

Research Article

Xiaoli Fu and Niwat Angkawisittpan*

Detecting surface defects of heritage buildings based on deep learning

<https://doi.org/10.1515/jisys-2023-0048>

received April 11, 2023; accepted December 13, 2023

Abstract: The present study examined the usage of deep convolutional neural networks (DCNNs) for the classification, segmentation, and detection of the images of surface defects in heritage buildings. A survey was conducted on the building surface defects in Gulang Island (a UNESCO World Cultural Heritage Site), which were subsequently classified into six categories according to relevant standards. A Swin Transformer- and YOLOv5-based model was built for the automated detection of surface defects. Experimental results suggested that the proposed model was 99.2% accurate at classifying plant penetration and achieved a mean intersection-over-union (mIoU) of over 92% in relation to moss, cracking, alkalization, staining, and deterioration, outperforming CNN-based semantic segmentation networks such as FCN, PSPNet, and DeepLabv3plus. The Swin Transformer-based approach for the segmentation of building surface defect images achieved the highest accuracy regardless of the evaluation metric (with an mIoU of 90.96% and an mAcc of 95.78%), when contrasted to mainstream DCNNs such as SegFormer, PSPNet, and DANet.

Keywords: deep learning, flipping, historical places, images, sustainability, thermography, transformer-based segmentation

1 Introduction

Heritage buildings are valuable historic structures that play an increasingly important role in reflecting history, promoting cultural inheritance, and displaying values, which makes their conservation attract increasing attention from the international community. Nevertheless, the situation of heritage building conservation in China looks grim. Statistics from the National Cultural Heritage Administration of the People's Republic of China (PRC) reveal that, over the recent 5 years, hundreds of projects have been submitted for approval which are related to the retrofit of heritage buildings on a national scale. The Ministry of Finance of the PRC invested a total of RMB six billion in the conservation and restoration of heritage buildings during the “12th Five-Year Plan” period alone. To date, heritage buildings of the first six batches have all experienced a round of restoration. It is demonstrated that the conservation and restoration of heritage buildings remain a frequent issue for which heavy investment is indispensable.

Building surface, the interface between the interior and exterior of a structure, retains the maximum possible degree of authenticity that each heritage conservation charter requires to preserve. It not only highlights the history of the architecture itself, but more importantly, it allows people to intuitively recognize the category and degree of a defect in a heritage masonry building and determine the corresponding

* **Corresponding author: Niwat Angkawisittpan**, Research Unit for Electrical and Computer Engineering Technology (RECENT), Faculty of Engineering, Mahasarakham University, Maha Sarakham, 44150, Thailand, e-mail: niwat.a@msu.ac.th, tel: +66-430-29665, fax: +66-437-54316

Xiaoli Fu: Department of Architecture, Xiamen Institute of Technology, 1251 SunBan South Road, Jimei District, Xiamen, 361021, Fujian Province, China, e-mail: 64010363005@o365.msu.ac.th

restoration required. Thus, the detection of building surface defects is a priority of architectural heritage survey and defect diagnosis. Traditional detection methods prove too lengthy and laborious to systematically analyze building surface defects effectively. Moreover, some spaces (e.g., roofs) are inaccessible for detection due to limitations posed by site conditions, potentially endangering the safety of surveyors [1].

For the past few years, with the development of both digital image processing algorithms and traditional machine learning algorithms, computer vision (CV) analysis and deep learning (DL) among other technologies have come into use in assessing the structural health of a building. Relevant research has covered the following scopes: detecting defects on the interior surfaces of buildings through CNN [2]; monitoring structural health through DL [3–5]; recognizing and evaluating cracks and spalls by applying Gaussian regression and support vector machine [6–8]; and detecting masonry wall defects through the logistic regression and point cloud [9,10]. A number of papers have also explored the classification of architectural heritage images [11–17]. However, very few of the aforementioned studies specifically examined defect image recognition on heritage building surfaces, while existing research methods could not meet the practical needs of detecting multiple defects on building surfaces. The major proposal of this disquisition is summarized as follows:

1. This study classifies the defects on heritage building surfaces into six categories as per their features such as plant penetration, moss, cracking, alkalization, staining, and deterioration.
2. This study proposes a model to recognize the images of surface defects in heritage building using the DL models of YOLOv5 and Swin Transformer with environmental differences affecting such defects that are considered in this research study.
3. Application tests are implemented to validate the feasibility and accuracy of the assumed model for detecting surface defects in heritage buildings.

The structure of the current work is organized as follows: Section 2 gives a retrospect of the interrelated studies on the research topic. Section 3 presents materials and methods of this study. Section 4 is focused on comparative tests and their results; Section 5 gives an illustration of practical applications of the proposed research, and Section 6 concludes the results and discussion of the proposed research.

2 Related work

The use of DL in the context of heritage building conservation has emerged as an essential area of research and innovation. This section presents a comprehensive investigation of the application of conventional and state-of-the-art DL-based approaches for classifying, segmenting, and detecting surface defects in heritage buildings. A comprehensive investigation of the literature review is summarized in Table 1.

2.1 Conventional approaches

To monitor historical buildings, IR thermography has been widely used for two decades. Hidden structures of walls, moisture status, and finishing status were investigated by using IR thermography [18]. Moreover, the same technique was applied to measure the porous material. The IR thermography technique is more suitable for investigating conserved, repaired, and restored structures [19]. Heritage as a latent structure of buildings requires conservation and preservation. Several social and natural factors have serious threats to deteriorate and damage the origin of the buildings. To make sure that heritage buildings are conserved and preserved, their visual inspection is of greater importance. Conventional practices are based on manual inspection that takes a lot of time and resources [20]. An innovative technique may replace manual inspection by using less human resources and much faster than conventional techniques. For tangible heritage conservation, planning practices were recommended for several digital technology companies [21].

Table 1: Summary of comprehensive investigation of literature review

Author	Approach	Major findings	Pros	Cons
Conventional approaches				
Lerones et al. [22]	3D Laser scanner surveying	Automatic moisture detection and area assessment in historic buildings	Non-intrusive, covers large areas quickly, does not interact with materials	Requires 3D laser scanner equipment, limited to visual assessment
Tavukçuoğlu [23]	QIRT and UPV measurements	<i>In situ</i> evaluation and monitoring of historical and contemporary structures	Useful for various diagnostics, non-destructive, supports material assessment	Need for proper equipment, data acquisition complexity
Biaszczak-Bąk et al. [24]	TLS	Detection and segmentation of wall defects in unlit sites using OptD method	Realistic visualization without external lighting, high defect detection effectiveness	Equipment-dependent, limited to unlit sites
Wong [25]	Various NDT techniques	Application of NDT techniques for identifying defects and moisture detection in historic buildings	Complements destructive tests, supports defect and moisture detection	Requires sound sampling plan, equipment, and expertise
Radnić et al. [26]	Restoration and strengthening	Case study on restoration and strengthening of historical masonry buildings	Provides insights into strengthening and restoration of similar structures	Dependent on specific case study, may not apply comprehensively
Biaszczak-Bąk et al. [27]	TLS	Automatic wall defect detection in unlit sites using modified OptD method	Reliable and convenient for data analysis, cost estimation for repairs	Equipment-dependent, limited to unlit sites
Wu et al. [28]	Rail surface defect detection	RBGNNet for accurate rail surface detection using DL	Addresses blurry rail edges, innovative hybrid loss, high detection rate	Limited accuracy in existing systems, equipment-dependent
Wood and Mohammadi [29]	Geometric Features in Point Cloud	Detecting surface damage and cracks in fresco walls using geometric features	Relies on geometric features, not influenced by color, non-destructive	Limited to geometric damage, requires specific equipment
Al-Sakkaf et al. [30]	GPR in Heritage Buildings	Review of methodologies for defect detection in heritage buildings using GPR	Non-invasive, effective for stone masonry structures, identifies common methodologies	Equipment-dependent, may not work for all heritage structures
Deep learning based approaches				
Perez et al. [2]	Convolutional neural networks (CNN)	Automated detection and localization of building defects from images	Fast, potentially scalable, works with mobile devices and drones	Challenges in real-life applications, model limitations
Wang et al. [35]	Faster R-CNN model	Automatic detection of efflorescence and spalling damage in historic masonry structures	High accuracy, real-time detection with webcams and smartphones	Limited to specific damage types, need for specialized equipment
Wenlong et al. [36]	Deep learning (ASPP module)	Bridge surface structural damage detection using DL	Addresses labeling challenges, improves minority class detection	Dependent on dataset quality, may require skilled professionals
Ye and Sun [37]	Machine vision	Surface defect detection in ceramic tableware using machine vision	Simple and high-precision, can be improved with deep neural networks	Potential for algorithm improvement, equipment-dependent
Stephen et al. [38]	CNN for surface defect classification	CNN-based method for surface defect classification in tile surface images	Efficient classification, potential for automated visual inspections	Equipment-dependent, may require optimization
Teng et al. [39]	YOLOv3 for bridge surface defect detection	Improved YOLOv3 for real-time bridge surface defect detection	High detection effect, real-time detection, anti-noise capabilities	Equipment-dependent, optimization required
Shao et al. [40]	Damage detection in concrete buildings	Two-stage method using point clouds and 3D neural network for damage detection	Effective damage detection, acceptable performance on aging concrete surfaces	Equipment-dependent, limited to point cloud data

(Continued)

Table 1: Continued

Author	Approach	Major findings	Pros	Cons
Bolourian [41]	Point cloud-based bridge surface defect detection	Point cloud-based DL method for bridge surface defect segmentation	Publicly available dataset, efficient data collection with LiDAR-equipped UAV	Equipment-dependent, specific focus on concrete surfaces
Meklati et al. [42]	Crowd-sensing for heritage wall damage detection	DL-based crowd-sensing for surface damage detection on walls	Automatic inspection, effective detection of common damages	Equipment-independent, instant diagnosis for users
Chen et al. [43]	DL for building facade cracks	Transfer learning approach for image classification in building facades	Effectiveness of DL, transfer learning improves accuracy	Equipment-independent, better performance with limited data
Bruno et al. [44]	CV for built heritage	Mask R-CNN model for decay morphology detection on historic buildings	Remote, non-invasive inspection, support for conservation efforts	Optimization required
Yang [45]	Surface defect detection overview	Overview of surface defect detection methods based on CNNs	Summarizes recent methods and application scenarios in industrial detection	High-level overview, limited to summarizing existing methods

Lerones et al. [22] presented an innovative method for detecting moisture in heritage buildings using 3D laser scanner surveying data. Moisture can lead to structural deterioration and aesthetic damage in historic buildings, making its detection crucial. This non-intrusive method analyzes laser reflectivity levels offline, covering large areas quickly without interacting with materials. This approach provides conservation professionals with objective and comprehensive information on moisture damage, aiding in decision-making. The effectiveness of this method is demonstrated through its application in the Cathedral of Ciudad Rodrigo, Spain. Tavukcuoğlu [23] discussed the significance of non-destructive testing (NDT) techniques for *in situ* building inspections. It highlights the value of quantitative infrared thermography (QIRT) and ultrasonic pulse velocity (UPV) measurements for assessing moisture, thermal, material, and structural issues in historical and contemporary structures. The joint use of QIRT and ultrasonic testing permits damage detection, assessment of materials, and thermal performance evaluation. The research highlights the need for a multi-disciplinary methodology to augment materials technology and conservation practices.

Błaszczak-Bąk et al. [24] explored the use of LiDAR technology for wall defect detection in unlit environments. In this work, the Terrestrial Laser Scanning (TLS) measurements are processed using the Optimum Dataset (OptD) method. This preserves more points of interest in imperfect surfaces (e.g., cracks) while removing redundant information in homogeneous areas. The improved OptD algorithm is effective in detecting and segmenting defects, aiding in estimating repair costs. Wong [25] proposed NDT method which includes infrared thermography, ground penetrating radar (GPR), microwave moisture tomography, and ultrasonic pulse echo tomography to evaluate historic building conditions. Case studies presented in this work demonstrate the identification of hidden details, defects, deterioration, and moisture detection in these structures.

Radnić et al. [26] deliberated the restoration and structural strengthening of historical masonry buildings using a case study of the Minceta fortress in Dubrovnik. This work emphasized the significance of non-destructive tests, including static and dynamic analysis in evaluating structural safety and deterioration. Błaszczak-Bąk et al. [27] presented an approach for automatic wall defect detection in unlit environments using LiDAR and the modified OptD method. TLS measurements are handled to identify and segment defects, facilitating cost estimation for repair and renovation of historic buildings.

Wu et al. [28] proposed RBGNet system for a new rail surface defect detection. RBGNet makes use of a novel architecture that combines rail surface and edge information to accurately identify rail surface defects. This work employs a hybrid loss function for network training and integrates edge features with rail surface features to improve detection precision. The system is verified on complex unmanned aerial vehicle (UAV) rail datasets, demonstrating high detection rates in challenging environments. Wood and Mohammadi [29] studied about detecting surface damage and cracks in historic fresco walls using geometric features in point cloud data. This work delivers a non-destructive and color-independent approach for identifying damage based on geometric descriptors. The method has been investigated on a diverse dataset of historic buildings, showcasing its potential for damage detection in heritage structures.

Al-Sakkaf et al. [30] did a review on the use of GPR technology for defect detection in heritage buildings. GPR suggests a non-invasive method for detecting internal features in structures, particularly stone masonry. This study identifies conventional methodologies and highlights the effectiveness of GPR in assessing heritage structures' condition.

2.2 DL approaches

Deep convolutional neural networks (DCNNs) were used to detect damages caused by various pathologies. The effects of pathologies were severe in damaging cultural heritage [31]. The main advantage of the proposed approach was to quantify the structural defects of buildings by using water infiltration, concrete carbonation, and efflorescence. To assess the conditions of heritage buildings, the study by Sharma et al. [32] was used to determine the dust amount deposited in buildings. The level of dust indicated the damage to a building and higher level of dust deposited generated a warning for maintenance.

To estimate the missing components of historical places, a recent work [33] proposed to use the Faster R-CNN model in forbidden cities. The proposed model has the capability to detect 2D images and position the missing components. Although the proposed technique laid a foundation for the intelligent inspection of heritage buildings, still a lot of research is needed for the comprehensive detection of missing parts of the buildings. Extended research was presented by Zou et al. [34] to inspect the distinctive patterns on the surfaces of ancient architectures. A DL approach based on inpainting, segmentation, and classification was proposed. Segmentation is aimed at gaining the mask for the defective parts. Afterwards, inpainting algorithm was applied to rebuild the damaged parts. Ultimately, Residual Neural Networks were applied for the classification of rebuilt images. Overall, the proposed approach improved the classification accuracy of reconstructed images. However, inpainting and segmentation were not presented for inspection in advance.

Wang et al. [35] facilitated the identification of damage in historic masonry structures by introducing an automatic detection technique. This utilizes the Faster R-CNN model based on ResNet101. This method successfully detects efflorescence and spalling damage with high precision. The research also presented IP webcam and smartphone-based real-time damage detection systems, contributing to the safeguarding and management of historic buildings. Wenlong et al. [36] concentrated on bridge surface damage detection using DL. This work addressed the challenges related to dataset size and class imbalance. This work presents an Atrous spatial pyramid pooling (ASPP) module and a weight-balanced Intersection over Union (IoU) loss function to enhance accuracy in detecting delamination and rebar exposure on bridges.

Ye and Sun [37] reviewed machine vision-based methods for detecting surface defects in ceramic tableware. This work recapitulated imaging methods, defect types, and mathematical modeling approaches. The research study identifies areas for improvement in feature extraction algorithms based on deep neural networks. Stephen et al. [38] presented a CNN model for classifying surface defects in tile surface images. The CNN learns discriminative feature representations and performs binary-class classification, distinguishing between cracked and non-cracked surfaces. This approach shows potential for automating visual inspections and achieving efficient classification of surface defects.

Teng et al. [39] introduced an improved YOLOv3 model for real-time bridge surface defect detection. This work addresses issues like blurry edges and noise, achieving high detection accuracy. This method has the ability to optimize bridge inspection processes and enhance defect detection in various bridge types. Shao et al. [40] presented a two-stage method for detecting surface defects in concrete buildings using point clouds and a 3D neural network. This method divides buildings into 3D grids and employs PointNet++ for damage classification. This method attains acceptable detection performance on aging concrete surfaces, yielding a non-destructive approach for damage assessment.

Bolourian [41] developed a point cloud-based DL method, SNEPointNet++ for semantic segmentation of concrete bridge surface defects. This work uses a publicly available dataset for defect detection and achieves high recall and precision rates for different types of defects. Furthermore, this work proposed an efficient path planning procedure for LiDAR-equipped UAVs during bridge inspections. Meklati et al. [42] introduced a crowd-sensing solution based on DL for automatically detecting common surface damages on heritage walls. This work makes use of a CNN integrated into a mobile application, enabling users to capture and diagnose wall damage instantly. The approach proves effective, providing rapid and objective damage assessments.

Chen et al. [43] employed a transfer learning based approach for image classification to detect cracks in building facades. Transfer learning enhances accuracy even with limited data. The research work highlights the potential of DL for efficient image classification in the context of building facade inspections. Bruno et al. [44] proposed a mask R-CNN model for detecting decay morphologies on built heritage, especially historic buildings. This method utilizes CV and artificial intelligence to remotely assess the conservation status of heritage structures. Experimental results prove effectiveness in identifying specific types of alterations and provides valuable support for heritage conservation efforts.

Yang [45] provided an overview of surface defect detection methods based on CNNs. This work summarizes modern methods and their application scenarios in industrial defect detection. The focus is on utilizing DL models for efficient and automated defect classification in various domains. Other than DL based approaches proposed in the literature, Kwon et al. [46] introduced a genetic algorithm based technique to predict the maintenance cost of the aging buildings in urban areas. The applicability of the proposed technique

was tested by performing a number of experiments. This technique devised a systematic method to forecast the maintenance cost and supported the management to decide about maintenance in long-term perspectives.

Functional deterioration of buildings is negatively impacted by the climate change conditions. Several works have been carried out to seek the reasons behind this negative relationship between climate change and building deterioration. One of these studies have been conducted to determine the climate change impacts on buildings in Chile where it was revealed that an increase in temperature could reduce the average annual precipitation [47]. The results of the above mentioned were surprising for researchers. However, the proposed technique's applicability to other issues of buildings remains to be conducted in future works. For example, prioritization of maintenance costs can be linked for buildings located in various parts of the world.

3 Materials and methods

3.1 Data collection for classification

Building defects normally arise due to the constant influence of external factors that are primarily comprised of five types: mechanical properties, electromagnetic radiation, climatic conditions, oxidation, and biological agents [48–52]. As shown in Figure 1, this study has classified surface defects into six categories, including plant penetration, moss, cracking, alkalization, staining, and deterioration. A total of 900 defect images were captured by a digital camera from the exterior walls of different heritage buildings, with each category accounting for 150 images. Out of the total captured images, 720 were used for training the model, 120 from each category, while the remaining 180 images were used for testing.

3.2 Image cropping and database creation

Due to their complex backgrounds, the sample images were randomly cropped and sifted through to highlight specific information on a certain category of the defect and minimize the influence of irrelevant factors. We selected the images with defect features, which were clearly visible and were over 15% of the entire cropped image in size.

Afterwards, the selected images were cut into sub-images of 512×512 pixels to generate the dataset. The training dataset contained 2,400 sub-images in total, which were labeled into six categories such as, plant penetration (400 sub-images), moss (400), cracking (400), alkalization (400), staining (400), and deterioration (400). Plant penetration images were classified and detected through a YOLOv5 model, whereas the remaining five categories (cracking, alkalization, staining, deterioration, and moss) adopted Swin Transformer for the segmentation and detection of their images. A transformer-based model totally relies on self-attention to compute the input and output representations regardless of using convolutional or RNN models. Thus, Swin Transformer has resolved the computational issue as well as costs that are linear to an image size [53]. Also, Swin Transformer model improves the performance efficiency by operating regionally and enhancing the respective fields that show a high correlation with the visual signals.

3.3 Image pre-processing

Data pre-processing matters in various DL algorithms. In practice, data normalization and whitening are essential to many algorithms in order to yield optimal results. Since the ambient light conditions exerted a considerable influence on the sample images when they were collected, this study has subtracted the mean from each image to minimize the effects on their overall luminance.

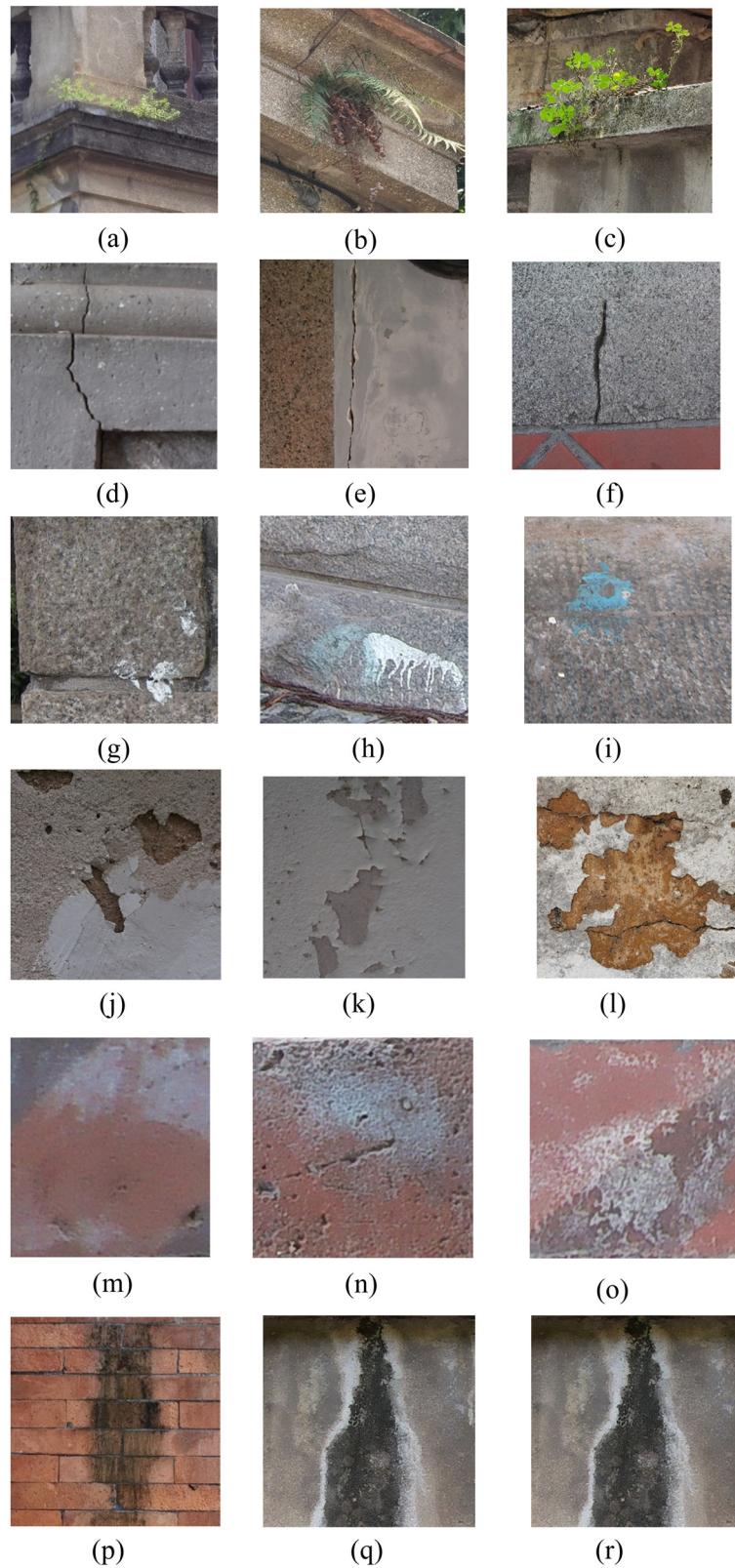


Figure 1: Samples of the dataset that were used to train our model showing diverse plant penetration images (a–c), cracking (d–f), staining (g–i), deterioration (j–l), alkalization (m–o), and moss (p–r).

Given that this study adopted the pre-trained model on the ImageNet dataset for image segmentation, the mean value and standard deviation were calculated based on the ImageNet dataset, which were respectively set to [123.675, 116.28, 103.53] and [58.395, 57.12, 57.375].

$$x_i^+ = x_i - \bar{X}, \quad (1)$$

where \bar{X} denotes the mean value of Sample X , x_i represents one of the sample data ($x_i \in X$), and x_i^+ refers to the sample obtained after mean subtraction.

3.4 Model training and comparative tests

This study utilized the YOLOv5 model for defect image classification and detection (plant penetration) while adopting Swin Transformer for defect image segmentation and detection (cracking, alkalization, staining, deterioration, and moss). The model in this research, when applied to detecting some building surface defects, produced the effects as illustrated in Figures 2 and 3)

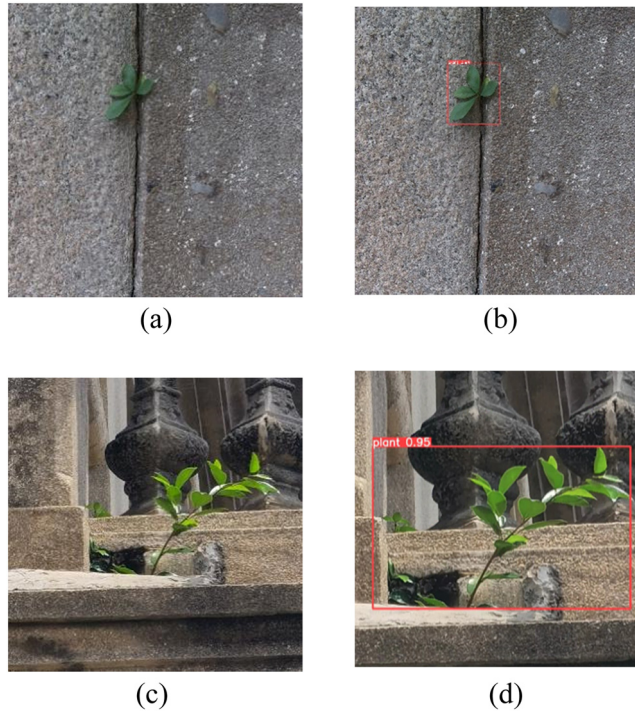


Figure 2: Effects of the model in plant penetration detection. (a and c) Original images and (b and d) classification results.

The proposed research used YOLOv5 as a surface defect classification and detection. The YOLOv5 also performs well based on speed and accuracy. Its accuracy is mainly based on three parts such as backbone, encoder, and decoder [54]. The strength of this network lies in the extraction of backbone features and using those features for prediction.

3.4.1 YOLOv5- and Swin transformer-based network models

YOLOv5, a typical one-stage object detection algorithm that combines image classification with localization, reframes object detection. Synchronously, generates the probability and bounding box coordinates of each

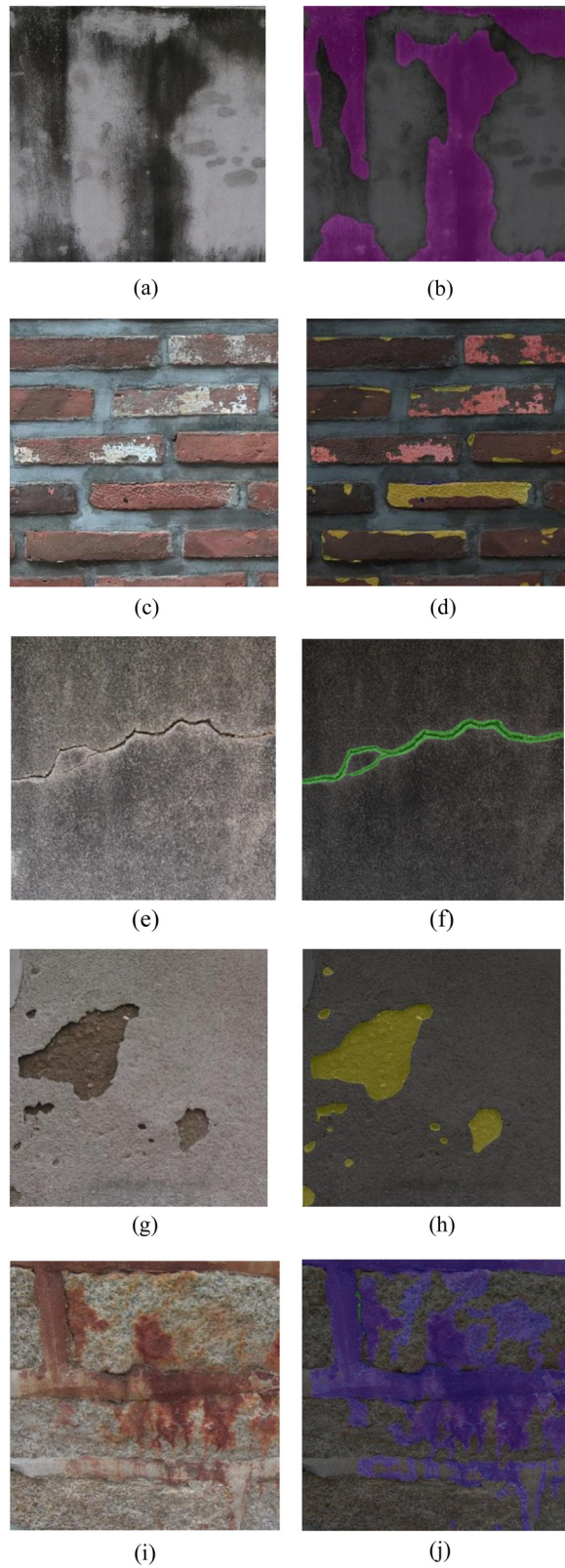


Figure 3: Effects of the model in the segmentation and detection of some building surface defects. (a, c, e, g, and i) Original images and (b, d, f, h, and j) segmentation results.

class by using the regression method. At its core is feature extraction where CNN underlies. It could identify the target class on an input image and output information about the position [55,56]. YOLOv5 is available in the form of four models, namely, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, which differ from each other in the width and depth of their backbone networks. In 2020, Ultralytics released the fifth YOLO variant version, which outperforms all the previous ones in speed and accuracy. Figure 4 shows the network architecture of YoLo which consists of the different modules.

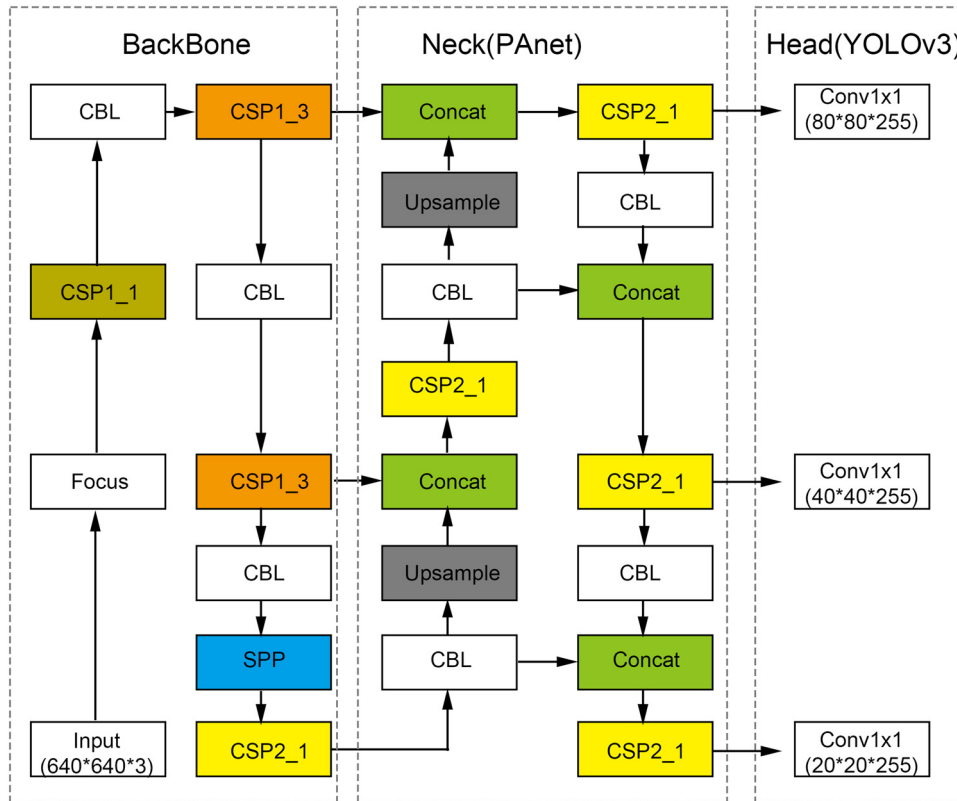


Figure 4: Overview of YoLo network architecture.

One of the major tasks in CV is semantic segmentation, which pertains to image classification because it is aimed at producing a pixel-level prediction (or predictions) for a class instead of an image-level one. Fully convolutional networks (FCNs), which perform semantic segmentation tasks through CNN, have inspired a variety of subsequent works since their appearance and have become a major design option for dense prediction tasks [57]. Segmentation transformer uses Vision Transformer (ViT) as a backbone, while incorporating several CNN decoders to increase feature resolution [58,59]. Despite its high performance, ViT also has limitations: (1) it can only produce single-scale representations of low resolution instead of multi-scale ones and (2) it incurs high computational costs when applied to large images. To address these limitations, Swin Transformer (Figure 5) splits an image into different patches (taking the form of windows) before computing self-attention efficiently within each window, while employing UPerNet the decoder produces segmentation results. Therefore, it is suitable for dense predictions [60].

The architecture of Swin Transformer is illustrated in Figure 5. The input image size is defined as $H \times W \times 3$. First of all, it splits an input RGB image into a number of non-overlapping patches. Each patch is considered to be a token and a concatenation of raw pixels such as RGB values. As seen in Figure 5, there is an additional linear embedding layer to an arbitrary dimension which is denoted by C . In the end, we have several Swin Transformer blocks with the self-attention computation that can be applied to patch tokens.

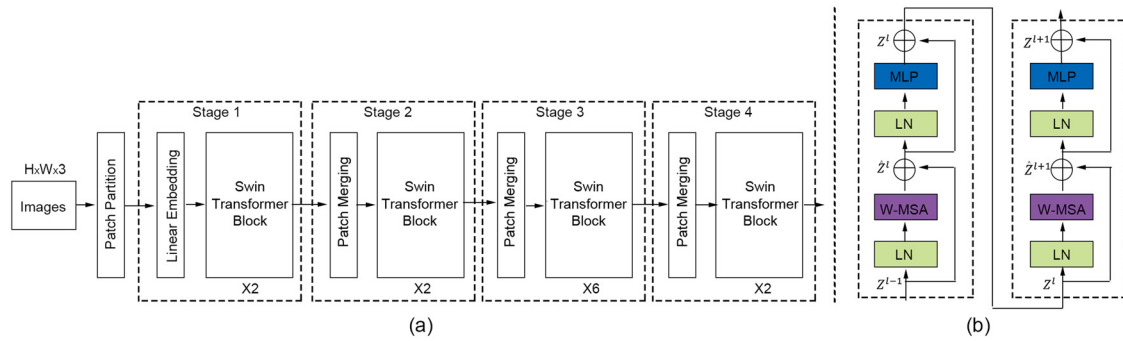


Figure 5: Overview architecture of Swin Transformer. (a) Architecture and (b) two successive swin transformer blocks.

3.4.2 Image augmentation and hyper-parameter settings of the segmentation algorithm

This study used random flipping as an image augmentation technique during training the model, while randomly adjusting the image brightness, hue, contrast, and saturation. Random flipping technique improves the image recognition accuracy and does not incur image loss during the process of augmentation. Random flipping is able to flip images horizontally [61].

The image recognition model for surface defects of heritage buildings in this research was trained on PyTorch DL framework. Adaptive moment estimation (Adam) and particularly its variant AdamW optimization algorithm was applied for model training, with the initial learning rate set to 10^{-5} , betas = (0.9, 0.999), and weight_decay = 0.01. Weight_decay [62] is computed as given in equation (2).

$$d_{\text{AdamW}} = -\eta\lambda\theta, \quad (2)$$

where η expresses the learning rate, λ is the parameter performing the parameter scaling and weight decaying, and θ is the parameter that is being optimized. These are subtracted from parameters during the update step.

The defect sample images were trained on Nvidia V100 32GB graphics cards, with the batch size equaling 4. The 400 epochs of training took around 15 h.

4 Comparative tests and results

The Swin Transformer-based method for the segmentation of building surface defect images achieved the highest accuracy regarding the evaluation metric, with an mIoU of 90.96% and a mean accuracy (mAcc) of 95.78%, outperforming mainstream DL networks such as SegFormer, PSPNet, and DANet. The comparative results when different segmentation algorithms were executed for some defect (alkalization and deterioration) images are shown in Figure 6, whereas the comparative test results (mIoU/mAcc metric) of each model are shown in Table 2.

Tables 2 and 3 show the comparison of the Swin-Base network with the state-of-the-art techniques used for detection of defects in buildings in the literature. Swin-Base algorithm has higher mIoU metric results for five defects as shown in Table 2. Among the defects, cracking defect detection accuracy is lower for Swin-Base algorithm compared with the other four defects. However, the accuracy rate for cracking defects and the rest of the defects from Swin-Base algorithm remained higher than the other algorithms. As illustrated in Table 3, higher mAcc metric results were gained by the Swin-Base algorithm in the current research.

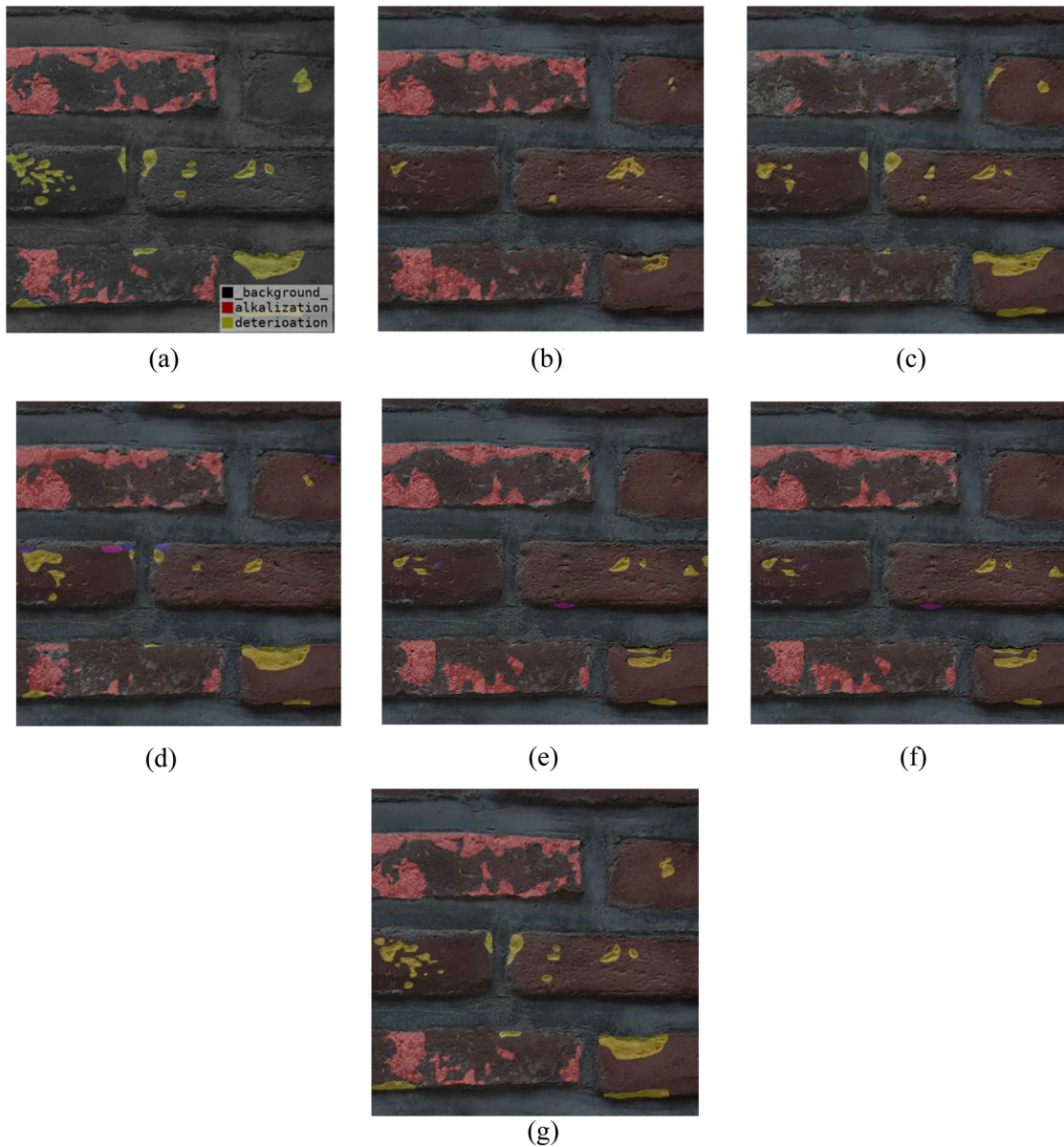


Figure 6: Swin Transformer surpassed the accuracy of its counterparts, with its output fitting the labeled image to a greater extent. (a) Swin Transformer. (b) FCN. (c) PSPNet. (d) DANet. (e) DeepLabv3-plus. (f) SegFormer. (g) Label.

Table 2: Comparative test results of different segmentation algorithms (mIoU metric), with Swin Transformer achieving the highest accuracy in detecting all five defects

Algorithms	Alkalization/mIoU	Cracking/mIoU	Deterioration/mIoU	Staining/mIoU	Moss/mIoU	mIoU (%)
FCN-ResNet101	23.14	40.19	63.48	27.74	49.88	40.89
DeepLabv3-plus ResNet101	53.82	42.62	78.39	36.21	17.65	45.74
DANet-ResNet101	10.36	45.45	58.04	19.46	3.25	27.31
PSPNet-ResNet101	32.65	36.93	56.13	60.29	48.14	46.83
SegFormer-B5	70.29	38.16	82.84	69.31	83.26	68.77
Swin-Base	91.04	81.24	95.38	92.28	94.86	90.96

The bold value represents Swin Transformer achieving the highest accuracy of pixel and mAcc in detecting all five defects

Table 3: Comparative test results of different segmentation algorithms (mAcc metric), with Swin Transformer achieving the highest accuracy in detecting all five defects

Algorithms	Alkalization/pixel accuracy	Cracking/pixel accuracy	Deterioration/pixel accuracy	Staining/pixel accuracy	Moss/pixel accuracy	mAcc (%)
FCN-ResNet101	25.07	58.23	65.27	49.58	61.47	60.10
DeepLabv3-plus ResNet101	65.57	74.35	89.88	87.15	58.56	75.10
DANet-ResNet101	10.38	49.27	58.62	19.68	3.25	28.24
PSPNet-ResNet101	69.75	41.44	93.33	71.87	68.68	69.01
SegFormer-B5	79.85	44.85	90.61	75.87	92.36	76.71
Swin-Base	96.24	91.34	97.82	95.84	97.64	95.78

The bold value represents Swin Transformer achieving the highest accuracy of pixel and mAcc in detecting all five defects

5 Practical applications

To meet practical needs, it is necessary to evaluate the generalization ability related to the previously trained model on the out-of-sample images. Practical testing consisted of two parts in this study: (1) validation test of the model on the remaining samples after the manual screening and (2) test of the model in recognizing the defects in the real test images.

5.1 Test on the remaining samples after the manual screening

The image datasets that were previously not used for model training (180 images in total, 30 from each category of defect) were sifted through and processed for the purpose of generating the test samples, which comprised a total of 550 sub-images: plant penetration (50), moss (100), cracking (100), alkalization (100), staining (100), and deterioration (100). Test results confirmed the good performance of the trained model, specific data of which are shown in Tables 4 and 5.

Table 4: Classification results of the remaining sample images of plant penetration

Type of test samples	Sample size	Recognition accuracy (%)
Plant penetration	50	100

Table 5: Segmentation results of the remaining sample images of building surface defects

Types of test samples	Sample size	mIoU (%)
Alkalization	100	91.02
Cracking	100	83.20
Deterioration	100	95.52
Staining	100	93.01
Moss	100	95.10

The bold value represents the model showed the good performance of the accuracy of mIoU in the real test images

5.2 Recognition of the defects in the real test images

Since field-collected images were normally large in size with a high pixel density and contained complex information, it was necessary to utilize sliding windows to scan those images in search of building surface

defects therein. As shown in Figure 7, each sliding window was moved over the image with a certain stride to identify any defect within this window. To ensure thorough coverage of the defect zones, overlapping scan areas were created by setting the window stride to half of the window size. The field-collected images differed in their scale factor, thus requiring different sizes of sliding windows to scan them, which were resized through the method of bilinear interpolation to fit the input size the model required (512×512). The specific recognition process is shown in Figure 8, while image recognition of some surface defects is demonstrated in Figure 9.

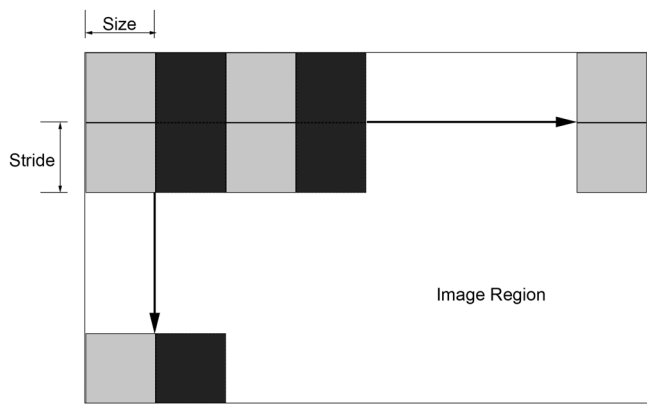


Figure 7: Segmentation results of the remaining samples of building surface defects (Sliding window).

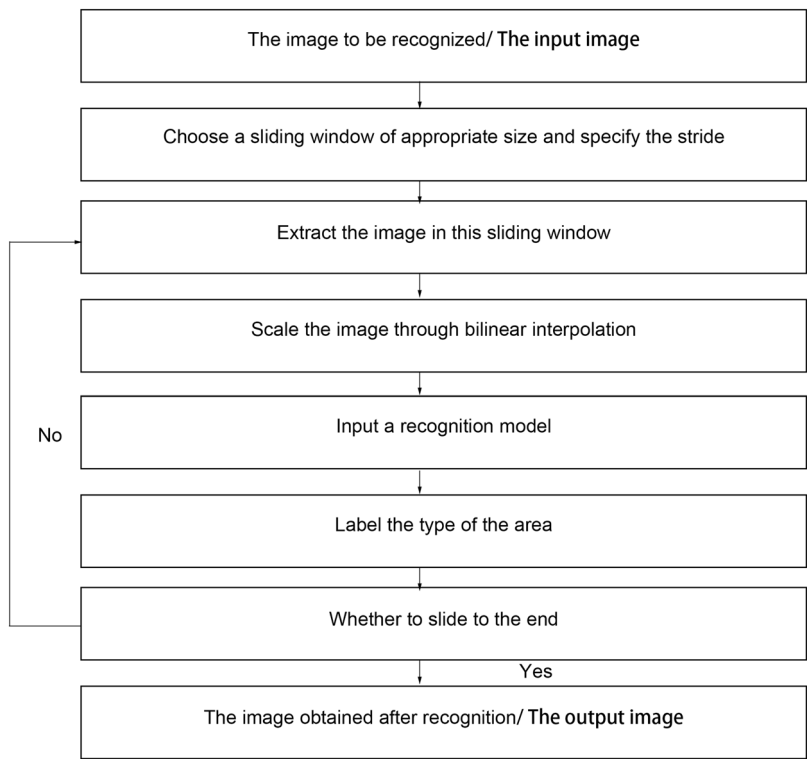


Figure 8: Flow chart of image recognition through sliding windows.

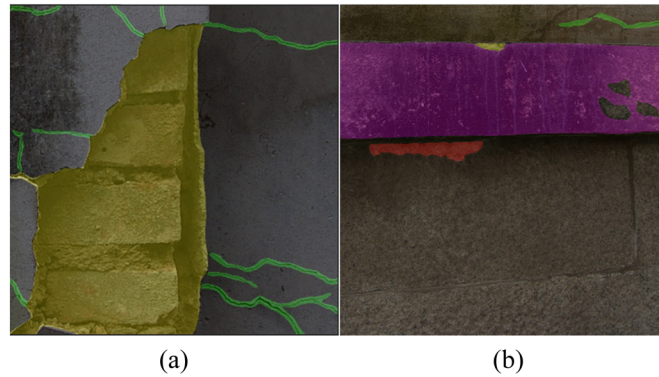


Figure 9: Results of building surface defect recognition. (a) Cracking and deterioration. (b) Moss and alkalization.

Application test results indicated that the trained model performed well in recognizing and localizing surface defects. Nevertheless, a number of detection errors also occurred, particularly in the following three cases: (1) surroundings in the images were complex, which easily affected the recognition; (2) defects in the same image abounded, varied, and strongly resembled each other; and (3) great disparities existed between the training samples and the defect images to be identified.

6 Conclusion

This study examined the application of the YOLOv5- and Swin Transformer-based DL model in recognizing images of surface defects in heritage buildings. This research conducted a comparative study on the mainstream models such as SegFormer, PSPNet, and DANet. Surface defects were classified into six categories according to their features, including plant penetration, moss, cracking, alkalization, staining, and deterioration. The proposed model for defect detection was practically tested for its reliability. The test results suggested that the YOLOv5-based image classification method and the Swin Transformer-based segmentation method could contribute to the rapid identification of surface defects in heritage buildings due to their high recognition accuracy.

Funding information: This research was supported by the National Natural Science Foundation of China (Grant number 52078154).

Authors contributions: Xiaoli Fu: conceptualization, methodology, investigation, software, data curation, writing – original draft preparation. Niwat Angkawisittpan: supervision, validation, writing – reviewing and editing.

Conflict of interest: The authors declare no conflict of interest.

Data availability statement: The data used to support the findings of the research are included within the article.

References

- [1] Mimura T, Mita A. Automatic estimation of natural frequencies and damping ratios of building structures. *Procedia Eng.* 2017;88:163–9.

- [2] Perez H, Tah JHM, Mosavi A. Deep learning for detecting building defects using convolutional neural networks. *Sens (Basel)*. 2019;19(16):3556.
- [3] Martin G, Selvakumaran S, Marinoni A, Sadeghi Z, Middleton C. Structural health monitoring on urban areas by using multi temporal insar and deep learning. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium IGARSS*; 2021. p. 176–9.
- [4] Dang HV, Tatipamula M, Nguyen HX. Cloud-based digital twinning for structural health monitoring using deep learning. *IEEE Trans Ind Inform*. 2022;18(6):3820–30.
- [5] Azimi M, Dadras A, Pekcan G. Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*. 2020;20(10):2778. doi: 10.3390/s20102778.
- [6] Davoudi R, Miller GR, Kutz JN. Structural load estimation using machine vision and surface crack patterns for shear-critical RC beams and slabs. *J Comput Civ Eng*. 2018;32(4):04018024. doi: 10.1061/(ASCE)CP.1943-5487.0000766.
- [7] Hoang ND. Image processing-based recognition of wall defects using machine learning approaches and steerable filters. *Comput Intell Neurosci*. 2018;2018:7913952. doi: 10.1155/2018/7913952.
- [8] Jo J, Jadidi Z, Stantic BA. Drone-based building inspection system using software-agents. In *Studies in computational intelligence*. Vol. 737. New York, NY, USA: Springer; 2017. p. 115–21.
- [9] 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. doi: 10.1016/j.autcon.2019.102846.
- [10] Valero E, Forster A, Bosché F, Renier C, Hyslop E, Wilson L. High level-of-detail BIM and machine learning for automated masonry wall defect surveying. In *Proceedings of the International Symposium on Automation and Robotics in Construction*. Berlin, Germany; 2018. p. 20–5.
- [11] Chu W-T, Tsai M-H. Visual pattern discovery for architecture image classification and product image search. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*. Hong Kong, China; 2012. p. 5–8.
- [12] Goel A, Juneja M, Jawahar CV. Are buildings only instances?: Exploration in architectural style categories. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*. Mumbai, India; 2012. p. 1–8.
- [13] Mathias M, Martinovic A, Weissenberg J, Haegler S, Van Gool L. Automatic architectural style recognition. In *Proceedings of the 4th ISPRS International Workshop 3D-ARCH 2011. XXXVIII-5/W16*, Trento, Italy; 2011. p. 171–6.
- [14] Shalunts G, Haxhimusa Y, Sablatni R. Architectural style classification of building facade windows. In *Advances in visual computing*. 6939, Las Vegas, NV, USA: Springer; 2011. p. 280–9.
- [15] Zhang L, Song M, Liu X, Sun L, Chen C, Bu J. Recognizing architecture styles by hierarchical sparse coding of blocklets. *Inf Sci*. 2014;254:141–54.
- [16] Xu Z, Tao D, Zhang Y, Wu J, Tsoi AC. Architectural style classification using multinomial latent logistic regression. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer vision – ECCV 2014*. ECCV 2014. Lecture notes in computer science. Vol. 8689. Cham: Springer; 2014. doi: 10.1007/978-3-319-10590-1_39.
- [17] Llamas J, Lerones PM, Medina R, Zalama E, Gómez-García-Bermejo J. Classification of architectural heritage images using deep learning techniques. *Appl Sci*. 2017;7:992. doi: 10.3390/app7100992.
- [18] Grinzato E, Bison PG, Marinetti S. Monitoring of ancient buildings by the thermal method. *J Cultural Herit*. 2002;3(1):21–9.
- [19] Avdelidis N, Moropoulou A. Applications of infrared thermography for the investigation of historic structures. *J Cultural Herit*. 2004;5(1):119–27.
- [20] Mansuri LE, Patel D. Artificial intelligence-based automatic visual inspection system for built heritage. *Smart Sustain Built Environ*. 2021;11(2):622–46.
- [21] Trillo C, Aburamadan R, Chikomborero B, Makore N, Udeaja C, Moustaka A, et al. Towards smart planning conservation of heritage cities: digital technologies and heritage conservation planning. In *International Conference on Human-Computer Interaction, Lecture Notes in Computer Science*. 12794, Cham: Springer; 2021. doi: 10.1007/978-3-030-77411-0_10.
- [22] Lerones PM, Vélez DO, Rojo FG, Gómez-García-Bermejo J, Casanova EZ. Moisture detection in heritage buildings by 3D laser scanning. *Stud Conserv*. 2016;61:46–54.
- [23] Tavukçuoğlu A. Non-destructive testing for building diagnostics and monitoring: Experience achieved with case studies. *MATEC Web Conf*. 2018;149:01015. doi: 10.1051/mateconf/201814901015.
- [24] Błaszczak-Bąk W, Suchocki C, Janicka J, Dumalski A, Duchnowski R. Defect detection of historic structures in dark places based on the point cloud analysis by modified Optd method. *Int Arch Photogram Remote Sens Spat Inf Sci*. 2019;XLII-3/W8:71–7.
- [25] Wong CW. Applications of non-destructive tests for diagnosis of heritage buildings: Case studies from Singapore and Malaysia. *Built Herit*. 2019;3:14–25.
- [26] Radnić J, Matešan D, Abaza A. Restoration and strengthening of historical buildings: The example of Minceta fortress in Dubrovnik. *Adv Civ Eng*. 2020;2020:8854397. doi: 10.1155/2020/8854397.
- [27] Błaszczak-Bąk W, Suchocki C, Janicka J, Dumalski A, Sobieraj-Żłobińska A. Automatic threat detection for historic buildings in dark places based on the modified OptD method. *ISPRS Int J Geo-Inf*. 2020;9(2):123. doi: 10.3390/ijgi9020123.
- [28] Wu Y, Qin Y, Qian Y, Guo F, Wang Z, Jia L. Hybrid deep learning architecture for rail surface segmentation and surface defect detection. *Comput Civ Infrastruct Eng*. 2022;37(2):227–44.
- [29] Wood RL, Mohammadi ME. Feature-based point cloud-based assessment of heritage structures for nondestructive and noncontact surface damage detection. *Heritage*. 2021;4:775–93.

- [30] Al-Sakkaf A, Ghodke S, An C, Bagchi A. Defect detection in heritage buildings using ground penetrating radar – A review. In proceedings of SHMII-11: 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure. Montreal, QC, Canada; 2022. p. 1–4.
- [31] Masrour T, El Hassani I, Bouchama MS. Deep convolutional neural networks with transfer learning for old buildings pathologies automatic detection. In Proceedings of International Conference on Advanced Intelligent Systems for Sustainable Development. Springer; 2019. p. 204–16.
- [32] Sharma T, Agrawal P, Verma NK. Detection of dust deposition using convolutional neural network for heritage images. In Proceedings of Computational Intelligence: Theories, Applications and Future Directions-Volume II. Springer; 2019. p. 347–59.
- [33] 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 Cultural Herit.* 2019;38:221–30.
- [34] Zou Z, Zhao P, Zhao X. Automatic segmentation, inpainting, and classification of defective patterns on ancient architecture using multiple deep learning algorithms. *Struct Contr Health Monit.* 2021;28(7):e2742. doi: 10.1002/stc.2742.
- [35] 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.
- [36] Wenlong D, Yongli M, Takahiro K, Sergio E, Kohei N, Kotaro N, et al. Vision based pixel-level bridge structural damage detection using a link ASPP network. *Autom Constr.* 2020;110:102973. doi: 10.1016/j.autcon.2019.102973.
- [37] Ye S, Sun L. Method for detecting surface defects of ceramic tableware based on deep learning. *J Phys: Conf Ser.* 2020;1650:032045. doi: 10.1088/1742-6596/1650/3/032045.
- [38] Stephen O, Maduh UJ, Sain MA. Machine learning method for detection of surface defects on ceramic tiles using convolutional neural networks. *Electronics.* 2022;11(1):55. doi: 10.3390/electronics11010055.
- [39] Teng S, Liu Z, Li X. Improved YOLOv3-based bridge surface defect detection by combining high- and low-resolution feature images. *Buildings.* 2022;12(8):1225. doi: 10.3390/buildings12081225.
- [40] Shao W, Kakizaki K, Araki S, Mukai T. Automated two-stage approach for damage detection of surface defects in historical buildings. In proceedings of 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC). Los Alamitos, CA, USA; 2022. p. 816–1821.
- [41] Bolourian N. Point cloud-based deep learning and UAV path planning for surface defect detection of concrete bridges. PhD Thesis. Canada: Concordia University; 2022.
- [42] Meklati S, Boussora K, Abdi MEH, Sid-Ahmed B. Surface damage identification for heritage site protection: A mobile crowd-sensing solution based on deep learning. *J Comput Cultural Herit.* 2023;16(2):25, 1–24.
- [43] Chen Y, Zhu Z, Lin Z, Zhou Y. Building surface crack detection using deep learning technology. *Buildings.* 2023;13:1814. doi: 10.3390/buildings13071814.
- [44] Bruno S, Galantucci RA, Musicco A. Decay detection in historic buildings through image-based deep learning. *VITRUVIO - Int J Archit Technol Sustain.* 2023;8:6–17.
- [45] Yang W. A survey of surface defect detection based on deep learning. In Proceedings of the 2022 7th International Conference on Modern Management and Education Technology (MMET 2022). Shanghai, China: 2022. doi: 10.2991/978-2-494069-51-0_51.
- [46] Kwon N, Song K, Ahn Y, Park M, Jang Y. Maintenance cost prediction for aging residential buildings based on case-based reasoning and genetic algorithm. *J Build Eng.* 2020;28:101006. doi: 10.1016/j.jobbe.2019.101006.
- [47] Prieto A, Verichev K, Silva A, de Brito J. On the impacts of climate change on the functional deterioration of heritage buildings in South Chile. *Build Environ.* 2020;183:107138. doi: 10.1016/j.buildenv.2020.107138.
- [48] ISO. ISO 19208:2016-Framework for specifying performance in buildings; ISO: Geneva, Switzerland; 2016.
- [49] CS Limited. Defects in buildings: Symptoms, investigation, diagnosis and cure. London, UK: Stationery Oce; 2001.
- [50] Seeley IH. Building maintenance. London, UK: Macmillan International Higher Education; 1987.
- [51] Richardson B. Defects and deterioration in buildings: A practical guide to the science and technology of material failure. London, UK: Routledge; 2002.
- [52] Wood BJ. Building maintenance. New York, NY, USA: JohnWiley & Sons; 2009.
- [53] Panboonyuen T, Jitkajornwanich K, Lawawirojwong S, Srestasathien P, Vateekul P. Transformer-based decoder designs for semantic segmentation on remotely sensed images. *Remote Sens.* 2021;13(24):5100. doi: 10.3390/rs13245100.
- [54] Wang K, Teng Z, Zou T. Metal defect detection based on Yolov5. *J Phys: Conf Ser.* 2022;2218(1):012050. doi: 10.1088/1742-6596/2218/1/012050.
- [55] Zhou J, Jiang P, Zou A, Chen X, Hu W. Ship target detection algorithm based on improved YOLOv5. *J Mar Sci Eng.* 2021;9(8):908. doi: 10.3390/jmse9080908.
- [56] Kasper-Eulaers M, Hahn N, Berger S, Sebulonsen T, Myrland Ø, Kummervold P. Short communication: Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5. *Algorithms.* 2021;14:114. doi: 10.3390/a14040114.
- [57] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(4):640–51.
- [58] Zheng S, Lu J, Zhao H, Zhu X, Luo Z, Wang Y, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. 2021;6877–86. doi: 10.1109/CVPR46437.2021.00681.
- [59] Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, et al. A survey on vision transformer. *IEEE Trans Pattern Anal Mach Intell.* 2023;45(1):87–110.

- [60] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, BC, Canada; 2021. p. 9992–10002
- [61] Zhong Z, Zheng L, Kang G, Li S, Yang Y. Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence; 2020. doi: 10.48550/arXiv.1708.04896.
- [62] Wright L, Demeure N. Ranger21: A synergistic deep learning optimizer. ArXiv, 2021. doi: 10.48550/arXiv.2106.13731.