Applied Geomatics (2026) 18:94
https://doi.org/10.1007/s12518-026-00743-w

RESEARCH

Crack detection in churches using deep learning

Matteo Bozzano[1] · **Mohamedelmustafa Omer Eid**[2,3] · **Serena Cattari**[1] · **Bianca Federici**[1]

Received: 28 February 2026 / Accepted: 14 May 2026

Abstract

Post-earthquake damage assessment of churches is a critical task, due to the architectural complexity and the safety risks associated with on-site inspections, which expose surveyors to potentially dangerous structurally compromised areas. Among visible indicators, cracks represent one of the most relevant features for evaluating structural damage. As part of the wider project RAISE (NRRP-ECS00000035), this study proposes an image-based deep learning approach for automatic crack detection in church buildings, with the aim of supporting traditional survey procedures and reducing both inspection time and operator risk. A convolutional neural network based on transfer learning is developed and evaluated using the Xception architecture. The model is first trained on a large benchmark dataset already available of concrete crack images. Image patches are extracted and classified in a binary framework (crack vs. non-crack), and model performance is assessed using standard metrics such as Precision, Recall, and F1-score. While initial training on concrete images yields satisfactory results (98% accuracy), performance drops significantly when applied to church images, highlighting a strong domain shift. Retraining with church-specific data, a custom dataset collected from churches in the Liguria region, including surfaces characterized by frescoes, plaster, and decorative elements, substantially improves performance. It achieves a Precision of 93.87%, a Recall of 88.45%, and an F1-score of 91.08% on frescoed surfaces of the vaults, a fundamental element in the damage survey phase. Additional tests investigate the influence of lighting conditions, image resolution, and capture distance, providing practical guidance for acquisition protocol design. The results demonstrate the potential of deep learning techniques for supporting crack detection in heritage contexts, while also highlighting current limitations related to data variability and patch-level classification. Future developments will focus on dataset expansion, pixel-level segmentation approaches, and integration with standardized damage assessment frameworks.

Keywords Image · Immovable artistic assets · Damage · Binary classification · Convolutional neural network

✉ Matteo Bozzano
matteo.bozzano@edu.unige.it

Mohamedelmustafa Omer Eid
meid@fbk.eu

Serena Cattari
serena.cattari@unige.it

Bianca Federici
bianca.federici@unige.it

[1] Department of Civil, Chemical and Environmental Engineering (DICCA), University of Genova, Via Montallegro 1, Genova 16145, Italy

[2] 3DOM Research Unit, Bruno Kessler Foundation, Via Sommarive 18, Trento 38123, Italy

[3] Department of Civil, Building, and Environmental Engineering (DICEA), Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy

Introduction

Churches represent a fundamental component of cultural heritage, hosting both immovable assets—such as architectural structures, frescoes, and decorative elements—and movable cultural goods of high historical and artistic value. Preserving these buildings is therefore crucial not only from a structural perspective, but also for safeguarding cultural identity. However, churches are widely recognized as being particularly vulnerable to seismic damage, primarily because of their architectural configuration, e.g. often associated to the presence of slender elements prone to the activation of out of plane mechanism (D'Ayala 2000; Lagomarsino 2006, 2012; Marotta et al. 2021). Past seismic events have repeatedly demonstrated that churches often suffer more severe damage than ordinary masonry buildings

Published online: 26 May 2026

(Guerreiro et al. 2000; Lagomarsino 2012; Sorrentino et al. 2014; Penna et al. 2019), as shown in Fig. 1.

Following an earthquake, the assessment of damage to churches is a critical activity to support emergency management, ensure public safety, and plan conservation and restoration interventions. In Italy, since 1997 the post-earthquake damage assessment of churches has been carried out using a specific form, which was formally approved in 2001 by the Italian Civil Protection. This form is known as the A-DC Church Damage Form (Lagomarsino and Podestà 2004). Being the most advanced tool available in the literature within this field, the Italian form has been widely used also internationally (Ceroni et al. 2022). More recently, an improved version of this form has been proposed (Lagomarsino et al. 2019). These forms require trained technical teams to evaluate and classify the level of damage observed, based on visual inspections conducted both externally and internally.

Damage levels are primarily assigned based on the identification and extent of visible damage, such as cracking and partial collapses, which represent key indicators of seismic distress in masonry churches. Damage levels are usually attributed to some recurring mechanisms associated to predefined macroelements (e.g. main facade, nave, apse, vaults,...) recognized as architectural portions of the church that may exhibit an autonomous behaviour (Doglioni 1998; Lagomarsino et al. 2019). The surveys aim to identify damage severity and overall structural conditions, providing a rapid but systematic picture of the building's state. Currently, this process relies on on-site inspections performed by teams of experts who must access all parts of the church, including potentially unsafe areas. While this approach represents the standard practice today, it presents significant limitations. First, it is time-consuming, especially when a large number of buildings must be assessed in a short period following a seismic event. Second, and more critically, it raises serious safety concerns for surveyors. Elements such as vaults, arches, and elevated portions of the structure may be severely damaged and prone to sudden collapse, posing a high risk during inspections. Vaults, in particular, are among the most hazardous components, as their failure mechanisms can lead to unexpected and catastrophic collapses. In this context, the availability of automatic tools capable of identifying damage indicators, such as cracks, from images can provide valuable support to traditional inspection procedures, improving both efficiency and safety.

To address the above-mentioned limitations, significant research has focused on developing automated crack detection methods, which minimize the need for manual inspection or human intervention in general (Munawar et al. 2021; Li et al. 2022). In this context, the RAISE project (Robotics and AI for Socio-economic Empowerment - ECS00000035), funded by European Union - NextGeneration EU - Ministry of University and Research (MUR) within the National Recovery and Resilience Plan (NRRP), aims to improve the post-earthquake damage assessment process by exploiting advanced technologies to support and enhance traditional inspection procedures. One of the main objectives of RAISE is to reduce inspection time and increase operator safety, while maintaining, or improving, the quality and reliability of damage evaluation. Within this broader framework, the present study focuses on the automatic identification of cracks in church buildings using image-based techniques. Crack detection plays a key role in damage assessment, as cracks are often the first visible indicators of structural distress. Automating their identification can significantly support experts by providing objective, repeatable, and rapid information, potentially reducing the need for risky on-site inspections.

This work develops a crack detection model specifically tailored to church buildings. Among the available techniques, a deep learning approach based on transfer learning was adopted, enabling high accuracy while limiting the need for large task-specific datasets. In particular, the Xception

Fig. 1 Example of damage that occurred to churches hit by past earthquakes: **a**) St. Maria Maggiore Church in Mirandola (Emilian earthquake, 2012); **b**) St. Maria Paganica Church (L'Aquila earthquake, 2009); **c**) St. Benedetto di Norcia (Norcia earthquake, 2016/2017)

architecture is fine-tuned on a custom dataset of church images to obtain a binary classification model dedicated to crack detection in church structures. The paper is structured as follows. Section "Methods" presents the methodological framework adopted for crack detection, while deep learning approach, data acquisition strategy, and model training procedure are presented in Section "Data acquisition, training and performance metrics". Section "Results" discusses the experimental results and provides a detailed analysis of the model performance under different acquisition conditions. Finally, Section "Conclusion" summarizes the main findings of the study, highlights the limitations of the proposed approach, and outlines directions for future research.

Methods

Automatic crack detection refers to identifying cracks using 2D and/or 3D data. This procedure provides essential data for crack diagnosis, such as type, severity, and extent, which are critical for informed maintenance decisions. This diagnosis informs the optimal timing and treatment strategies in the intervention process (Hsieh and Tsai 2020). Methods such as laser, infrared, radiographic, and thermal testing have historically been employed to automate crack detection (Sánchez-Aparicio et al. 2023; Shrestha et al. 2025). Recently, however, there has been an increasing shift toward image-based techniques due to their processing efficiency (Munawar et al. 2021). Automatic damage detection from images in civil engineering is used in a variety of applications, including: bridge inspection, pavement crack detection, building inspection, in addition to other applications (Mathavan et al. 2015; Marchewka et al. 2020; Munawar et al. 2022; Shahbazi et al. 2024). In seismic engineering, it has applications in post-earthquake damage assessment (Xu et al. 2023), in monitoring of structural health (Belloni et al. 2023), and in damage prediction and forecasting. Processing algorithms have been designed to detect cracks (Hampel and Maas 2009; Barazzetti and Scaioni 2009) or to characterize their properties (Liebold et al. 2020).

Numerous methods can be used to process acquired photos, including: Image Processing (IP), Digital Image Correlation (DIC), classical Machine Learning (ML), and Deep Learning (DL). Image Processing is often paired with ML or deep learning for better results (Kim et al. 2018). Previous studies have successfully implemented these methods, with their corresponding datasets made publicly accessible for further research and validation (Kaveh and Alhajj 2024). However, none of the existing works have specifically focused on crack detection in churches. Churches present unique challenges for crack detection due to their distinct construction materials and features. Unlike concrete walls, church walls often include frescoes, plaster, and stucco, which can affect both the appearance of cracks and the performance of detection methods.

Traditional ML algorithms enable object detection and segmentation using low-dimensional data based on simple, manually selected features (Hsieh and Tsai 2020; Tahir et al. 2022). Conversely, DL provides advanced functionalities such as automatic feature detection, robust generalization, and high precision, with reported classification accuracies exceeding 90%, for crack detection on materials such as concrete (Alipour et al. 2019; Dorafshan et al. 2019; Golding et al. 2022). There is no general best option between ML and DL, as the go-to option depends on the objective as well as on the dataset (Matrone et al. 2020). Within DL, tasks can range from simple binary classification, determining whether an image contains a crack or not, to more complex segmentation, which identifies the pixels belonging to individual cracks, as shown in Fig. 2. Intermediate tasks include object detection, which localizes cracks using bounding boxes, as visible in Fig. 2c. These bounding boxes can then be further analyzed for type-specific crack classification (Cha et al. 2017; Mandal et al. 2018; Xue and Li 2018).

Neural networks have already been extensively employed for crack identification. In some cases, entirely new neural networks with multiple convolutional or fully connected layers have been designed for the detection of cracks in pavements and asphalt (Zhang et al. 2016; Wang et al. 2017) and in concrete (Rezaie et al. 2020; Zadeh et al. 2024). Neural networks are sometimes integrated with other techniques, such as the sliding window technique, to enhance performance, particularly when working with datasets containing few images. However, it is generally challenging to achieve good results with a neural network designed from scratch using fewer than 10,000 images (Cha et al. 2017). Another critical aspect is the number of layers in the network: while increasing the number of layers can improve performance, it may compromise the network's ability to generalize effectively (Pauly et al. 2017). An important alternative is the use of transfer learning, which reduces the need for large datasets and expensive hardware. This approach leverages pretrained models that are further fine-tuned for a specific task using only a small number of example images. This technique allows for successful task execution even with fewer than 1,000 training images (Gopalakrishnan et al. 2017).

The identification of cracks is necessary for the assessment of damage at the different scales, at which structures can be analyzed: (a) the single element or macro-element scale, which refers to a portion of an architectural asset that allows for the analysis of its seismic behavior independently of the rest of the structure; (b) the single building scale; and (c) the urban scale. At the macro-element scale, laboratory tests have been conducted on masonry walls to compare

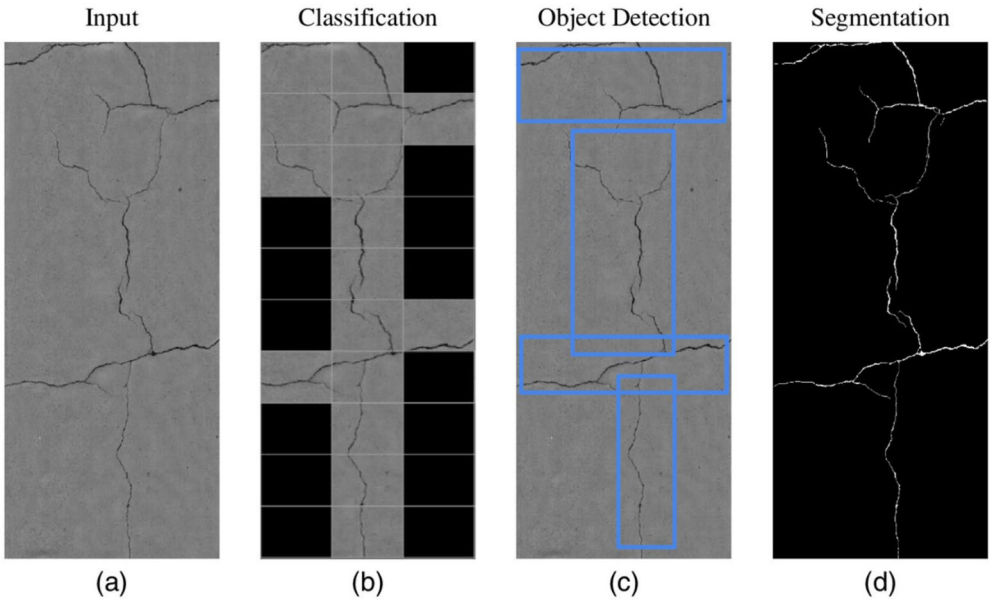


Fig. 2 (**a**) Input image; (**b**) Classification; (**c**) Object Detection; and (**d**) Segmentation (Hsieh and Tsai 2020)

different crack identification methodologies, showing promising results by deep learning (Rezaie et al. 2020). On the single building scale, the combined use of UAVs, deep learning, and photogrammetry has already been proposed for crack identification in two different buildings within a simulated environment (Ko et al. 2021). A recent paper proposed an integrated methodology relying on photogrammetric survey and AI-driven recognition of masonry texture, for non-linear mechanical analysis (Chura et al. 2024). Another example of crack identification is an end-to-end pipeline for automatically generating Damage Augmented Digital Twins of freestanding buildings, which integrates 3D simplified models and cracks (Pantoja-Rosero et al. 2023). At the urban scale, the topological loss (TOPO) has been employed to train a deep learning network, yielding convincing results when applied to images of cracked buildings in urban scenes (Pantoja-Rosero et al. 2022).

This study utilizes Convolutional Neural Networks (CNNs) for patch-level binary classification, where each image patch is classified as a crack or non-crack. The used model is based on Google's Xception, a deep convolutional neural network architecture that involves Depthwise Separable Convolutions, and that was developed by Google researchers (Chollet 2017). Using patches is advantageous because it substantially increases the number of training samples, effectively expanding the dataset. Moreover, it allows to quickly determine, albeit approximately, the position of cracks within the structure through the automatic analysis of many images, such as those derived from a photogrammetric survey. The output of the classification pipeline consists of a spatial map of labeled patches superimposed on the original input image. Each patch is assigned a binary label (crack or non-crack) and the reassembled result allows the operator to visually identify which regions of the surveyed surface contain cracks. The patch-level output does not provide metric information about individual cracks (length, width, orientation, continuity...) nor does it distinguish between crack typologies. The output is therefore best interpreted as a screening tool: a rapid, automated first-pass indicator of damaged regions, rather than a quantitative geometric characterization of individual cracks.

The methodology, designed for the development and evaluation of an image-based crack detection model, is structured in four main phases: data acquisition, image preprocessing, model training, and model performance evaluation. An overview of this approach is presented in Fig. 3. The process begins with the acquisition of visual data in the form of standard photographs and 360° videos of church interiors and exteriors. This phase, in the RAISE project, is supported by the usage of robots (tracked and legged) and drones. The 360° videos are subsequently converted into conventional video formats, from which image frames are extracted to ensure comprehensive coverage of structural surfaces.

The collected images are then pre-processed by converting them to grayscale and subdividing into smaller image patches. This step aims to reduce computational complexity while enhancing the visibility of local features relevant to crack detection. The resulting dataset of image patches is divided into two subsets, with 75% allocated for model training and 25% reserved for validation and testing purposes.

Following data preparation, the training phase is carried out using the selected deep learning architecture, during which the model learns to discriminate between crack

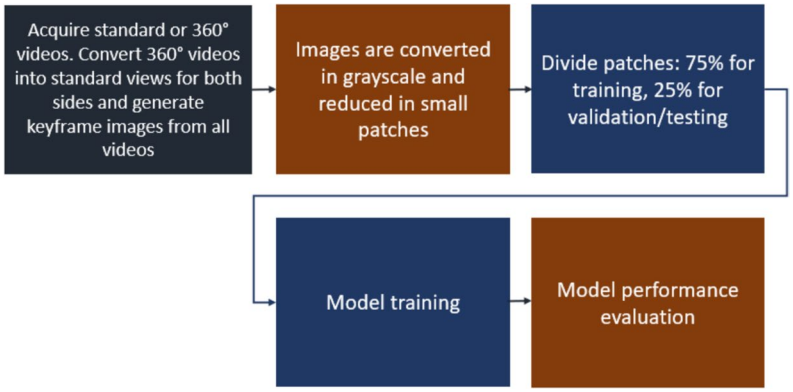


Fig. 3 Methodology flowchart

Table 1 Google Xception model summary

Layer (type)	Output Shape	Param#
Gray image (InputLayer)	(None, None, None, 1)	0
Graylevel to RGB (Conv2D)	(None, None, None, 3)	30
Xception	(None, None, None, 2048)	20,861,480
Pooling (MaxPooling2D)	(None, None, None, 2048)	0
Dropout_2 (Dropout)	(None, None, None, 2048)	0
Pooling_2 (GlobalAveragePooling2D)	(None, 2048)	0
Fully connected (Dense)	(None, 256)	524,544
Output layer (Dense)	(None, 1)	257
Total params:		**64,049,879**
Trainable params:		**21,331,783**
Non-trainable params:		**54,528**
Optimizer params:		**42,663,568**

and non-crack patterns. Finally, the trained model is evaluated using the validation dataset to assess its performance through standard classification metrics.

Data acquisition, training and performance metrics

To create a deep learning model capable of detecting cracks in complex settings, such as frescoed church walls, a pre-trained model is used, supported by a specifically designed training dataset to help it complete the task. Pre-trained models generally require fewer training images and are often more effective in tasks with limited data (Han et al. 2021). In this case, the selected model was Google's Xception (Chollet 2017). This was selected due to its reliability and high accuracy in crack detection, outperforming other pre-trained models in similar contexts. A summary of the characteristics of the selected model and the layers composing the Xception network can be seen in Table 1. A 1-channel grayscale input is first converted to 3 channels with a small Conv2D layer so it can be fed into the Xception backbone, which extracts high-level features (ending with 2048 feature channels). The subsequent pooling, dropout, and global average pooling reduce spatial dimensions and help regularize. Then a dense layer maps the 2048-D feature vector to 256 units, and the final dense layer outputs a single value for binary classification, usually with a sigmoid function. The table is concluded by the parameters number and their subdivision.

At first, training was conducted with an initial dataset of cracked concrete images, known as Concrete Crack Images for Classification dataset (Özgenel 2019). This dataset, compiled from images of various buildings on the Middle East Technical University (METU) Campus, is divided into two categories: positive and negative crack images (Özgenel and Sorguç 2018), examples of which are shown in Figs. 4 and 5. Each category contains 20,000 images, for a total of 40,000 images at a resolution of 227x227 pixels in RGB format. The dataset originates from 458 high-resolution images

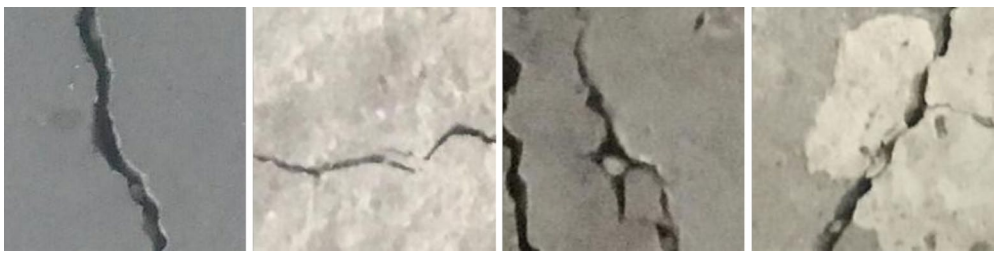


Fig. 4 Positive crack images from the Concrete Crack Images for Classification dataset (Özgenel 2019)

Fig. 5 Negative crack images from Concrete Crack Images for Classification dataset (Özgenel 2019)

Fig. 6 Sites of the first in-situ surveys: the San Francesco d'Albaro church (on the left), the Oratory of Nostra Signora Assunta in Genova Coronata (in the middle) and the meeting room of the Dean of the Polytechnic School at the University of Genova, in Villa Giustiniani Cambiaso (on the right)

Fig. 7 Sites of the second in-situ surveys: Aggio (on the left), Montoggio (in the middle) and Vobbia (on the right) churches

(4032x3024 pixels) created using the method proposed by Zhang et al. (2016), with diverse surface finishes and lighting conditions, though without any data augmentation techniques. From this dataset, 28,000 images were used to train the model, equally divided into 14,000 labeled as "crack" and 14,000 labeled as "non-crack." The remaining 12,000 images were used for validation and test, in 50:50 ratio.

Subsequently, a second training phase was conducted to enable the model to identify cracks on surfaces with additional visual complexity, such as frescoes and decorative elements. This was achieved by creating a custom dataset, hereinafter referred to as the *University of Genova (UniGe) dataset*, gathered from churches and buildings located in the Liguria region. Specifically, two rounds of data collection were carried out. The first round, conducted between September and October 2023, employed a DJI Avata drone and an insta360 ONE RS spherical camera to survey the walls and vaults of the meeting room of the Dean of the Polytechnic School at Villa Giustiniani Cambiaso, the Nostra Signora Assunta Oratory in Genova Coronata, and the San Francesco d'Albaro Church (Fig. 6). All three buildings are located within the municipality of Genova. A second round of surveys took place in January 2024 in the churches of San Giovanni Battista in Aggio, San Giovanni Decollato in Montoggio, and Nostra Signora delle Grazie in Vobbia (Fig. 7). Aggio is a district within the municipality of Genova, while Montoggio and Vobbia are small municipalities in the province of Genova. The churches were selected with the assistance of the Genoese Curia, based on their diverse floor plan typologies and the presence of cracks on vaults and/or walls. A sketch outlining the layout of the three churches is shown in Fig. 8. For this second round of surveys, the same instruments were used. They are visible in Fig. 9.

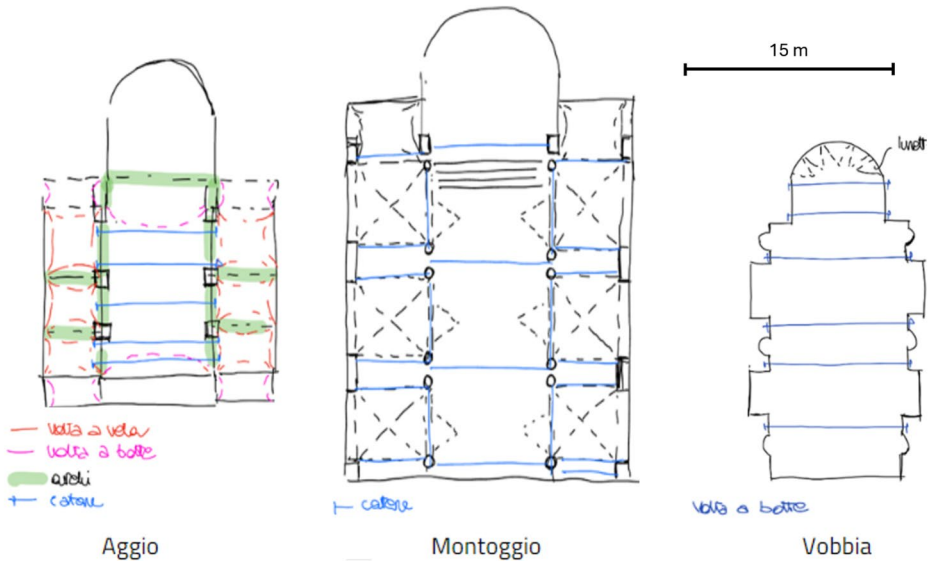


Fig. 8 In-plan configuration of Aggio, Montoggio and Vobbia churches. The drawing was scaled based on laser point cloud data for each church

Fig. 9 Used instruments: a DJI Avata drone equipped with an Insta360 spherical camera

The spherical camera was chosen for its ability to capture both side walls and the ceiling in a single flight. This capability is especially advantageous in post-earthquake assessments, where rapid data collection and operational efficiency are critical for damage evaluation and recovery planning. The 360-degree field of view provided by the camera minimizes the need for multiple passes, thereby accelerating structural assessments and enhancing workflow efficiency. From these surveys, a set of images was obtained, from which the best-quality images were selected. Prior to training, the images were converted to grayscale. These images were then divided into 1,858 patches, each measuring 227x227 pixels. Of the resulting dataset, 75% was allocated for training, while the remaining 25% was used for model validation. Within this training set, 539 patches were labeled as "crack" and 854 as "non-crack." This disparity ensures adherence to a maximum 1:3 ratio between crack and non-crack patches, which is essential for preventing model bias (Fan et al. 2018). The validation set was split into 180 patches labeled as "crack" and 285 as "non-crack."

In evaluating the model, the selected classifier was assessed based on its ability to predict classes, where positive denotes a crack and negative denotes a non-crack. An example is visible in Figs. 10 and 11. A false positive (FP) is a non-crack (negative) mistakenly predicted as a crack (positive). A false negative (FN) is a crack (positive) mistakenly predicted as a non-crack (negative). True positives (TP) are correctly identified cracks, and true negatives (TN) are correctly identified non-cracks. This can be clarified by the confusion matrix in Fig. 12.

The model's classification performance was evaluated using three metrics: Precision, Recall, and F1-score. These are standard evaluation metrics widely adopted in the machine learning and computer vision literature (Sokolova and Lapalme 2009), as in recent works on crack detection (Golding et al. 2022; Zadeh et al. 2024). Precision measures how many predicted positives were correct, Recall assesses how well the model identified actual positives, and the F1-score combines both metrics into a harmonic mean, providing a balanced performance measure, especially in cases

Fig. 10 Examples of positive crack patches in the UniGe Dataset

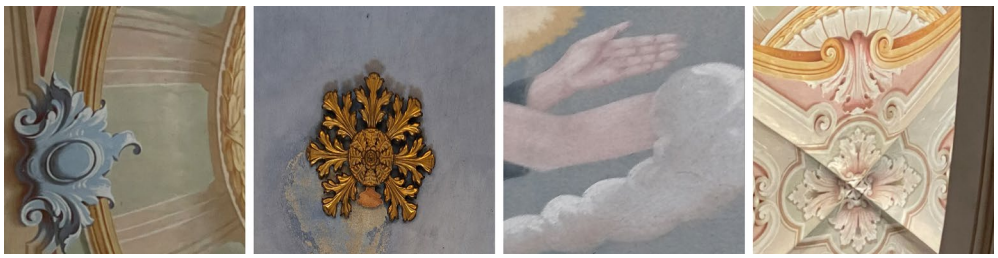


Fig. 11 Examples of negative crack patches in the UniGe Dataset

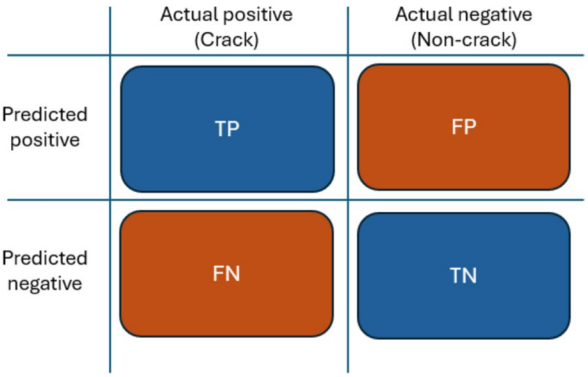


Fig. 12 Confusion matrix

of imbalanced datasets. Their formulas are respectively shown in Eqs. 1, 2 and 3.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

In the context of crack detection, Precision measures the proportion of predicted cracks that were actual cracks, Recall evaluates the model's success in identifying all true cracks among the images that should have been classified as such, and F1-score balances Precision and Recall to provide an overall assessment of the model's accuracy in crack detection.

Beyond the initial evaluations, the model was further tested using images with distinct characteristics to assess the effect of different acquisition conditions. Categories were established, grouping images by specific attributes: lighting conditions, resolution, and capture distance. For lighting conditions, bright images were categorized separately from darker ones. To assess the impact of resolution, images were acquired at different resolution levels. For comparison, 4K images were captured in standard mode (3840 $\times$ 2160 pixels, 16:9), and 3K spherical images (3008 $\times$ 1504 pixels) were resized to standard frame dimensions, yielding a resolution of 1920 $\times$ 1080 pixels. The last categorization was based on capture distance, distinguishing images taken close to the walls from those taken further away. An example of the last mentioned group is visible in Fig. 13.

To evaluate model performance under each acquisition condition, a dedicated set of image patches was reserved exclusively for category-specific testing and kept strictly separate from both the training and validation sets used in the main model development pipeline. In total, 3,800 test patches were withheld across the three acquisition categories: 2,365 patches for the lighting condition tests (389 crack, 1,976 non-crack), 823 patches for the resolution tests (321 crack, 502 non-crack), and 612 patches for the acquisition distance tests (219 crack, 393 non-crack). For each category, the available images were split into two equal subsets: the first was incorporated into the training set alongside the existing UniGe data to ensure prior model exposure to each acquisition condition, while the second was withheld and used exclusively for category-specific evaluation. This design ensures that the observed performance variations reflect the intrinsic difficulty of each acquisition condition,

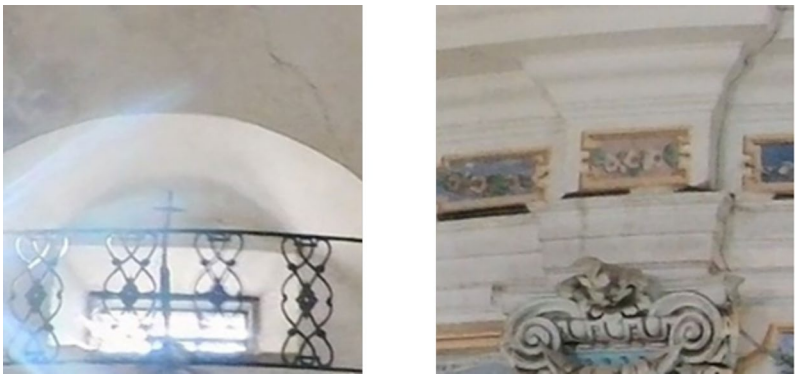


Fig. 13 Example for crack images (on the left) and non-crack images (on the right) in the Far distance category

rather than the model's unfamiliarity with that image type, thereby providing objective and condition-specific guidance for acquisition protocol design. Without this exposure, it would not be possible to distinguish between performance loss caused by the condition itself and that caused by the model never having encountered such images during training. The results of this evaluation are therefore intended to provide objective, condition-specific guidance for acquisition protocol design, informing decisions such as the need for artificial lighting, the minimum acceptable resolution, the maximum operational distance, and the most suitable sensor or platform.

Results

Before presenting the results, it should be noted that the validation patches referenced in all summary tables represent held-out subsets that were not included in any training phase and were used exclusively for the evaluation of the specific test they refer to. In each case, the reported patches constitute an independent test set, ensuring that the performance metrics reflect the model's ability to generalize to unseen data rather than its fit to the training distribution.

The Xception model was initially trained using the 28,000 images from the Concrete Crack Images for Classification dataset. This training process yielded the training and validation plots presented in Fig. 14. The metrics used in the plot are accuracy (left in Fig. 14) and loss (right in Fig. 14), computed for both the training and validation sets. Accuracy is dimensionless and expressed as a ratio between 0 and 1 (or 0-100%), representing the percentage of correctly classified samples. The epochs, displayed on the x-axis, are complete training cycles in which the model processes the entire training dataset. During each epoch, the neural network updates its weights through backpropagation after analyzing all image batches. In the presented graph, the Xception model completes 7 epochs. Convergence does not have a fixed predefined value but is evaluated empirically by observing when training and validation loss stabilize and cease to improve significantly. In this specific case, the model achieves practical convergence around epoch 5-6, where validation accuracy settles at approximately 0.98 (98%) and the loss remains stable at around 0.1. The stopping criterion can terminate training when validation loss fails to improve for a predetermined number of consecutive epochs, thereby preventing overfitting. In fact, overfitting becomes apparent when training accuracy continues to increase while validation performance stagnates. The model demonstrates proper generalization performance on the validation data set. The plots suggest that the model has effectively converged and does not show clear indications of high bias or variance. Further training epochs are unlikely to yield significant improvements in the model's predictive accuracy.

The trained model was evaluated on a held-out test set consisting of 6,000 image patches (3,000 positive and 3,000 negative), drawn from the same Concrete Crack Images for Classification dataset, and kept strictly separate from both the training and validation splits. The test set was balanced,

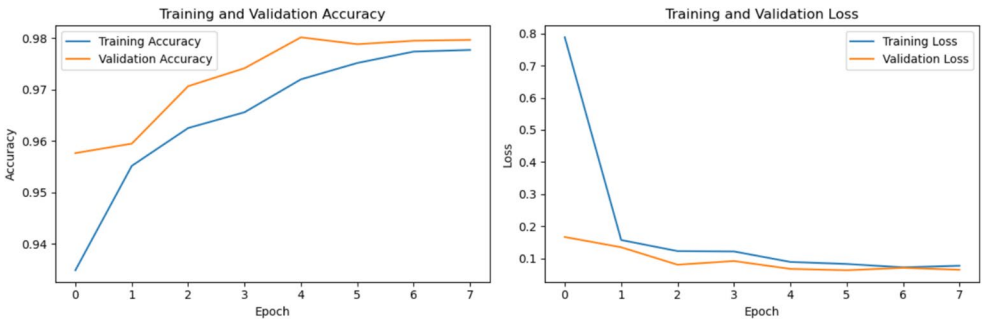


Fig. 14 Xception model trained using the Concrete Crack Images for Classification data set

Table 2 Concrete Crack Images for Classification dataset split and model performance metrics

	Train	Validation	Test
Positive (crack)	14,000	3,000	3,000
Negative (no-crack)	14,000	3,000	3,000
Total	**28,000**	**6,000**	**6,000**
Ratio	70%	15%	15%
Input shape	227 × 227		
Loss	0.0771	0.0646	0.0671
Accuracy	97.77%	97.97%	**98.03%**

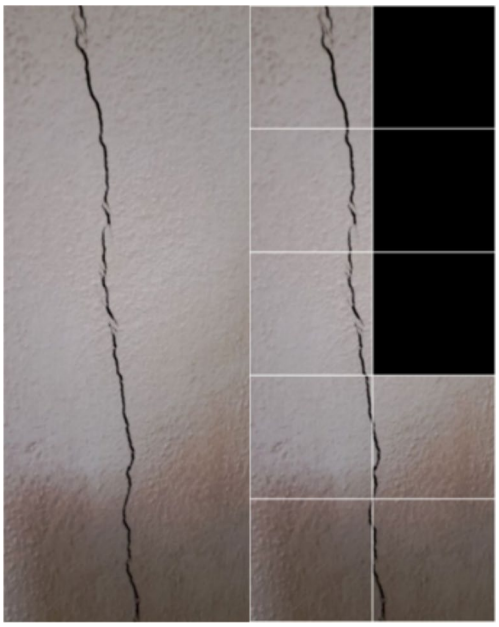


Fig. 15 Concrete image (on the left) and the corresponding image classified in crack patches and non-crack one, the latter represented in black (on the right)

with an equal number of crack and non-crack samples, and no preprocessing beyond grayscale conversion was applied. The model achieved a final accuracy of 98.03% with a cross-entropy loss of 0.067, consistent with the validation accuracy of 97.97% recorded at the last training epoch, confirming the absence of significant overfitting. All data referring to the usage of the Concrete Crack Images for Classification dataset are shown in Table 2.

Table 3 Model performance on original church images prior to domain-specific retraining

Class	Accuracy	Correctly Classified
Crack	44.6 %	37/83
Non-crack	52.6 %	131/249

An example of a crack identified by the model in a concrete image is shown in Fig. 15. The single patches are rearranged, indicating the location of the crack in the image.

The model was then applied to a part of the images acquired along church walls and vaults, characterized by frescoes and decorations as showed in Figs. 10 and 11, to evaluate its ability to identify cracks in so complex pictures even if it was not trained yet for this kind of complexity. The result is graphically shown in Fig. 16. The model performed modestly when tested on original church images, specifically 332 test patches: out of 83 crack patches, the model was able to identify only 37 (44.6% accuracy). In non-crack patches, the model's performance was slightly better, identifying 131 out of 249 non-crack patches (52.6% accuracy). These values are summarized in Table 3.

The model was subsequently retrained using images collected from the local churches, portraying decorated surfaces. The result was that the model accurately identified 245 out of 277 crack images, achieving an accuracy of 88.45%, and correctly classified 369 out of 385 non-crack images, yielding an accuracy of 95.84%. The histogram binning is visible in Fig. 17. The evaluation metrics for this model were as follows: Precision was 93.87%, Recall was 88.45%, and the F1-score was 91.08%, as they are shown in Table 4.

Since the model's results showed strong confidence in predicting non-crack images, as can be seen by the histogram binning in Fig. 17(b), this allowed for flexibility in adjusting the threshold to improve crack detection. In fact, it is possible to choose the threshold value on the x-axis of the histogram, at which to set the boundary between crack and

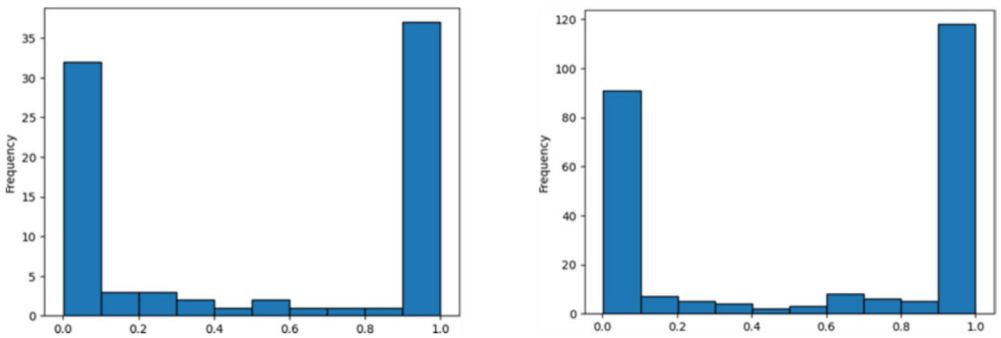
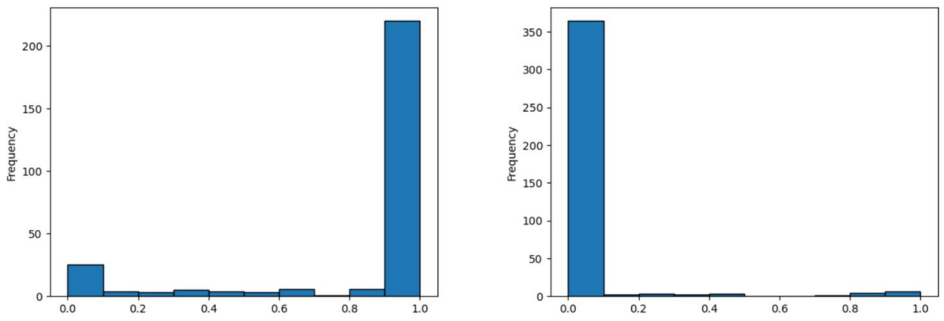


Fig. 16 Ability to identify crack patches (on the left) and non-crack patches (on the right), using the model trained exclusively on concrete crack imagery



(a) Crack patches (b) Non-crack patches

Fig. 17 Ability to identify crack images (**a**) and non-crack images (**b**), after training on church-specific imagery from the UniGe dataset

Table 4 Model performance on the UniGe dataset after domain-specific retraining

Model	Validation patches	Crack accuracy	Non-crack accuracy	Accuracy metrics
Domain-specific Xception	662	245/277	369/385	Precision: 93.87%
	(277 cracks, 385 non-cracks)	(88.45%)	(95.84%)	Recall: 88.45%
				F1-score: 91.08%

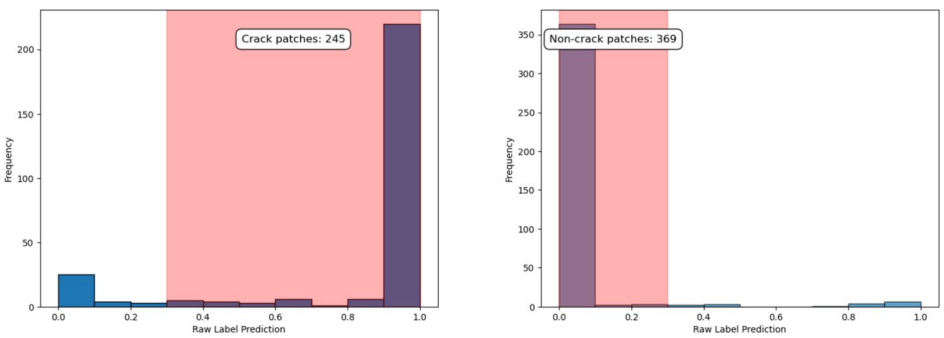
no crack. To maximize Recall, the threshold was lowered to 0.3, rather than the conventional 0.5. This approach aimed to reduce the likelihood of overlooking actual cracks. This resulted in obtaining a new clustering, the pink area showed in Fig. 18.

Figure 19 shows the model's ability to detect cracks on a decorated wall. In fact, most of the patches characterized by cracks are highlighted in the image.

However, the analysis of model accuracy during the training and validation steps, shown in Fig. 20, highlights the model's good performance but suggests the presence of overfitting, evidenced by the discrepancy between training and validation accuracy. Overfitting occurs when a model learns the training data too well, memorizing specific patterns and noise rather than generalizing underlying relationships. This phenomenon is evident when training accuracy continues to improve while validation accuracy plateaus or degrades, indicating that the model has adapted excessively to the training set's peculiarities. Overfitting significantly affects the model's usability by reducing its ability to perform accurately on unseen data in real-world applications. While the model may achieve excellent results on the training dataset, it will likely produce poor predictions when deployed for structural crack detection on new church images, limiting its practical value for autonomous inspection systems where reliable generalization is essential.

Following the initial retraining on the UniGe dataset, the model was evaluated under the acquisition categories described in Section "Data acquisition, training and performance metrics", with the aim of quantifying performance degradation under each constrained condition and providing practical guidance for acquisition protocol design. The category-specific test results are presented in Tables 5, 6, and 7, while the validation results of the final model are reported in Table 8.

The results referring to images acquired under different lighting conditions are shown in Table 5. A significant impact of lighting conditions on crack detection performance is evident, particularly for the crack class. Under dark lighting conditions, the model achieves only 64% accuracy for crack identification, compared to 85% under bright lighting, a substantial 21% performance gap. This discrepancy indicates that insufficient illumination severely



(a) Crack patches (b) Non-crack patches

Fig. 18 Impact of adjusting the threshold on the detection of crack patches. Plots are referred to training on the UniGe dataset

Fig. 19 Image of a decorated wall (on the left) and the corresponding image classified in crack patches and non-crack one, the latter represented in black (on the right)

Fig. 20 Training and validation accuracy of the trained on decorated images model

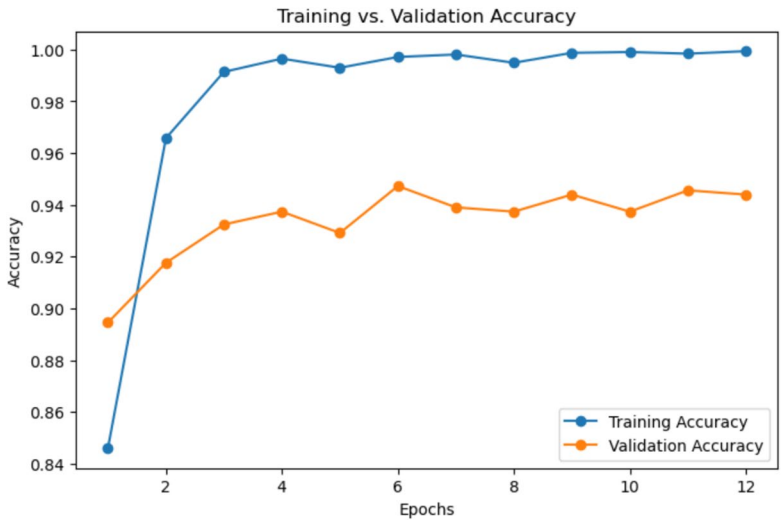


Table 5 Classification accuracy under different lighting conditions

Lighting	Class	Accuracy	Discussion
Dark	Crack	103/162 (64%)	59 patches incorrectly classified; 36 patches non-visible.
	Non-crack	887/949 (93%)	62 patches incorrectly classified; 27 patches non-visible.
Bright	Crack	193/227 (85%)	21% difference in accuracy between bright and dark images.
	Non-crack	941/1027 (92%)	Non-crack patches are seemingly unaffected by lighting conditions.

hampers the model's ability to correctly identify cracks, with 36 non-visible patches further complicating detection. Conversely, the non-crack class maintains consistently high accuracy across both lighting scenarios (93% dark, 92% bright). These values suggest that the use of an artificial light source is advisable when acquiring images in poorly illuminated environments, such as church interiors, as dark conditions result in a 21% drop in crack detection accuracy, potentially compromising the reliability of the inspection.

To investigate the effect of image resolution, the mixed-resolution basic model was fine-tuned separately on 3K-only

Table 6 Performance metrics at different image resolutions

Resolution	Crack acc.	Non-crack acc.	Metric	Basic model	New model	Variation
3K	62/72 (86.1%)	134/159 (84.3%)	Precision	93.5%	71.3%	−22%
			Recall	87.8%	86.1%	−2%
			F1-Score	90.5%	78%	−13%
4K	210/249 (84.3%)	313/343 (93.7%)	Precision	93.5%	87.5%	−6%
			Recall	87.8%	84.3%	−4%
			F1-Score	90.5%	85.9%	−5%

Table 7 Performance metrics at different acquisition distances

Distance	Crack acc.	Non-crack acc.	Metric	Basic model	New Model	Variation
Close	180/212 (84.9%)	314/366 (85.8%)	Precision	93.5%	77.6%	−16%
			Recall	87.8%	84.9%	−3%
			F1-Score	90.5%	81%	−10%
Far	6/7 (85.7%)	14/27 (51.8%)	Precision	93.5%	31.6%	−62%
			Recall	87.8%	85.7%	−2%
			F1-Score	90.5%	45%	−46%

Table 8 Xception model performance after fine-tuning with patches from the acquisition tests

Model	Validation patches	Crack accuracy	Non-crack accuracy	Accuracy metrics
Fine-tuned Xception	669	252/284	372/385	Precision: 95.1%
	(284 cracks, 385 non-cracks)	(88.7%)	(96.6%)	Recall: 88.7%
				F1-score: 91.8%

and 4K-only subsets of the training data, and evaluated on the corresponding held-out test partitions. This approach assesses whether specializing the model on a single resolution source improves performance relative to the mixed-resolution basic model, simulating scenarios where only one camera type is available during acquisition. Table 6 presents the comparative results. When applied to spherical 3K images (equivalent to Full HD resolution when converted to standard format), the fine-tuned model correctly identified 62/72 crack patches (Recall: 86.1%) and 134/159 non-crack patches (84.3%), with a notable drop in Precision of 22% and F1-score of 13% relative to the basic model. For standard 4K images, the degradation was more modest, with Precision dropping by 6% and F1-score by 5%, correctly classifying 210/249 crack patches (84.3%) and 313/343 non-crack patches (91.2%). These results suggest that 4K resolution represents an acceptable trade-off between image quality and acquisition flexibility, whereas 3K resolution introduces a significant loss in Precision that may compromise the reliability of crack detection.

Finally, the effect of acquisition distance was evaluated by testing the model separately on close and far-distance image subsets. Table 7 presents the comparative results. For close images, the model correctly identified 180/212 crack patches (84.9%) and 314/366 non-crack patches (85.8%), indicating robust and balanced performance at shorter distances. For far-distance images, crack detection remained relatively stable at 85.7% (6/7 patches), whereas non-crack classification dropped dramatically to 51.8% (14/27 patches). It should be noted that the far-distance subset contains a very limited number of samples, which reduces the statistical reliability of these estimates. Nevertheless, the sharp decline in non-crack accuracy is consistent with the reduced spatial resolution at greater distances, which may cause textured backgrounds or shadows to be misclassified as cracks. These results suggest that close acquisition is preferable for reliable crack detection, particularly to avoid false positives on non-crack surfaces.

The fine-tuned model, thanks to the addition of images relating to tests carried out by changing the distance, resolution, and light, improved the domain-specific model, initially trained on the Unige dataset relating to cracks in churches. Table 8 presents the validation results obtained after fine-tuning the model, in order to identify the acquisition conditions most suitable for reliable crack detection. As shown, the resulting metric values remain largely consistent with those of the domain-specific model reported in Table 4.

To summarize, the analysis of model performance under different acquisition conditions reveals significant limitations related to resolution, lighting, and shooting distance. Lower resolution images (3K) resulted in precision reductions of 22%, while lighting conditions demonstrated an asymmetric impact particularly critical for crack detection: dark images achieved only 64% accuracy compared to 85% under optimal illumination conditions. Images acquired from a distance presented the most severe limitation, with non-crack accuracy dropping drastically to 51.8%, primarily due to reduced spatial resolution and limited sample

size. These results underscore the necessity of ensuring controlled acquisition conditions consistent with the original training set, or alternatively implementing more robust data augmentation strategies to improve model generalization in real-world operational scenarios characterized by variability in lighting and shooting distance.

Conclusion

This study presented a deep learning approach for automatic crack detection in church buildings, a domain largely unexplored in the existing literature despite its relevance for post-earthquake damage assessment. The Xception architecture was adopted through a two-stage transfer learning strategy: first trained on a large-scale concrete crack dataset, then fine-tuned on a dedicated dataset collected from churches in the Liguria region, comprising surfaces characterized by frescoes, plaster, and decorative elements.

The initial training on concrete imagery confirmed the effectiveness of the transfer learning approach, achieving near-perfect performance on the source domain (98% accuracy). However, direct application to church images revealed a substantial domain shift, with crack accuracy dropping to 44.6%. This finding underscores a fundamental challenge: visual features learned on plain concrete surfaces do not generalize to heritage contexts, where pictorial patterns, non-uniform illumination, and heterogeneous materials radically alter the appearance of both cracks and background regions.

Domain-specific retraining on the UniGe dataset substantially mitigated this gap, achieving 88.45% crack accuracy and 95.84% non-crack accuracy, with a well-balanced F1-score of 91.08%. While these results confirm the viability of the proposed approach for decorated surfaces, the discrepancy between training and validation accuracy indicates a tendency toward overfitting, attributable to the limited size and site-specific nature of the training set. Further validation on a broader and more diverse set of churches is therefore necessary to quantify the model's generalization capability in new architectural settings.

The acquisition condition tests provided actionable guidance for field deployment. Illumination was the most impactful factor, with crack accuracy dropping by 21 percentage points under dark conditions, suggesting that artificial lighting should be considered mandatory for interior surveys. Resolution analysis showed that 4K imagery represents an acceptable trade-off, while 3K spherical images introduce a significant precision loss of 22%. The threshold adjustment from 0.5 to 0.3, aimed at maximizing Recall, proved effective for safety-first screening scenarios, though at the cost of increased false positives; threshold selection should therefore be adapted to the operational objective.

The final output consists of a binary spatial map of crack/non-crack patches overlaid on the original image (right image in Fig. 19). When integrated within a photogrammetric pipeline, the binary patch map can be overlaid onto georeferenced orthophotos of the surveyed surfaces, enabling approximate spatial localization of crack-affected regions in metric coordinates.

While approximate localization of damaged regions at the patch scale is enabled, a key limitation of this formulation is the inability to automatically characterize crack geometry, including length, width, and continuity, or to distinguish between crack typologies, all of which are relevant for engineering-grade damage assessment. Hence, the natural extension of this work is the integration of pixel-level segmentation models, which would enable quantitative crack measurements and more informative outputs. Furthermore, linking detected crack patterns to the logic of standardized post-earthquake church assessment forms, such as the A-DC form, represents a fundamental step toward operational deployment within digital damage assessment workflows. In this perspective, the proposed tool paired with robotics and drones surveys, which is one of the goals of the aforementioned RAISE project, has the potential to contribute in reducing the time required for post-earthquake inspections compared to traditional on-site surveys, while also improving operator safety by minimizing the need for direct access to structurally compromised areas, such as damaged vaults and elevated portions of the building where the risk of sudden collapse is highest.

Acknowledgements This work was supported by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project "RAISE -Robotics and AI for Socio-economic Empowerment" (ECS00000035). Serena Cattari and Bianca Federici are part of RAISE Innovation Ecosystem. The authors thank the Curia arcivescovile di Genova for the assistance in selecting the churches and providing access to them.

Author Contributions Conceptualization: S.C., B.F.; Methodology: M.M.O.E., S.C., B.F.; Data acquisition and field survey: M.B., M.M.O.E.; Investigation and validation: M.M.O.E.; Writing - original draft preparation: M.B.; Writing - review and editing: S.C., B.F.; Supervision: S.C., B.F.; Funding acquisition: S.C., B.F. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement. Funded by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project "RAISE -Robotics and AI for Socio-economic Empowerment" (ECS00000035).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

Alipour M, Harris DK, Miller GR (2019) Robust pixel-level crack detection using deep fully convolutional neural networks. J Comput Civ Eng 33(6):04019040. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000854

Barazzetti L, Scaioni M (2009) Crack measurement: Development, testing and applications of an automatic image-based algorithm. ISPRS J Photogrammet Remote Sens 64(3):285–296. https://doi.org/10.1016/j.isprsjprs.2009.02.004. Theme Issue: Image Analysis and Image Engineering in Close Range Photogrammetry

Belloni V, Sjölander A, Ravanelli R, Crespi M, Nascetti A (2023) Crack monitoring from motion (cmfm): Crack detection and measurement using cameras with non-fixed positions. Autom Constr 156:105072. https://doi.org/10.1016/j.autcon.2023.105072

Cha Y-J, Choi W, Büyüköztürk O (2017) Deep learning-based crack damage detection using convolutional neural networks. Comput-Aided Civil Infrastruct Eng 32(5):361–378. https://doi.org/10.1111/mice.12263. https://onlinelibrary.wiley.com/doi/pdf/10.1111/mice.12263

Ceroni F, Casapulla C, Cescatti E, Follador V, Prota A, Porto F (2022) Damage assessment in single-nave churches and analysis of the most recurring mechanisms after the 2016–2017 central italy earthquakes. Bull Earthq Eng 20(15):8031–8059

Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. arXiv preprint arXiv:1610.02357

Chura RY, Meriggi P, Choueiri C, de Felice G (2024) Seismic assessment of churches through integration of digital survey, texture recognition and distinct element modelling. J Build Eng 98:111375. https://doi.org/10.1016/j.jobe.2024.111375

D'Ayala D (2000) Establishing correlation between vulnerability and damage survey for churches, p. 2237. https://www.iitk.ac.in/nicee/wcee/article/2237.pdf

Doglioni F (1998) Models for vulnerability analysis of monuments and strengthening criteria. In: Proceedings of the eleventh european conference on earthquake engineering: 6-11 September 1998, Paris, France, vol 1, p 179. AA Balkema. https://search.library.wisc.edu/catalog/9910298023702121

Dorafshan S, Thomas RJ, Maguire M (2019) Benchmarking image processing algorithms for unmanned aerial system-assisted crack detection in concrete structures. Infrastructures 4(2). https://doi.org/10.3390/infrastructures4020019

Fan Z, Wu Y, Lu J, Li W (2018) Automatic pavement crack detection based on structured prediction with the convolutional neural network. arxiv:1802.02208

Guerreiro LMC, de JJRT, Proença JMSFM, Bento RMNLP, MMPd (2000) Damage in ancient churches during the 9th of July 1998 Azores Earthquake. In: XII World conference on earthquake engineering

Golding VP, Gharineiat Z, Munawar HS, Ullah F (2022) Crack detection in concrete structures using deep learning. Sustainability 14(13):8117. https://doi.org/10.3390/su14138117

Gopalakrishnan K, Khaitan SK, Choudhary A, Agrawal A (2017) Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr Build Mater 157:322–330. https://doi.org/10.1016/j.conbuildmat.2017.09.110

Hampel U, Maas H-G (2009) Cascaded image analysis for dynamic crack detection in material testing. ISPRS J Photogramm Remote Sens 64(4):345–350. https://doi.org/10.1016/j.isprsjprs.2008.12.006

Hsieh Y-A, Tsai Y (2020) Machine learning for crack detection: Review and model performance comparison. J Comput Civ Eng 34:04020038. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000918

Han X, Zhang Z, Ding N, Gu Y, Liu X, Huo Y, Qiu J, Yao Y, Zhang A, Zhang L, Han W, Huang M, Jin Q, Lan Y, Liu Y, Liu Z, Lu Z, Qiu X, Song R, Tang J, Wen J-R, Yuan J, Zhao WX, Zhu J (2021) Pre-trained models: Past, present and future. AI Open 2:225–250. https://doi.org/10.1016/j.aiopen.2021.08.002

Kaveh H, Alhajj R (2024) Recent advances in crack detection technologies for structures: a survey of 2022-2023 literature. Front Built Environm 10. https://doi.org/10.3389/fbuil.2024.1321634

Kim I-H, Jeon H, Baek S-C, Hong W-H, Jung H-J (2018) Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle. Sensors 18(6):1881

Ko P, Prieto SA, Soto B (2021) Abecis: an automated building exterior crack inspection system using uavs, open-source deep learning and photogrammetry. In: Feng C, Linner T, Brilakis I, Castro D, Chen P-H, Cho Y, Du J, Ergan S, Soto B, Gašparík J, Habbal F, Hammad A, Iturralde K, Bock T, Kwon S, Lafhaj Z, Li N, Liang C-J, Mantha B, Ng MS, Hall D, Pan M, Pan W, Rahimian F, Raphael B, Sattineni A, Schlette C, Shabtai I, Shen X, Tang P, Teizer J, Turkan Y, Valero E, Zhu Z (eds) Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC), pp 637–644. International Association for Automation and Robotics in Construction (IAARC), Dubai, UAE. https://doi.org/10.22260/ISARC2021/0086

Lagomarsino S (2006) On the vulnerability assessment of monumental buildings. Bull Earthq Eng 4(4):445–463. https://doi.org/10.1007/s10518-006-9025-y

Lagomarsino S (2012) Damage assessment of churches after l'aquila earthquake (2009). Bull Earthq Eng 10(1):73–92. https://doi.org/10.1007/s10518-011-9307-x

Lagomarsino S, Cattari S, Ottonelli D et al (2019) Earthquake damage assessment of masonry churches: proposal for rapid and detailed forms and derivation of empirical vulnerability curves. Bull Earthq Eng 17(7):3327–3364. https://doi.org/10.1007/s10518-018-00542-8

Liebold F, Maas H-G, Deutsch J (2020) Photogrammetric determination of 3d crack opening vectors from 3d displacement fields. ISPRS J Photogramm Remote Sens 164:1–10. https://doi.org/10.1016/j.isprsjprs.2020.03.019

Lagomarsino S, Podestà S (2004) Seismic vulnerability of ancient churches: I. damage assessment and emergency planning. Earthq Spectra 20(2):377–394. https://doi.org/10.1193/1.1737735

Li H, Wang W, Wang M, Li L, Vimlund V (2022) A review of deep learning methods for pixel-level crack detection. J Traffic Transport Eng (English Edition) 9(6):945–968. https://doi.org/10.1016/j.jtte.2022.11.003

Matrone F, Grilli E, Martini M, Paolanti M, Pierdicca R, Remondino F (2020) Comparing machine and deep learning methods for large 3d heritage semantic segmentation. ISPRS Intern J Geo-Inf 9(9). https://doi.org/10.3390/ijgi9090535

Munawar H, Hammad A, Haddad A, Soares C, Waller S (2021) Image-based crack detection methods: A review. Infrastructures 6:115. https://doi.org/10.3390/infrastructures6080115

Marotta A, Liberatore D, Sorrentino L (2021) Development of parametric seismic fragility curves for historical churches. Bull Earthq Eng 19(13):5609–5641. https://doi.org/10.1007/s10518-021-01174-1

Mathavan S, Rahman M, Kamal K (2015) Use of a self-organizing map for crack detection in highly textured pavement images. J Infrastruct Syst 21(3):1–11

Mandal V, Uong L, Adu-Gyamfi Y (2018) Automated road crack detection using deep convolutional neural networks, pp 5212–5215. https://doi.org/10.1109/BigData.2018.8622327

Munawar HS, Ullah F, Heravi A, Thaheem MJ, Maqsoom A (2022) Inspecting buildings using drones and computer vision: A machine learning approach to detect cracks and damages. Drones 6(1):5. https://doi.org/10.3390/drones6010005

Marchewka A, Ziółkowski P, Aguilar-Vidal V (2020) Framework for structural health monitoring of steel bridges by computer vision. Sensors 20(3):700. https://doi.org/10.3390/s20030700

Özgenel (2019) Concrete crack images for classification. Mendeley Data. https://doi.org/10.17632/5y9wdsg2zt.2

Özgenel ÇF, Sorguç AG (2018) Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In: Isarc. Proceedings of the international symposium on automation and robotics in construction, vol 35, pp 1–8. IAARC Publications

Penna A, Calderini C, Sorrentino L et al (2019) Damage to churches in the 2016 central italy earthquakes. Bull Earthq Eng 17(12):5763–5790. https://doi.org/10.1007/s10518-019-00594-4

Pauly L, Hogg D, Fuentes R, Peel H (2017) Deeper Networks for Pavement Crack Detection. IAARC. This is an author produced version of a paper published in 34th international symposium on automation and robotics in construction. https://eprints.whiterose.ac.uk/120380/

Pantoja-Rosero BG, Achanta R, Beyer K (2023) Damage-augmented digital twins towards the automated inspection of buildings. Autom Constr 150:104842

Pantoja-Rosero BG, Oner D, Kozinski M, Achanta R, Fua P, Perez-Cruz F, Beyer K (2022) Topo-loss for continuity-preserving crack detection using deep learning. Constr Build Mater 344:128264

Rezaie A, Achanta R, Godio M, Beyer K (2020) Comparison of crack segmentation using digital image correlation measurements and deep learning. Constr Build Mater 261:120474

Sánchez-Aparicio LJ, Blanco-García FL, Mencías-Carrizosa D, Villanueva-Llauradó P, Aira-Zunzunegui JR, Sanz-Arauz D, Pierdicca R, Pinilla-Melo J, Garcia-Gago J (2023) Detection of damage in heritage constructions based on 3d point clouds. a systematic review. J Build Eng 77:107440. https://doi.org/10.1016/j.jobe.2023.107440

Shrestha P, Avci O, Rifai S, Abla F, Seek M, Barth K, Halabe U (2025) A review of infrared thermography applications for civil infrastructure. Struct Durabil Health Monitor 19(2):193–231. https://doi.org/10.32604/sdhm.2024.049530

Shahbazi S, Barra A, Gao Q, Crosetto M (2024) Detection of buildings with potential damage using differential deformation maps. ISPRS J Photogramm Remote Sens 218:57–69. https://doi.org/10.1016/j.isprsjprs.2024.10.008

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manage 45(4):427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Sorrentino L, Liberatore L, Decanini LD et al (2014) The performance of churches in the 2012 emilia earthquakes. Bull Earthq Eng 12(6):2299–2331. https://doi.org/10.1007/s10518-013-9519-3

Tahir A, Munawar HS, Akram J, Adil M, Ali S, Kouzani AZ, Mahmud MAP (2022) Automatic target detection from satellite imagery using machine learning. Sensors 22(4):1147

Wang KC, Zhang A, Li JQ, Fei Y, Chen C, Li B (2017) Deep learning for asphalt pavement cracking recognition using convolutional neural network. In: Airfield and highway pavements 2017, pp 166–177

Xue Y, Li Y (2018) A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. Comput-Aided Civil Infrastruct Eng 33(8):638–654. https://doi.org/10.1111/mice.12367

Xu Y, Li Y, Zheng X, Zheng X, Zhang Q (2023) Computer-vision and machine-learning-based seismic damage assessment of reinforced concrete structures. Buildings 13:1258

Zadeh SS, Khorshidi M, Kooban F et al (2024) Concrete surface crack detection with convolutional-based deep learning models. arXiv preprint arXiv:2401.07124. https://doi.org/10.5281/zenodo.10061654

Zhang L, Yang F, Zhang YD, Zhu YJ (2016) Road crack detection using deep convolutional neural network. In: 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp 3708–3712. https://doi.org/10.1109/ICIP.2016.7533052