The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium "Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation", 25–29 August 2025, Seoul, Republic of Korea

Ground Sampling Distance as a Key Parameter for Automatic Crack Detection in Built Heritage: A Practical Framework With YOLOv5

Simon Boutet[1], Pierre Hallot[2]

[1] Faculty of Architecture, University of Liège, Belgium – simon.boutet@student.uliege.be
2 DIVA, Art Archaeology Heritage University of Liège, 4020 Liège, Belgium – p.hallot@uliege.be

Keywords: Artificial intelligence, Deep Learning, Computer Vision, Built heritage, Pathology detection, YOLOv5.

Abstract

This study presents a practical approach to applying deep learning for the conservation of built heritage, focusing on automatic crack detection in historic masonry using the YOLOv5 object detection model. While most existing research emphasizes model precision under controlled conditions, this work evaluates YOLOv5's performance in real-world scenarios, accounting for variations in image acquisition conditions. The study contributes a qualitative comparison of deep learning models relevant to automatic surface pathology detection in built heritage and introduces a field-oriented framework to guide experts in selecting and deploying those tools. A key innovation is the investigation of Ground Sampling Distance (GSD), already used in actual inspection methods like photogrammetry, as a critical parameter influencing detection accuracy and model usability. Results show that YOLOv5 can effectively detect both large cracks and microcracks across varied GSD values, and reinforce the value of interdisciplinary practices that combine Deep Learning technologies with established heritage documentation practices.

1. Introduction

The conservation of built heritage is a fundamental challenge in maintaining culturally and historically significant structures. Among these, masonry constructions represent a substantial portion of global heritage architecture. However, many of these structures continue to serve well beyond their originally intended lifespan, rendering them increasingly vulnerable to structural deterioration due to the progressive aging of construction materials (Saviano et al., 2022). One of the most prevalent and concerning manifestation of this degradation is the appearance of cracks within masonry elements, which, if not identified and addressed in time, can significantly affect structural integrity, leading to long-term damage (Philipparie, 2019).

Traditional visual inspection methods, which rely on expert evaluation, are time-intensive and subject to human interpretation, potentially leading to inconsistencies in diagnosis (Watt & Swallow, 1995) while the use of existing digital documentation techniques based on photogrammetry and lasergrammetry already enhance the accuracy and repeatability of visual surveys (Hallot et al., 2022). Moreover, the integration of artificial intelligence and deep learning presents an opportunity to automate and further improve the accuracy of pathology detection in heritage buildings, enhancing expert analysis (Mishra & Lourenço, 2024).

Yet, most existing research regarding automatic pathology detection in built heritage emphasizes the precision and technical performance of deep learning models, and often overlooks image acquisition in real-world conditions or operational deployment for professionals in the heritage field - an aspect this work seeks to address. Studies by Hallée et al. (2021), Marín-García et al. (2023), and Pratibha et al. (2024), all focusing on brick walls, employed ideal conditions during dataset preparation. Marín-García et al. maintained consistent image distance and angle, while Pratibha et al. excluded images with unrelated pathologies, heterogeneous colors, deformed brick surfaces, or irregular mortar joints. Hallée et al. constructed brick wall segments in the laboratory and manually created cracks for controlled imaging. Although such controlled image acquisition often leads to high-performing models, they limit applicability in real-world scenarios, which are inherently more variable. In contrast, Zou et al. (2019) introduced variation into image capture conditions to better prepare models for environmental fluctuations, and Yang et al. (2023) accounted for diverse brick types, joint sizes, and textures to improve results across different masonry walls.

This paper extends research originally conducted as part of a master's thesis in architecture. It aims to assess the relevance of deep learning in the context of built heritage conservation by focusing on its real-world application - particularly regarding its accessibility to field operators and the robustness of model predictions under varying image acquisition conditions and parameter configurations. The paper is structured as follows:

First, a state-of-the-art review of the various DL tasks and models was conducted to develop a practical guide that helps field operators understand the distinctions between some of the more commonly used DL models in surface pathology detection of built heritage, up to December 2024.

Next, the model deployment methodology was developed as well as image acquisition parameters and conditions. Then, a detailed comparison of local and cloud-based execution environments suitable for running the model was conducted including key training parameters and their impact on model performance.

Finally, we evaluated the results and robustness of YOLOv5's predictions and assessed the accessibility of the technology for its potential users during the image acquisition phase. To do so, we conducted a study on various buildings, manually verifying detected cracks against AI predictions. In this paper, we focused specifically on the impact of Ground Sampling Distance (GSD) on crack detection performance. Although not detailed in this article, we also considered several other image acquisition parameters to better reflect field conditions. To simulate spatial limitations or obstructions around a surveyed building, images were captured at various incidence angles, introducing perspective distortions. Nighttime images with artificial lighting and varying ISO levels simulated low-light conditions, assessing the impact of illumination artifacts and reduced contrast. Some scenes also included visual noise such as vegetation or architectural elements to evaluate the model's robustness. Complete and detailed results of those experiments can be found in Boutet (2025).

2. Deep Learning for Pathology Detection

2.1 State-of-the-art review

Since the first documented studies of automatic surface detection around 2017 (Guo et al., 2024), deep learning has emerged as a powerful tool for RGB image analysis in built heritage conservation. While the most documented and often more complicated research lies in surface pathology detection, many studies have demonstrated the capacity of DL to support diverse tasks, including damage classification, architectural element recognition, quantification of missing features, and real-time detection of pathologies.

For example, Dini et al. (2023) explored the classification of façade conditions. They employed a convolutional neural network to assign a severity level—ranging from intact to severely damaged—to historic building facades, enabling conservation teams to prioritize interventions. Similarly, Kumar et al. (2020) explored post-disaster buildings by applying CNNs to social media images, allowing for rapid evaluation of whether the shown buildings were heritage structures and whether they were damaged or intact. Kwon and Yu (2019) applied DL to identify missing parts in Korean stone structures while Zou et al. (2019) proposed a method to not only detect but also quantify missing components within heritage buildings in Beijing's Forbidden City.

Beyond basic object detection, some studies quantified damage severity. For example, Hatır et al. (2021) integrated crack width measurement into a pixel-wise segmentation model. This enabled not just localization of the pathology, but also the estimation of its seriousness and variation over time. Finally, studies such as Wei et al. (2023) and Pratibha et al. (2024) introduced real-time detection using lightweight models and mobile platforms. These approaches were compatible with smartphones, surveillance systems, or even augmented reality via drones, focusing on more agile, on-site diagnostics.

Together, these studies reflect a wide range of applications in which deep learning technologies increasingly support not only the detection and interpretation of pathologies but also the documentation and intervention planning for cultural heritage buildings. While some reviews of surface defect detection using deep learning already address data limitations (Guo et al., 2024) and explore various applications in cultural heritage (Mishra et al., 2024), model performance is typically assessed using quantitative metrics such as mean Average Precision (mAP), Intersection over Union (IoU) and inference speed (Padilla et al., 2021). However, these indicators alone do not fully capture the models' specifications for diverse tasks. In fact, the most widely employed models are often designed and reviewed for specific computer vision tasks, such as image classification (IC), object detection (OD), semantic segmentation (SS), or instance segmentation (IS). As a result, it is unclear in which applications those models might perform better, and when an alternative should be selected.

Consequently, there seems to be a lack of a comprehensive and task-oriented framework that clearly outlines the specialization, strengths, and limitations of deep learning models in the field of built heritage conservation. This gap makes it challenging for field experts to determine which models best suit their needs, which ultimately hinders the broader adoption and democratization of automatic pathology detection processes.

To address this issue, this section proposes a qualitative approach as an alternative to mAP-based benchmarking, aiming to support experts in identifying the most relevant models and tasks for their specific applications.

2.2 State-of-the-art framework

The evaluation framework is based on six criteria tailored to the specific needs of automatic pathology detection in built heritage. These include *Precision, Speed, Efficiency, Accessibility*, *Automation Potential* and *Versatility*. These criteria were chosen to align with both technical performance benchmarks and field constraints observed in heritage conservation. Scores were derived from reported results in literature and qualitative analysis (Boutet, 2025).

- ***Precision*** reflects the model's ability to accurately localize pathologies within an image. It covers both the mAP performance relative to similar models and the prediction accuracy, with pixel-wise predictions being considered more precise than bounding box predictions. The image classification task is excluded from this criterion, since it does not involve any localization of the defect.

- ***Inference speed*** applies to the number of images the model can process in a given amount of time (often in frames per second), which is crucial for real-time or large-scale inspection. While many state-of-the-art models today can produce near-instantaneous predictions, this parameter remains essential for identifying models best suited for continuous video-feed analysis or on-site deployment scenarios, and is already commonly used in many studies.

- ***Efficiency*** concerns the model's training time, recommended dataset size, and number of epochs needed to reach convergence during training. A more efficient model needs fewer epochs to reach convergence and thus requires fewer computational resources, while non-efficient models require huge amounts of training data and iterations to reach acceptable mAP values.

- ***Accessibility*** indicates the computational demands of the model. Higher accessibility typically means lower GPU and RAM usage, making it suitable for standard or lower-end hardware. Conversely, low accessibility indicates that the model is resource-intensive and may require high-performance computing infrastructure.

- ***Automation potential*** describes how easily experts can use and interpret the model's prediction outputs for heritage conservation tasks, ideally without requiring extensive post-processing. It is closely associated with the model's precision and recall metrics, as lower values typically indicate a greater need for prediction cleaning and additional labeling, respectively. Additionally, automation potential considers the compatibility of the model's outputs with other digital documentation tools, such as photogrammetry and lasergrammetry, enabling better integration of predictions into 3D heritage models.

- ***Versatility*** indicates the model's ability to perform multiple deep learning tasks (image classification, object detection and image segmentation), increasing its adaptability across various use cases. This parameter is evaluated solely based on scientific literature related to pathology detection in built heritage up until December 2024 and may not accurately represent all the tasks DL models can perform in other domains.

Based on these criteria, nine commonly used models and architectures across all four computer vision tasks were analyzed to give a qualitative comparison of their strengths and weaknesses (Table 1). We compared Convolutional Neural Network (CNN) for image classification and Region-based Convolutional Neural Network (R-CNN), Fast R-CNN, Faster R-CNN, Single Shot MultiBox Detector (SSD), and You Only Look Once (YOLO) for object detection. For segmentation tasks, Mask R-CNN, Fully Convolutional Network (FCN) and DeepLab (based on FCN) were chosen.

Task	Model	Precision	Inference Speed	Efficiency	Accessibility	Automation Potential	Versatility
IC	CNN	/	4	4	6	1	3
OD	R-CNN	3	1	2	3	3	4
	Fast R-CNN	4	2	4	3	3	4
	Faster R-CNN	5	4	4	4	4	4
	SSD	4	5	5	5	4	4
	YOLO	3	6	6	5	4	4
SS & IS	Mask R-CNN	6	2	2	2	6	6
	FCN	5	3	4	3	5	2
	DeepLab (FCN)	5	2	3	2	5	2

Table 1. Performance comparison of DL models and architectures across computer vision tasks (scale from 1 [poor] to 6 [best])

This table provides a clear classification of the computer vision tasks and their specifications, as well as a visual overview of the performances of the deep learning models. This diversity allows for the selection of a model according to project needs, whether that involves real-time results, high precision, use on low-performance computing infrastructure or minimal additional operations by the user.

Regarding precision, each task type addresses a specific level of spatial understanding and annotation complexity: image classification assigns a single label to an entire image, object detection identifies and localizes elements via bounding boxes, semantic segmentation classifies each pixel, and instance segmentation differentiates between multiple objects of the same class on a pixel level. This means that in terms of precision, segmentation models like FCN, DeepLab and Mask R-CNN dominate, as they are capable of precisely outlining object contours. FCN is known for its simplicity and speed in semantic segmentation but lacks fine edge precision due to resolution loss across successive convolutional layers. DeepLab builds on this by introducing the Atrous Spatial Pyramid Pooling (ASPP) module, which enhances its ability to detect objects at multiple scales. This architectural improvement comes at the cost of increased training time and computational load.

Mask R-CNN further refines segmentation by performing instance segmentation, allowing the model to distinguish between separate instances of the same pathology. While this model offers the highest precision among the compared approaches, it also demands significant processing resources. It is, however, built from the architecture of Faster R-CNN and its predecessors, which allows it to achieve other tasks like object detection, making it rather versatile. Despite its complexity, Mask R-CNN allows for near-perfect automation potential with its high accuracy and by being capable of instance segmentation, treating images on the pixel level like other digital survey tools.

If the emphasis lies in higher model accessibility or if pixel wise predictions are not necessary for a selected application, object detection focused models like Faster R-CNN, SSD and YOLO become better choices. Being one stage detectors, the latter two are frequently selected for real-time detection tasks due to their high inference speed and lightweight architecture, making them well suited for mobile or field-based applications. According to Bharati and Pramanik (2020), architectures like SSD and FCN offer high inference speed but fall short of the precision and accuracy achieved by Faster R-CNN. SSD demonstrates strong performance in detecting large objects and maintains a balance between speed and accuracy, outperforming YOLO in precision while remaining faster than Faster R-CNN. While the first iterations of both YOLO and SSD used to encounter difficulties when detecting small objects, especially in images containing larger elements, these limitations have been progressively addressed over the years.

Study from Li et al. (2019) follows these observations by presenting encouraging results for YOLOv5 in pathology detection, while also comparing its efficiency with Faster R-CNN. Although both models achieve comparable accuracy, Faster R-CNN requires nearly four times more training time on the same dataset and 160 times more epochs, making YOLO architecture more efficient in practice.

Image classification seems less pertinent for pathology detection due to the lack of localization of defects, but its simplicity usually allows for more approachable and accessible models following a CNN architecture.

3. YOLOv5 Development Methodology

3.1 Model and case study selection

The study deliberately limits itself to a single category of pathology, one type of material, and one model to focus observations on the model's behavior under varying image acquisition conditions and to obtain sufficiently precise results within the research timeframe.

Based on the scientific literature review by Guo et al. (2024) and observations presented in other studies (Cha et al., 2018; Rao et al., 2020; Li et al., 2023), cracks were identified as the most relevant pathology due to their complexity, their diverse characteristics, and the limitations of bounding boxes in object detection models like.

Following a similar line of reasoning, brick masonry was selected as the target material, as it is observed that most research on automatic damage detection has been concentrated on modern and simpler construction materials like concrete, asphalt, and metal. Karimi et al. (2024) states that there is a growing interest in applying similar techniques to historic masonry structures within the field of built heritage conservation. Also, masonry presents greater challenges, as brick and stone exhibit variations in color, bonding patterns, and shapes depending on context—factors that make it difficult to represent comprehensively within a dataset (Ye et al., 2024). These difficulties are further compounded by the limited availability of annotated datasets focused on masonry-related pathologies, which hinders the development of effective deep learning models (Katsigiannis et al., 2022).

Regarding the deep learning model, YOLOv5 was selected due to its high computational efficiency and accessibility. Recent studies by Pratibha et al. (2023) and Guo et al. (2024) have

further highlighted YOLOv5’s advantages in real-time object detection and its effectiveness during training, making it especially well-suited for the precise localization of cracks in historic masonry. Based on these observations, this specific version of the YOLO architecture was adopted for the study.

3.2 Dataset selection

To realistically evaluate the model's robustness in a context different from its original training environment, we employed a dataset created and annotated by Karimi et al. (2024), which is publicly available on Kaggle (https://www.kaggle.com/datasets/nargeskarimii/various-materials-from-historic-buildings). It features different types of cracks under varying lighting conditions and masonry textures from historical bridges in Isfahan, Iran (fig. 1).

Figure 1. Brick dataset sample from Karimi et al. (2024).

To create the brick dataset, Karimi et al. (2024) acquired high-resolution images using a Samsung Galaxy A32 (64 MP), under diverse weather and lighting conditions. After manually filtering these images to remove non-relevant or low-quality data, brightness and contrast processing were applied to improve damage visibility. Also, 45° image rotation was employed as a data augmentation technique to prevent the model from confusing cracks and mortar joints and to enhance the model’s robustness.

Images were then downscaled from 3456 × 3456 pixels to 416 × 416 pixels to match the input requirements of the YOLOv5 architecture and to optimize GPU resource usage during training. This resolution adjustment was not related to ground sampling distance (GSD) evaluation but rather meant to comply with the model’s standardized input format. Finally, the dataset was partitioned into training (70%), validation (20%), and testing (10%) subsets, resulting in a total of 861 annotated images available for training.

3.3 Model preparation and training

After retrieving the YOLOv5 source code, which is publicly available on GitHub, we chose Google colab, a cloud-based execution environment to run and train the model. It allowed for the execution of the model with minimal setup on the browser using remote GPU servers, significantly accelerating processing time. The model achieved a mAP50 of 96.8% and a mAP50-95 of 68.3% after training for 100 epochs with a batch size of 16. This configuration was identified as the most balanced during the training phase, offering an optimal compromise between training time, hardware resource consumption, and prediction accuracy.

An analysis of epoch variation showed that increasing the number of training epochs led to a linear increase in training time. Although training the model for 500 epochs yielded a near-perfect mAP of 99.5%, prediction quality in unseen data was suboptimal, likely due to overfitting—where the model learns training data too precisely and fails to generalize to unseen images. Conversely, training for 50 epochs proved insufficient for model convergence, with the mAP only reaching 87%. Therefore, 100 epochs were identified as the optimal choice, offering above 96% mAP while maintaining a reasonable training duration of 21 minutes and 12 seconds.

The impact of batch size on training efficiency and memory usage was equally significant. A batch size of 16 was identified as optimal, requiring only 2.3 GB of GPU memory and approximately 4 GB of system RAM, thus making it suitable for standard cloud computing platforms such as Google Colab. Larger batch sizes, such as 128 and 192, slightly reduced training time (to around 19 minutes) but led to a substantial increase in memory usage, up to 15.1 GB of GPU memory and more than 8.5 GB of RAM, approaching the limits of typical hardware and providing only marginal performance gains.

YOLOv5’s training output also revealed a balanced spatial distribution of defects within the dataset (Fig. 2a). However, it simultaneously highlighted a typological imbalance with a disproportionate number of small, vertically oriented cracks dominating the dataset (Fig. 2b). This imbalance and lack of representation of horizontal and stepped cracks could outline some of the model’s prediction and performance issues when faced with these kinds of cracks.

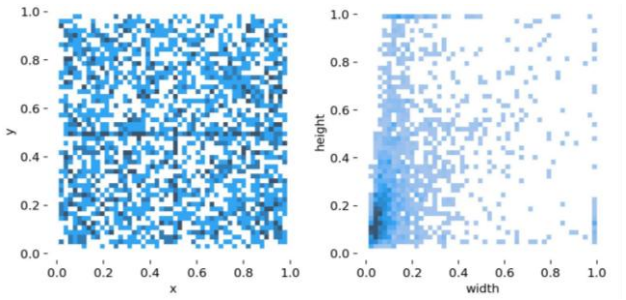

Figure 2. Spatial representation of cracks from the dataset: a) position relative to the image, b) height and width of defects

3.4 Camera settings

All other images presented in this study were used to evaluate YOLOv5’s performance and were captured using a Nikon Z6 camera with a native resolution of 6048 × 4024 pixels in various masonry facades in the city of Liège, Belgium (fig. 3). A Nikon Speedlight SB-700 flash was employed for nighttime photography and a Vanguard tripod was used to mark and replicate camera positions across day-and-night sessions, allowing for consistent framing and enabling the comparison of lighting conditions under near-identical setups.

Figure 3. Masonry façades used for testing YOLOv5’s robustness in the Faculty of Architecture of the University of Liège, Belgium.

To ensure optimal image quality and consistency across all image acquisition conditions, the experimental campaign was conducted over a short period in winter. This limited timeframe allowed for controlled natural lighting conditions, thereby reducing external interference in the evaluation of model performance across the parameters studied. Specifically, the choice of season and case study location excluded facades exposed to extreme sunlight or deep shadows, which could be addressed in a further study to better evaluate YOLOv5 robustness under varied natural lighting conditions.

Similarly, camera settings were adjusted manually in response to varying field conditions to allow for ideal image exposure. By default, images were captured with a shutter speed of 1/30 s and an aperture of f/5. For daytime images, ISO sensitivity was maintained at 100. These parameters were selected in reference to commonly used guidelines in architectural photogrammetry to ensure compatibility with documentation standards. For consistency, the white balance was fixed to “cloudy” throughout the study.

3.5 Image acquisition parameters

The study systematically varied several image acquisition parameters to reflect challenges commonly encountered in heritage visual inspections. This helped assess their potential impact on model performance and offered insights into its robustness beyond controlled laboratory conditions. The first set of experiments focused on validating the trained model in semi-controlled conditions, designed to closely resemble the training environment with similar crack proportions. The goal was to establish a performance baseline for further experiments and to identify early signs of the model's limitations due to training imbalances. While this process allows for rapid verification of a DL model’s effectiveness, it holds little value for architects and experts aiming to diagnose an entire facade or building. Pre-identifying pathologies or cropping them outside their original context before applying the model is ultimately less efficient than manually annotating the original images. Therefore, it was essential to verify whether pathology detection could be performed from greater distances.

Following this objective, one of the most critical parameters tested was GSD variation. The GSD (Ground Sampling Distance) corresponds to the distance between the centers of two consecutive pixels. It varies depending on the size of the camera sensor, the image resolution, the focal length used, and the distance to the photographed surface. It is equivalent to the spatial resolution of the image captured and can be calculated using the following formula (Hallot et al., 2022):

$$GSD = \frac{distance\ between\ the\ image\ and\ the\ object\ \times\ sensor\ size}{focal\ length\ \times\ image\ size}$$

While the focal length and distance can only be adjusted during field analysis, the image resolution is modified during predictions to comply with the model’s input requirements. The sensor size depends on the camera used and measures 36×24 mm for the Nikon Z6. This means that the operator’s distance to the defect, the focal length of the camera and the model’s detection command settings directly affect the number of pixels representing each crack and, by extension, their minimum visible detail. Since pathology datasets typically rely on low-resolution, highly zoomed-in images of defects to reduce training time, it is crucial for field operators to understand how the model responds to varying defect proportions and sizes within the image. By extension, we examined the maximum distance at which YOLOv5 could accurately locate cracks to determine its relevance in visual inspections of cultural heritage.

4. Predictions and Results Analysis

4.1 Training validation

Initial trials using cracks framed similarly to those in the training dataset confirmed the model’s baseline ability to detect real pathologies in unprocessed and never seen field images (fig. 4). Results were satisfactory for vertical cracks; however, YOLOv5 demonstrated limitations in detecting oblique defects, likely due to their low representation in the dataset. The model's tendency to fragment a single crack into multiple smaller bounding boxes also appears to originate from the training data with a high concentration of small vertical instances of defects, suggesting a training imbalance.

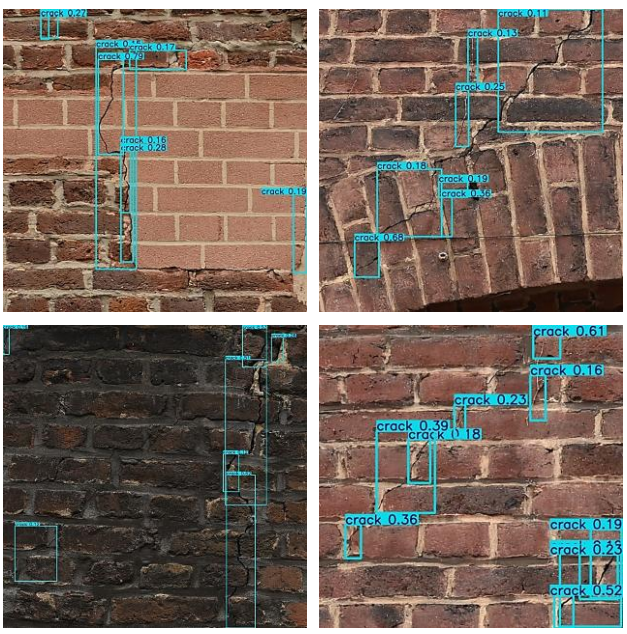

Figure 4. First predictions of YOLOv5

To further explore spatial sensitivity, peripheral pathologies were introduced by capturing scenes where target cracks appeared near the image edges instead of being centered (fig. 5). This aimed to test whether the position of defects within the image affected detection accuracy. In accordance with the graph shown in fig. 2a, results indicated that defect location had no significant impact on performance.

Figure 5. Predictions on peripheral pathologies

Next, rotated images were used to assess the model’s orientation invariance. Cracks were either photographed or digitally rotated to a 45° angle to verify whether background brick alignment interfered with prediction, and whether the detection gap between vertical and oblique cracks was due to limited representation in training. It was confirmed that background orientation did not affect performance (fig. 6) but YOLOv5’s reduced accuracy for oblique cracks is attributable to insufficient representation during training, as represented earlier in fig. 2b.

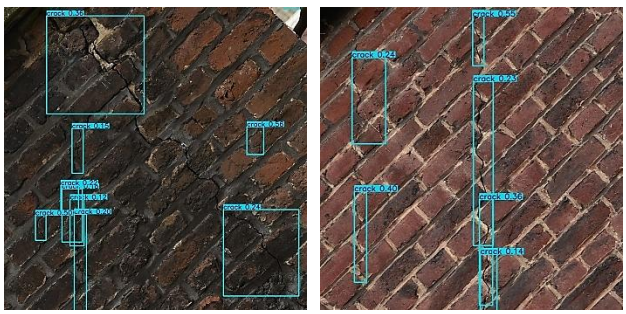

Figure 6. Predictions on images rotated at a 45° angle

4.2 Influence of GSD variation on model performance

Since it was not possible to calculate the ground sampling distance (GSD) of the training images due to the unknown distance between the camera and the object, subsequent experiments were conducted using different image resolutions, thereby generating different GSD values, to determine which range of GSD yielded the most accurate results. The presented results all originate from a specific image which offered optimal conditions for visualizing various crack scales, captured 7 meters from the facade. For clarity, representative results were selectively cropped and enlarged to facilitate the readability of predictions at different input resolutions. Resolution steps were chosen arbitrarily, primarily as multiples of 416 pixels, which is the default input resolution for YOLOv5.

The evaluation of YOLOv5 performance across varying input resolutions revealed that resolution (and GSD) significantly impacts both detection accuracy and operational usability. At lower resolutions such as 208 and 416 pixels, the model failed to detect the two main cracks of the image and only predicted false positives (fig. 7a). However, at 624 pixels, two of the four primary defects were correctly identified. Increasing the resolution to 832 pixels resulted in improved confidence scores and enabled the detection of a third crack, while predicted areas accurately covered most of the crack surfaces, with limited overlap and a manageable number of false positives. Starting from 1040 pixels, the fourth and last main crack became visible, and previously identified cracks remained consistently detected. At 1664 pixels, the model achieved its first successful detections of microcracks, particularly those confined to individual bricks or along mortar joints, without a significant rise in false positives and while keeping confidence scores (fig. 7b).

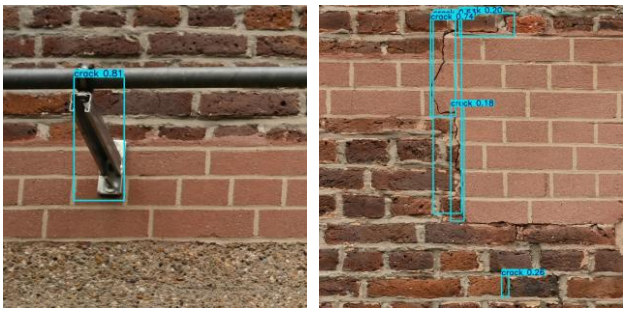

Figure 7. Results at various GSD values:
a) False positive at 8.65 mm/pix, b) Primary crack and first visible micro-crack at 2.16 mm/pix

Beyond this point, however, detection quality began to degrade. At very high resolutions such as 4160, 6240 pixels, and above, the confidence scores for the main cracks decreased, and although the model identified a greater number of microcracks, the volume of predictions and false positives rose sharply. This increase hindered the image output’s readability and interpretability, making it more challenging for human operators to distinguish relevant defects. Table 2 shows a quantitative analysis of those results, while figure 8 allows for better visual representation. True positive predictions were subdivided into *main cracks* and *microcracks*, representing respectively the primary targeted defects and smaller, less critical cracks. False positives were categorized into *mortar joints and brick textures* and *noise*. The first category reflects cases where the model confuses background patterns with actual cracks, while the second is due to unexpected architectural elements which are absent from the training data.

GSD (mm/pix)	Resolution (pix)	True positive		False positive		Total of predictions	% of true positive	% of main cracks
		Main cracks	Microcracks	Mort. joints and brick text.	Noise			
17,31	208	0	0	0	7	7	0,0%	0,0%
8,65	416	0	0	0	6	6	0,0%	0,0%
5,77	624	3	0	0	5	8	37,5%	37,5%
4,33	832	8	0	0	4	12	66,7%	66,7%
2,88	1248	10	0	2	9	21	47,6%	47,6%
2,16	1664	11	3	2	9	25	56,0%	44,0%
1,73	2080	16	11	3	9	39	69,2%	41,0%
1,44	2496	19	12	2	15	48	64,6%	39,6%
1,24	2912	18	16	3	12	49	69,4%	36,7%
1,08	3328	21	17	3	22	63	60,3%	33,3%
0,96	3744	19	16	5	25	65	53,8%	29,2%
0,87	4160	23	17	2	19	61	65,6%	37,7%
0,58	6240	24	20	14	26	84	52,4%	28,6%
0,43	8320	14	27	18	41	100	41,0%	14,0%
0,35	10400	14	19	31	39	103	32,0%	13,6%
0,29	12480	16	20	56	28	120	30,0%	13,3%
0,25	14560	8	18	92	27	145	17,9%	5,5%

Table 2. Analysis of predictions for specific GSD values

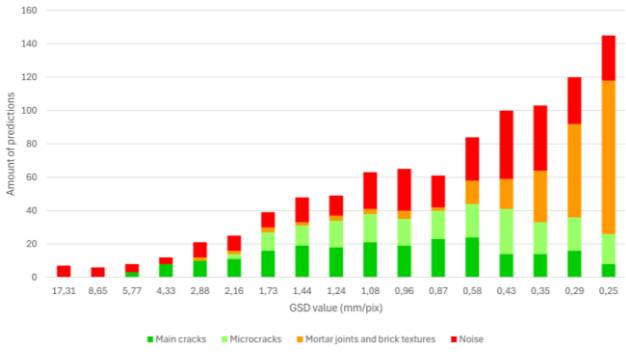

Figure 8. Visual graph of predictions for specific GSD values

Based on these findings, the study concluded that a selected range of input resolutions yielded better results in maximizing true positives while minimizing false positives. In this case, a resolution of 832 pixels provided an optimal balance between accuracy and speed for identifying primary damage. Higher resolutions, such as 1664 and 2080 pixels, showed an improved sensitivity to the shape of major defects and enabled the detection of finer cracks. At resolutions up to 6240 pixels, the model produced finer and more detailed predictions at the cost of an increase in false positives. Further increases in resolution lowered performance meaningfully and even resulted in reduced detection clarity and longer inference time, while resolutions under 624 did not detect any of the primary cracks. Figure 9 further illustrates the impact of GSD and image resolution on prediction accuracy. Annotated colors correspond to earlier tables and figures.

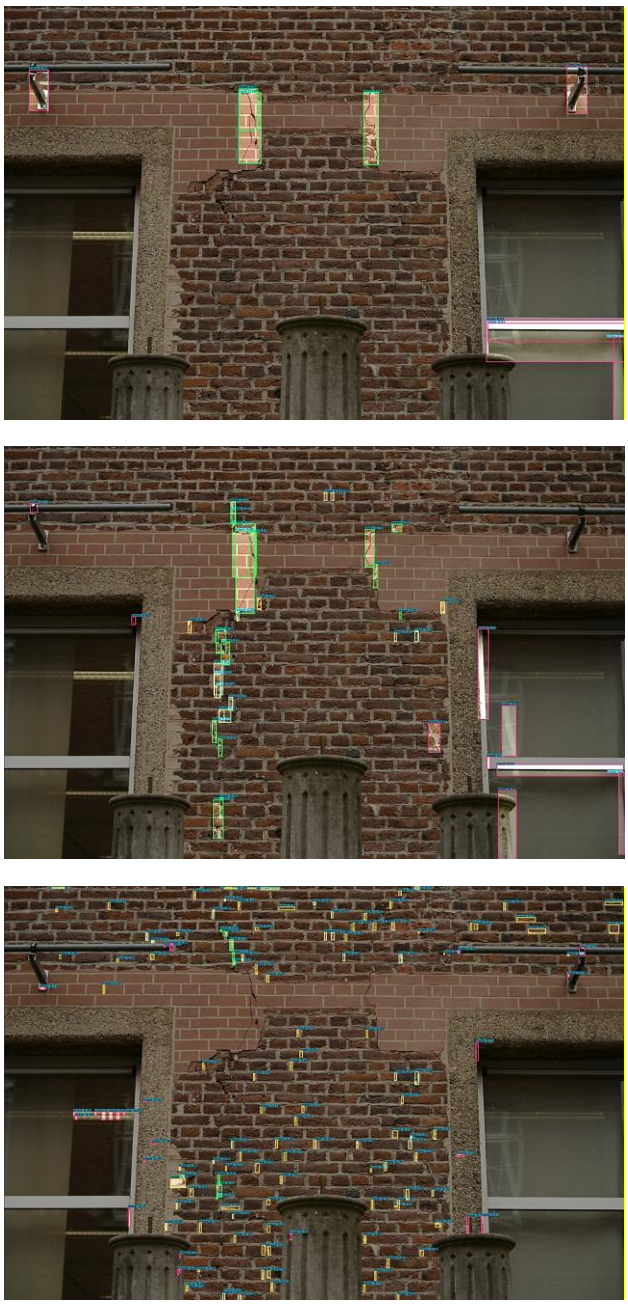

Figure 9. Predictions of YOLOv5 at various GSD values: a) 5.77 mm/pix, b) 1.73 mm/pix and c) 0.25 mm/pix

Similarly to photogrammetry principles, the experiment underscored the importance of maintaining consistent focal length and distance to the object across all images, as excessive variation in ground sampling distance (GSD) negatively impacts the model's performance. However, if field conditions do not allow constant image acquisition settings, manually adjusting the GSD through YOLOv5 detection parameters shows great potential.

5. Conclusion

The integration of Deep Learning models like YOLOv5 into built heritage conservation already represents a significant advancement in automated crack detection, assisting manual inspections while improving diagnostic accuracy. Nonetheless, automatic pathology detection still requires refinements to address environmental and architectural variations.

Despite training on a rather small sample of images from Iran, YOLOv5 demonstrated strong generalizations in detecting previously unseen cracks on varied brick types in Belgium. Performance remained robust under challenging conditions, including high incidence angles, low lighting, and unwanted elements like vegetation. However, the model struggled with certain crack typologies, particularly horizontal and stepped cracks, due to insufficient representation during training. Image rotation and peripheral detection experiments confirmed that these limitations originated from insufficient training rather than model deficiencies.

This study highlighted the critical influence of Ground Sampling Distance (GSD) and input image resolution on the performance and interpretability of deep learning models applied to heritage pathology detection. Through the deployment and testing of YOLOv5 across multiple resolutions, it was demonstrated that specific resolution thresholds enhanced the detection of fine-scale features, such as microcracks, and improved model confidence. However, detection performance declined heavily when the resolution and consequently the GSD value did not meet the specific range. When GSD was significantly lower than training data, it caused YOLOv5 to realize an excessive number of predictions, increasing false positives and causing longer inference times because of the images' high resolution, all of which reduce operational clarity and usability in conservation workflows.

These findings underscore the need to define an optimal GSD not only as a technical parameter, but also as a strategic variable that must be adapted to show various scales of the pathology. In this context, GSD also serves as a bridge between automatic detection methods and traditional photogrammetric standards, offering a quantifiable reference that can guide image acquisition protocols and ensure consistency in field conditions. Future studies should take this into consideration when developing datasets featuring diverse pathologies. This would enable a more comprehensive analysis and comparison of GSD values between training data and field-acquired images, leading to a better understanding of their impact on model predictions.

By framing image resolution within a GSD-based logic, this study provides a foundation for improving both the robustness and efficiency of DL-based detection systems in heritage contexts. Ultimately, the calibration of GSD parameters, alongside appropriate acquisition angles, lighting, and framing, can substantially enhance the interpretability of automated predictions and facilitate their integration into existing documentation and monitoring tools. This reinforces the value of interdisciplinary practices that combine Deep Learning technologies with actual visual inspection methods.

References

Boutet, S. (2025). L'intégration des nouvelles technologies et de l'intelligence artificielle dans la conservation du patrimoine bâti pour détecter la présence de pathologies. (Unpublished master's thesis). Université de Liège, Liège, Belgique. Retrieved from https://matheo.uliege.be/handle/2268.2/22348

Cha, Y.-J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. Computer-Aided Civil and Infrastructure Engineering, 33, 731–747. https://doi.org/10.1111/mice.12334

Dini, M., Fauzi, M. F., & Subekti, D. R. (2023). Facade damage classification of historic buildings in Lasem using deep learning. Journal of Cultural Heritage, 61, 234–242. https://doi.org/10.1016/j.culher.2023.01.009

Guo, J., Liu, P., Xiao, B., Deng, L., & Wang, Q. (2024). Surface defect detection of civil structures using images: Review from data perspective. Automation in Construction, 158, 105186. https://doi.org/10.1016/j.autcon.2023.105186

Hallée, M. J., Napolitano, R. K., Reinhart, W. F., & Glisic, B. (2021). Crack detection in images of masonry using CNNs. Sensors, 21(14), 4929. https://doi.org/10.3390/s21144929

Hallot, P., Mathys, A., & Jouan, P. (2022). Cours de documentation et modélisation du patrimoine tangible, Documentation et modélisation du patrimoine. Université de Liège.

Hatır, M. E., Demirel, M., & Kalkan, S. (2021). Pixel-wise quantification of cracks in concrete using instance segmentation. Computers, Materials & Continua, 66(2), 1941–1960. https://doi.org/10.32604/cmc.2021.014646

Karimi, N., Mishra, M., & Lourenço, P. B. (2024). Automated surface crack detection in historical constructions with various materials using deep learning-based YOLO network. International Journal of Architectural Heritage, 1–17. https://doi.org/10.1080/15583058.2024.2376177

Katsigiannis, S., Seyedzadeh, S., Agapiou, A., & Ramzan, N. (2023). Deep learning for crack detection on masonry façades using limited data and transfer learning. Journal of Building Engineering, 76, 107105.

Kumar, S., Lakhmani, M., & Sharma, S. (2020). Crowdsourced post-disaster condition assessment of cultural heritage using CNNs. Remote Sensing, 12(17), 2781. https://doi.org/10.3390/rs12172781

Kwon, Y., & Yu, I. (2019). Detection of damage in Korean stone heritage using object detection. Heritage Science, 7(1), 12. https://doi.org/10.1186/s40494-019-0260-2

Li, B.-L., Qi, Y., Fan, J.-S., Liu, Y.-F., & Liu, C. (2023). A grid-based classification and box-based detection fusion model for asphalt pavement crack. Computer-Aided Civil and Infrastructure Engineering, 38, 2279–2299. https://doi.org/10.1111/mice.12962

Li, S., Zhao, X., & Zhou, G. (2019). Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Computer-Aided Civil and Infrastructure Engineering, 34(7), 616–634. https://doi.org/10.1111/mice.12433

Marín-García, D., Bienvenido-Huertas, D., Carretero-Ayuso, M. J., & De La Torre, S. (2023). Deep learning model for automated detection of efflorescence and its possible treatment in images of brick facades. Automation in Construction, 145, 104658. https://doi.org/10.1016/j.autcon.2022.104658

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. https://doi.org/10.1016/j.culher.2024.01.005

Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & da Silva, E. A. B. (2021). A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. Electronics, 10(3), 279. https://doi.org/10.3390/electronics10030279

Philipparie, P. (2019). Pathologie générale du bâtiment : Diagnostic, remèdes & prévention. Editions Eyrolles.

Pratibha, K., Mishra, M., Ramana, G. V., & Lourenço, P. B. (2024). Deep learning-based YOLO network model for detecting surface cracks during structural health monitoring. In Y. Endo & T. Hanazato (Eds.), Structural Analysis of Historical Constructions. SAHC 2023. RILEM Bookseries (Vol. 47). Springer. https://doi.org/10.1007/978-3-031-39603-8_15

Rao, A. S., Nguyen, T., Palaniswami, M., & Ngo, T. (2020). Vision-based automated crack detection using convolutional neural networks for condition assessment of infrastructure. Structural Health Monitoring, 20(4), 2124–2142. https://doi.org/10.1177/1475921720965445

Saviano, F., Parisi, F., & Lignola, G. P. (2022). Material aging effects on the in-plane lateral capacity of tuff stone masonry walls: a numerical investigation. Materials and Structures, 55(7), 198.

Watt, D., & Swallow, P. (1995). Surveying historic buildings. Donhead.

Wei, X., Wang, Y., & Zhou, X. (2023). Lightweight YOLOv7 model for real-time defect detection in cultural heritage facades. Sensors, 23(2), 456. https://doi.org/10.3390/s23020456

Yang, X., Zheng, L., Chen, Y., Feng, J., & Zheng, J. (2023). Recognition of Damage Types of Chinese Gray-Brick Ancient Buildings Based on Machine Learning—Taking the Macau World Heritage Buffer Zone as an Example. Atmosphere, 14(2), Article 2. https://doi.org/10.3390/atmos14020346

Ye, Z., Lovell, L., Faramarzi, A., & Ninić, J. (2024). Sam-based instance segmentation models for the automation of structural damage detection. Advanced Engineering Informatics, 62, 102826.

Zou, Q., Ni, L., Zhang, T., Wang, Q., & Wang, S. (2019). Quantitative detection of missing architectural elements in historic sites using Faster R-CNN. Journal of Cultural Heritage, 38, 144–155. https://doi.org/10.1016/j.culher.2019.01.009

Zou, Z., Zhao, X., Zhao, P., Qi, F., & Wang, N. (2019). CNN-based statistics and location estimation of missing components in routine inspection of historic buildings. Journal of Cultural Heritage, 38, 221–230. https://doi.org/10.1016/j.culher.2019.02.002