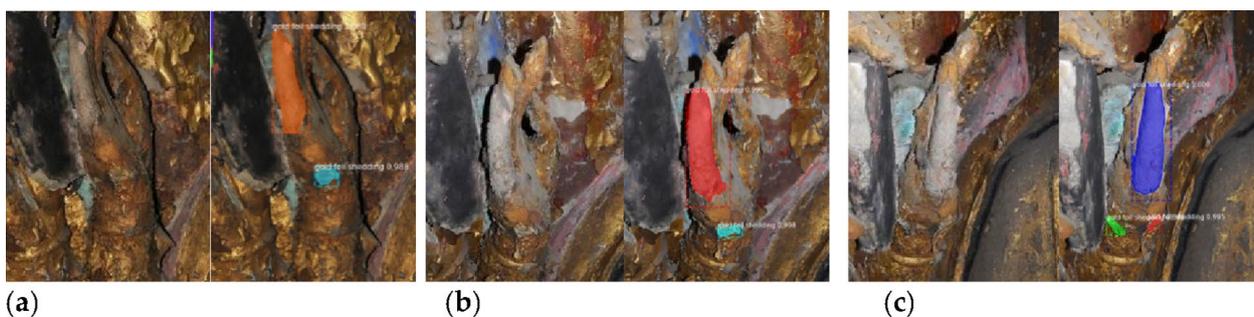


Fig. 10 The test results of 4–5-S5 from different directions (Areas of different colors all represent the disease of gold foil shedding)

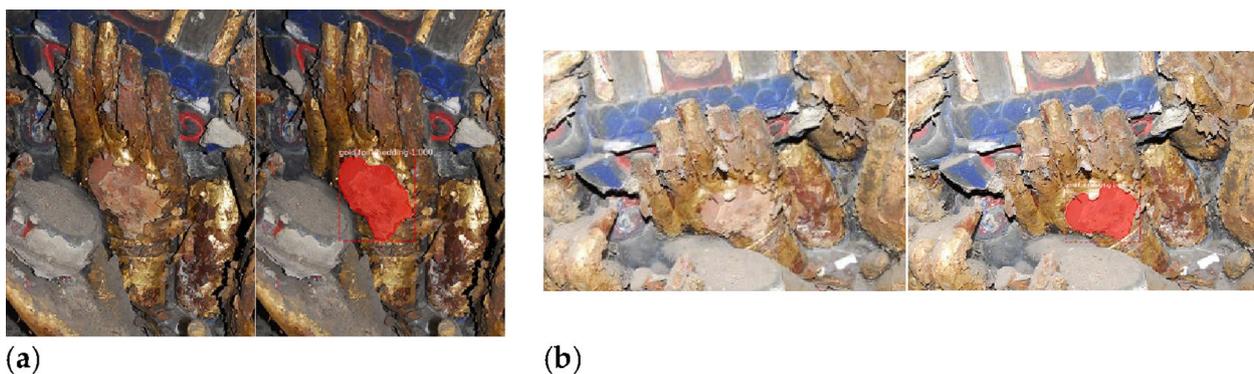


Fig. 11 The test results of 7–2-S8 from different directions

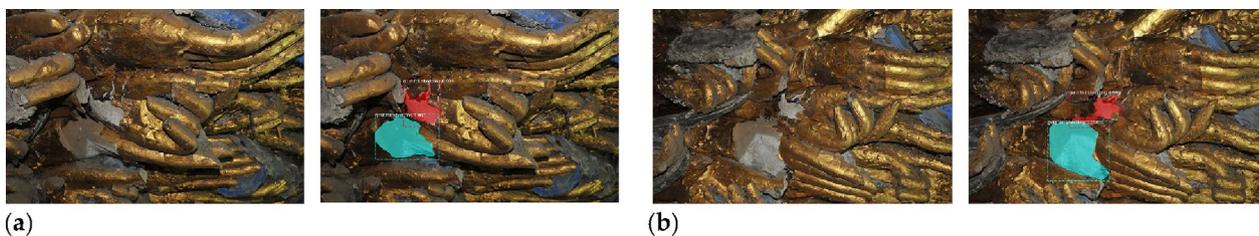


Fig. 12 The test results of 5–8-S1 and 4–8-S11 from different directions (Areas of different colors all represent the disease of gold foil shedding)

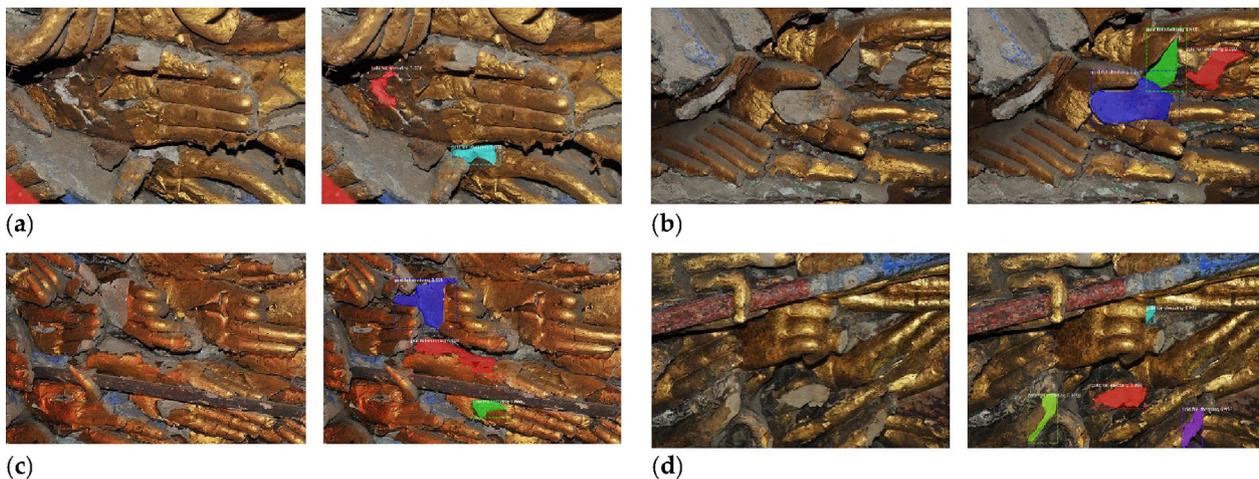


Fig. 13 The test results with complex backgrounds (Areas of different colors all represent the disease of gold foil shedding)

images include gold foil warping, gold foil plasters shedding, gold foil dust, stone incompleteness, dharma instruments, and Tian Yan (“The Eye of Heaven”), which have similar characteristics to gold foil shedding.

As shown in Fig. 13, the test results are good for images with complex backgrounds and topologies. Figure 13(a) is the detection result of arms numbered 5-8-S11 (upper side) and 5-8-S12 (lower side). There are two gold foil shedding in the image. The thumb of 5-8-S11 has a large area of gold foil warping, and its palm has a Tian Yan, but the model does not detect them as gold foil shedding. Figure 13(b) shows the detection results of 5-10-S7 (lower side) and 5-10-S11 (upper side). The image has three gold foil shedding. There are gold foil plasters shedding and dharma instruments on the left side, and the model is also not affected by these interference factors.

Figure 13 (c) shows the detection results of 5-11-S5 (upper side) and 5-11-S6 (lower side). The image has three gold foil shedding. There is gold foil warping, gold foil dust, and Tian Yan on the left and lower sides of the image. The model also performs well in this case. Figure 13 (d) shows the detection results of 9-3-S3 (upper side), 9-3-S5 (lower left side), 9-3-S6 (lower side) and 9–3-S7 (lower right side). The image has four gold foil

shedding. Although the damage area is generally small, they are accurately detected by the model. In addition, the thumb of 9-3-S3 is broken and the stone part is exposed, but the model does not detect it as gold foil shedding.

Additionally, this article also conducted detection of two gold foil-related diseases in Thousand-Hand Bodhisattva Statue: gold foil dust and gold foil warping. As shown in Fig. 14, the results are from the detection of arms in areas 2-12-S1 and 3-2-S13. In Fig. 14(a) and (b) display the detection results of the gold foil dust disease, achieving an AP value of 0.921, while (c) and (d) show the results for the gold foil warping disease, with an AP value of 0.903. These results indicate that the model has high accuracy in recognizing the three diseases: gold foil shedding, gold foil dust, and gold foil warping. Therefore, Mask R-CNN demonstrates good universality for stone relics with gold foil-related diseases.

In a word, Mask R-CNN can play an effective role in the range of gold foil shedding topology from regular to complex and damage area from small to large. This shows that Mask R-CNN can achieve high-precision detection results under various conditions.

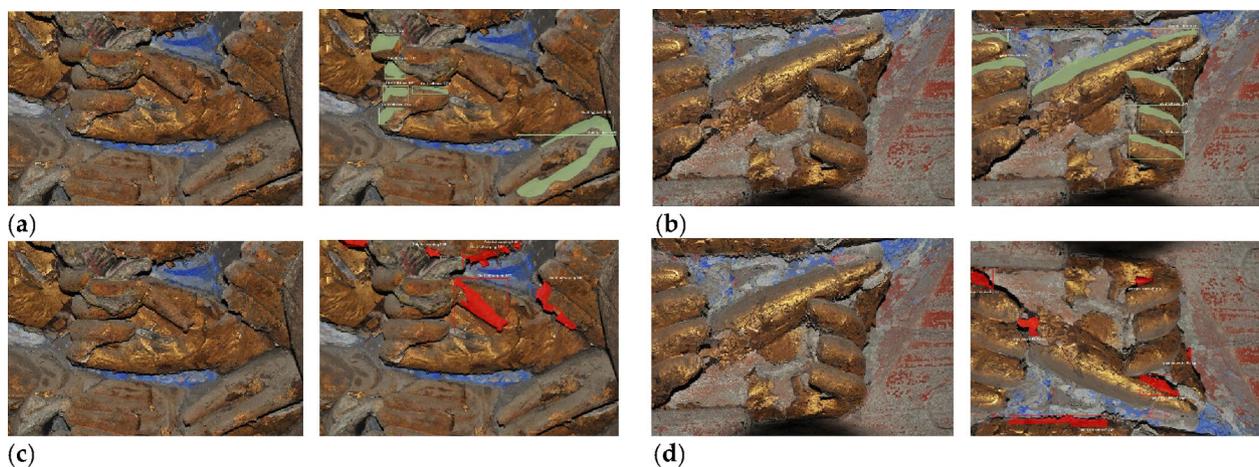


Fig. 14 Detection results of gold foil dust and gold foil warping disease in 2-12-S1 and 3-2-S13

Conclusion

At present, the detection of cultural heritage damages is mainly based on an expert's visual capacity, which is always time-consuming and laborious. These problems can be solved by using a deep learning algorithm to assist damage detection. In this paper, taking Dazu Thousand-Hand Bodhisattva Statue as an example, the Mask R-CNN algorithm is used to detect the gold foil shedding. The algorithm can automatically detect the gold foil shedding with high accuracy. After 40 epochs of training, the AP value for the segmentation is 0.967.

The research data of this paper are from the high-precision images of the Dazu Thousand-Hand Bodhisattva Statue. During model training, the validation dataset was used to verify the model performance. After the model was trained, 380 new images were used to evaluate the performance of the trained model. The results show that the trained model can accurately detect gold foil shedding in different directions, different brightness, and different distances. In addition, the model will not mistakenly detect other damages (such as gold foil dust, gold foil warping, stone incompleteness and gold foil plasters shedding) and arm decoration (such as Tian Yan and dharma instruments) as gold foil shedding. This method can be applied to the gold-overlaid stone cultural heritage in other areas, and new images need to be added to the dataset to enhance the applicability of the model.

This research utilizes the Mask R-CNN algorithm to detect three types of disease in the Thousand-Hand Bodhisattva Statue: gold foil shedding, gold foil dust, and gold foil warping. It achieves high accuracy and practical value, supplementing the types of disease detection in stone relics as proposed by Hatir et al. [35]. However, there are still some limitations in this study. For example,

due to the limitation of stone relic disease benchmark, the research object of this paper is mainly limited to the Thousand-Hand Bodhisattva Statue, more cultural relic objects are still needed for more extensive verification in the future. In addition, it is necessary to integrate the characteristics of stone cultural relics disease (including gold foil disease) with the design and architecture of the neural network model sub-module to optimize the model, which is expected to improve the detection accuracy and clearer boundaries. In future works, it is necessary to further expand the quantities and categories of datasets and use a state-of-the-art deep learning algorithm to improve the accuracy of automatic damage detection for gold-overlaid stone cultural heritage.

Acknowledgements

The authors would like to acknowledge and thank the Dazu Museum for its support of this work. The authors would also like to thank Prof. Changfa Zhan who always give the authors the novel suggestion. Thanks to Fangyin Li and Huili Chen for their work on the capturing and processing of the data. Thanks to Songnian Li from Toronto Metropolitan University put forward constructive suggestions for the revision of this paper.

Author contributions

Conceptualization, MH and SY; methodology, DH, YY; software, YY and DH; resources, MH; writing—original draft preparation, YY, DH, HC; writing—review and editing, All authors; supervision, MH and SY; project administration, MH; funding acquisition, MH All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by National Natural Science Foundation of China (Grant Number 42171444) and Beijing Municipal Natural Science Foundation (Grant Number KZ202110016021).

Availability of data and materials

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to confidentiality.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 14 October 2023 Accepted: 23 January 2024

Published online: 06 February 2024

References

- Bromblet P, Vallet J, Verges-Belmin V. Illustrated glossary on stone deterioration patterns. *Monuments Sites*. 2008;5:60.
- Hou M, Li S, Jiang L, Wu Y, Hu Y, Yang S, Zhang X. A new method of gold foil damage detection in stone carving relics based on multi-temporal 3D LiDAR point clouds. *ISPRS Int J Geo Inf*. 2016;5:60.
- Wang H, Luo Y, An C, Chu S, Shen Z, Huang L, Zhang D. Application of imaging polarimeters to enhanced detection of stone carving. *J Cult Herit*. 2019;40:92–8.
- Wang K, Xu G, Li S, Ge C. Geo-environmental characteristics of weathering deterioration of red sandstone relics: a case study in *Tongtianyuan* Grottoes, Southern China. *Bull Eng Geol Env*. 2018;77:1515–27.
- LeiLei Z, XueZhi F. Study on mechanism of water condensation and field experiments of Thousand-Hand Guanyin in Dazu Rock Carvings. *Proc IOP Conf Series Earth Environ Sci*. 2018. <https://doi.org/10.1088/1755-1315/186/2/012007>.
- Elmasry MI, Johnson EA. Health monitoring of structures under ambient vibrations using semiactive devices. *Proc Proc Am Control Conf*. 2004. <https://doi.org/10.1111/j.1467-8667.2005.00379.x>.
- Gattulli V, Chiaramonte L. Condition assessment by visual inspection for a bridge management system. *Computer-Aided Civil Infrastruct Eng*. 2005;20:95–107.
- O'Byrne M, Schoefs F, Ghosh B, Pakrashi V. Texture analysis based damage detection of ageing infrastructural elements. *Computer-Aided Civil Infrastruct Eng*. 2013;28:162–77.
- Amin HU, Malik AS, Ahmad RF, Badruddin N, Kamel N, Hussain M, Chooi W-T. Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Australas Phys Eng Sci Med*. 2015;38:139–49.
- Ortega-Zamorano F, Jerez JM, Gómez I, Franco L. Layer multiplexing FPGA implementation for deep back-propagation learning. *Integr Computer-Aided Eng*. 2017;24:171–85.
- Abed M, Mohammed KH, Abdulkareem G-Z, Begonya M, Salama A, Maashi MS, Al-Waisy AS, Ahmed M, Subhi AA, Mutlag L. A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of COVID-19 based on X-ray images. *Computers Mater Continua*. 2021. <https://doi.org/10.3604/cmc.2021.012874>.
- Hou M, Hu Y, Wu Y, Zhao X. 3D documentation and data management in the Dazu thousand-hand bodhisattva statue in China. *International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*. 2016. 5.
- Tian XL, Li NS, Zhang ZG, Hu Y, Zhan CF. Analysis of the stones of thousand hand Buddhism by the combination of on-site and laboratory techniques. *Proc Adv Mater Res*. 2012. <https://doi.org/10.4028/www.scientific.net/AMR.557-559.1461>.
- Gao F, Zhou X, Zhou H, Li M, Tong H, Liu S. Characterization and analysis of sandstone substrate, mortar layers, gold foils, and paintings of the Avalokitesvara Statues in Dazu County (China). *J Cult Herit*. 2016;21:881–8.
- Changfa, Z. Inheritance and innovation of lacquer and gold foil coating. *China and Italy* 2017, 13.
- Yu-hua, W.; Miao-le, H. The Conservation and 3D Virtual restoration on the Thousands hands Bodhisattva of Dazu Rock carvings world heritage. *China and Italy* 2017, 69.
- Angra, S.; Ahuja, S. Machine learning and its applications: A review. In *Proceedings of the 2017 international conference on big data analytics and computational intelligence (ICBDAC)*, 2017; pp. 57–60.
- Wäldchen J, Mäder P. Machine learning for image based species identification. *Methods Ecol Evol*. 2018;9:2216–25.
- Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria. *Transl Res*. 2018;194:36–55.
- Nazarian E, Taylor T, Weifeng T, Ansari F. Machine-learning-based approach for post event assessment of damage in a turn-of-the-century building structure. *J Civ Struct Heal Monit*. 2018;8:237–51.
- Valero E, Forster A, Bosché F, Hyslop E, Wilson L, Turmel A. Automated defect detection and classification in ashlar masonry walls using machine learning. *Autom Constr*. 2019;106: 102846.
- Meng T, Huang R, Lu Y, Liu H, Ren J, Zhao G, Hu W. Highly sensitive terahertz non-destructive testing technology for stone relics deterioration prediction using SVM-based machine learning models. *Heritage Science*. 2021;9:1–9.
- Adamopoulos E. Learning-based classification of multispectral images for deterioration mapping of historic structures. *J Build Pathol Rehabil*. 2021;6:1–15.
- Chaiyasarn, K.; Khan, W.; Ali, L.; Sharma, M.; Brackenbury, D.; DeJong, M. Crack detection in masonry structures using convolutional neural networks and support vector machines. In *Proceedings of the ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2018; pp. 1–8.
- Kwon, D.; Yu, J. 2019. Automatic damage detection of stone cultural property based on deep learning algorithm. *The International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*. 42: 639–6.
- Wang N, Zhao X, Zou Z, Zhao P, Qi F. Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning. *Computer-Aided Civil Infrastruct Eng*. 2020;35:277–91.
- Zou Z, Zhao X, Zhao P, Qi F, Wang N. CNN-based statistics and location estimation of missing components in routine inspection of historic buildings. *J Cult Herit*. 2019;38:221–30.
- Angheluta L, Chiroşca A. Physical degradation detection on artwork surface polychromies using deep learning models. *Romanian Rep Phys*. 2020;72:805.
- Hatir E, Korkaç M, Schachner A, Ince İ. The deep learning method applied to the detection and mapping of stone deterioration in open-air sanctuaries of the Hittite period in Anatolia. *J Cult Herit*. 2021;51:37–49.
- He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2017; pp. 2961–2969. Zhou C, Li Y. Influence of rainfall infiltration on stability of Forbidden City wall. *Journal of Building Structures*. 2020. 286–296.
- He K, Zhang X, Ren S, Sun J. End-to-end training of object class detectors for mean average precision. Berlin: Springer; 2017.
- Henderson P, Ferrari V. End-to-end training of object class detectors for mean average precision. Berlin: Springer International Publishing; 2016.
- Everingham M, Eslami S, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes challenge: a retrospective. *Int J Comput Vision*. 2015;111:98–136.
- 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.
- Ergün Hatir M, Ince I, Korkaç M. Intelligent detection of deterioration in cultural stone heritage. *J Build Eng*. 2021;44:102690.
- Angheluta LM, Chiroşca A. Physical degradation detection on artwork surface polychromies using deep learning models. *Rom. Rep. Phys*. 2020;72(3):805.

Publisher's Note

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