

Deep Learning-Based AI-Assisted Visual Inspection Systems for Historic Buildings and their Comparative Performance with ChatGPT-4O

Mayank Mishra ^{1,2}, Kai Zhang ², Chiara Mea ², Luigi Barazzetti ², Francesco Fassi ², Fausta Fiorillo ², Mattia Previtali ²

¹University of Minho, ISE, ARISE, Department of Civil Engineering, Guimarães, 4800-058, Portugal
mayank@civil.uminho.pt

²Dept. of Architecture, Built Environment and Construction engineering (ABC)
Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, Italy
(kai.zhang, chiara.mea, luigi.barazzetti, francesco.fassi, fausta.fiorillo, mattia.previtali)@polimi.it

Keywords: Artificial intelligence, ChatGPT, Computer Vision, Damage identification, Deep Learning

Abstract

Historical buildings and monuments are typically subject to degradation over time due to the passage of time and constant exposure to external agents. The use of artificial intelligence (AI) to support the work of conservation and restoration specialists in identifying surface decay is a research topic of considerable interest at present. This study presents two approaches: ChatGPT and an object detection architecture (YOLOv5). Specifically, this investigation sought to evaluate the ChatGPT's ability to identify and describe surface degradation pathologies by exploiting its pre-trained models for image analysis. The ICOMOS-ISCS: Illustrated Glossary on Stone Deterioration Patterns (2008) was provided as a reference to guide the use of specific terminology. In the first test phase, to verify the accuracy of the ChatGPT results, benchmark images (depicting different types of damage) extracted from the UNI 11182 (2006) standard referring to the definition of degradation types were used. Only later were images from literature studies and other photographic datasets also used. In general, the results of the analysis were validated with the conclusions of professionals and with the conclusions of other AI techniques, as well as with the descriptions provided by reference manuals in the literature. In particular, the decay annotations predicted by the pre-trained object detection model were compared with those made by human experts. The capabilities and limitations of both approaches as tools for identifying deterioration pathologies are illustrated.

1. Introduction

1.1 Overview

Cultural heritage (CH) structures are typically prone to degradation because of their age and exposure to various external factors. Detecting and identifying surface-related damages to CH structures can enable the timely monitoring of their progression and the implementation of appropriate monitoring strategies and interventions if required.

In general, professionals in the field of conservation and restoration rely on visual inspections. Although these manual inspections have been proven to be reliable and can be performed by “eyes,” eliminating the need for sophisticated equipment, they could be supported by automatic procedures. Indeed, several recent studies have focused on augmenting manual visual inspections with computer vision (CV) techniques that rely on object detection algorithms to identify and locate elements in CH images, such as “surface deterioration patterns” (e.g., cracks, mortar removal, joint damages), “surface damages in façades,” and “missing components.”

An artificial intelligence (AI)-assisted visual inspection system based on “object detection” models could be used to detect and identify damages in the images of CH buildings. Object detection models are trained by an expert using supervised learning in order to identify decay on a new dataset of photos. These models, once trained, have demonstrated good performance in detecting damage in CH structures. Mishra et al. (2024) deployed deep learning (DL) you only look once (YOLO) framework as an object detection model for their Dadi Poti case study in New Delhi, India; they were able to identify four types of defects— cracks, exposed bricks, discoloration,

and spalling—with a maximum average precision (mAP) of approximately 93.7%. Numerous applications exist where DL models have been deployed for damage detection in CH sites, such as damage to Portuguese tiles azulejos (Karimi et al., 2024), three types of defects from English and Dutch cemeteries in Surat (Mansuri and Patel, 2022), efflorescence and spalling damages in historic masonry structures (Wang et al., 2019), four types of damages to CH façades (sabotage, erosion, change of color, material loss) in the city of “Al-Salt” in Jordan (Samhoury et al., 2022), cracks in timber elements of CH structures (Ma et al., 2022), detecting deterioration patterns following ICOMOS glossary for CH fields (Zhang et al., 2024), and weathering in stone CHs (Hatir et al., 2020).

Mishra and Lourenco (2024) summarized various applications of DL models to ease the inspection process of CH buildings and help inspection professionals in their fieldwork via automatic identification of damage types for implementing suitable measures. Thus, AI-assisted visual inspection systems have found various applications in the inspection of CH structures. However, CV models are not sufficiently mature for professional use in field activities. The CV approach requires an important level of expert training and familiarity, as well as considerable time and money, to be effectively implemented.

1.2 Related work based on ChatGPT

ChatGPT has been used in several civil engineering applications (Kim et al., 2024). Hostetter et al. (2024) evaluated the performance of ChatGPT for handling questions related to fire safety and obtained an accuracy of 88%. In addition to applications in CV, GPT models have been deployed in the construction industry for material selection and optimization (Saka et al., 2024).

ChatGPT-assisted programming has been used to solve problems in geotechnical engineering such as stability of slopes, seepage, and X-ray image processing (Kim et al., 2024); safety management in construction through proper hazard recognition (Uddin et al., 2023); scheduling for simple construction projects (Prieto et al., 2023); risk management in construction projects (Aladağ, 2023); and seismic engineering by communicating complex concepts in simple terms (Ray 2024).

Some ChatGPT applications are also suitable for image-processing tasks. Cai et al. (2024) assessed the performance of ChatGPT in cementitious composites by asking forty-five questions related to the cementitious composites sector; the module could identify cracks and successfully render the distribution chart of their widths. However, to draw rendered images of cracks using ChatGPT, an example image having cracks that were drawn manually by red lines by authors was given as input to ChatGPT 4.0, and then the following question was asked: “How many cracks are there similar” to the image inputted now? “Is it possible to create a distribution chart of crack widths based on the scale?”. Osco et al. (2024) utilized Visual ChatGPT, which combines ChatGPT’s large language model (LLM) capabilities with visual computation, for image processing related to the remote sensing domain. The model performed various image processing tasks, such as edge/straight line detection, scene classification, and image segmentation, and it could be incorporated into CH images. In a recent study by Ciccone (2024), ChatGPT was used as an information technology assistant to analyze aerial and satellite images, mainly multispectral orthophotos from drone surveys, to detect vegetation and anthropogenic anomalies.

2. The Proposed Method

This study evaluated the ability of ChatGPT to identify decay in CH structures, leveraging its pre-trained models to process and describe images. The ICOMOS (ICOMOS-ISCS: Illustrated Glossary on Stone Deterioration Patterns, 2008) glossary was provided as a reference to guide the use of specific terminology. The model was implemented to limit results to a description of damage without suggesting possible interventions.

Benchmark images depicting several types of damages sourced from manuals, images analyzed using CV reported in the literature (e.g., Mishra et al. 2024) and other photographic datasets were all used to assess the accuracy of ChatGPT in this context. Some sample pictures used for testing ChatGPT were taken from the UNI 11182 standard, referring to the definition of decay types (UNI – Ente Italiano Di Normazione, 2006). The analysis result was validated with the conclusions of professional experts and conclusions from other AI techniques, apart from descriptions provided by benchmark manuals in the literature.

The results, summarized in Table 1, include damage pathology descriptions generated by ChatGPT. The results of the benchmark images (Figure 1) and their descriptions can be compared by the reader with benchmarked descriptions, like those available in the UNI 11182 standard. The model was evaluated in various scenarios to estimate its ability to describe damages to different CH structures.

The ChatGPT potential for practitioners lies mainly in the absence of the necessity to build custom models. Overall, ChatGPT was able to describe the deterioration patterns (Table 1 and Figure 1) and locate them on the image, providing a description of the position and some information about the extension. On the other hand, the description appears to follow a predefined checklist rather than addressing the specific problem in detail. This approach could be useful, but it should be regarded as a supporting tool rather than a definitive solution.

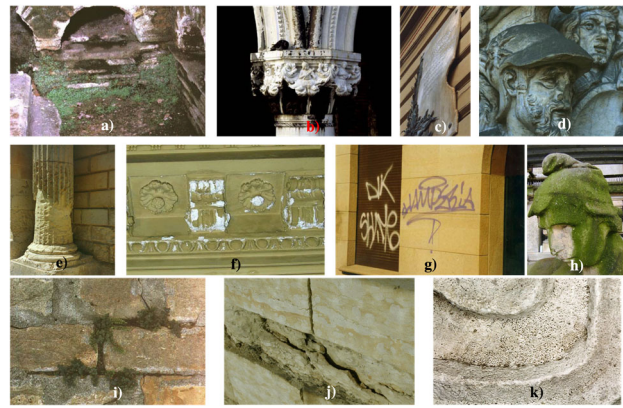


Figure 1. Test images for ChatGPT chosen from UNI 11182 (2006) showing various types of damages: a) biological colonization, b) crust deposit, c) deformation, d) superficial deposits, e) disintegration, f) efflorescence, g) graffiti, h) biological colonization, i) presence of vegetation, j) detachment, and k) pitting.

Image (Figure 1)	Description of damages by ChatGPT
a	<p>Biological Colonization</p> <ul style="list-style-type: none"> - Moss: Extensive growth of moss is visible on the stone surfaces, particularly on the lower sections and near the ground. The moss covers large areas and is likely contributing to moisture retention and further stone deterioration. - Algae: The greenish tint on the stone surfaces indicates the presence of algae, which thrives in moist environments. - Plant: Small plants and possibly lichens are growing in crevices and on the surface of the stone, contributing to biological colonization and potential further deterioration of the stone material.
b	<p>Discoloration and Deposit</p> <ul style="list-style-type: none"> - Black Crust: Significant black crust deposits are visible, particularly on the upper sections of the capital and along the archway. This crust is a result of atmospheric pollution and the accumulation of particulate matter.
c	<p>Crack and Deformation</p> <ul style="list-style-type: none"> - Deformation: The plaque shows noticeable deformation, evident in its warped and bent shape. This deformation suggests physical stress or thermal expansion and contraction over time.
d	<p>Discoloration and Deposit</p> <ul style="list-style-type: none"> - Black Crust: A noticeable black crust has formed on the helmet and other parts of the sculpture, indicative of atmospheric pollution and the accumulation of particulate matter. - Staining: Dark staining is visible on the lower sections of the helmet and on the face, likely caused by water runoff and the deposition of pollutants.
e	<p>Features Induced by Material Loss</p> <ul style="list-style-type: none"> - Erosion: The column exhibits significant erosion, particularly on the lower part. The stone surface is heavily worn and pitted, indicating the gradual removal of material due to weathering and possibly human interaction. - Mechanical Damage: There are areas of mechanical damage, particularly around the base of

	the column, where pieces of stone appear to have been chipped or knocked off.
f	Discoloration and Deposit - Efflorescence : White efflorescence is prominently visible in the recessed areas and on the surface of the decorative elements. This indicates the presence of soluble salts that have migrated to the surface and crystallized, likely due to moisture issues.
g	Discoloration and Deposit - Graffiti : The surface of the stone wall has been vandalized with graffiti, with both white and purple spray paint applied to the surface. This graffiti represents a form of discoloration and visual pollution, detracting from the aesthetic and potentially causing chemical damage to the stone.
h	Biological Colonization - Algae and Moss : The statue is extensively covered with green algae and moss, particularly on the head and upper sections. This biological colonization is dense and widespread, leading to significant discoloration and potential damage to the stone surface by retaining moisture and promoting further deterioration.
i	Biological Colonization - Plants : There is significant biological colonization by plants, particularly ferns and other small vegetation, growing within the cracks and mortar joints. This growth is likely contributing to further deterioration by retaining moisture and exerting physical pressure on the stone and mortar. - Algae and Moss : In addition to the plants, there is evidence of algae and moss on the stone surface, especially in the areas around the cracks and where the stones meet. These biological elements thrive in the moist conditions provided by the cracks and contribute to the staining and further breakdown of the material.
j	Detachment - Scaling : Scaling is evident on the stone surfaces, where thin layers of material have detached, particularly around the cracks and along the horizontal joints. - Exfoliation : Some areas exhibit exfoliation, where multiple thin layers of stone have peeled away from the underlying material. This is especially noticeable near the cracks.
k	Features Induced by Material Loss - Pitting : The stone surface exhibits significant pitting, with numerous small, shallow depressions scattered across the area. This pitting indicates localized material loss, likely due to chemical or physical weathering processes.

Table 1: Response of ChatGPT in determining the damages in the test images of Figure 1 (some responses are shortened due to space requirements and damage descriptions pertaining to that damage typology).

Additional images (captured by the authors) were used to test ChatGPT (Figure 2). In many instances, the ChatGPT model provided additional interesting information. For example, it counted the number of visible birds (Figure 2a), unlike the model presented by Mishra et al. (2022), as it was not trained for identifying birds and other objects such as vegetation and missing components.



Figure 2. Additional images chosen from a) Mishra et al. (2024), b) Photo by L. Barazzetti, c) Limonaia case study (equirectangular image by F. Fassi and K. Zhang) for ChatGPT showing various types of damages.

In fact, Figure 2a from Mishra et al. (2024) was assessed using a YOLOv5 defect detection model, which could detect spalling and cracks at two locations. The ChatGPT model still correctly predicted that the structure is a dome, identifying cracks and vegetation as surface degradation. However, the ChatGPT model lacked precision in pinpointing the damaged locations, so generic responses were provided instead. As shown in Figure 2b, the ChatGPT model described the exact location of cracks and provided some additional information. The response was exact, *“The crack starts at the top left corner of the archway, just below the edge of the wall. It extends downward, following a roughly vertical path, slightly curving to the right. The crack passes through the painted area with the decorative elements and the two crests. It continues downward until it reaches the white stone frame of the archway, where it passes through the corner of the stone”*. This description was accurate, as confirmed through a naked-eye inspection.

For Figure 2c, the ChatGPT model pinpointed various deterioration patterns, such as material loss, discoloration and deposit, and biological colonization. It also described the presence of potted plants along the right side of a corridor; this, however, is not a form of deterioration but human activity. However, when the exact locations were pointed out by drawing bounding boxes (object detection phase) over the deteriorations, the ChatGPT model failed. This result can be expected because ChatGPT is primarily an AI language processing model rather than an image processing model, typically using CV algorithms for object detection.

We prompted a related query, and ChatGPT clarified that it lacks direct integration with specific machine learning frameworks for object detection tasks, such as YOLO or region-based convolutional neural networks (R-CNNs), which are already used for damage detection in CH structures (Mansuri and Patel, 2022M; Zhang et al. 2024; Mishra and Lourenco, 2024).

ChatGPT’s limitations became evident when asked to precisely pinpoint damage locations, as shown in test images (Figure 3 a). Even after the query prompts were modified, only slight improvement was observed in crack locations (Figure 3 b), but it was still not sufficiently accurate to match the standards of CV-identified images. Similarly, for Figures 3c and 3d the damage was described correctly by the ChatGPT tool, but when asked to pinpoint the location of crack (Figure 3c) and peelings (Figure 3d), it failed.

Furthermore, when the images of random CH monuments were used with the implemented GPT, it was able to recognize the architectural style and elements (Llamas et al., 2017; Siountri and Anagnostopoulos, 2023) quite accurately, which has been the subject of research in many papers related to CH and object detection area. Additionally, when compared to the images from previous research papers on tile-related damages (Karimi et al. 2024), the GPT model was able to correctly describe visible deterioration patterns, with details such as the noticeable fracture runs across the yellow tile pattern (please refer to research paper of Karimi et al. 2024, Figure 2b where a yellow tile has a crack as a damage typology). The GPT model describes the crack location by following the description “The crack appears to be a large fracture cutting through several tiles”. Also, in the case of another image (i.e. Figure 2a) in Karimi et al. (2024), the GPT model description category was “detachment”, and the description as follows “we observe peeling or scaling in the lower portion of the blue and white tiles. The ceramic glaze or outer layer of the tiles seems to be detaching, leaving a rough, exposed underlying surface”, which is quite accurate.

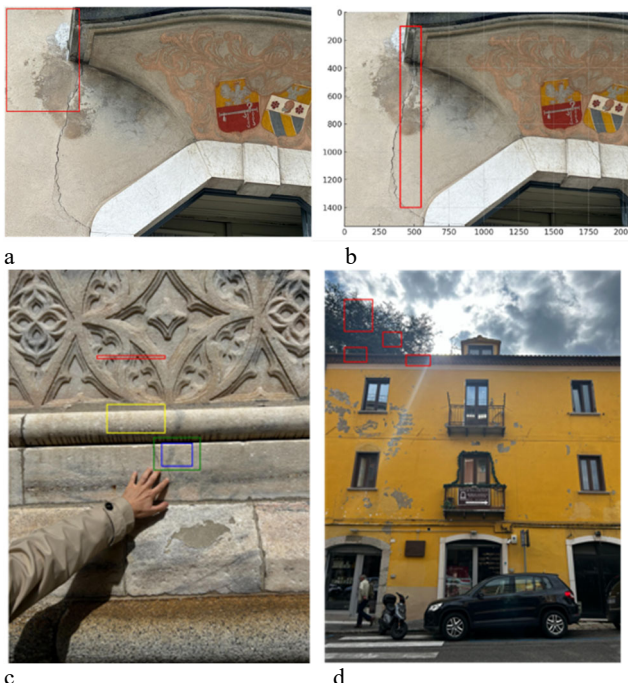


Figure 3: Limitations in identifying damages using ChatGPT (bounding boxes are not at exact locations).

ChatGPT model was tested to describe damages, but it occasionally suggested possible causes and intervention remedies for repair. Interventions are not considered here in the comparison. Moreover, the GPT model used was instructed not to propose interventions.

Tailored applications can be developed specifically to detect damage pathologies. For example, Spennemann (2023) developed a smart anomaly detection assistant (S.A.D.A version 0.9) for detecting archaeological anomalies. This is a step ahead of using object detection models, which only pinpoint the damage locations; in some cases, this model could also segment damages. Inspection professionals can decide what to do with the damage and suggest remedies. Another aspect brought forth by the tests suggests that, in many instances, the description and the causes of damages were too generic, with multiple damage descriptions likely to be inaccurate. Therefore, ChatGPT may

not be completely reliable in some cases and cannot replace a human specialist; however, it can be considered a useful aid. The nature of ChatGPT as a language model rather than an image-processing model is the main limitation analyzed in this paragraph. The tests emphasized that custom CV models better perform drawing bounding boxes over decay on architectural surfaces. After acknowledging this gap, the next chapter (Chapter 3) shows how a CV algorithm for object detection operates, allowing readers to compare the two approaches.

3. Comparison with Object Detection Experiences

In previous chapters, we evaluated ChatGPT’s ability to identify and describe decay. In the following one, we compare the annotations made by an object detection architecture, YOLOv5, with those made by human experts.

3.1 Existing Object Detection Models for Architectural Pathology

A lot of efforts have been made recently to develop a robust object detection model for architectural pathology. For example, Chen et al. (2019) used a convolutional neural network to train an image dataset of 40,000 images of cracked and uncracked concrete spaces. The model achieved an accuracy of 99.71%. Also, Deng et al. (2020) used a YOLOv2 network to detect concrete cracks with complex backgrounds, achieving a maximum mAP of 77%, while the Faster RCNN model achieved 76%. Even if complete and with quite high accuracy in detecting cracks, these works struggle under conditions where cracks and background are not easily distinguishable.

Moreover, they focus on one material and one decay type. For existing models, it is difficult to distinguish decay when trained over a various set of photographs with different kinds of pathologies. Indeed, all show different exterior appearances, varying depending on the material, color, lighting, etc.

Set against this complex background, the experience of the authors (Mishra et al., 2024a; Zhang et al., 2024) led to acquiring new considerations in relation to how superficial CV architectures identify pathology.

3.2 Challenging outcomes

Unlike the off-the-shelf ChatGPT approach, using the object detection model requires datasets expressly and consciously annotated for the application scene. It used predefined categories extracted from the ICOMOS glossary of stone pathology (see Table 2). Some decays included in the ICOMOS were excluded from this work (too difficult to detect by visual analysis alone, i.e. deformation, blistering, scaling, fragmentation, bursting, perforation, pitting, mechanical damage, missing parts, microkarst, encrustation, patina, soiling, film, glossy aspect, algae, mould, lichen). In certain instances, two decay types have been amalgamated for visual coherence, with both names kept, to allow the experts to deepen their analysis. Table 2 shows the categories that serve as the foundation for our classification and labeling.

Used categories for detection	Definition
Crack	Crack - Individual fissure, clearly visible by the naked eye, resulting from the separation of one part from another.

Peeling Delamination Disintegration	<p>Detachment</p> <ul style="list-style-type: none"> - Peeling: Shedding, coming off, or partial detachment of a superficial layer (thickness: sub-millimetric to millimetric) having the aspect of a film or coating which has been applied on the stone surface. - Delamination: (...) It corresponds to a physical separation into one or several layers following the stone laminae. The thickness and the shape of the layers are variable. (...) -Disintegration: Detachment of single grains or aggregates of grains.
Alveolization Erosion	<p>Features induced by material loss</p> <ul style="list-style-type: none"> - Alveolization: Formation, on the stone surface, of cavities (alveoli), which may be interconnected and may have variable shapes and sizes (generally centimetric, sometimes metric). - Erosion: Loss of original surface, leading to smoothed shapes.
Discoloration Crust and Deposit Subflorescence and Efflorescence Graffiti	<p>Discoloration and Deposit</p> <ul style="list-style-type: none"> - Discoloration: Change of the stone colour in one to three of the colour parameters: hue, value and chroma. (...) -Crust: Generally coherent accumulation of materials on the surface. A crust may include exogenic deposits in combination with materials derived from the stone. (...) -Deposit: Accumulation of exogenic material of variable thickness. (...)
Biological Colonization Plant	<p>Biological Colonization</p> <ul style="list-style-type: none"> -Biological Colonization: Colonization of the stone by plants and micro-organisms such as bacteria, cyanobacteria, algae, fungi, and lichen (symbioses of the latter three). (...) -Plant: Vegetal living being, having, when complete, root, stem, and leaves, though sometimes consisting only of a single leafy expansion (e.g., Tree, fern, herb).

Table 2: Predefined categories used for training the object detection model (description by ICOMOS-ISCS, 2008).

The Yolov5 model was trained on a dataset of 1,600 images collected and annotated manually, representing 21,000 instances (decays) labelled across five different materials (stone, cement, plaster, wood, and ceramic). Each image contains an average of 13 instances. Furthermore, the architecture employs data augmentation techniques to prepare the neural network for robust training. Unfortunately, in this case, as in many examples, a balance problem exists in sample distribution across categories. The categories that show the highest number of instances (over 3,000) are biological colonization, peeling, and chromatic alteration. Other categories, like graffiti, are instead underrepresented. Examples of the predicted decays are reported in Figures 4 and 5.

Little and adjacent boxes were used to manually mark the pathology, rather than a single box labelling the whole damaged area at once. The annotation boxes mainly cover under 10% of the entire image area.

The model performance was evaluated using a t-SNE (T-distributed Stochastic Neighbour Embedding) graph (please refer to Zhang et al., 2024), a visualization that analyses the confusion between each category within the provided images. It points out visual clusters of related classes that are easily mistaken for each other, indicating that the dataset annotation is under criticism. The more isolated clusters are those that are

best defined by the machine without many misclassification issues. For example, biological colonization, crust & deposit are in this work the most differentiable from the other categories, while the areas related to plants, discoloration, disintegration, and crack show overlapped features. This behaviour reflects shared visual characteristics of categories, particularly if examining the pathologies of the same material (Figure 6).



Figure 4. YOLOv5 model prediction (Zhang et al. 2024) on an image of Casa Parrocchiale in Civate (photo by F. Fassi): green squares correctly detect plants, purple ones detect black crust.



Figure 5. YOLOv5 model (Zhang et al. 2024) prediction performed on an image of Casa Parrocchiale in Civate (photo by F.Fassi). The model also identifies plants far away or not damaging the building in the outer courtyard (green squares).

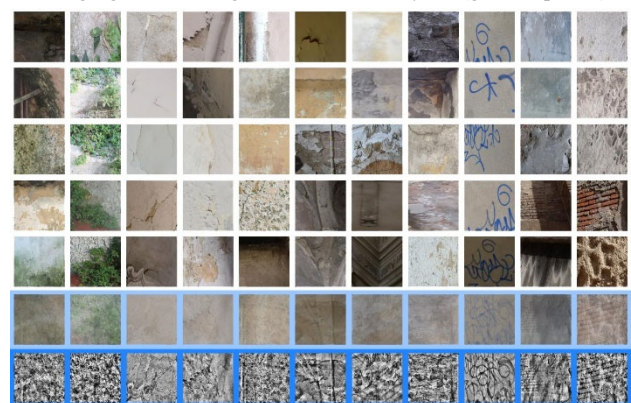


Figure 6. Every column represents a category used to train the model, from the left: Biological Colonization, Plant, Crack, Delamination, Disintegration, Peeling, Discoloration, Crust and Deposit, Graffiti, Subflorescence and Efflorescence. The first 5 rows are sample images, the light blue row is an Average Image, and the last one is the Strengthened Pattern Image derived from it. (Adapted and modified from Zhang., et al., 2024)

Taking plaster as an example (Figure 6), its pathologies (cracks, peeling, etc.) might share similar visual features in images. Average images were generated by averaging pixel values from all the example images provided. A strengthened pattern, instead, is an image that shows a modified or enhanced version of the average image, where certain patterns or features are accentuated or intensified.

However, there are instances where understanding how computers distinguish between similar types of decay can be enlightening. For example, in distinguishing between cracks and detachment (Figure 7), the trained model discerned subtle differences, such as the presence of shadows around detached areas. By digging into various aspects of model performance, the reader can better understand how far the machine can go in detecting surface anomalies and, eventually, pave the way for bridging the gap between AI expertise and human specialists in architectural heritage analysis.

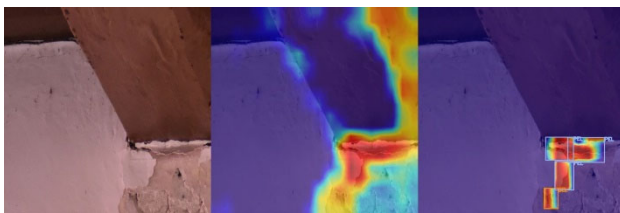


Figure 7. Object Detection Model Visualization with Eigen-CAM (Muhammad and Yeasin, 2020). Left to right: Test image with peeling pathology; Eigen CAM extracted from the 7th layer; Scaled CAM within the bounding boxes.

3.3 Discussion: Human and Machine Annotation

Architectural pathology is indeed a difficult and multi-faceted subject to be analyzed with the aid of visual data. So, to prove the behaviour of the model trained by Zhang et al. (2024), the authors enlisted the expertise of Prof. Sonia Pistidda, a seasoned professional working with architectural preservation in Politecnico di Milano. The model (as described in previous Chapter 3.2) comes from research focusing on practical tests on architectural pathological issues using image classification and object detection methods. Prof. Pistidda performed an image annotation on cube face images with a pixel size of 640*640. This data was labelled using ICOMOS categories, just like the training set of the model. We then used a YOLO v5 model, trained on 1,621 samples containing 21,008 instances of pathology, to make predictions on images of the same size and compare annotations. The comparison (Figures 8 and 9) reveals clear distinctions between human and machine annotation approaches.

Figures 8 and 9 allow us to compare the annotation done by Prof. Pistidda and the YOLOv5 framework used by Zhang et al. (2024). Note that labelling is associated with a three-letter code. “DEL” stands for delamination, “COL” for biological colonization, “PLT” for the presence of vegetation, “CHR” for discoloration, “ERO” for erosion, “CRU” for crust and deposit. In Figure 8, human annotations focus on biological colonization in the main area of the picture, and plants and vegetation are totally ignored. Instead, the YOLO model annotates also plants in vases at the bottom, even if they do not pose any threat to the state of the building.

In the second image (Figure 9), an opposite result is visible: the model didn't detect most of the decay phenomena that the human annotator identified. This result probably depends on the fact that pathologies like erosion, crust, and deposit coexist in the pavement represented in the picture. There is no way to map properly this surface using a bounding box. Maybe a

segmentation model could behave better in a nuanced and complex task like this one.

Some considerations must be made if the aim is to integrate this automatic, fast and low-cost technology into future professional life.

Concerning the subjectivity of human decision-making, several conclusions can be drawn:

- First, machines excel at making exhaustive predictions across the entire image compared to humans (obviously when the degradation analysis is done on a face and not on site). Human attention is mostly focused on the central, clear parts of the image, considering difficulties in recognizing objects in blurry, dusky, distant areas. On the contrary, machine learning detection systems, though limited by the receptive field affected by the kernel size, dilation, and pooling operations, still the model prediction is always exhaustive and analyses the full image.
- Second, human predictions are more context-dependent, both an advantage and a disadvantage. In the tested cases, expert annotations focused on pathological objects on architectural surfaces, ignoring ‘irrelevant’ elements (such as the plant pots in Figure 8). However, the machine made a more accurate and timely identification of the plants among the pavement tiles (Figure 9). Instead, a human analysis could overlook the annotation of some of these.
- Third and last, deep learning-based networks for detecting decays in photographs can offer a solution to overcome physical limitations, aid experts in their work, and offer the prospect of a non-destructive testing technique. This approach can provide information about the health condition of a structure without compromising its integrity.

Instead, concerning the (sometimes problematic) objectivity of machine decision-making, other conclusions can be drawn:

- First, the accuracy of the trained model is not comparable to the human decision-making process, as experts draw on extensive prior experience and factors such as materiality, estimated geometric shape, lighting conditions, and tactile perception.
- Second, architectural pathology detection is more complex than identifying objects like traffic lights or people due to the absence of neat boundaries. Decay phenomena blend one into the other, and their appearance depends also on their interaction. This complicates labelling. In this work, both the human expert and machine used a rectangular bounding box to delimit the pathology, but a single box can have difficulty embracing the full decay without including other unaffected areas of the wall. Continuous and adjacent boxes are used to mark the pathology in the YOLOv5 training, while the human expert used a single rectangular box including the whole damaged area. A more “natural” way to annotate.
- Third, the evident advantage of using CV architectures for object detection is its ability to learn what an image of a pathology entails—through an internal representation—enabling it to correctly predict and classify a crack, for example, in new images. Thus, creating a good set of training data with a good definition of categories is essential. In the work with ICOMOS, for example, a stage of selecting detectable patterns and categorizing them was inserted (Table 2). This is a step that has been revealed to be less time-expensive for tools like ChatGPT, for example. This makes it faster but potentially less accurate.

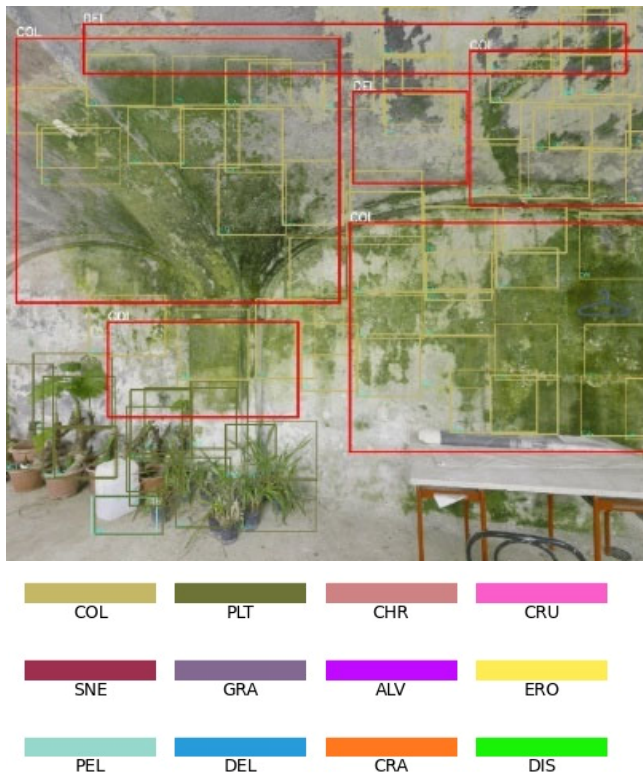


Figure 8. YOLOv5 object detection (cyan-labeled little boxes) vs. human annotation (red-labeled bigger boxes). Limonaia, Lainate (photo by F. Fassi).

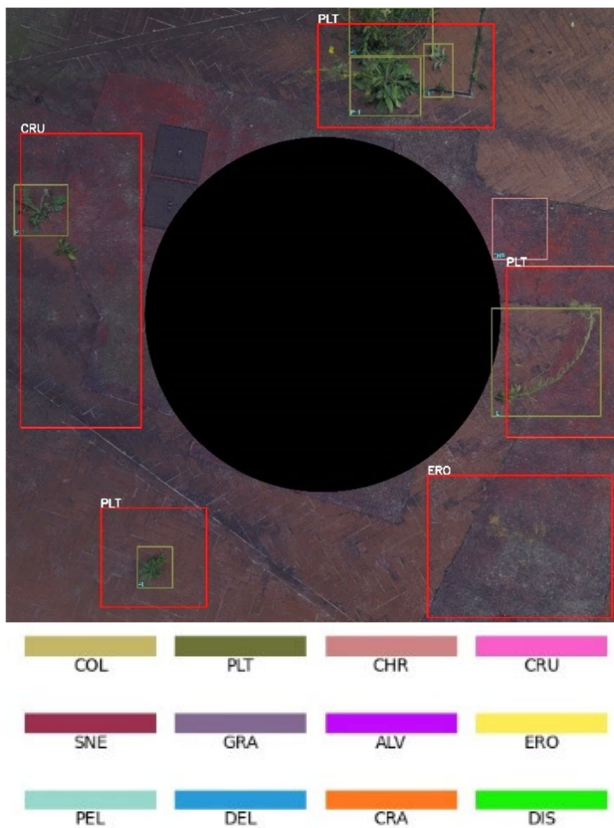


Figure 9. YOLOv5 object detection (cyan-labeled little boxes) vs. human annotation (red-labeled bigger boxes). Casa Parrocchiale, Civate (photo by F. Fassi).

4. Conclusion

Based on the literature review and analysis, we concluded that most ChatGPT applications are currently focused on language processing tasks (answering queries, as reported by Hostetter et al., 2024 and Cai et al., 2024) rather than CV tasks.

Various technologies associated with AI can be broadly categorized into three primary domains: Sense, Comprehend, and Act (Prasad and Choudhary, 2021). Sensing involves the perception of surroundings, including acquiring and processing sounds, images, and speech. Comprehending entails analyzing the collected information while acting, representing the physical execution of computer language instructions. AI technologies align with corresponding human capabilities, connecting CV and sensing, machine learning and action and natural language processing and comprehension.

However, when the AI is used for architectural pathology detection (requiring a deeper understanding of the matter), there could be ambiguous, incorrect, or inaccurate statements, mostly when asking it to point out the decay phenomena in the picture. This is because ChatGPT has a better grasp of text-based conclusions (Cai et al., 2024). Professionals can deploy ChatGPT to detect damage; however, the accuracy of the description would depend on the accuracy of the prompt. Also, they should be aware that all information given by ChatGPT must be double-checked. We do not advocate the replacement of inspection professionals by generative AI, but as suggested by Spennemann (2023), “AI will not replace workers, but it is the people who work with AI who will replace the people who do not use it”.

In general, minor training is required to develop a model capable of analyzing similar case studies. In many instances, the ChatGPT model performed better if we provided some clues, such as what to look for in the images (such as cracks and missing components); thus, in some cases, a pre-trained model could recognize similar damages via supervision.

An important aspect of ChatGPT is that it does not require training to yield results and can be used instantly by inspection professionals on-site because it only requires the images to be inputted along with the appropriate prompts. It is already integrated with mobile phones, delivering real-time results.

Finally, in many cases where it was tested, the description of damage was quite good, and it also suggested the possible causes and interventions (which were not evaluated in this work and were excluded from the implemented GPT). This feature provides additional information compared to the object detection models, which allows the damage locations instead. Overall, the GPT description can be seen as a supplementary aid for the specialist, without substituting their expertise.

REFERENCES

- Aladağ, H., 2023. Assessing the accuracy of ChatGPT use for risk management in construction projects. *Sustainability*, 15(22), 16071.
- Cai, J., Yuan, Y., Sui, X., Lin, Y., Zhuang, K., Xu, Y., Zhang, Q., Ukrainczyk, N., Xie, T., 2024. Chatting about ChatGPT: How does ChatGPT 4.0 perform on the understanding and design of cementitious composite?. *Construction and Building Materials*, 425, 135965.
- Chen, K., Yadav, A., Khan, A., Meng, Y., Zhu, K., 2019. Improved Crack Detection and Recognition Based on Convolutional Neural Network. *Modelling and Simulation in Engineering*, 2019, 8796743.

- Deng, L., Hong-Hu, C., Peng, S., Wei, W., Xuan, K., 2020. Region-Based CNN Method with Deformable Modules for Visually Classifying Concrete Cracks. *Applied Sciences*, 10(7), 2528. doi.org/10.3390/app10072528.
- Ciccone G., 2024. ChatGPT as a Digital Assistant for Archaeology: Insights from the Smart Anomaly Detection Assistant Development. *Heritage*, 7(10), 5428-5445. doi.org/10.3390/heritage7100256
- Hostetter, H., Naser, M. Z., Huang, X., Gales, J., 2024. The role of large language models (AI chatbots) in fire engineering: An examination of technical questions against domain knowledge, *Natural Hazards Research*, 2024, ISSN 2666-5921, doi.org/10.1016/j.nhres.2024.06.003.
- Hatir, M. E., Barstuğan, M., İnce, İ., 2020. Deep learning-based weathering type recognition in historical stone monuments. *Journal of Cultural Heritage*, 45, 193-203.
- ICOMOS-ISCS. Illustrated Glossary on Stone Deterioration Patterns. 2008 Available online: https://www.icomos.org/public/publications/monuments_and_sites/15/pdf/Monuments_and_Sites_15_ISCS_Glossary_Stone.pdf (accessed on 12 September 2024).
- Karimi, N., Mishra, M., Lourenço, P.B., 2024. Deep learning-based automated tile defect detection system for Portuguese cultural heritage buildings. *Journal of Cultural Heritage*, 68, 86-98.
- Kim, D., Kim, T., Kim, Y., Byun, Y. H., Yun, T.S., 2024. A ChatGPT-MATLAB framework for numerical modeling in geotechnical engineering applications. *Computers and Geotechnics*, 169, 106237.
- Llamas, J. M., Lerones, P., Medina, R., Zalama, E., Gómez-García-Bermejo, J., 2017. Classification of architectural heritage images using deep learning techniques. *Applied Sciences*, 7(10), 992.
- Oscó, L. P., Lemos, E. L. D., Gonçalves, W. N., Ramos, A. P. M., Marcato Junior, J., 2023. The potential of visual ChatGPT for remote sensing. *Remote Sensing*, 15(13), 3232.
- Ma, J., Yan, W., Liu, G., Xing, S., Niu, S., Wei, T., 2022. Complex texture contour feature extraction of cracks in timber structures of ancient architecture based on YOLO algorithm. *Advances in Civil Engineering*, 1, 7879302.
- Mansuri, L.E., Patel, D.A., 2022. Artificial intelligence-based automatic visual inspection system for built heritage. *Smart and Sustainable Built Environment*, 11(3), 622-646.
- Mishra, M., Lourenço, P.B., 2024. Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review. *Journal of Cultural Heritage*, 66, 536-550.
- Mishra, M., Barman, T., Ramana, G. V., 2024. Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage. *Journal of Civil Structural Health Monitoring*, 14(1), 103-120.
- Muhammad, M. B., Yeasin, M., 2020. Eigen-cam: Class activation map using principal components. *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 1-7, doi: 10.1109/IJCNN48605.2020.9206626. 1-7.
- Prasad, R., Choudhary, P., 2021. State of-the-art of artificial intelligence. *Journal of Mobile Multimedia*, 17, 427-454.
- Prieto, S. A., Mengiste, E. T., García de Soto, B., 2023. Investigating the use of ChatGPT for the scheduling of construction projects. *Buildings*, 13(4), 857.
- Ray, P. P., 2024. ChatGPT in transforming communication in seismic engineering: Case studies, implications, key challenges and future directions. *Earthquake Science*, 37(4), 352-367.
- Saka, A., Taiwo, R., Saka, N., Salami, B.A., Ajayi, S., Akande, K. and Kazemi, H., 2023. GPT models in construction industry: Opportunities, limitations, and a use case validation. *Developments in the Built Environment*, 100300.
- Samhouri, M., Al-Arabi, L. Al-Atrash, F., 2022. Prediction and measurement of damage to architectural heritages facades using convolutional neural networks. *Neural Computing and Applications*, 34(20), 18125-18141.
- Siountri, K., Anagnostopoulos, C. N., 2023. The classification of cultural heritage buildings in Athens using deep learning techniques. *Heritage*, 6(4), 3673-3705.
- Spennemann, D. H., 2023. Exploring ethical boundaries: Can ChatGPT be prompted to give advice on how to cheat in university assignments?. Preprints. doi.org/10.20944/preprints202308.1271.v1
- Uddin, SMJ, Albert, A., Ovid, A., Alsharif, A., 2023. Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training. *Sustainability*, 15(9), 7121. doi.org/10.3390/su15097121
- UNI – Ente Italiano di Normazione, 2006. UNI 11182: Materiali lapidei naturali ed artificiali. Descrizione della forma di alterazione. Termini e definizioni. Milano.
- Wang, N., Zhao, X., Zhao, P., Zhang, Y., Zou, Z., Ou, J., 2019. Automatic damage detection of historic masonry buildings based on mobile deep learning. *Automation in Construction*, 103, 53-66.
- Zhang, K., Mea, C., Fiorillo, F., Fassi, F., 2024. Classification and object detection for architectural pathology: practical tests with training set. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 48, 477-484.