Heritage Science

RESEARCH Open Access

Automatic defect detection in infrared thermal images of ancient polyptychs based on numerical simulation and a new efficient channel attention mechanism aided Faster R-CNN model

Xin Wang[1], Guimin Jiang[1,2*], Jue Hu[2*], Stefano Sfarra[3], Miranda Mostacci[4], Dimitrios Kouis[5], Dazhi Yang[6], Henrique Fernandes[7,8], Nicolas P. Avdelidis[7,9], Xavier Maldague[9], Yonggang Gai[1] and Hai Zhang[2*]

Abstract

In recent years, the preservation and conservation of ancient cultural heritage necessitate the advancement of sophisticated non-destructive testing methodologies to minimize potential damage to artworks. Therefore, this study aims to develop an advanced method for detecting defects in ancient polyptychs using infrared thermography. The test subjects are two polyptych samples replicating a 14th-century artwork by Pietro Lorenzetti (1280/85–1348) with varied pigments and artificially induced defects. To address these challenges, an automatic defect detection model is proposed, integrating numerical simulation and image processing within the Faster R-CNN architecture, utilizing VGG16 as the backbone network for feature extraction. Meanwhile, the model innovatively incorporates the efficient channel attention mechanism after the feature extraction stage, which significantly improves the feature characterization performance of the model in identifying small defects in ancient polyptychs. During training, numerical simulation is utilized to augment the infrared thermal image dataset, ensuring the accuracy of subsequent experimental sample testing. Empirical results demonstrate a substantial improvement in detection performance, compared with the original Faster R-CNN model, with the average precision at the intersection over union = 0.5 increasing to 87.3% and the average precision for small objects improving to 54.8%. These results highlight the practicality and effectiveness of the model, marking a significant progress in defect detection capability, providing a strong technical guarantee for the continuous conservation of cultural heritage, and offering directions for future studies.

Keywords Efficient channel attention, Faster R-CNN, Defect detection, Deep learning, Infrared thermography

*Correspondence:
Guimin Jiang
2502818383@qq.com
Jue Hu
juehundt@hit.edu.cn
Hai Zhang
hai.zhang@hit.edu.cn
Full list of author information is available at the end of the article

Springer Open

© 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.

Introduction

Cultural heritage, with its unique societal and historical value, has become one of the key drivers of social development and the progression of civilization. Respecting and conserving the cultural heritage of different ethnic regions are the foundation for the coexistence of humanity. The preservation of cultural heritage is essential for maintaining the diversity of human culture, representing significant wealth of human civilization. Polyptychs hold a significant position within cultural heritage due to their historical, artistic, and cultural significance [1]. These polyptychs, which are integral parts of conventional art forms in many cultures, often depict scenes from mythology, history, or religion, serving as both decorative elements and narrative mediums. However, over time, defects, such as gaps, cavities or splits, may exist inside ancient polyptychs [2]. It is thus crucial to detect different types of defects during the polyptych's conservation.

Non-destructive testing (NDT) has emerged as an essential technique for material safety assurance in many fields due to its non-destructive nature, safety and structural adaptability [3]. In addition, NDT techniques are extensively employed to guarantee the effectiveness and integrity of manufacturing processes [4]. Notably in the realm of cultural heritage, due to the irreplaceable and historical nature of the objects of study, the maturity of NDT techniques is of particular importance in order to conserve the objects of study from further damage after the restoration procedure [5]. The primary NDT techniques include digital imaging [6], X-radiography [7], ultraviolet imaging [8], terahertz imaging [9], and infrared (IR) photography [10], which can be conducted under ambient, cross-polarized, or raking light.

Infrared thermography (IRT), as an NDT technique, has attracted significant attention in cultural heritage inspection because of its high resolution, non-contact nature, and ability to scan extensive regions in a brief timeframe [11–16]. In recent years, IRT technology has seen significant advances in cultural heritage conservation. Ovadia and Brook [17] showed in their study that IRT is very effective in detecting wall painting damages, but attention must be paid to the potential effects of heat on different materials. Similarly, Attas et al. [18] used near-infrared spectroscopic imaging to detect painting components in artworks, successfully identifying defects in a 15th-century painting "Untitled (The Holy Trinity)." Delaney et al. [19] identified major pigments and pigment mixtures in artworks using visible and infrared imaging spectroscopy. These studies highlight the widespread application and importance of IRT technology in cultural heritage preservation.

IRT is particularly effective in assessing defects and damages in ancient cultural artifacts. In addition, IRT techniques have other important uses in heritage conservation, such as monitoring environmental conditions, detecting small changes or assisting in restoration. In monitoring environmental conditions, it can detect temperature and humidity changes in the preservation environment of artifacts, helping to identify potential environmental risks and take appropriate precautions. This role of IRT technology has been confirmed in the article [20]. IRT technology can also detect small changes on the surface of artifacts, such as cracks or flaking, aiding in the preservation and restoration of artifacts through high-resolution thermal images. Reference [21] highlighted the application of IRT technology in detecting minor surface changes on artifacts.

Although IRT techniques are very useful in detecting artworks, high-frequency use may cause cumulative damage to experimental objects [22]. Reference [23] stated that reducing the frequency of IRT experiments can be effective in avoiding secondary damage to artworks. To that end, parallel to the ability to detect defects being critical in artwork testing, it is equally important to avoid any secondary damage that may be done to the artwork by the IRT technique used. In most cases, it is not recommended to repeat the experiment on the test object to maximize the preservation of the artwork.

Two strategies are effective to reduce the number of IRT experiments for artworks: (1) reducing the input energy, and (2) employing numerical modelling to optimize and standardize the test procedure. The focus of this study is on numerical modelling to enhance the experimental design of IRT defect detection by predicting results and understanding thermal mechanisms in complex materials.

IRT defect detection experiments are conducted in two main phases. First, IRT experiments record temperature changes on the object's surface. Next, defects are identified by differential analysis of thermophysical properties, producing a clear infrared contrast. During the second phase, defects need to be identified by analyzing the reflected near-infrared and short-wave infrared spectra, which can be challenging for detecting deep internal defects and often requires visual inspection, leading to potential errors.

Challenges remain in dealing with visual inspections. To reduce errors in defect assessment, the author has conducted extensive research on deep learning. Over the past few years, deep learning has made remarkable strides in the domain of image defection. For instance, Hu et al. [24] successfully applied the YOLOv4 network detect automatically mural shedding disease. Hatır et al. [25] used a Mask R-CNN based algorithm model to train and test eight types of deterioration in open-air sanctuaries from the Hittite period, achieving

satisfactory results. Wang et al. [26] developed the so-called GreatWatcher system for expedited defect identification in the Great Wall, combining Fast R-CNN and mobile crowd sensing technology.

Drawing on the aforementioned studies, this work constructs an automatic defect detection system using numerical simulations with deep learning networks and machine learning algorithms. This system aims to reduce errors in actual defect detection and optimize the detection process. The primary focus of this study is on polyptychs, which typically have anisotropic structures as they are usually composed of multiple materials. To achieve faster and more accurate automatic defect detection in IRT on polyptychs, this work first expands the defect dataset using numerical modeling techniques, then trains deep learning networks, and finally validates the model's effectiveness through actual experiments.

Indeed, the novelty of this work lies in the comprehensive approach that involves expanding the defect dataset using numerical modeling techniques, which enhances the training of deep learning networks. This integration is particularly effective in capturing and representing the complex anisotropic structures and material compositions of polyptychs. By utilizing advanced numerical simulations and machine learning algorithms, our method significantly improves the precision and reliability of defect detection. This study validates the model's effectiveness through actual experiments, ensuring robust and scalable solutions for the conservation of cultural heritage.

Description of the sample under test and geometric modeling

Description of the replicas (samples under test)

To verify the ability of the proposed numerical-simulation-aided method for NDT of polyptychs, replicas were realized by a professional restorer and then investigated. It should be highlighted that this replica has been thoroughly described in [35]. The key information is reiterated in what follows.

The original polyptych, see Fig. 1a, was painted in 1320 by Pietro Lorenzetti. This artwork is currently under preservation in the Church of *Santa Maria della Pieve* in Arezzo, Italy. Figure 1b shows a detail of a magnified portion of the polyptych. To assess the performance of the constructed automatic defect detection system on infrared thermal images, two replicas (i.e., samples under test) were prepared. The samples under test were built on support boards. A typical 14th-century tempera technique was used. This technique employs mainly eggs, with a protein-based or plant-derived adhesive as a binder for the pigments, and is executed on a wooden substrate as shown in Fig. 2a.

For the sake of clarity, the samples are hereafter referred to as samples A and B, respectively. One of the main differences among these samples was the technique used to realize the halo. In that, a yellow pigment was indeed used in the sample A, while a gold leaf was applied in the sample B. The sizes of the samples are both $200 \times 300 \times 15$ mm (cf. Figure 2a). Rabbit skin glue, soaked overnight and then melted, was spread onto the panels and left to dry (cf. Figure 2b). Defects were simulated with folded Teflon pieces placed between the panel and the canvas layer—linen for

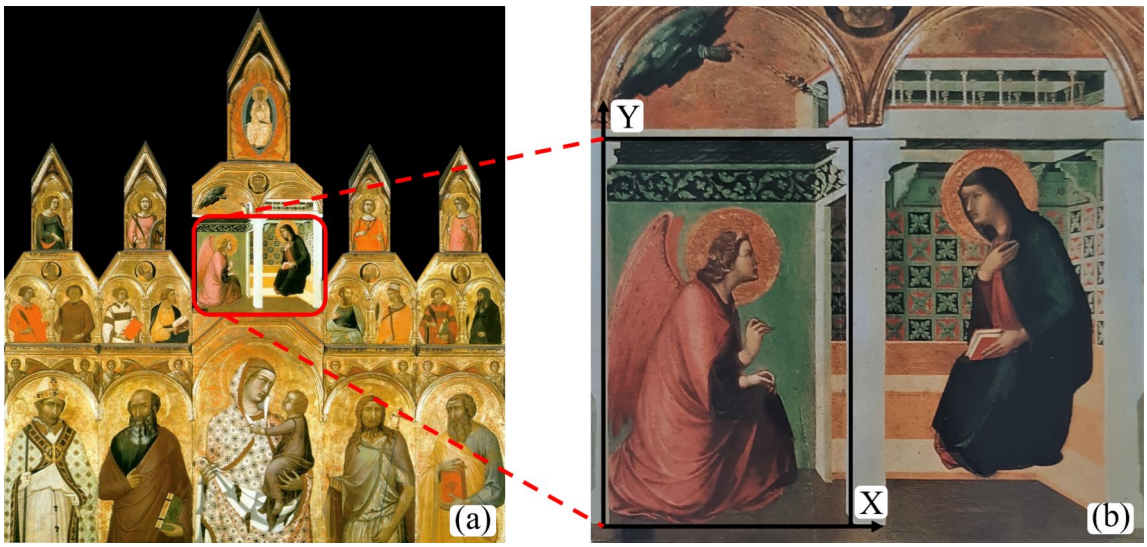


Fig. 1 **a** A photograph of the polyptych, **b** an enlarged view of the replicated portion, with the box indicating the research area and providing reference dimensions (X: horizontal, Y: vertical) for numerical modeling



Fig. 2 Fabrication of the replicas: **a** The two panels are used as the substrate, **b** coating adhesive, **c** inserting the initial Teflon piece (defect 1), **d** applying adhesive to the linen canvas, **e** allowing the linen canvas layer to dry, **f** inserting the second Teflon piece (defect 2), **g** coating a plaster-glue mixture, **h** making the plaster layer smooth, **i** inserting the third Teflon piece (defect 3) **j** making the second plaster layer smooth, **k** sketching the outline of the figure, **l** placing the gold foil **(m)** employing a moistened sponge to affix the gold foil **(n)** applying pigment to sample B, **o** applying different colors to sample A, and **(p)** drying of the final samples

sample A, and flax for sample B (cf. Figure 2c–e). After that the canvas layers dried, a second Teflon insert was added (cf. Figure 2f). Next, a Bologna plaster and rabbit glue mixture was applied as a first layer (cf. Figure 2g), sanded post-drying (cf. Figure 2h), followed by a third Teflon insert (cf. Figure 2i). A second layer of plaster was then applied and sanded for a painting-ready surface (cf. Figure 2j). The pictorial representation began with charcoal tracing (cf. Figure 2k), followed by creating a gilded halo using red bolus and gold leaf for sample A, and yellow pigments for sample B (cf. Figure 2l–m). The tempera paint, made with egg yolk and a drop of vinegar, was layered with marten hair brushes (cf. Figure 2n–o), finalizing the mock-ups (cf. Figure 2p).

Geometric modeling

In this section, the process of constructing a geometric model for numerical simulation of the temperature distribution on the sample surface is discussed. In this study, the geometric model was created using Autodesk AutoCAD 2024, a computer-aided design software, and then imported into the COMSOL Multiphysics software for numerical simulation. During the modeling process, the location (cf. Figure 3a), size (cf. Figure 3b), thickness, and depth (cf. Figure 3c) of the defects were made consistent with the artificially created samples A and B from the previous section. The modeling method is based on cubes corresponding to the geometrical dimensions of the model parts. Thus, the structure is reconstructed in in a reversed way. The position of the cube within the spatial framework is fine-tuned by altering the starting point coordinates. Once the positions are determined, the contours of the model surface are drawn using the work plane function in Autodesk AutoCAD 2024 along the X- and Y-axis. The depth of the work plane is then extended along the Z-axis. This creates a three-dimensional model, which subsequently needs to be numerically simulated using COMSOL Multiphysics software. The geometric modeling is showed in Fig. 3. For more detailed information on the modeling and numerical simulation process, readers are encouraged to refer to the sections "Geometric modeling" and "Simulation setting" in the article [35].

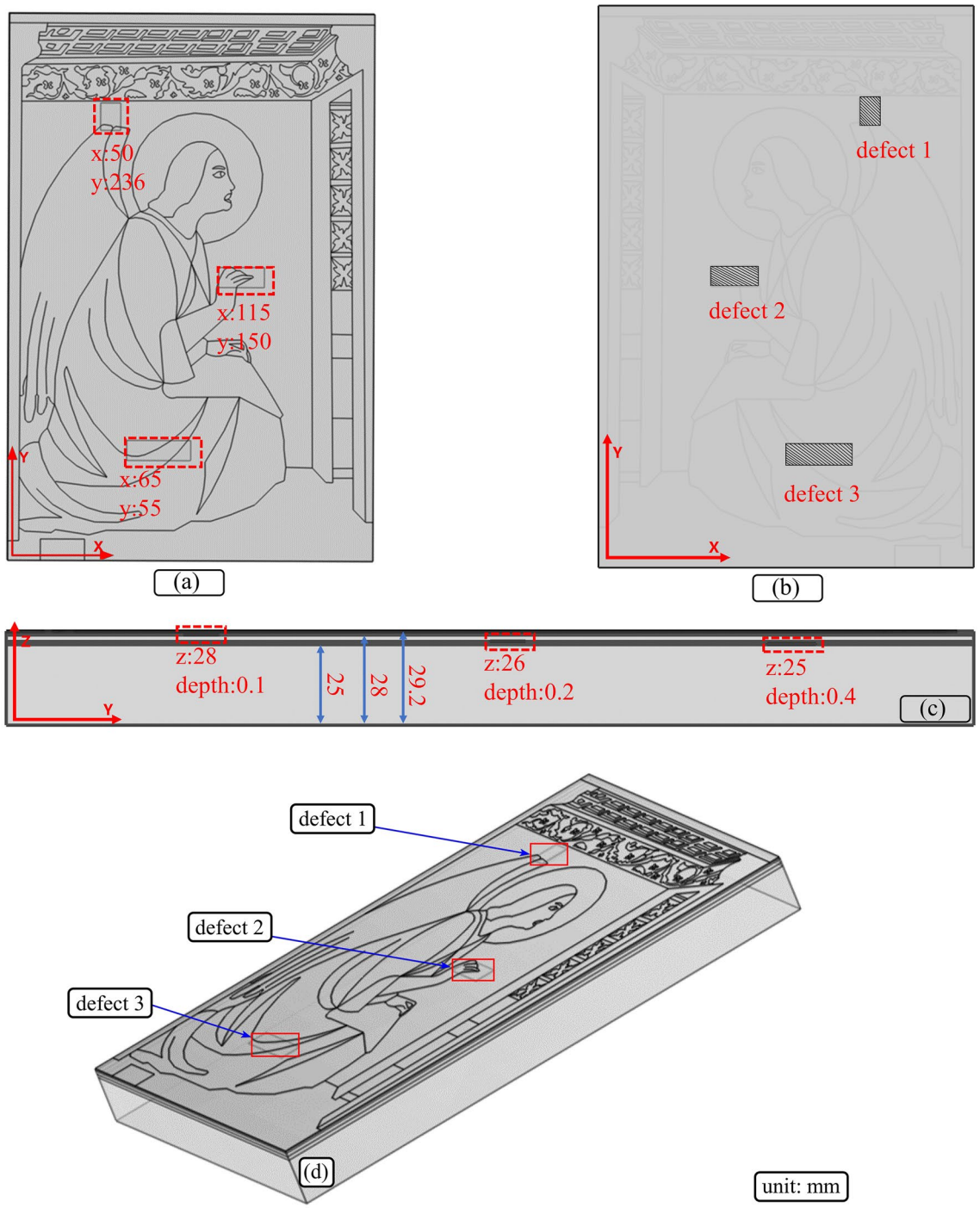


Fig. 3 Geometric modeling of the sample: **a** Top-view photograph displaying the XY coordinates of the defects; **b** bottom-view diagram indicating the positions of the defects; **c** side-view diagram illustrating the Z coordinates and depths of the defects; **d** three-dimensional representation highlighting the locations of the defects. (All values are in mm; the coordinates refer to the positions of the top points of the defects.)

Faster-RCNN

In 2016, Ren et al. [27] proposed the Faster R-CNN architecture. In the initial stage of the Faster R-CNN network, data preprocessing steps such as normalization were performed on infrared thermal images to reduce variance between images and ensure the model adapts well to variable inputs. The structure of the Faster R-CNN network, depicted in Fig. 4, is delineated into three sequential

steps in its operation: (1) extracting features from the input image using a pre-trained network; (2) passing the extracted features through the region proposal network (RPN) to generate a specified number of candidate frames; and (3) inputting the predicted classification and regression results into region of interest (RoI) pooling, along with the candidate frames and image features, to classify the candidate regions, ascertain their categories, and adjust their coordinates.

The following sections provide a detailed introduction to the three steps: the feature extraction network, RPN, and the final detection network.

Feature extraction network

In the Faster R-CNN target detection framework, the feature extraction network is particularly important for the defect detection task because it directly determines the model's ability to recognize defects in infrared thermal images. In this study, the visual geometry group16 (VGG16) network is chosen as the feature extraction network due to its capability to learn local features of the image through multilayer filters.

The VGG16 network structure consists of several layers, each playing a critical role in feature extraction. The first convolutional (Conv1) layer uses two sets of 64 filters of size 3×3 to extract basic edge and texture features from the image, followed by a rectified linear unit (ReLU) activation function. This is succeeded by the first max pooling (MaxPool) layer with a 2×2 filter and stride of 2, performing spatial down-sampling to reduce the feature map size and increase translation invariance. The second convolutional (Conv2) layer with two sets of 128 filters of size 3×3 further extracts more complex edge and texture features, followed by another ReLU activation function and a second MaxPool layer with similar down sampling properties. The third convolutional (Conv3) layer, consisting of three sets of 256 filters of size 3×3, extracts higher-level features, such as shapes or contours, again followed by ReLU activation and a third MaxPool layer for additional down-sampling. The fourth convolutional (Conv4) layer, also consisting of three sets of 512 filters of size 3×3, captures even more complex features, enabling the network to recognize a broader range of object characteristics. This is followed by another ReLU activation function and a fourth MaxPool layer. The fifth convolutional (Conv5) layer, with three sets of 512 filters of size 3×3, extracts the highest-level features, recognizing complex shapes and objects. This layer is followed by a final MaxPool before transitioning to the fully connected (FC) layers. The FC layers, with 4096 nodes in the first two layers and 1000 nodes in the last layer, integrate the extracted features for advanced feature extraction and classification, utilizing ReLU activation

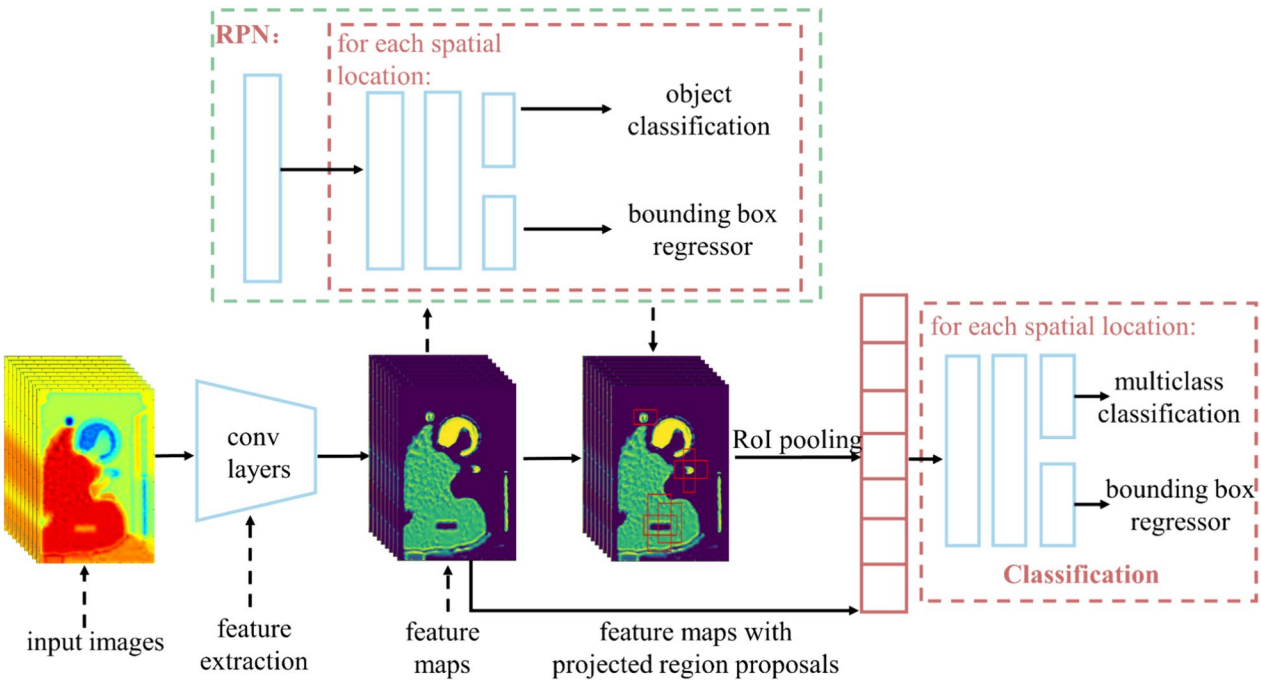


Fig. 4 Schematic diagram of the basic Faster R-CNN network structure

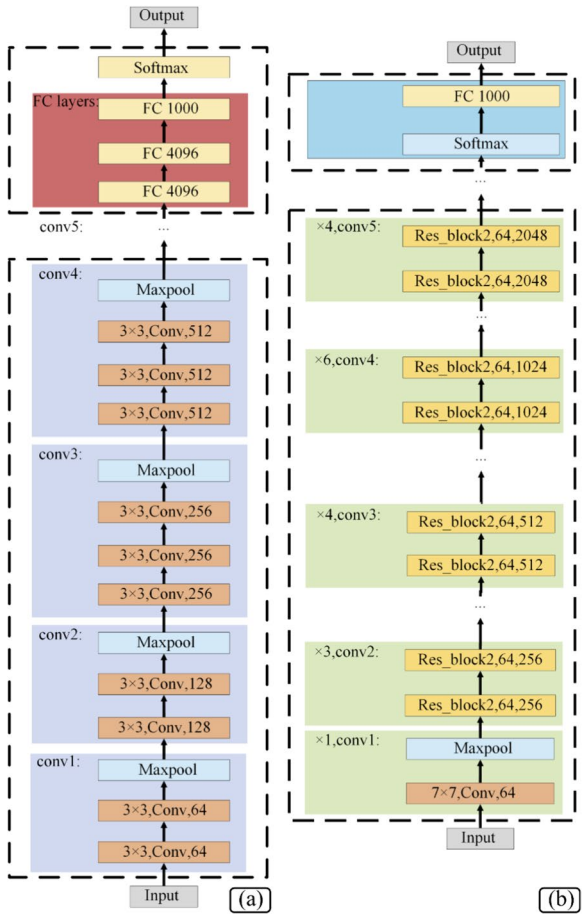


Fig. 5 The two backbone network structures: **a** VGG16, **b** ResNet50

functions. Finally, the output layer uses a soft maximum (SoftMax) function to map the features to the final classification labels.

Through these layers, VGG16 can extract fundamental features, such as edges or corners, from images. These features are crucial for detecting subtle defects on the surface of infrared thermal images. Research [28] also indicated that the residual network 50 (ResNet50) was capable of extracting similar features, as shown in Fig. 5. However, to minimize the risk of overfitting when handling smaller datasets, this study has selected VGG16 as the backbone network for the follow experiments.

Regional proposal network

Similarly, the RPN plays a significant role in infrared thermal defect detection, as shown in Fig. 6. After using the VGG16 feature extraction network in the previous section, a 512-channel feature map is obtained. Based on the feature map, RPN utilizes nine anchor frame structures (three scales: 128×128, 256×256, and 512×512

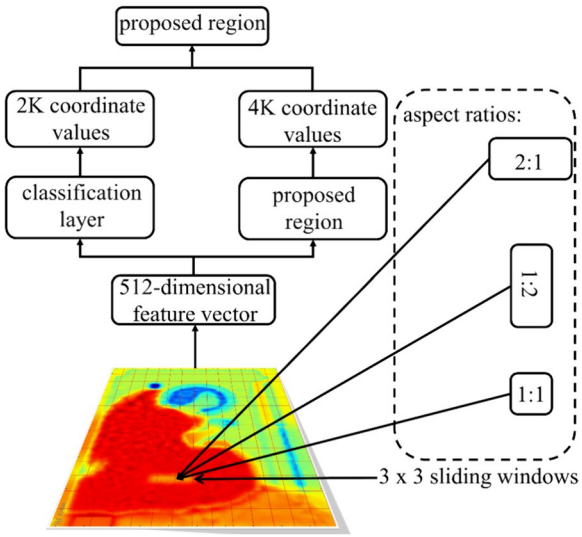


Fig. 6 RPN anchor frame structure diagram

combinations, and three aspect ratios: 2:1, 1:2, and 1:1) to cover defects of infrared thermal images of various sizes and shapes, thus enabling the detection of defective images of different sizes and shapes. Thus, this approach allows the model to generate candidate regions that cover a variety of potential defect areas, both subtle and blurred. Each generated candidate region is subsequently processed through a series of convolutional layers to evaluate its overlap with the true target frame. Those regions with a high degree of overlap with the true defect location will be labeled as targets, a step that is critical in identifying the region's most likely to contain defects from a large number of candidate regions.

In addition, RPN's anchor framing mechanism enhances the model's ability to adapt to defects of different sizes and shapes, and further refines the candidate region selection process by generating multiple anchors (total number of H×W×K) at each location on the feature map to ensure comprehensive coverage of defects in the infrared thermal image.

Detection network

The detection network consists of two key components: an RoI-pooling layer and a FC layer. The network is designed to ensure that every defect in the sample under test is accurately identified and localized.

First, the RoI-pooling layer efficiently maps numerous defects in the ancient polyptych of varying sizes, extracted by the RPN network, into fixed-dimension feature vectors. This transformation standardizes data format, allowing the network to learn with greater efficiency, and eventually improves both the accuracy and processing efficiency of defect detection.

Second, the role of the FC layer is to combine the features output from the RoI-pooling layer with the information from the classification and regression layers to achieve accurate classification and localization of defects. The classification layer enables the model to distinguish whether or not defects are present in candidate regions, while the regression layer is responsible for fine-tuning the boundaries of these regions to ensure high accuracy in defect localization.

Improved faster R-CNN

Efficient channel attention model

The squeeze and excite network (SENet) [29] is pioneering work that applies attentional mechanisms to channel dimensionality. This approach fundamentally enhances the representational capabilities of deep convolutional neural networks by explicitly modeling the interdependencies between channels. However, the feature transformation in SENet relies on FC layers, which may lead to the loss of some feature expressiveness in the process of dimensionality reduction and expansion. To address this problem, Wang et al. [30] introduced the efficient channel attention (ECA) mechanism, which improved the feature transformation aspect of SENet. It circumvented the FC method, preserving the advantages of SENet while reducing the model complexity and computational cost, thereby refining the channel interactions.

This work proposes an innovative design utilizing the ECA mechanism to enhance Faster R-CNN's capability for defect detection in infrared images of ancient polyptychs. By replacing the FC layer with a one-dimensional convolutional (Conv1D) layer, the ECA mechanism enables the local interactive capture of each channel and its K neighboring channels. This approach not only reduces the number of the model parameters but also enhances the model's sensitivity to the local channel information, thereby improving the localization accuracy of defects in infrared thermal images. Consequently, defect detections in the complex infrared thermal images becomes more accurate and efficient.

The ECA mechanism's enhancement of defect detection capability, achieved through local interactive capture, can be broken down into the following steps:

1. Adaptive selection of convolution kernel size

To improve the overall accuracy of defect detection in infrared thermal images, it is necessary to appropriately adjust the kernel size, denoted as K. This value can be dynamically calculated as a function of the total number of channels, denoted as C, that is,

$$C = \phi(K) \tag{1}$$

The most straightforward mapping method is to use a linear function:

$$\phi(K) = \gamma * K - b \tag{2}$$

where the parameter γ in Eq. (2) is used to adjust the scaling of the logarithmic function, which affects the sensitivity between the size of the convolutional kernel and the number of channels. The translation factor b is used to fine-tune the kernel size obtained from the logarithmic function K.However, the number of channels C is generally set to a power of 2. Therefore, with the number of channels C automatically known, the value of the kernel size K can be determined by Eq. (3). Specifically, the function $\varphi(C)$, which is associated with the channel-related dimension C, is determined as follows:

$$K = \varphi(C) = \left| \frac{\log_2 C}{\gamma} + b \right|_{\text{odd}} \tag{3}$$

In the equation, $|\bullet|_{\text{odd}}$ denotes that the closest odd integer to the calculated is taken. In Eq. (3), the result is ensured to be odd by scaling and translating the logarithm of the C and then ensuring that the result is odd for different convolutional neural networks (CNN) structures.

2. Generation of channel attention weights

Subsequently, channel descriptors are derived through the application of global average pooling (GAP) on the input feature maps. This process transforms a feature map of dimensions $H \times W \times C$ into a $1 \times 1 \times C$ vector. Following this, the vector is subjected to Conv1D, facilitating the computation of the final attention weight, denoted as α_i.

Based on the above specific analysis, the ECA mechanism proposed in this study uses Conv1D to complete the information interaction between channels:

$$\alpha_i = \sigma\left(\text{Conv1D}_K(y)\right) \tag{4}$$

where Conv1D related only to the K, which determines the extent of the local interaction captured by the convolution; y denotes the feature value associated with the given channel; $\sigma(\bullet)$ represents the sigmoid nonlinear activation function.

Thus, Eq. (5) can be derived from the computational formula associated with Conv1D:

$$\alpha_i = \sigma\left(\sum_{j=1}^{K} \omega_i^j y_i^j\right), y_i^j \varepsilon \Omega_i^K \tag{5}$$

Where Ω_i^K denotes the set of the i^{th} channel and its K neighboring channels; y_i^j denotes the feature of the

j^{th} neighboring channel of the i^{th} channel; ω_i^j denotes the learning parameter of the j^{th} neighboring channel of the i^{th} channel [31]; The final attention weight for the i^{th} channel, denoted as α_i, is obtained by applying a weighted sum to the feature values y_i^j of the channel and its K neighboring channels, using the weight parameters ω_i^j. Subsequently, the sigmoid function $\sigma(\bullet)$ is employed to activate and compute the resulting values.

3. Application of channel attention weights

Ultimately, the channel attention weights obtained (α_i) are re-integrated into the initial input feature map (X) via a channel-specific multiplication process, thereby facilitating the recalibration of features:

$$X' = X \otimes \alpha_i \tag{6}$$

In this context, X' represents the resulting feature map. The α_i is effectively mapping the attention weights within the interval from 0 to 1. Furthermore, symbol $\otimes$ is employed to denote a multiplication operation that is conducted across channels.

In defect detection, the ECA mechanism calculates channels with high thermal contrast between defects and the background, assigning higher weights to these channels based on the aforementioned formulas. To illustrate this process, a frame from the infrared image dataset was used, as shown in Fig. 7. Here, the capital letter X represents the feature map extracted by the VGG16 network, whereas $X\prime$ denotes the re-weighted feature map obtained through the ECA mechanism. By comparing the visualizations of X and X' in Fig. 7b, it is evident that the ECA mechanism effectively emphasizes channels critical for defect detection, thereby enhancing detection accuracy. This viewpoint is further corroborated in Fig. 7c, where the X' proposals more accurately highlight the defects, as evidenced by their closer match with the sample image showing the ground truths of the defects, compared to the proposals derived from X.

Improved faster R-CNN base on ECA mechanism

In the context of the Faster R-CNN- framework, the integration of the ECA mechanism enhances the overall architecture, as illustrated in Fig. 8.

In the improved Faster R-CNN model in Fig. 8, the VGG16 network is used as a feature extractor (cf. Figure 8) to extract features from the input image. These features are then weighted by an ECA module, which adjusts them according to the importance of the features, emphasizing the feature channels that are more critical to the task of defect detection in infrared thermal images.

The weighted feature maps are fed into the RPN, which is a FC network that slides over the feature maps and proposes target regions that may contain defects. The extracted information is generated based on the probability of target presence and bounding box location information. The extracted information is then passed through the RoI-pooling layer (see Fig. 8) to ensure that all proposal areas are of the same size and provide standardized inputs for subsequent steps.

In the final step, the RoI-pooling features are passed to the FC layers, including a SoftMax classifier and a bounding box regressor (cf. Figure 8). The classifier is responsible for recognizing the class of each proposal region, whereas the regressor adjusts the position and size of the proposal regions to accurately cover each defect. Such

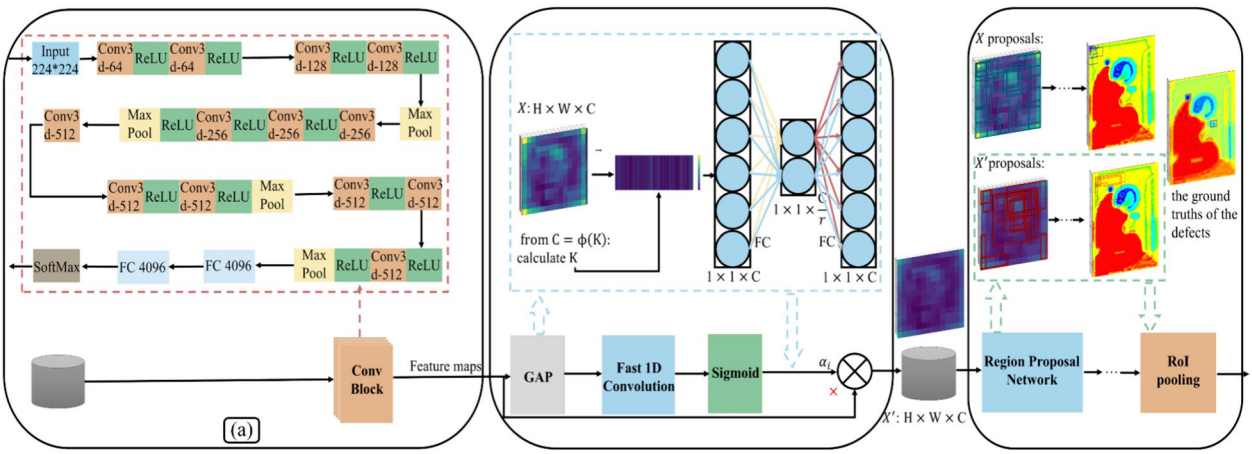


Fig. 7 **a** Original image input forward VGG16 network structure, **b** ECA module structure diagram, visualizing input feature map X and re-weighted output feature map X′ **c** comparing the visualizations of X and X′ proposals

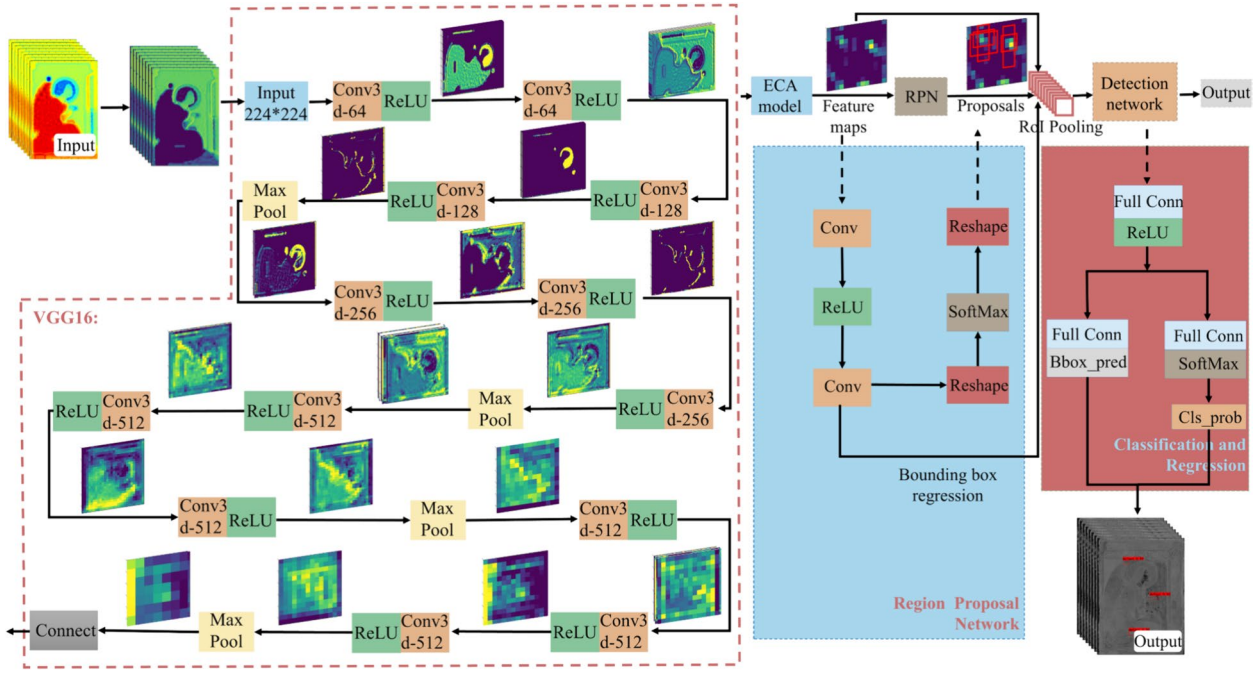


Fig. 8 Structural diagram of the improved Faster R-CNN network model

structural improvements, especially through the introduction of the ECA mechanism, make the model more accurate and efficient in defect detection.

To help readers better understand the convolution operations, the authors have visualized the convolutional process using input images from the training stage, as shown in Fig. 9:

Experimental results and analysis

In order to verify the effectiveness of the Faster R-CNN-ECA model for detecting defects in polyptychs, this study experimentally analyzes infrared thermal images collected on a realistic sample and compares the detection performance of the improved Faster R-CNN model with that of the traditional Faster R-CNN model. Following this, the comparison experiments were conducted on an identical comprehensive home-made dataset.

Experimental environment

These experiments were conducted on Windows 10 operating system and the deep learning framework used was PyTorch. For hardware, the experiments used Intel Xeon E5-2670 CPUs and NVIDIA GTX 3060 GPUs, as well as hard disks of 2 TB capacity [32]. The defect detection and analysis experiments followed a couple of open-source code and mechanisms, including the ECA mechanism [30] and Faster R-CNN network structure [27].

Dataset and parameterization

In this experiment, thermal images of polyptych samples were collected using an infrared thermal camera. The infrared camera employed is the FLIR X8501sc with a resolution of 1280×1024 pixels, utilizing an InSb detector and operating in the 3–5 μm mid-wave infrared range. For samples A and B, five sets of valid infrared datasets were collected respectively. Each set contains 1000 frames of infrared thermal images, covering the heating phase, cooling phase, and other relevant stages. From these datasets, selected the most representative 28 thermal images of the polyptych samples for model testing. The experimental setup and the actual photograph are shown in Fig. 10.

The limited variety of defect types and locations in the experimental samples, combined with the potential risk of secondary damage to the artworks from repeated IRT experiments, necessitates the use of numerical simulations to expand the dataset. Numerical simulations were conducted using the geometric model described in Sect. 2.2 with the aim of expanding the dataset. Specific details on parameter settings for numerical modeling using COMSOL Multiphysics software, the reader is referred to [35].

To enhance the generalizability and robustness of the experiment, the defective infrared thermal images obtained from numerical simulation were therefore

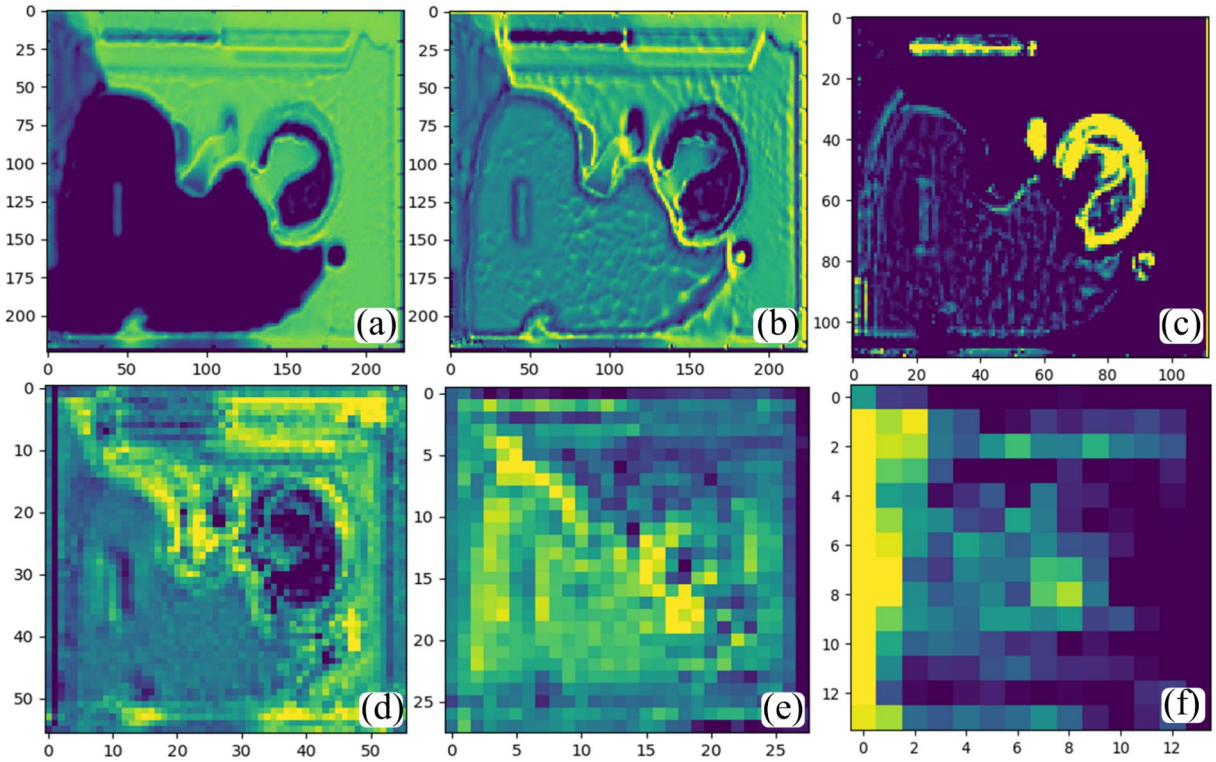


Fig. 9 Input image for convolution operation in training phase: **a** Input image (256*256 dimensions), **b** Performs 64 3*3 convolutional kernel operations (256*256 dimensions), **c** Performs 128 3*3 convolutional kernel operations (128*128 dimensions), **d** Performs 256 3*3 convolutional kernel operations (64*64 dimensions), **e** Performs 512 3*3 convolutional kernel operations (28*28 dimensions), **f** Performs 512 3*3 convolutional kernel operations (14*14 dimensions)

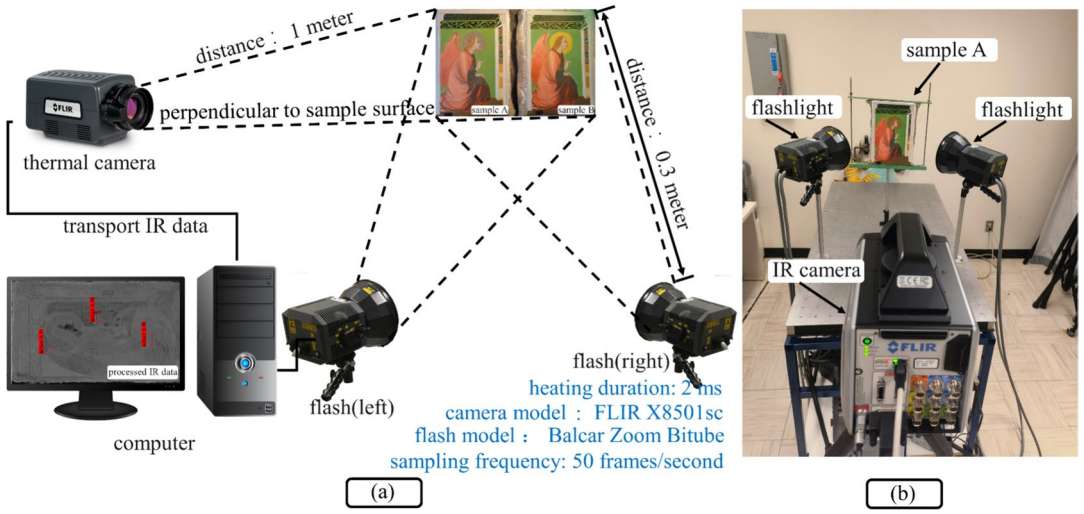


Fig. 10 **a** Schematic diagram of the infrared thermal imaging experimental acquisition device with detailed equipment information and distances, **b** the actual Photograph of the experimental setup

scaled and distorted, and the number of infrared thermal images in the dataset was increased to 784. Therefore, there are 812 infrared thermal images in the experimental dataset, of which 28 are homemade datasets collected by infrared camera and 784 are numerically simulated infrared thermal images by using COMSOL Multiphysics software. Among the 812 images, 90% is used as training set and remaining 10% is used for testing set.

In response to concerns regarding the disproportionate number of simulated images, it should be noted that the inclusion of a substantial number of high-fidelity simulated images is essential. This strategy allows the model to be exposed to a diverse array of defect types and conditions, thereby enhancing its generalization and robustness.

The training optimization strategy of the model uses the stochastic gradient descent (SGD) method, and VGG16 is chosen for the backbone network. The specific settings of each parameter in the experiment are detailed in Table 1.

Assessment of indicators

To evaluate the detection capability of the model, the following key metrics are used: average precision (AP), average recall rate (AR), time required for detection, and the model's memory consumption [33]. Together, these metrics reflect the model's performance in terms of accuracy, efficiency, and resource utilization, providing a comprehensive basis for evaluating the model's performance. The explanation of the model evaluation metrics is detailed in Table 2.

Comparison of experimental results

To further enhance the accuracy of the defect detection study of polyptychs, this study introduces the ECA mechanism model, which enhances the model's ability to identify and localize defects. This measure helps to improve the performance of the model in the defect detection task.

To observe the differences of the two models more intuitively, experiments were conducted on the same test dataset using both the original Faster R-CNN model and the improved Faster R-CNN with both models employing VGG16 as backbone network. The relevant performance metrics were measured, and the detailed values are provided in Table 3.

Table 1 Parameterizations of the improved Faster R-CNN based ECA mechanism

Parameter name	**Total training steps**	**Basic learning rate**
Value	37000	0.0001
Parameter name	**Parameters** ϵ	**Impulse setting**
Value	10^{-8}	0.9

Table 2 Explanation of the meaning of the model evaluation indicators

Indicators	**Meaning**
AP_{50}	When the Intersection over Union (IoU) threshold is set to 0.5, measures the average accuracy of the model in to identify the targets [34]
AP_{75}	When the IoU threshold is set to 0.75, measures the average accuracy of the model to identify the targets
AP_S	The average precision for small objects (less than 32^2 pixels in area), measures the model's capacity to identify small targets
AP_L	The average precision foy large objects (more than 96^2 pixels in area), measures of the model's capacity to identify large targets
AR_1	The average recall when processing one image at a time, indicates the proportion of the number of targets detected by the model from a single image to the actual number of targets
AR_{10}	The average recall when processing 10 images at a time, indicates the proportion of targets detected by the model from 10 images to the actual number of targets
AR_{100}	The average recall rate when processing 100 images at a time, indicates the proportion of the number of targets detected by the model from 100 images to the actual number of targets
Memory consumption	The amount of memory used by the model during the detection process

Table 3 Comparison of model testing performance metrics

Model	AP_{50}	AP_{75}	AP_S	AP_L	AR_1	AR_{10}	AR_{100}	**Detection time**	**Memory consumption**
Faster R-CNN	81.6	70.2	46.3	64.7	50.1	54.8	56.4	0.31	3200–7000
Improved modelling	87.3	72.2	54.8	66.5	52.2	57.7	57.8	0.33	3400–7500

AP is in percent; time is in seconds; memory consumption is in MB

Results and discussion

To provide readers with a more intuitive visual comparison of the differences between the two models, four representative infrared thermal frames from the cooling phase of the IRT experiment were selected and detected using both models. The experimental results are shown in Fig. 11: The images in each row are, in order, the infrared thermal image of experimental sample A, the infrared thermal image of sample A after a total-variation regularized low-rank tensor decomposition (LRTDTV) de-noise and Fourier transform, the infrared thermal image of experimental sample B, and the infrared thermal image of sample B after LRTDTV de-noise and Fourier transform.

Among them, the infrared thermal image in Fig. 11b is obtained from sample A after LRTDTV noise removal and Fourier transform, and Fig. 11d is obtained from

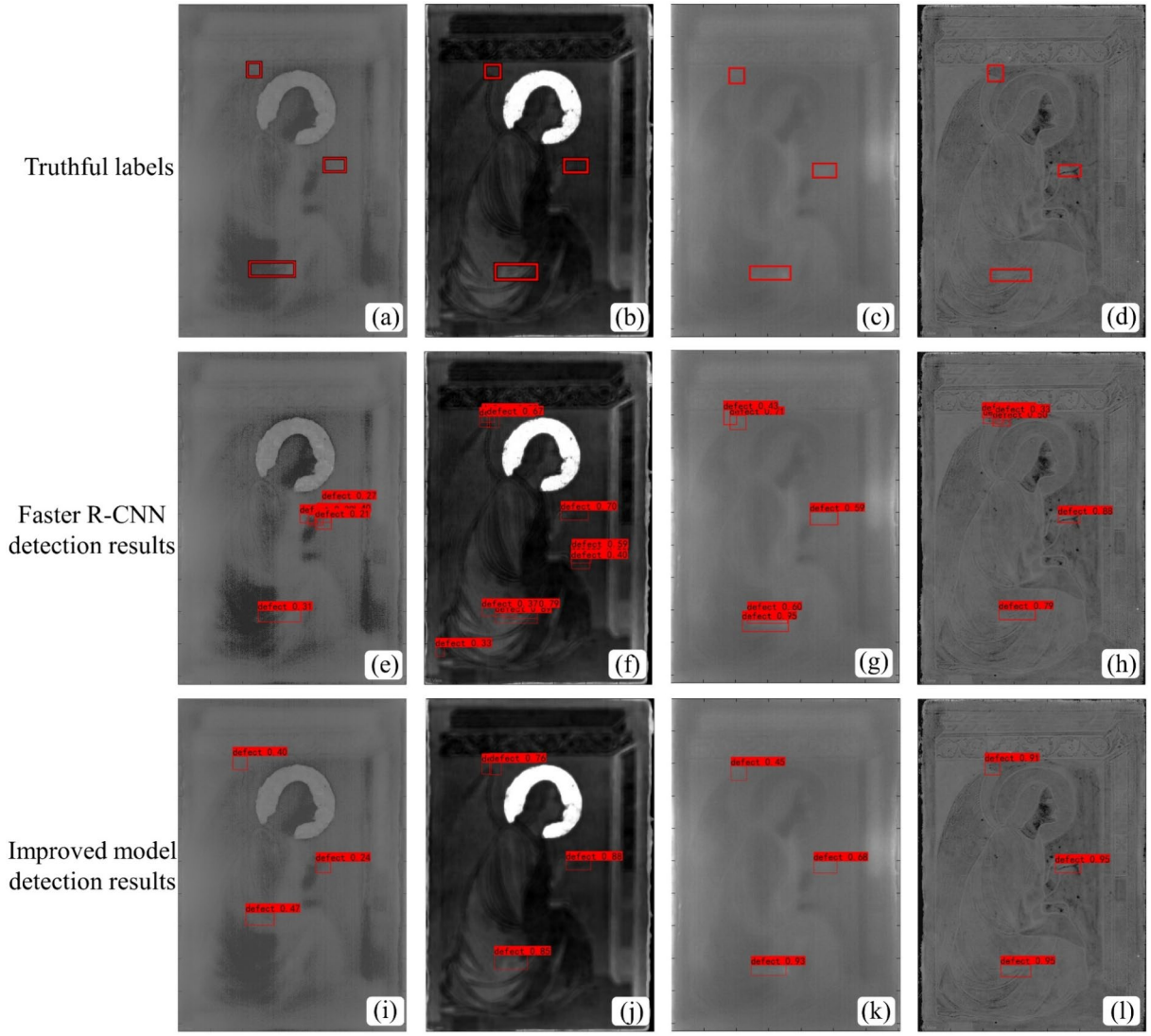


Fig. 11 Results of the IRT experiments: **a** the original thermal image of sample A marking three real labels, **b** the thermal image of sample A after processing and marking three real labels, **c** the original thermal image of sample B marking three real labels, **d** the thermal image of sample B after processing and marking three real labels, **e** the original thermal image of sample A at the 4.96 s moment after applying the Faster R-CNN model, **f** the thermal image of sample A after processing and applying the Faster R-CNN model, **g** the original thermal image at the 5.88 s moment of sample B after applying the Faster R-CNN model, **h** the thermal image of sample B after applying processing and the Faster R-CNN model, **i** the original thermal image of sample A after applying the improved model, **j** the thermal image of sample A after processing and applying the improved model, **k** the original thermal image of sample B after applying the improved model and **l** the thermal image of sample B after processing and applying the improved model. (The term of processing refers to LRTDTV de-noising and Fourier transformation; the improved model is the improved Faster R-CNN model base on ECA mechanism.)

sample B for the same image processing. After the LRTDTV denoising process, the Fourier transform is applied to optimize the contrast and sharpness of the image even more. The specific implementation about the aforementioned algorithm, the reader can refer to [35].

By comparing Fig. 11b, Fig. 11f, j, it can be found that the constructed automatic defect detection network based on the improved Faster R-CNN can be better detected for the tiny defects that are difficult to be found by human eyes.

By comparing the first set: Figs. 11c, d, the second set: Figs. 11g, k, and the third set: Figs. 11h, l, it can be observed that the noise interference in infrared thermal imaging is better removed after applying the LRTDTV denoising method. More data supporting this viewpoint can be found in [35].

Conclusion

In this study, an improved Faster R-CNN model integrated with the ECA mechanism is proposed, aiming at detect potential defects in infrared thermal images at high-resolution acquired on ancient polyptychs. By integrating the ECA mechanism into the VGG16 architecture, this model significantly improves the recognition accuracy of tiny defects, which is experimentally validated and demonstrates its remarkable efficacy in improving the detection accuracy.

1. This study innovatively combines the VGG16 network with the ECA mechanism to optimize the model for the detection of subtle defects in cultural heritage via IRT technique, which effectively improves the accuracy of defect detection in polyptychs.
2. The introduced ECA mechanism optimizes the feature extraction process and enhances the model's ability to identify and characterize the defective features of polyptychs, thus improving the model's sensitivity to and speed of identification of small defects.
3. This method combines IRT, numerical simulation and image de-noise technique to construct an improved automatic defect detection network, which is capable of accurately identifying surface and subsurface defects in artifacts, while maximizing their preservation. The visualization of the thermal images is improved by the LRTDTV image denoising step, which enables the detection of imperceptible damages and defects in cultural heritage samples.

To summarize, the improved Faster R-CNN model combined with the ECA mechanism not only significantly improves the accuracy of IRT defect detection for polyptychs, but also maintains rapid processing speed without substantially increasing the computational cost. These results are valuable for the real-time cultural heritage preservation and automated defect recognition system, offering a novel technical strategy for quality monitoring and safety assessment in the field of cultural heritage. Future research will focus on extending this method to a wider range of cultural heritage artifacts to evaluate its robustness and generalizability, as well as exploring other deep learning models to further enhance detection performance.

Acknowledgements

This work is supported by the Ministry of Science and Technology of China (MOST) through the National Key R&D Program (Grant No. 2023YFE0197800). This work is also supported by the Ministry of University and Research of Italy (Grant No. PGR02016).

Author contributions

X.W. and J.H. analyzed the data and provided the algorithm. G.J., Y.G., S.S. and H.Z. supervised the research. S.S. and M.M. prepared the samples. H.Z. conducted the experiments. X.W. wrote the original manuscript. S.S., D.K., D.Y., N.A., H.F., X.M. and H.Z. revised the manuscript.

Funding

National Key Research and Development Program of China: 2023YFE0197800. Italian Ministry Affairs and International Cooperation: PGR02016.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China. [2]Centre for Composite Materials and Structures (CCMS), Harbin Institute of Technology, Harbin 150001, China. [3]Department of Industrial and Information Engineering and Economics, University of L'Aquila, Monteluco di Roio, 67100 L'Aquila, (AQ), Italy. [4]Professional Restorer, Via Muranuove 64, 67043 Celano, AQ, Italy. [5]Department of Archival, Library and Information Studies, University of West Attica, Egaleo, Greece. [6]School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China. [7]Integrated Vehicle Health Management Centre (IVHM), Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield, Bedford MK43 0AL, UK. [8]Faculty of Computing, Federal University of Uberlandia, Uberlandia, 38.408-100, Santa Monica, Brazil. [9]Computer Vision and System Laboratory (CVSL), Department of Electrical and Computer Engineering, Laval University, Quebec City G1V 0A6, Canada.

Received: 31 March 2024 Accepted: 3 September 2024
Published online: 16 September 2024

References

1. Fragasso L, Masini N. Postprocessing of infrared Reflectography to support the study of a painting: the case of Vivarini's Polyptych. Int J Geophys. 2011. https://doi.org/10.1155/2011/738279.

2. Grifoni E, Vannini E, Lunghi I, Faraioli P, Ginanni M, Santacesarea A, et al. 3D multi-modal point clouds data fusion for metrological analysis and restoration assessment of a panel painting. J Cult Herit. 2024;66:356–66. https://doi.org/10.1016/j.culher.2023.12.007.
3. Mix PE. Introduction to nondestructive testing: a training guide. Sons: John Wiley; 2005.
4. Pehlivan GF. Condition and characterization analysis of a twentieth century cultural heritage through non-destructive testing (NDT) methods: the case of the Sivas industry school ironworking atelier in Turkey. Herit Sci. 2023;11:1–19. https://doi.org/10.1186/s40494-023-00889-5.
5. Hu J, Zhang H, Sfarra S, Pivarčiová E, Yao Y, Duan Y, et al. Autonomous dynamic line-scan continuous-wave terahertz non-destructive inspection system combined with unsupervised exposure fusion. NDT E Int. 2022;132: 102705. https://doi.org/10.1016/j.ndteint.2022.102705.
6. Pavlidis G, Koutsoudis A, Arnaoutoglou F, Tsioukas V, Chamzas C. Methods for 3D digitization of cultural heritage. J Cult Herit. 2007;8:93–8. https://doi.org/10.1016/j.culher.2006.10.007.
7. Reinhardt J, Tischer M, Schmid S, Kollofrath J, Burger R, Jatzlau P, et al. X-ray-based examination of artworks by Cy Twombly: art technology and condition of the "Original Sculptures." Herit Sci. 2023;11:1–11. https://doi.org/10.1186/s40494-023-01073-5.
8. Cosentino A. Identification of pigments by multispectral imaging; a flowchart method. Herit Sci. 2014;2:1–12. https://doi.org/10.1186/2050-7445-2-8.
9. Li X, Li J, Li Y, Ozcan A, Jarrahi M. High-throughput terahertz imaging: progress and challenges. Light Sci Appl. 2023;12:233. https://doi.org/10.1038/s41377-023-01278-0.
10. Bodnar JL, Mouhoubi K, Vallet JM. Examples of SVD decomposition contributions to the non-destructive testing of cultural heritage mural paintings using stimulated infrared thermography. Eur Phys J Appl Phys. 2022. https://doi.org/10.1051/epjap/2022220088.
11. Paoloni S, Orazi N, Zammit U, Bison P, Mercuri F. A note on the early thermographic approaches for the investigation of the Cultural Heritage. Quant Infrared Thermogr J. 2023. https://doi.org/10.1080/17686733.2023.2243575.
12. Liu Y, Wang F, Liu K, Mostacci M, Yao Y, Sfarra S. Deep convolutional autoencoder thermography for artwork defect detection. Quant Infrared Thermogr J. 2023. https://doi.org/10.1080/17686733.2023.2225246.
13. Melada J, Arosio P, Gargano M, Ludwig N. Automatic thermograms segmentation, preliminary insight into spilling drop test. Quant Infrared Thermogr J. 2023. https://doi.org/10.1080/17686733.2023.2213555.
14. Vavilov VP, Bison PG, Burleigh DD. Ermanno Grinzato's contribution to infrared diagnostics and nondestructive testing: in memory of an outstanding researcher. Quant Infrared Thermogr J. 2023. https://doi.org/10.1080/17686733.2023.2170647.
15. Liu K, Huang K-L, Sfarra S, Yang J, Yi L, Yuan Y. Factor analysis thermography for defect detection of panel paintings. Quant Infrared Thermogr J. 2023;20:25–37. https://doi.org/10.1080/17686733.2021.2019658.
16. Mouhoubi K, Detalle V, Vallet J-M, Bodnar J-L. Improvement of the non-destructive testing of heritage mural paintings using stimulated infrared thermography and frequency image processing. J Imaging Sci Technol. 2019. https://doi.org/10.3390/jimaging5090072.
17. Ovadia M, Brook A. ADAPTION OF INDUSTRIAL NDT PROTOCOLS BASED ON ACTIVE INFRARED THERMOGRAPHY TO THE ART CONSERVATION WORLD: THE CASE OF THE WALL PAINTING AT HERODIUM. 29th CIPA Symposium "Documenting, Understanding, Preserving Cultural Heritage Humanities and Digital Technologies for Shaping the Future" - 25–30 June 2023, Florence, Italy. Copernicus GmbH; 2023. p. 223–8. https://doi.org/10.5194/isprs-annals-X-M-1-2023-223-2023.
18. Attas M, Cloutis E, Collins C, Goltz D, Majzels C, Mansfield JR, et al. Near-infrared spectroscopic imaging in art conservation: investigation of drawing constituents. J Cult Herit. 2003;4:127–36. https://doi.org/10.1016/S1296-2074(03)00024-4.
19. Delaney JK, Thoury M, Zeibel JG, Ricciardi P, Morales KM, Dooley KA. Visible and infrared imaging spectroscopy of paintings and improved reflectography. Herit Sci. 2016. https://doi.org/10.1186/s40494-016-0075-4.
20. Pappalardo G, Mineo S, Caliò D, Bognandi A. Evaluation of natural stone weathering in heritage building by infrared thermography. Herit Rev. 2022;5:2594–614. https://doi.org/10.3390/heritage5030135.
21. Bodnar JL, Candoré JC, Nicolas JL, Szatanik G, Detalle V, Vallet JM. Stimulated infrared thermography applied to help restoring mural paintings. NDT E Int. 2012;49:40–6. https://doi.org/10.1016/j.ndteint.2012.03.007.
22. Candoré JC, Bodnar JL, Detalle V, Grossel P. Non-destructive testing of works of art by stimulated infrared thermography. Eur Phys J Appl Phys. 2012;57:21002. https://doi.org/10.1051/epjap/2011110266.
23. Dritsa V, Orazi N, Yao Y, Paoloni S, Koui M, Sfarra S. Thermographic imaging in cultural heritage: a short review. Sensors. 2022. https://doi.org/10.3390/s22239076.
24. Hu C, Dong Y, Xia G, Liu X. An automatic detection method of the mural shedding disease using YOLOv4. International Conference on Environmental Remote Sensing and Big Data (ERSBD 2021). SPIE; 2021. p. 183–92. https://doi.org/10.1117/12.2625707.
25. Hatır E, Korkanç M, Schachner A, İnce 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. https://doi.org/10.1016/j.culher.2021.07.004.
26. Niannian W, Xuefeng Z, Linan W, Zheng Z. Novel system for rapid investigation and damage detection in cultural heritage conservation based on deep learning. J Infrastruct Syst. 2019;25:04019020. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000499.
27. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2017;39:1137–49. https://doi.org/10.1109/TPAMI.2016.2577031.
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2016. p. 770–778.https://doi.org/10.1109/CVPR.2016.90.
29. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2018. p. 7132–41. https://doi.org/10.1109/CVPR.2018.00745.
30. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2020. p. 11531–9. https://doi.org/10.1109/CVPR42600.2020.01155.
31. Zhu L, Geng X, Li Z, Liu C. Improving YOLOv5 with attention mechanism for detecting boulders from planetary images. Remote Sens. 2021;13:3776. https://doi.org/10.3390/rs13183776.
32. Hurtik P, Molek V, Hula J, Vajgl M, Vlasanek P, Nejezchleba T. Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLOv3. Neural Comput Appl. 2022;34:8275–90. https://doi.org/10.1007/s00521-021-05978-9.
33. Hsu W-Y, Lin W-Y. Ratio-and-scale-aware YOLO for pedestrian detection. IEEE Trans Image Process. 2021;30:934–47. https://doi.org/10.1109/TIP.2020.3039574.
34. Zhu L, Xie Z, Liu L, Tao B, Tao W. IoU-uniform R-CNN: breaking through the limitations of RPN. Pattern Recognit. 2021;112: 107816. https://doi.org/10.1016/j.patcog.2021.107816.
35. Jiang G, Wang X, Hu J, Wang Y, Li X, Yang D, et al. Simulation-aided infrared thermography with decomposition-based noise reduction for detecting defects in ancient polyptychs. Herit Sci. 2023;11:223. https://doi.org/10.1186/s40494-023-01040-0.

Publisher's Note

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



https://dspace.org/

Automatic defect detection in infrared thermal images of ancient polyptychs based on numerical simulation and a new efficient channel attention mechanism aided Faster R-CNN model

Wang, Xin; Jiang, Guimin; Hu, Jue; Sfarra, Stefano; Mostacci, Miranda ...

2024-09-16
Springer
https://doi.org/10.1186/s40494-024-01441-9
Wang X, Jiang G, Hu J, et al., (2024) Automatic defect detection in infrared thermal images of ancient polyptychs based on numerical simulation and a new efficient channel attention mechanism aided Faster R-CNN model. Heritage Science, Volume 12, September 2024, Article number 329

Downloaded from a DSpace repository. https://dspace.org/