COMPDYN 2025
10th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15-18 June 2025

AUTOMATIC RECOGNITION OF DAMAGES IN CULTURAL HERITAGE BUILDING IMAGES USING AN ENCODER-DECODER NETWORK

Yunpeng Yue[1], Hai Liu[1,*], Xiaoyu Liu[1], Francesca da Porto[2], and Marco Donà[2,3,4]

[1]School of Civil Engineering and Transportation, Guangzhou University, Guangzhou, 510006, China
e-mail: yueyunpeng@gzhu.edu.cn; hliu@gzhu.edu.cn; xiaoyuliu@gzhu.edu.cn

[2]Department of Geosciences, University of Padova, Via Gradenigo 6, 35131 Padova, Italy
marco.dona.1@unipd.it; francesca.daporto@unipd.it

[3]Department of Cultural Heritage, University of Padova, Piazza Capitaniato 7, 35131 Pa-dova, Italy
e-mail: marco.dona.1@unipd.it

[4]Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy
e-mail: marco.dona.1@unipd.it

*Correspondence: hliu@gzhu.edu.cn

Abstract

The widespread presence of damaged cultural heritage buildings highlights the urgent need for timely detection and the implementation of appropriate protective measures. However, manually analyzing a large volume of damage in cultural heritage buildings is not only time-consuming but also heavily dependent on the expertise of the practitioner. To address this challenge, this paper proposes an automatic recognition algorithm based on an encoder-decoder network (DeepLabV3+) to segment various types of damage in cultural heritage buildings, including plant, cracks/discontinuous, spalling, efflorescence, erosion, etc. To enhance the recognition capability of the proposed algorithm, a Grid-Mask approach is applied for data augmentation on the established cultural heritage damage dataset. Preliminary results from two field experiments demonstrate that the proposed recognition algorithm can accurately identify and classify damage in cultural heritage buildings.

Keywords: Cultural heritage building; Deep learning; Encoder-decoder network; Damage segmentation.

1 INTRODUCTION

The preservation of cultural heritage buildings is crucial for ensuring the continuity and transmission of human civilization [1, 2]. However, various forms of damage, e.g., plant, micro-organism, crack, spalling, efflorescence, and erosion, can significantly affect the structural integrity of walls [3]. These damages typically result from aging, poor construction quality, and inadequate maintenance [4, 5]. If not addressed in a timely manner, they may lead to deformation, wall collapse, and other structural failures that threaten the longevity of cultural heritage buildings [6, 7]. Therefore, early detection and effective remediation are essential for safeguarding their structural stability and long-term preservation.

Conventional methods for damage detection at historic heritage sites primarily rely on in-situ visual inspections, often supported by specialized equipment [8, 9]. While this approach offers practical advantages, its accuracy is largely dependent on the expertise of the inspector. Less experienced inspectors may misjudge the severity of damage, potentially affecting subsequent structural safety evaluations and repair planning [10]. An alternative method involves sensor-based structural damage detection techniques tailored for historic heritage sites. However, analyzing images of Renaissance architecture remains a labor-intensive task [11]. For example, processing and interpreting UAV-captured images from a single day of surveying can take up to a month or more [12-14]. Additionally, limited expertise among practitioners can further compromise the accuracy of inspection results [15].

In recent years, deep learning methods have thus found broad applications across civil engineering fields, including structural health monitoring [16], construction process automation [17], and predictive maintenance of infrastructure systems [18]. Importantly, there have been efforts to apply deep learning techniques to masonry building detections [19]. Marin et al. [20] applied a deep learning model for detecting crack damage in masonry structures using Faster R-CNN models, capable of accurately identifying crack within complex backgrounds. Brackenbury et al. [21] developed a CNN-based automatic masonry bridge monitoring system to detect underlying faults and classify defects in masonry structures. Wang et al. [22] proposed a method combining mobile crowd sensing and deep learning for rapid damage detection in masonry structures, successfully identifying and locating damage on the Great Wall. Compared to object detection, semantic segmentation offers precise, pixel-level defect identification, enabling detailed classification and localization in structural damage detection [23]. Zhang et al [24] proposed a two-step automatic segmentation method for detection the mortar loss based on a U-Net algorithm, shows strong generalization and robustness to various types of noise and can detect small damages. Loverdos and Sarhosis [11] proposed a deep learning model for brick segmentation and crack detection in masonry walls, achieving high accuracy in segmenting both cracks and masonry units. However, applying deep learning techniques to segment different types of damage in cultural heritage buildings remains relatively unexplored.

In this paper, an automatic recognition algorithm is proposed to segment the damages in cultural heritage buildings. The algorithm is based on an encoder-decode network, namely Deeplabv3+[25]. This paper is organized as follows. Architecture of the proposed Deeplabv3+ algorithm is introduced in Section 2. In Section 3, a field imaging dataset of damages in cultural heritage buildings is established for training and testing of the Deeplabv3+ model. The results of the field experiment are given in Section 4, and conclusions are presented in Section 5.

2 DEEPLABV3+ RECOGNITION MODEL

The DeepLabV3+ model utilizes an optimized encoder-decoder network for semantic segmentation tasks[26], as shown in Fig. 1. In the encoder module, a convolutional neural network (CNN) processes the entire input image to extract object features. Xception serves as the backbone, leveraging atrous convolution to generate dense feature maps. The model's weights are pre-trained on the VOC image dataset [27] and used to initialize parameters, replacing randomly assigned weights for improved performance. To capture multi-scale information, convolutional operations adjust feature resolution using diverse kernel sizes, which are resized through various methods to optimize computational efficiency. Atrous spatial pyramid pooling (ASPP) [29] is then applied to reduce dimensionality and progressively refine the feature maps, facilitating the extraction of high-level semantic information. This technique enhances the model's capability to accurately identify and segment damage across different scales. Finally, a 1×1 convolutional kernel compresses the extracted features, further improving computational efficiency.

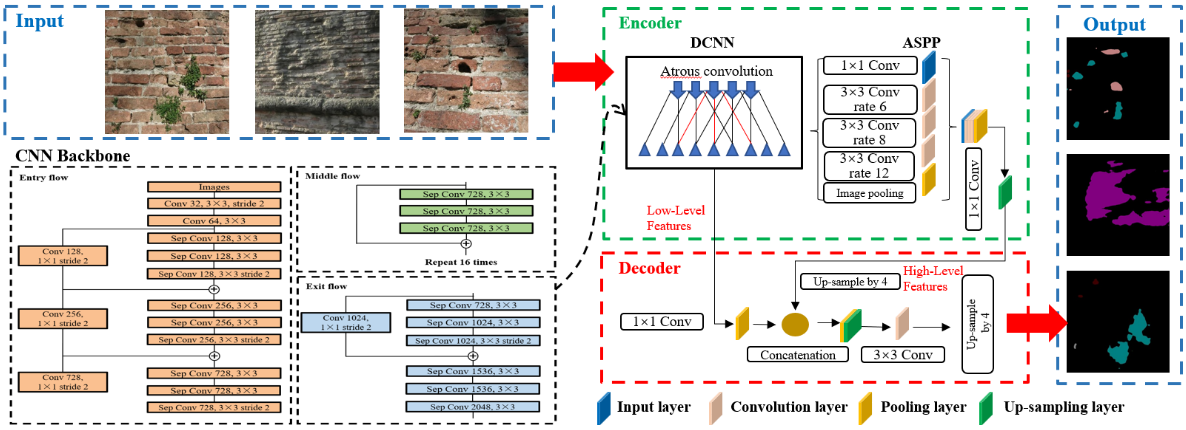


Fig. 1 Workflow of the encoder-decoder network architecture.

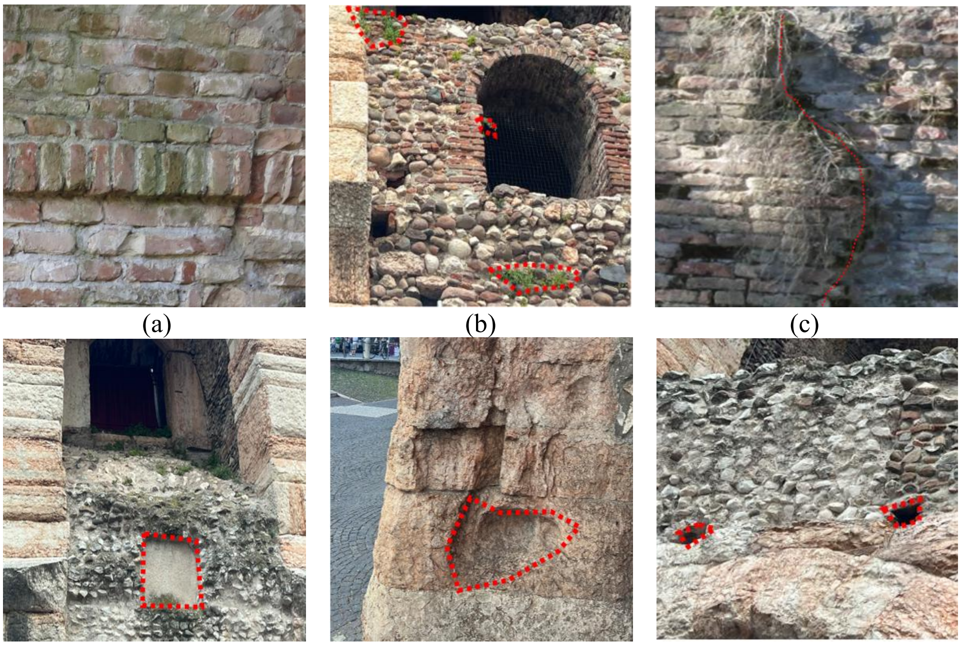
In the decoder module, a bilinear up-sampling operation is performed iteratively four times to reconstruct the target feature map extracted from the encoder module [26]. This process enhances the identification of shallow-layer features within the image. The up-sampled features are then fused with the corresponding lower-level features extracted by the network's backbone. To optimize computational efficiency, a 1×1 convolutional kernel is applied to reduce the number of channels in the feature map generated by the encoder's lower layers. This module plays a crucial role in detecting higher-order features within the images. Once the features are combined, the predicted feature map is restored to its original spatial dimensions and further refined to achieve precise semantic segmentation of the damaged areas.

3 DATASET SETUP

Field images of cultural heritage buildings were captured at various heritage architectural walls and arenas in Italy, as illustrated in Fig. 2. The dataset was acquired using a professional camera with a resolution of 180 dpi in both horizontal and vertical directions. To ensure diverse illumination conditions, optical images were collected under different lighting and weather scenarios, spanning various times of the day. The dataset includes images taken in sunny, cloudy, and overcast conditions, ensuring a comprehensive representation of the heritage architectural surfaces.

Fig. 2 Data acquisition in historical building wall

Figure 3 presents different examples of damage observed in the cultural heritage building images. To standardize the dataset, all images were resized to 512×512 pixels. The open-source tool Labelme was utilized for both semantic and instance-level annotation of various types of damage. During manual annotation, damage in each image was carefully identified and labeled, with the results stored in files containing both the image and its corresponding annotations. The generated JSON files were later converted into a simplified text format to facilitate label visualization and their transformation into binary images. To enhance the segmentation model's robustness, the GridMask algorithm [28] was applied for data augmentation after labeling. Following this process, a total of 612 cultural heritage building images were collected, encompassing 816 instances of damage. These include 412 cases of higher plant growth, 126 occurrences of spalling, 173 instances of efflorescence and erosion, and 105 cases of cracks or discontinuities. The resulting dataset serves as the training set for the model developed in Section 4.



(a) (b) (c)

Fig. 3 Examples of damage with different types in cultural heritage buildings. (a) Background, (b) higher plant, (c) cracks/discontinuous, (d) spalling, (e) efflorescence, and (f) erosion.

4 RESULTS

The proposed DeepLabv3+ model is trained and evaluated on a standard desktop computer equipped with a 3.0 GHz Intel Xeon Gold 5217 CPU, 256 GB of RAM, an NVIDIA GeForce RTX 2080 Ti GPU, and running Windows 10. A total of 428 building images are used for training, while 122 images are allocated for validation and 62 for testing. The model is trained with an initial learning rate of 5×10^{-4}, a weight decay of 5×10^{-4}, and a momentum of 0.94. After 800 training epochs, the final DeepLabv3+ model is obtained.

The effectiveness of the proposed algorithm in segmentation is demonstrated in Fig. 4. Two case studies were conducted for testing: one involving the Renaissance wall in Padova, Italy, and the other focusing on the Arena in Verona, Italy. The results indicate that the trained DeepLabv3+ model successfully differentiates between efflorescence and erosion, two types of damage that are often challenging to distinguish through manual inspection. Moreover, higher plants within the wall are clearly identified, and extensive area of falling can be accurately localized. Additionally, the model effectively detects discontinuities in the Arena case. These findings highlight the robustness of the DeepLabv3+ approach in segmenting damage within cultural heritage structures, even in the presence of significant environmental interferences.

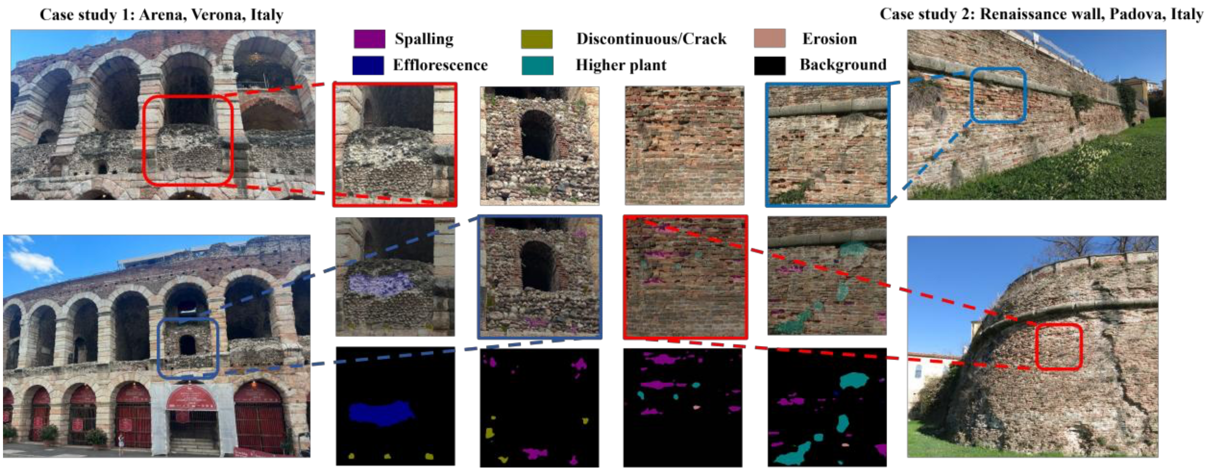


Fig. 4 Detection examples of damages in wall images by the established Deeplabv3+ model.

5 CONCLUSIONS

This paper presents a DeepLabv3+-based approach, leveraging an encoder-decoder network, for the segmentation of various types of damage in cultural heritage buildings, including spalling, discontinuities, erosion, efflorescence, and higher plants. To enhance the recognition capability of the proposed DeepLabv3+ model, the GridMask algorithm is employed for data augmentation. Preliminary results demonstrate that the trained DeepLabv3+ model effectively segments damages in cultural heritage structures. Two case studies further validate the accuracy and reliability of the proposed method. Future research will focus on improving the precision of damage segmentation. Additionally, efforts will be directed toward the automatic localization of damages, enabling the quantitative characterization of their dimensions and affected areas within cultural heritage buildings.

ACKNOWLEDGMENTS

This work was carried out in the framework of the partnership agreement CIPAR (China–Italy International Research Centre for Protection of Historical Architectures and Cultural Relics), supported by the University of Padua and Guangzhou University. China Postdoctoral Science Foundation (No. 2024M760619).

REFERENCES

[1] E. Valero, A. Forster, F. Bosché, E. Hyslop, L. Wilson, A. Turmel, Automated defect detection and classification in ashlar masonry walls using machine learning, Automation in Construction 106 (2019).
[2] A.Y. Yiğit, M. Uysal, Automatic crack detection and structural inspection of cultural heritage buildings using UAV photogrammetry and digital twin technology, Journal of Building Engineering 94 (2024).
[3] M.E. Hatır, İ. İnce, M. Korkanç, Intelligent detection of deterioration in cultural stone heritage, Journal of Building Engineering 44 (2021).
[4] K. Idjaton, X. Desquesnes, S. Treuillet, X. Brunetaud, Transformers with YOLO Network for Damage Detection in Limestone Wall Images, 21st International Conference on Image Analysis and Processing (ICIAP), Lecce, ITALY, 2022, pp. 302-313.
[5] Y. Yue, S. Zhang, H. Liu, F. Hong, C. Liu, J. Cui, Y. Du, Damage mechanism of a shield tunnel with cavities behind the concrete lining: An insight from a scaled model test, Tunnelling and Underground Space Technology 153 (2024).
[6] İ. Kocaman, Effect of restoration interventions on the seismic behavior of historical masonry buildings: The case of Molla Siyah mosque, Engineering Failure Analysis 148 (2023) 107206.
[7] N. Karimi, N. Valibeig, H.R. Rabiee, Deterioration detection in historical buildings with different materials based on novel deep learning methods with focusing on Isfahan historical bridges, International Journal of Architectural Heritage 18(6) (2024) 981-993.
[8] N. Wang, X. Zhao, P. Zhao, Y. Zhang, Z. Zou, J. Ou, Automatic damage detection of historic masonry buildings based on mobile deep learning, Automation in Construction 103 (2019) 53-66.
[9] F. Lorenzoni, F. Casarin, M. Caldon, K. Islami, C. Modena, Uncertainty quantification in structural health monitoring: Applications on cultural heritage buildings, Mechanical Systems and Signal Processing 66-67 (2016) 268-281.
[10] N. Karimi, M. Mishra, P.B. Lourenço, Automated surface crack detection in historical constructions with various materials using deep learning-based YOLO network, International Journal of Architectural Heritage (2024) 1-17.
[11] D. Loverdos, V. Sarhosis, Automatic image-based brick segmentation and crack detection of masonry walls using machine learning, Automation in Construction 140 (2022).
[12] H. Shakhatreh, A.H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N.S. Othman, A. Khreishah, M. Guizani, Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges, Ieee Access 7 (2019) 48572-48634.
[13] Y. Bai, B. Zha, H. Sezen, A. Yilmaz, Engineering deep learning methods on automatic detection of damage in infrastructure due to extreme events, Structural Health Monitoring 22(1) (2022) 338-352.
[14] P. Wang, J. Xiao, X. Qiang, R. Xiao, Y. Liu, C. Sun, J. Hu, S. Liu, An automatic building façade deterioration detection system using infrared-visible image fusion and deep learning, Journal of Building Engineering 95 (2024).
[15] I.O. Agyemang, X. Zhang, D. Acheampong, I. Adjei-Mensah, G.A. Kusi, B.C. Mawuli, B.L.Y. Agbley, Autonomous health assessment of civil infrastructure using deep learning and smart devices, Automation in Construction 141 (2022).
[16] M. Azimi, G. Pekcan, Structural health monitoring using extremely compressed data through deep learning, Computer‐Aided Civil and Infrastructure Engineering 35(6) (2020) 597-614.
[17] M. Watfa, A. Bykovski, K. Jafar, Testing automation adoption influencers in construction using light deep learning, Automation in Construction 141 (2022) 104448.

[18] P. Bhattarai, Implementing Deep Learning for Autonomous Infrastructure Management: From Predictive Analytics to Proactive Maintenance, Tensorgate Journal of Sustainable Technology and Infrastructure for Developing Countries 7(6) (2024) 1-10.
[19] C. Wang, S.E. Antos, L.M. Triveno, Automatic detection of unreinforced masonry buildings from street view images using deep learning-based image segmentation, Automation in Construction 132 (2021) 103968.
[20] B. Marin, K. Brown, M.S. Erden, Automated masonry crack detection with faster R-CNN, 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), IEEE, 2021, pp. 333-340.
[21] D. Brackenbury, I. Brilakis, M. DeJong, Automated defect detection for masonry arch bridges, International Conference on Smart Infrastructure and Construction 2019 (ICSIC) Driving data-informed decision-making, ICE Publishing, 2019, pp. 3-9.
[22] N. Wang, X. Zhao, L. Wang, Z. Zou, Novel System for Rapid Investigation and Damage Detection in Cultural Heritage Conservation Based on Deep Learning, Journal of Infrastructure Systems 25(3) (2019).
[23] K. Hacıefendioğlu, A.C. Altunışık, T. Abdioğlu, Deep learning-based automated detection of cracks in historical masonry structures, Buildings 13(12) (2023) 3113.
[24] J. Zhang, H. Qiu, J. Sun, Automatic detection of mortar loss on masonry building facades based on deep learning, Journal of Civil Structural Health Monitoring (2025) 1-12.
[25] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, Computer Vision and Pattern Recognition (2018).
[26] S. Cuypers, M. Bassier, M. Vergauwen, Deep Learning on Construction Sites: A Case Study of Sparse Data Learning Techniques for Rebar Segmentation, Sensors (Basel) 21(16) (2021).
[27] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The Pascal Visual Object Classes (VOC) Challenge, International Journal of Computer Vision 88(2) (2009) 303-338.
[28] P. Chen, S. Liu, H. Zhao, J. Jia, Gridmask data augmentation, arXiv preprint arXiv:2001.04086 (2020).