

Utilizzo di reti neurali convoluzionali nella tecnica Structure from Motion

Carlo Battini
Matilde Ridella

Abstract

Il lavoro propone un workflow integrato che combina YOLOv11 con Agisoft Metashape per automatizzare il pre-processing fotogrammetrico e il riconoscimento di fessurazioni su superfici murarie ad alta complessità visiva, tipiche del campo architettonico e ingegneristico. A partire dalle evidenze dello stato dell'arte sull'uso di YOLO per il degrado e sulle pipeline ibride rilevamento-segmentazione-misurazione, viene sviluppata una pipeline operativa in cui YOLOv11 genera bounding box e maschere successivamente impiegate in Metashape tramite API Python. Il modello è addestrato su un dataset eterogeneo comprendente immagini con fessure annotate e immagini di sfondo. Le prestazioni mostrano come sia possibile utilizzare queste tecnologie anche nel campo della diagnostica all'interno di sistemi che impiegano algoritmi di Structure from Motion. I risultati del rilevamento vengono post-processati in immagini tematiche ad alta leggibilità e reintegrati nel workflow fotogrammetrico per generare texture diagnostiche, trasferendo l'informazione sulle fessurazioni dal 2D alla superficie 3D del manufatto.

Parole chiave

Rilievo Architettonico, Fotogrammetria Digitale, Agisoft Metashape, Crack Detection, YOLOv11.



Dettaglio dell'abside
dell'Oratorio di San
Giovanni Battista
a Bussana Vecchia
(fotografia degli autori, 22
aprile 2023).

Di più direi; ma di men dir bisogna.

Ludovico Ariosto, *Orlando Furioso*, XXVI, 22.

Introduzione

Negli ultimi anni, la fotogrammetria digitale – con la diffusione di piattaforme UAV (Unnamed Aerial Vehicles), sensori ad alta risoluzione e software avanzati di elaborazione tridimensionale, come Agisoft Metashape [1] – ha conosciuto una rapida evoluzione trovando applicazione nel rilievo topografico, nell'ingegneria civile e nella documentazione dei beni culturali.

Nonostante i significativi progressi, i workflow fotogrammetrici tradizionali presentano ancora diverse criticità, soprattutto nella gestione di grandi volumi di dati e nella necessità di interventi manuali durante le fasi di pre-processing. Operazioni come la mascheratura delle immagini, la selezione delle aree di interesse e la rimozione di elementi non rilevanti richiedono tempo e sono fortemente legati all'operatore, incidendo sulla soggettività del processo, sull'efficienza e sulla riproducibilità dei risultati.

Parallelamente, i recenti sviluppi nel campo del Deep Learning [2], e in particolare delle Convolutional Neural Networks (CNNs), hanno portato a notevoli miglioramenti nell'object detection [3] e nell'image segmentation [4]. Tra queste, l'insieme di algoritmi YOLO [5] (You Only Look Once) si è affermata come uno standard nel riconoscimento automatico di oggetti in tempo reale. La versione più recente, YOLOv11, introduce ulteriori miglioramenti in termini di precisione, robustezza e capacità di generalizzazione, risultando particolarmente adatta all'analisi di immagini ad alta risoluzione tipiche della fotogrammetria.

In tempi recenti, diversi studi hanno indagato l'impiego di tecniche di Deep Learning a supporto dei processi fotogrammetrici, ad esempio per la classificazione delle nuvole di punti, la segmentazione semantica o la mascheratura automatica delle immagini. Tuttavia, l'integrazione operativa di modelli di object detection avanzati all'interno di software commerciali come Metashape risulta ancora limitata e poco esplorata, soprattutto in relazione all'utilizzo delle versioni più recenti degli algoritmi YOLO.

In questo contesto, il presente lavoro propone un workflow integrato che combina YOLOv11 con Agisoft Metashape al fine di automatizzare il riconoscimento di oggetti e il pre-processing delle immagini fotogrammetriche. In particolare, YOLOv11 viene impiegato per l'individuazione automatica di specifiche classi di oggetti, come le fessurazioni, all'interno delle immagini, generando bounding box e maschere successivamente utilizzate all'interno di Metashape tramite API Python. L'obiettivo è migliorare l'efficienza del processo e la qualità dei prodotti finali, riducendo al contempo l'intervento manuale. In ambito architettonico e ingegneristico il monitoraggio del quadro fessurativo è molto spesso cruciale, ma viene svolto ancora oggi in modo analogico, individuando manualmente le fessurazioni sulle ortofoto. Sensibile, dunque, sarebbe il vantaggio apportato dall'applicazione dell'IA in questo campo. Il contributo principale dello studio consiste nella definizione e validazione di una pipeline operativa che integra Deep Learning nei tradizionali workflow fotogrammetrici per individuare automaticamente e con precisione le fessurazioni presenti sulla muratura. I risultati dimostrano come l'uso di YOLOv11 possa rappresentare una soluzione efficace per l'automazione dei processi di elaborazione, aprendo nuove prospettive nel campo della fotogrammetria digitale.

Stato dell'Arte

L'integrazione di tecniche di computer vision e Deep Learning nei processi di rilievo digitale rappresenta uno degli sviluppi più rilevanti degli ultimi anni nel campo della fotogrammetria applicata al costruito. In particolare, le reti neurali convoluzionali hanno dimostrato un'elevata efficacia nell'automatizzare operazioni di riconoscimento e classificazione visiva che, nei workflow fotogrammetrici tradizionali, risultano ancora fortemente dipendenti dall'intervento umano.

Nel contesto della documentazione del patrimonio architettonico e decorativo, diversi studi hanno evidenziato come modelli di object detection basati su YOLO consentano di individuare con precisione fenomeni di degrado superficiale quali lacune pittoriche, distacchi e discontinuità materiche su superfici murarie complesse [Chen 2025].

In particolare, l'approccio PLDS-YOLO, basato su una versione ottimizzata di YOLOv8 con strutture di feature fusion multiscala e backbone duali, ha dimostrato come sia possibile ottenere risultati accurati mantenendo una ridotta complessità computazionale,

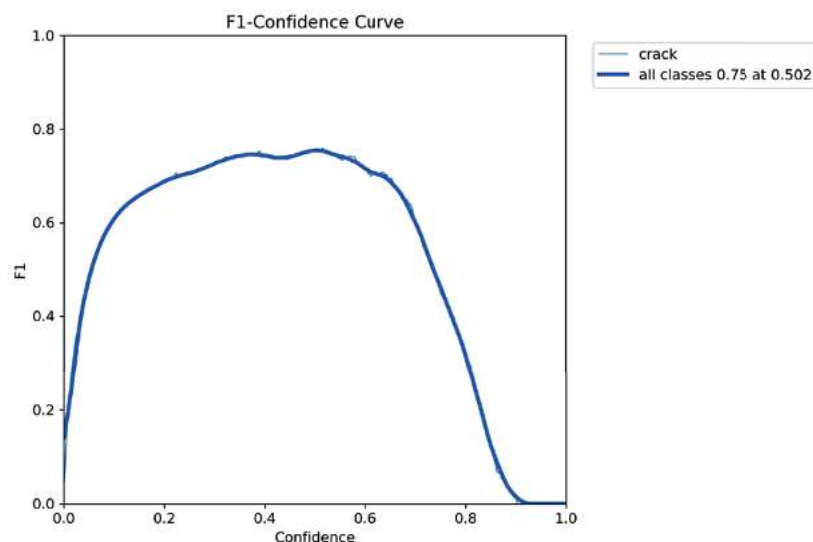


Fig. 1. Curva F1-Confidence per la classe "crack". Il grafico evidenzia il bilanciamento tra precisione e sensibilità del modello YOLOv11, che raggiunge picco $F1=0.75$ a soglia di confidenza 0.502. Il plateau stabile nell'intervallo 0.3-0.6 dimostra la robustezza nel discriminare le lesioni da elementi di disturbo visivo.

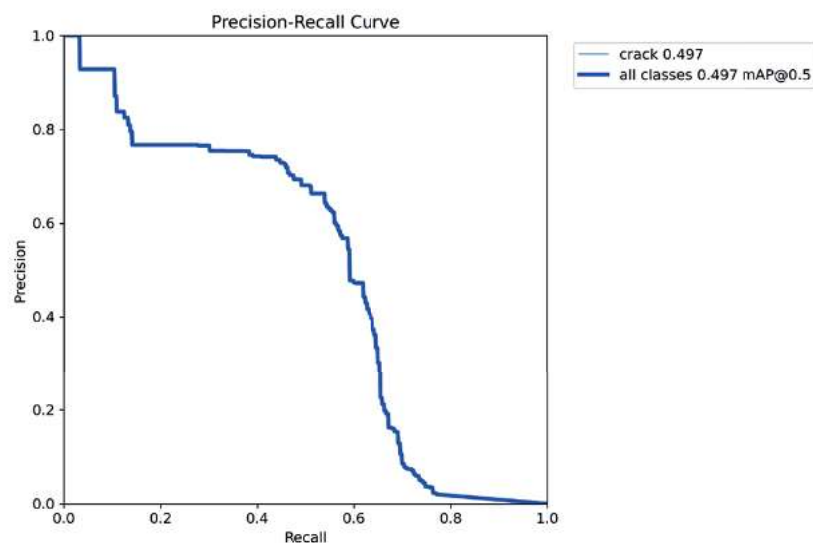
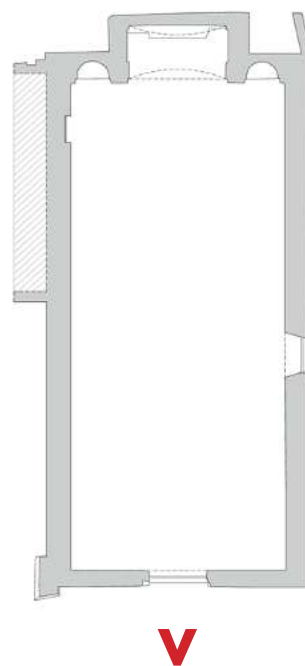


Fig. 2. Grafico della curva Precision-Recall (PR) per la classe "crack". La pendenza della curva indica una priorità assegnata alla precisione rispetto alla sensibilità totale: una scelta metodologica finalizzata a garantire la pulizia dei keypoints per la successiva elaborazione fotogrammetrica.



Fig. 3. A sinistra, una foto della facciata principale dell'Oratorio di San Giovanni Battista a Bussana Vecchia (fotografia degli autori, 22 aprile 2023). A destra, una pianta schematica del manufatto.



caratteristica fondamentale per applicazioni su dataset fotogrammetrici estesi. Tali strategie sono state ulteriormente sviluppate da Guo e Kumar (2025), che propongono DKR-YOLO, un modello orientato alla digitalizzazione sostenibile delle tessiture murarie antiche, introducendo convoluzioni adattive e strutture dinamiche in grado di gestire morfologie allungate e variabilità stilistica tipiche delle superfici storiche.

Il framework YOLO Mural integra la fase di rilevamento automatico del degrado con la ricostruzione digitale basata su reti generative avversarie (GAN), dimostrando come analisi diagnostica e restituzione virtuale possano essere concepite come un unico processo end-to-end [Chen 2025]. Questo approccio risulta particolarmente rilevante nei contesti in cui la documentazione digitale non è finalizzata esclusivamente alla misurazione geometrica, ma anche alla conservazione, interpretazione e valorizzazione del bene culturale.

Parallelamente, ambiti affini quali il rilievo di edifici di nuova costruzione e delle infrastrutture hanno fornito contributi metodologici trasferibili ai workflow fotogrammetrici. Studi recenti mostrano come insiemi collaborativi di reti encoder-decoder migliorino la segmentazione pixel-level delle fessurazioni rispetto ai modelli singoli, aumentando la robustezza in presenza di texture complesse e rumore visivo [Alkayem 2024]. Altri lavori propongono pipeline ibride che combinano modelli YOLO per il rilevamento delle regioni di interesse con tecniche di post-processing geometrico, consentendo la misurazione accurata della larghezza delle discontinuità a partire da immagini ad alta risoluzione [Tran 2023].

Su scala più ampia, valutazioni sistematiche dei modelli YOLOv5 applicati a pavimentazioni asfaltate confermano accuratezze superiori all'85%, evidenziando un chiaro compromesso tra velocità di inferenza e precisione metrica, aspetto centrale nei flussi di lavoro fotogrammetrici che prevedono l'elaborazione di grandi volumi di immagini [Hu 2021]. Una sintesi organica di tali contributi è fornita dall'indagine di Nguyen et al. (2023), che analizza 26 modelli di Deep Learning per il rilevamento delle fessurazioni, individuando ResNet e DenseNet come backbone efficaci per la classificazione, Faster R-CNN per il rilevamento ad alta accuratezza e pix2pix come riferimento per la segmentazione semantica, sottolineando al contempo il ruolo crescente di approcci semi-supervisionati per ridurre i costi di annotazione.

In questa traiettoria evolutiva si collocano i più recenti sviluppi basati su architetture ibride CNN-Transformer. Yu e Zhou (2023) propongono un modello YOLOv5-CBoT che integra



Fig. 4. A sinistra, l'immagine originale (fotografia degli autori, 15 gennaio 2023). A destra, l'immagine processata all'interno di YOLOv7. Si noti come la maggior parte delle fessurazioni, in rosso, siano state identificate.

Bottleneck Transformer e modulo C2f, migliorando la modellazione delle dipendenze spaziali a lungo raggio tipiche delle fessurazioni lineari e riducendo al contempo il numero di parametri e il costo computazionale. Tali risultati confermano la crescente importanza di modelli capaci di operare efficacemente su immagini ad alta risoluzione, come quelle comunemente impiegate nei rilievi fotogrammetrici UAV.

Infine, studi applicativi su larga scala relativi ai ponti dimostrano la maturità operativa di pipeline Deep Learning a due stadi. Confronti estesi indicano che YOLOv7 rappresenta una soluzione particolarmente efficace per il rilevamento automatico delle discontinuità in

immagini ad altissima risoluzione, mentre architetture U-Net opportunamente ottimizzate superano pix2pix nella segmentazione e nella stima metrica della larghezza delle fessure, raggiungendo accuratèzze superiori al 92% in applicazioni reali di rilievo e monitoraggio [Tran 2023]. Nel complesso, lo stato dell'arte evidenzia una chiara convergenza tra fotogrammetria digitale e deep learning, ma mette in luce anche una lacuna operativa ancora poco esplorata: l'assenza di workflow pienamente integrati che incorporino modelli avanzati di object detection direttamente nelle fasi di pre-processing dei software fotogrammetrici.

Sviluppo e applicazione del modello CNNs in SfM

Alla luce delle considerazioni emerse dallo stato dell'arte, il presente lavoro propone un workflow integrato che combina tecniche di object detection basate su YOLOv11 con i processi di elaborazione fotogrammetrica implementati in Agisoft Metashape, con specifico riferimento all'individuazione delle fessurazioni su superfici murarie affrescate.



Fig. 5. In alto, l'immagine originale (fotografia degli autori, 15 gennaio 2023). In basso, l'immagine processata all'interno di YOLOv11.

L'obiettivo è quello di sviluppare una pipeline operativa in grado di operare efficacemente in contesti caratterizzati da elevata complessità visiva, riducendo l'incidenza dei falsi positivi e minimizzando l'intervento manuale nelle fasi di pre-processing.

La scelta di focalizzare l'analisi su pareti affrescate deriva dalla particolare criticità di tali superfici, che presentano una forte variabilità cromatica e materica, oltre alla frequente presenza di elementi architettonici e decorativi quali cornici, modanature e arredi. Questi elementi, non essendo riconducibili a fenomeni di degrado strutturale, costituiscono una fonte significativa di ambiguità per i sistemi automatici di rilevamento, rendendo necessaria una strategia di addestramento mirata.

In questo contesto, YOLOv11 è stato addestrato su un dataset eterogeneo includendo non

solo immagini contenenti fessurazioni annotate [6] (3717 immagini con le corrispondenti annotazioni per il training e 200 immagini con le annotazioni per la validazione), ma anche immagini di background prive di difetti strutturali (1763 per training e 442 per la validazione). L'inclusione sistematica di tali immagini nel processo di training consente al modello di apprendere esplicitamente i pattern visivi associati ai falsi positivi, migliorando la capacità di discriminazione tra fessure reali ed elementi lineari o discontinuità apparenti legate alla decorazione pittorica o alla morfologia architettonica.

La validazione metrica dell'architettura YOLOv11 ha confermato l'elevata specificità del sistema nel filtraggio delle componenti non strutturali, un requisito fondamentale per la successiva fase di ricostruzione tridimensionale. L'analisi della Matrice di Confusione ha



Fig. 6. Elaborazione delle immagini costituenti il dataset necessario alla realizzazione del modello tridimensionale della Chiesa di San Carlo Borromeo a Cassinghen. In alto, l'immagine originale. In basso, l'immagine processata all'interno di YOLOv11.

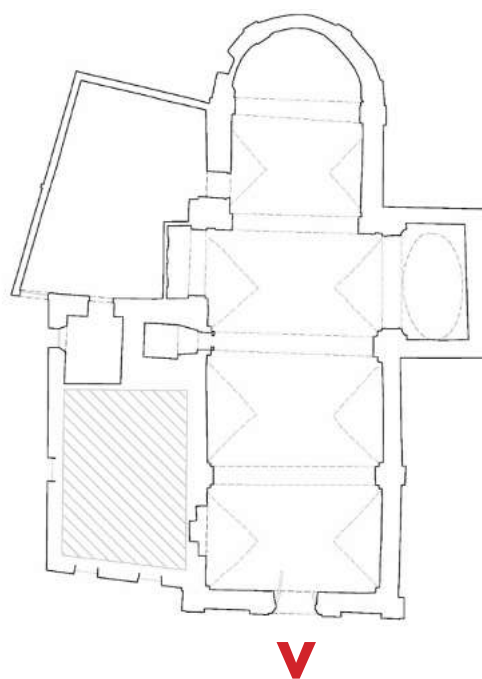


Fig. 7. A sinistra, una foto della facciata principale dell'Chiesa di San Carlo Borromeo a Cassinigheno (fotografia di G. Sacco, 3 ottobre 2020). A destra, una pianta schematica del manufatto.



Fig. 8. Elaborazione SfM delle immagini originali e loro ulteriore sviluppo in ambiente Metashape per mezzo dell'algoritmo.

evidenziato come il modello sia stato addestrato con un approccio conservativo: la capacità di discriminare con successo le patologie fessurative dal complesso background decorativo, come affreschi, modanature e arredi, ha permesso di minimizzare i falsi positivi, garantendo che solo le reali discontinuità murarie venissero processate nella pipeline Structure from Motion (SfM).

Dal punto di vista delle metriche prestazionali, il modello ha raggiunto un picco di FI-Confidence Score pari a 0.75 con una soglia di confidenza di 0.502 (fig. 1).

La stabilità di tale curva dimostra che il classificatore mantiene un bilanciamento ottimale tra precisione e capacità di richiamo (recall) anche in presenza di elevata entropia visiva.

Il valore di mAP@0.5 di 0.497, osservato nella curva Precision-Recall (PR) (fig. 2), riflette la complessità intrinseca del dataset: in contesti monumentali, dove le lesioni possono presentare spessori millimetrici e confondersi con i pigmenti pittorici, tale valore indica un modello estremamente selettivo. Tale selettività è cruciale per la qualità del rilievo, poiché evita la generazione di tie-points errati derivanti da artefatti visuali o decorazioni bidimensionali.

Una volta completata la fase di addestramento, le immagini fotogrammetriche sono state analizzate automaticamente dal modello YOLOv11, restituendo per ciascuna immagine le regioni di interesse associate alla classe "fessura". I primi test hanno visto l'applicazione del modello addestrato ad alcune immagini ricche di fessure ed altre patologie come lacune, vegetazione ecc. Le immagini impiegate fanno parte di un vasto lavoro di studio ed analisi dell'Oratorio di San Giovanni Battista a Bussana Vecchia, Sanremo (Imperia) (fig. 3).

L'Oratorio, costruito in muratura portante, presenta un'unica navata di forma rettangolare e un'abside coperta da una volta a botte lunettata, quasi completamente crollata. La muratura perimetrale è realizzata in pietra a spacco, principalmente calcare marnoso e arenaria. Un piccolo campanile svetta sopra la parte absidale, posizionata a sud dell'Oratorio [Valentini 2023]. In questo caso, l'applicazione dell'algoritmo mostra come anche le fessure più leggere possano essere evidenziate ed impiegate per analizzare il degrado ed i cedimenti presenti nell'Oratorio (figg. 4, 5).

Il secondo test è stato eseguito utilizzando il dataset acquisito durante i rilievi compiuti sulla Chiesa di San Carlo Borromeo a Casingheno, frazione di Fascia - Genova (fig. 6).

La chiesa, di cui qui si prende in considerazione solo un dettaglio, presenta una navata unica, un presbiterio, un'abside semicircolare, due cappelle laterali e un campanile. La navata e il presbiterio sono coperti da volte a botte, mentre l'abside è sovrastato da una cupola emisferica. L'intero edificio è realizzato in muratura di pietra ed è coperto da un tetto ligneo recentemente ricostruito [Sacco 2023] (fig. 7).

In questa fase, lo script Python è stato integrato direttamente all'interno di Agisoft Metashape per automatizzare il processo di individuazione delle fessure e la creazione di un apposito chunk dedicato. Tale chunk, rispetto a quello originale, utilizza immagini fotografiche pre-processate dall'algoritmo YOLO per la generazione delle texture del modello tridimensionale, con le fessurazioni opportunamente evidenziate. Le immagini così elaborate sono state quindi reintegrate nel workflow fotogrammetrico di Metashape – semplicemente modificando il percorso delle immagini - e utilizzate per la generazione delle texture del modello 3D (fig. 8). In questo modo, l'informazione diagnostica relativa alla presenza e alla distribuzione delle fessurazioni viene trasferita dal dominio bidimensionale alla superficie tridimensionale del manufatto, consentendo una rappresentazione spazialmente coerente e direttamente interpretabile del fenomeno (fig. 9).

Il risultato ottenuto, a meno di qualche errore nelle situazioni più dubbie, è molto fedele rispetto a quello evidenziato dall'analisi condotta manualmente.

Conclusioni

Il lavoro dimostra come l'integrazione di modelli di object detection basati su YOLOv11 all'interno dei workflow fotogrammetrici consenta di automatizzare in modo efficace le fasi di pre-processing, riducendo l'intervento manuale e la soggettività dell'operatore. L'applicazione al rilievo di superfici murarie affrescate evidenzia la capacità del sistema di discriminare, con una buona percentuale, le fessurazioni, anche in caso di complesso background decorativo. In particolare, risultano ancora difficili da individuare piccole fessurazioni su pareti uniformi e, in special modo, su pareti affrescate.

La reintroduzione delle informazioni diagnostiche nel modello tridimensionale permette infine una lettura spazialmente coerente del degrado, aprendo nuove prospettive per il monitoraggio e la documentazione digitale del patrimonio architettonico.

Allo stesso tempo ci sono ancora molti aspetti critici che influenzano il risultato finale del

processo quali la complessità dell'immagine, l'illuminazione ambientale, la distanza di presa, il contrasto dei colori e la presenza di oggetti altamente contrastanti con l'architettura investigata. Tra le possibili applicazioni è pensabile l'impiego nel campo della costruzione di ambienti HBIM dove l'implementazione automatizzata può garantire una maggiore interazione con il modello e la creazione di nuove relazioni per la diagnostica.

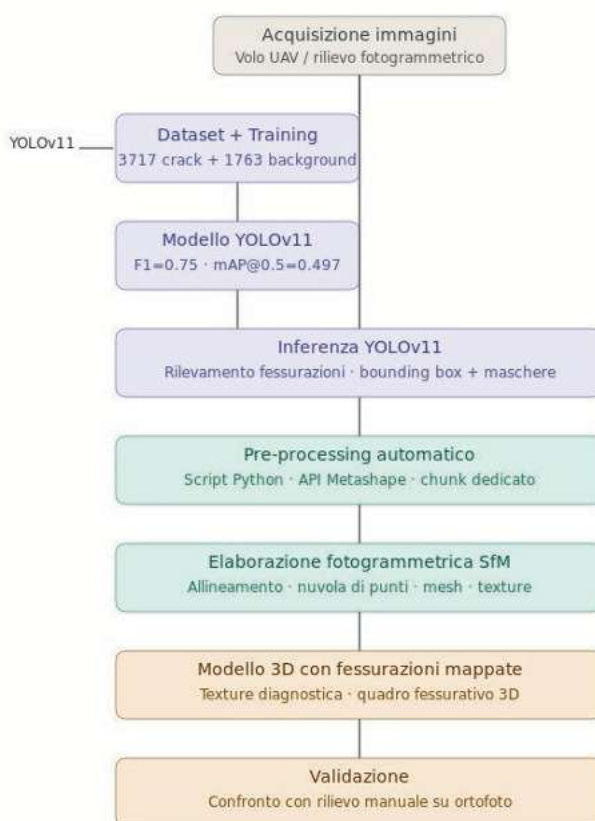


Fig. 9. Grafico riassuntivo della pipeline metodologica adottata.

Note

[1] Agisoft Metashape è un software autonomo che esegue l'elaborazione fotogrammetrica di immagini digitali e genera dati spaziali 3D da utilizzare in applicazioni GIS, documentazione del patrimonio culturale e produzione di effetti visivi, nonché per misurazioni indirette di oggetti di varie scale.

[2] Il Deep Learning è un sottoinsieme del Machine Learning che utilizza reti neurali multistrato (reti profonde) per apprendere automaticamente rappresentazioni e modelli complessi di dati.

[3] Il rilevamento di oggetti è una tecnica di computer vision che utilizza le reti neurali per localizzare e classificare gli oggetti nelle immagini digitali. In quanto tale, è un'istanza di Intelligenza Artificiale che consiste nell'addestrare i computer a riconoscere e classificare gli oggetti in base a categorie predefinite.

[4] La segmentazione è il processo con il quale si classificano i pixel di un'immagine che abbiano caratteristiche comuni, pertanto, ciascun pixel afferente ad una regione è simile, per una qualche proprietà o caratteristica, agli altri della medesima regione.

[5] Il link al progetto: <https://pjreddie.com/darknet/yolo/>.

[6] Immagini acquisite dal repository "crack Dataset" [University 2024], un'Open Source Dataset pubblicato da Roboflow su Roboflow Universe nel 2024. <https://universe.roboflow.com/university-bswxt/crack-bphdr>.

Riferimenti bibliografici

- Alkayem, N. F., Mayya, A., Shen, L., Zhang, X., Asteris, P. G., Wang, Q., Cao, M. (2024). Co-CrackSegment: a New Collaborative Deep Learning Framework for Pixel-Level Semantic Segmentation of Concrete Cracks. In *Mathematics*, 12(19), 3105. <https://doi.org/10.3390/math12193105>.
- Chen, L., Wu, L., Wan, J. (2025). Damage Detection and Digital Reconstruction Method for Grotto Murals Based on YOLOv10. In *npj Heritage Science*, vol. 102. <https://doi.org/10.1038/s40494-025-01607-z>.
- Chen, Y., Zhang, A., Shi, J., Gao, F., Guo, J., Wang, R. (2025). Paint Loss Detection and Segmentation Based on YOLO: an Improved Model for Ancient Murals and Color Painting. In *Heritage*, 8(4), 136. <https://doi.org/10.3390/heritage8040136>.
- Guo, Z., Kumar, S. (2025). AI-Driven Recognition and Sustainable Preservation of Ancient Murals: the DKR-YOLO Framework. In *Heritage*, 8(10), 402. <https://doi.org/10.3390/heritage8100402>.
- Hu, G. X., Hu, B. L., Yang, Z., Huang, L., Li, P. (2021). Pavement Crack Detection Method Based on Deep Learning Models. In *Wireless Communication and Mobile Computing*. <https://doi.org/10.1155/2021/5573590>.
- Nguyen, S. D., Tran, T. S., Tran, V. P., Lee, H. J., Piran, M. J., Le, V. P. (2023). Deep Learning-Based Crack Detection: a Survey. In *International Journal of Pavement Research and Technology*, n. 16, pp. 943-967. <https://doi.org/10.1007/s42947-022-00172-z>.
- Sacco, G. L. S., Ferrero, C., Battini, C., Calderini, C. (2023). Combined Use of Deformation and Structural Analysis for the Structural Damage Assessment of Heritage Buildings: A Case Study in the Liguria Region (Italy). In *Engineering Failure Analysis*, n. 147 <https://doi.org/10.1016/j.engfailanal.2023.107154>.
- Sapkota, R., Flores-Calero, M., Qureshi, R., Badgujar, C., Nepal, U., Poullose, A., Zeno, P., Vaddevolu, U. B. P., Khan, S., Shoman, M., Yan, H., Karkee, M. (2025). YOLO Advances to Its Genesis: A Decadal and Comprehensive Review of the You Only Look Once (YOLO) Series. In *Artificial Intelligence Review*, vol. 58, n. 274. <https://doi.org/10.1007/s10462-025-11253-3>.
- Tran, T. S., Nguyen, S. D., Lee, H. J., Tran, V. P. (2023). Advanced Crack Detection and Segmentation on Bridge Decks Using Deep Learning. In *Construction and Building Materials*, n. 400. <https://doi.org/10.1016/j.conbuildmat.2023.132839>.
- Valentini, M., Battini, C., Vecchiattini, R. (2023). HBIM to Support the Executive Design of a Restoration. Critical Issues Related to Geometric and Semantic Modeling. In *SCIRES-IT - SCientific REsearch and Information Technology*, vol. 13, n. 2, pp. 125-136. <https://doi.org/10.2423/i22394303v13n2p125>.
- Wang, P., Fan, X., Yang, Q., Tian, S., Yu, L. (2025). Object Detection of Mural Images Based on Improved YOLOv8. In *Multimedia Systems*, vol. 31, n. 93. <https://doi.org/10.1007/s00530-025-01687-8>.
- Wu, Y., Li, S., Li, J., Yu, Y., Li, J., Li, Y. (2025). Deep Learning in Crack Detection: A Comprehensive Scientometric Review. In *Journal of Infrastructure Intelligence and Resilience*, vol. 4, n. 3. <https://doi.org/10.1016/j.jintel.2025.100144>.
- Yu, G., Zhou, X. (2023). An Improved YOLOv5 Crack Detection Method Combined with a Bottleneck Transformer. In *Mathematics*, n. 11. <https://doi.org/10.3390/math11102377>.
- Zhang, K., Tekin, A., Mea, C., Younan, S., Chalhoub, M., Fassi, F. (2026). From Survey to Action: AI-Driven Severe Damage Mapping. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII, pp. 519-526. <https://doi.org/10.5194/isprs-archives-XLVIII-2-W12-2026-519-2026>.

Autori

Carlo Battini, Università di Genova, carlo.battini@unige.it

Matilde Ridella, Università di Genova, matilde.ridella@edu.unige.it

Per citare questo capitolo: Carlo Battini, Matilde Ridella (2026). Utilizzo di reti neurali convoluzionali nella tecnica Structure from Motion. In M. Balzani, M. Incerti, F. Maietti, L. Rossato, F. Raco (a cura di) *Disegno furioso. Atti del 47° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione/Furious Drawing. Proceedings of the 47th International Conference of Representation Disciplines Teachers*. Roma: UID Press, pp. 1867-1888.

The Use of Convolutional Neural Networks in the Structure from Motion Technique

Carlo Battini
Matilde Ridella

Abstract

This study proposes an integrated workflow that combines YOLOv11 with Agisoft Metashape to automate photogrammetric pre-processing and the detection of cracks on masonry surfaces with high visual complexity, typical of the architectural and engineering fields. Building on state-of-the-art research into the use of YOLO for degradation detection and hybrid detection-segmentation-measurement pipelines, an operational pipeline is developed in which YOLOv11 generates bounding boxes and masks that are subsequently utilised in Metashape via the Python API. The model is trained on a heterogeneous dataset comprising images with annotated cracks and background images. The performance results demonstrate how these technologies can also be utilised in the field of diagnostics within systems employing Structure from Motion algorithms. The detection results are post-processed into highly readable thematic images and reintegrated into the photogrammetric workflow to generate diagnostic textures, transferring information on cracks from 2D to the 3D surface of the structure.

Keywords

Architectural Surveying, Digital Photogrammetry, Agisoft Metashape, Crack Detection, YOLOv11.



Detail of the apse of the Oratory of St John the Baptist in Bussana Vecchia (photograph by the authors, 22 April 2023).

Di più direi; ma di men dir bisogna.

Ludovico Ariosto, *Orlando Furioso*, XXVI, 22.

Introduction

In recent years, digital photogrammetry – driven by the widespread adoption of UAV (Unmanned Aerial Vehicle) platforms, high-resolution sensors and advanced 3D processing software such as Agisoft Metashape [1] – has undergone rapid development, finding applications in topographic surveying, civil engineering and the documentation of cultural heritage.

Despite significant progress, traditional photogrammetric workflows still present several challenges, particularly in managing large volumes of data and the need for manual intervention during pre-processing stages. Operations such as image masking, the selection of areas of interest and the removal of irrelevant elements are time-consuming and heavily operator-dependent, affecting the subjectivity of the process, efficiency and the reproducibility of results. At the same time, recent developments in the field of Deep Learning [2], and in particular Convolutional Neural Networks (CNNs), have led to significant improvements in object detection [3] and image segmentation [4]. Among these, the YOLO [5] (You Only Look Once) algorithm suite has established itself as a standard in real-time automatic object recognition. The latest version, YOLOv11, introduces further improvements in terms of accuracy, robustness and generalisation ability, making it particularly well-suited to the analysis of high-resolution images typical of photogrammetry.

In recent times, several studies have investigated the use of deep learning techniques to support photogrammetric processes, for example for point cloud classification, semantic segmentation or automatic image masking. However, the operational integration of advanced object detection models within commercial software such as Metashape remains limited and under-explored, particularly in relation to the use of the latest versions of YOLO algorithms. In this context, this paper proposes an integrated workflow that combines YOLOv11 with Agisoft Metashape in order to automate object recognition and the pre-processing of photogrammetric images. In particular, YOLOv11 is used for the automatic detection of specific object classes, such as cracks, within images, generating bounding boxes and masks that are subsequently utilised within Metashape via the Python API. The aim is to improve the efficiency of the process and the quality of the final products, whilst reducing the need for manual intervention. In the fields of architecture and engineering, monitoring the cracking pattern is very often crucial, but is still carried out today using analogue methods, with cracks being identified manually on orthophotos. The benefits of applying AI in this field would therefore be significant.

The main contribution of this study lies in the definition and validation of an operational pipeline that integrates deep learning into traditional photogrammetric workflows to automatically and accurately detect cracks in masonry. The results demonstrate that the use of YOLOv11 can provide an effective solution for automating processing workflows, opening up new possibilities in the field of digital photogrammetry.

State of the Arts

The integration of computer vision and deep learning techniques into digital surveying processes represents one of the most significant developments in recent years in the field of photogrammetry applied to the built environment. In particular, convolutional neural networks have proven highly effective in automating visual recognition and classification tasks which, in traditional photogrammetric workflows, remain heavily reliant on human intervention. In the context of documenting architectural and decorative heritage, several studies have highlighted how YOLO-based object detection models enable the precise identification of surface degradation such as paint loss, flaking and material discontinuities on complex wall surfaces [Chen Y. 2025]. In particular, the PLDS-YOLO approach, based on an optimised version of YOLOv8 with multi-scale feature fusion structures and dual backbones, has demonstrated that it is possible to achieve accurate results whilst maintaining low computational complexity, a fundamental characteristic for applications on large photogrammetric datasets.

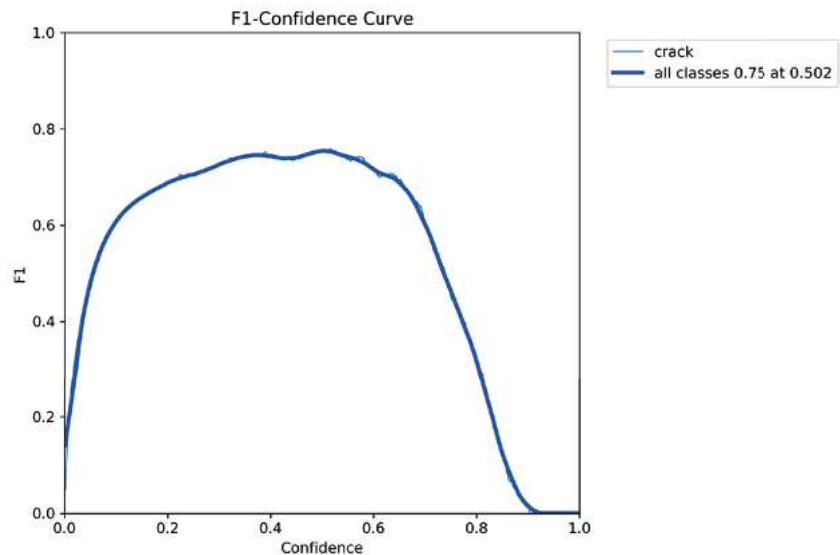


Fig. 1. F1-Confidence curve for the "crack" class. The graph highlights the balance between accuracy and sensitivity of the YOLOv11 model, which achieves a peak F1 score of 0.75 at a confidence threshold of 0.502. The stable plateau in the 0.3–0.6 range demonstrates the model's robustness in distinguishing lesions from visual noise.

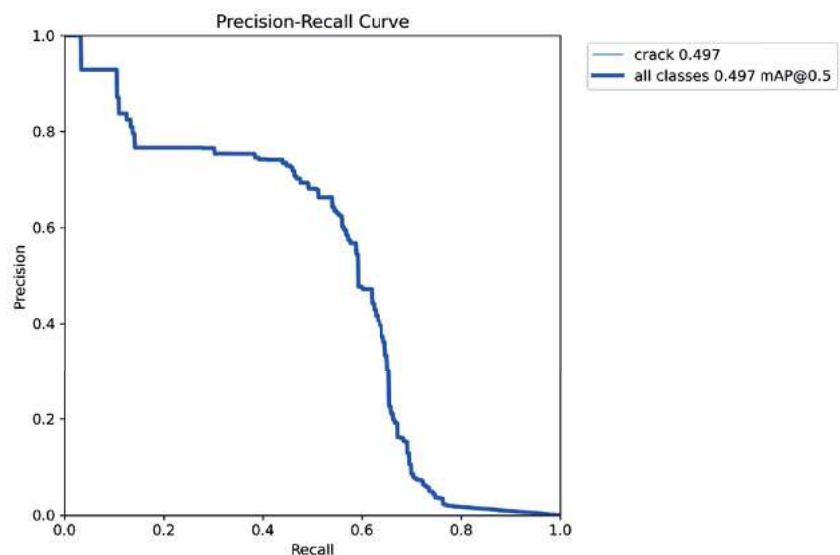
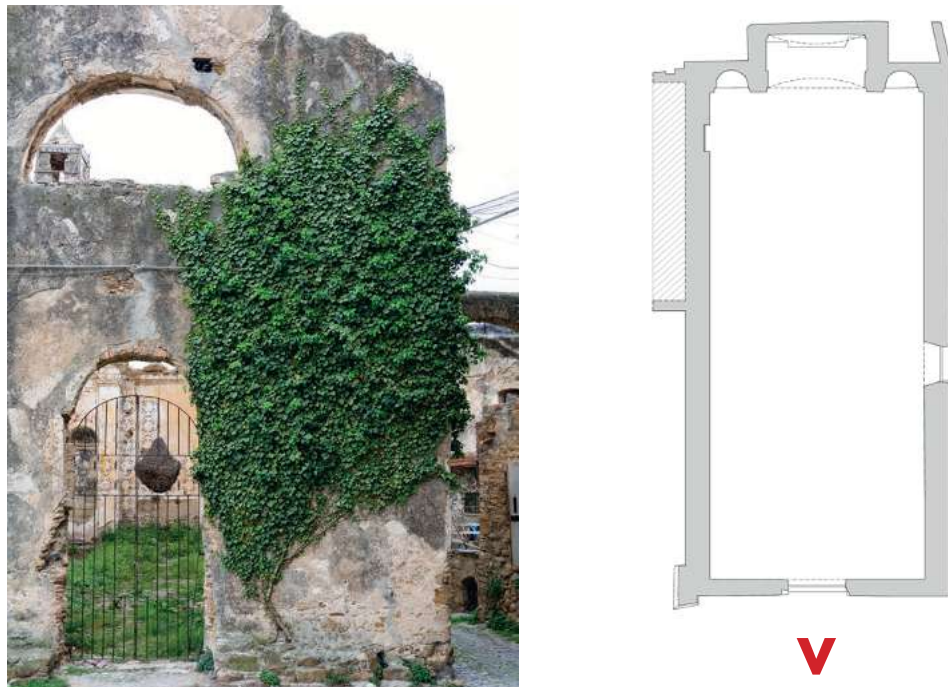


Fig. 2. Precision-Recall (PR) curve for the "crack" class. The slope of the curve indicates that priority has been given to precision over overall sensitivity: a methodological choice aimed at ensuring the quality of the keypoints for subsequent photogrammetric processing.

Fig. 3. On the left, a photograph of the main façade of the Oratory of San Giovanni Battista in Bussana Vecchia (photograph by the authors, 22 April 2023). On the right, a schematic plan of the building.



These strategies were further developed by Guo and Kumar (2025), who propose DKR-YOLO, a model geared towards the sustainable digitisation of ancient wall textures, introducing adaptive convolutions and dynamic structures capable of handling the elongated morphologies and stylistic variability typical of historic surfaces.

The YOLO Mural framework integrates the automatic detection of deterioration with digital reconstruction based on generative adversarial networks (GANs), demonstrating how diagnostic analysis and virtual reconstruction can be conceived as a single end-to-end process [Chen 2025]. This approach is particularly relevant in contexts where digital documentation is not aimed solely at geometric measurement, but also at the conservation, interpretation and enhancement of cultural heritage.

At the same time, related fields such as the surveying of new buildings and infrastructure have provided methodological insights that can be applied to photogrammetric workflows. Recent studies show that collaborative ensembles of encoder–decoder networks improve pixel-level segmentation of cracks compared to single models, increasing robustness in the presence of complex textures and visual noise [Alkayem 2024]. Other works propose hybrid pipelines that combine YOLO models for the detection of regions of interest with geometric post-processing techniques, enabling the accurate measurement of crack widths from high-resolution images [Tran 2023].

On a broader scale, systematic evaluations of YOLOv5 models applied to asphalt surfaces confirm accuracies exceeding 85%, highlighting a clear trade-off between inference speed and metric accuracy – a key consideration in photogrammetric workflows involving the processing of large volumes of images [Hu 2021]. A comprehensive summary of these contributions is provided by the study by Nguyen et al. (2023), which analyses 26 deep learning models for crack detection, identifying ResNet and DenseNet as effective backbones for classification, Faster R-CNN for high-accuracy detection and pix2pix as a benchmark for semantic segmentation, whilst emphasising the growing role of semi-supervised approaches in reducing annotation costs.

The most recent developments based on hybrid CNN–Transformer architectures fall within this evolutionary trajectory. Yu and Zhou (2023) propose a YOLOv5-CBoT model that integrates a Bottleneck Transformer and a C2f module, improving the modelling of long-



Fig. 4. On the left, the original image (photograph by the authors, 15 January 2023). On the right, the image processed using YOLOv11. Note how most of the cracks, shown in red, have been identified.

range spatial dependencies typical of linear cracks whilst reducing the number of parameters and computational cost. These results confirm the growing importance of models capable of operating effectively on high-resolution images, such as those commonly used in UAV photogrammetric surveys.

Finally, large-scale applied studies on bridges demonstrate the operational maturity of two-stage Deep Learning pipelines. Extensive comparisons indicate that YOLOv7 represents a particularly effective solution for the automatic detection of discontinuities in very high-resolution images, whilst suitably optimised U-Net architectures outperform pix2pix in

segmentation and in the metric estimation of crack widths, achieving accuracies exceeding 92% in real-world surveying and monitoring applications [Tran 2023]. Overall, the current state of the art reveals a clear convergence between digital photogrammetry and deep learning, but it also highlights an operational gap that has yet to be fully explored: the lack of fully integrated workflows that incorporate advanced object detection models directly into the pre-processing stages of photogrammetric software.



Fig. 5. Above, the original image (photograph by the authors, 15 January 2023). Below, the image processed using YOLOv11.

Development and Application of the CNN Model in SfM

In light of the findings from the state-of-the-art review, this study proposes an integrated workflow that combines object detection techniques based on YOLOv11 with photogrammetric processing implemented in Agisoft Metashape, with a specific focus on identifying cracks on frescoed wall surfaces. The aim is to develop an operational pipeline capable of functioning effectively in contexts characterised by high visual complexity, reducing the incidence of false positives and minimising manual intervention during the pre-processing stages. The decision to focus the analysis on frescoed walls stems from the particular challenges posed by such surfaces, which exhibit significant chromatic and textural variability, as well as the frequent presence of architectural and decorative elements such as cornices, mouldings and furnishings. As these elements are not attributable to structural deterioration, they constitute a significant source of ambiguity for automatic detection systems, necessitating a targeted training strategy. In this context, YOLOv11 was trained on a heterogeneous dataset comprising not only images containing annotated cracks [6] (3,717 images with corresponding annotations for training and 200 images with annotations for validation), but also background images free of structural defects (1,763 for training and 442 for validation).

The systematic inclusion of these images in the training process allows the model to explicitly learn the visual patterns associated with false positives, improving its ability to distinguish between real cracks and linear features or apparent discontinuities related to pictorial decoration or architectural morphology. Metrical validation of the YOLOv11 architecture confirmed the system's high specificity in filtering out non-structural components, a fundamental requirement for the subsequent three-dimensional reconstruction phase. Analysis of the confusion matrix highlighted that the model was trained using a conservative approach: its ability to successfully distinguish crack-related pathologies from complex



Fig. 6. Processing of the images comprising the dataset required to create the 3D model of the Church of San Carlo Borromeo in Cassingheno. Above, the original image. Below, the image processed using YOLOv11.



Fig. 7. On the left, a photograph of the main façade of the Church of San Carlo Borromeo in Cassingheno (photograph by G. Sacco, 3 October 2020). On the right, a schematic plan of the building.



Fig. 8. SfM processing of the original images and their further processing in the Metashape environment using the algorithm.

decorative backgrounds, such as frescoes, mouldings and furnishings, made it possible to minimise false positives, ensuring that only genuine wall discontinuities were processed in the Structure from Motion (SfM) pipeline. In terms of performance metrics, the model achieved a peak F1-Confidence Score of 0.75 with a confidence threshold of 0.502 (fig. 1). The stability of this curve demonstrates that the classifier maintains an optimal balance between precision and recall even in the presence of high visual entropy. The F1-score of 0.497, observed in the Precision-Recall (PR) curve (fig. 2), reflects the intrinsic complexity of the dataset: in monumental contexts, where lesions may be only a

few millimetres thick and blend in with the paint pigments, this value indicates an extremely selective model. This selectivity is crucial for the quality of the survey, as it prevents the generation of erroneous tie-points resulting from visual artefacts or two-dimensional decorations. Once the training phase was complete, the photogrammetric images were automatically analysed by the YOLOv11 model, which identified the regions of interest associated with the 'crack' class for each image. Initial tests involved applying the trained model to a number of images containing cracks and other defects such as gaps, vegetation, etc. The images used form part of a comprehensive study and analysis of the Oratory of St John the Baptist in Bussana Vecchia, Sanremo (Imperia) (fig. 3). The Oratory, built of load-bearing masonry, features a single rectangular nave and an apse covered by a barrel vault with lunettes, which has almost completely collapsed. The perimeter walls are constructed of split stone, mainly marly limestone and sandstone. A small bell tower rises above the apse, situated to the south of the Oratory [Valentini 2023].

In this case, the application of the algorithm shows how even the slightest cracks can be highlighted and used to analyse the deterioration and subsidence present in the Oratory (figs. 4, 5).

The second test was carried out using the dataset acquired during the surveys conducted on the Church of San Carlo Borromeo in Cassingheno, a hamlet of Fascia (Genoa) (fig. 6). The church, of which only a detail is considered here, features a single nave, a chancel, a semicircular apse, two side chapels and a bell tower. The nave and chancel are covered by barrel vaults, whilst the apse is surmounted by a hemispherical dome. The entire building is constructed of stone masonry and is covered by a recently reconstructed wooden roof [Sacco 2023] (fig. 7).

At this stage, the Python script was integrated directly into Agisoft Metashape to automate the process of identifying cracks and creating a dedicated chunk. Unlike the original chunk, this one uses photographic images pre-processed by the YOLO algorithm to generate the textures for the 3D model, with the cracks clearly highlighted. The images processed in this way were then reintegrated into Metashape's photogrammetric workflow – simply by changing the image path – and used to generate the textures for the 3D model (fig. 8). In this way, diagnostic information regarding the presence and distribution of cracks is transferred from the two-dimensional domain to the three-dimensional surface of the structure, enabling a spatially consistent and directly interpretable representation of the phenomenon (fig. 9).

The result obtained, barring a few errors in the most ambiguous cases, is very accurate compared to that revealed by the manual analysis.

Conclusions

The study demonstrates how the integration of YOLOv11-based object detection models into photogrammetric workflows enables the effective automation of pre-processing stages, reducing the need for manual intervention and operator subjectivity. The application to the survey of frescoed walls highlights the system's ability to distinguish, with a high degree of accuracy, cracks, even in the presence of complex decorative backgrounds. In particular, small cracks on uniform walls, and especially on frescoed walls, remain difficult to detect.

Finally, the reintegration of diagnostic information into the three-dimensional model allows for a spatially coherent interpretation of the deterioration, opening up new prospects for the monitoring and digital documentation of architectural heritage.

At the same time, there are still many critical factors that influence the final outcome of the process, such as image complexity, ambient lighting, shooting distance, colour contrast, and the presence of objects that contrast sharply with the architecture under investigation. Among the possible applications, one can envisage its use in the field of HBIM environment construction, where automated implementation can ensure greater interaction with the model and the creation of new relationships for diagnostics.

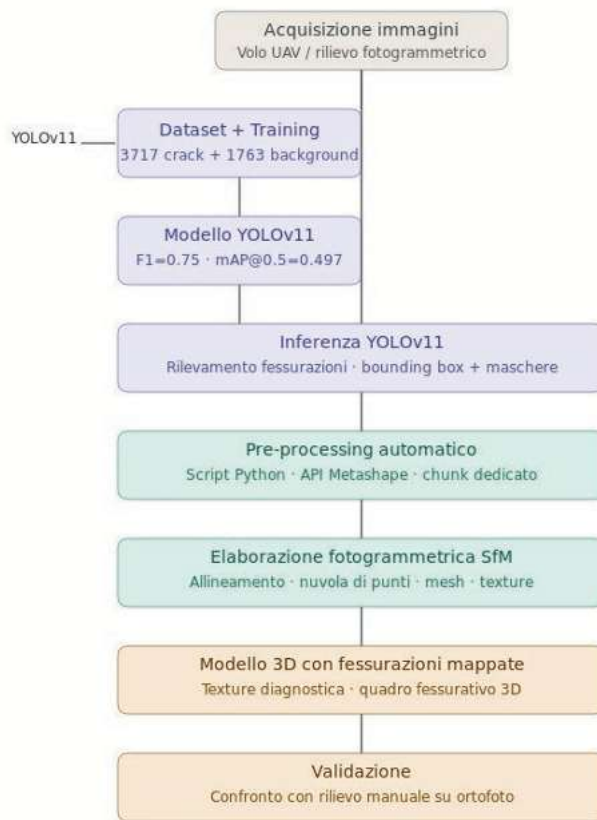


Fig. 9. Summary diagram of the methodological approach adopted.

Notes

[1] Agisoft Metashape is a standalone software programme that performs photogrammetric processing of digital images and generates 3D spatial data for use in GIS applications, cultural heritage documentation and visual effects production, as well as for indirect measurements of objects of various scales.

[2] Deep learning is a subset of machine learning that uses multi-layer neural networks (deep networks) to automatically learn complex representations and models of data.

[3] Object detection is a computer vision technique that uses neural networks to locate and classify objects in digital images. As such, it is a form of artificial intelligence that involves training computers to recognise and classify objects according to predefined categories.

[4] Segmentation is the process of classifying pixels in an image that share common characteristics; consequently, each pixel within a region is similar, in terms of some property or characteristic, to the others in the same region.

[5] Link to the project: <https://pjreddie.com/darknet/yolo/>.

[6] Images sourced from the “crack Dataset” repository [University 2024], an open-source dataset published by Roboflow on Roboflow Universe in 2024. <https://universe.roboflow.com/university-bswxt/crack-bphdr>.

Reference List

- Alkayem, N. F., Mayya, A., Shen, L., Zhang, X., Asteris, P. G., Wang, Q., Cao, M. (2024). Co-CrackSegment: a New Collaborative Deep Learning Framework for Pixel-Level Semantic Segmentation of Concrete Cracks. In *Mathematics*, 12(19), 3105. <https://doi.org/10.3390/math12193105>.
- Chen, L., Wu, L., Wan, J. (2025). Damage Detection and Digital Reconstruction Method for Grotto Murals Based on YOLOv10. In *npj Heritage Science*, vol. 102. <https://doi.org/10.1038/s40494-025-01607-z>.
- Chen, Y., Zhang, A., Shi, J., Gao, F., Guo, J., Wang, R. (2025). Paint Loss Detection and Segmentation Based on YOLO: an Improved Model for Ancient Murals and Color Painting. In *Heritage*, 8(4), 136. <https://doi.org/10.3390/heritage8040136>.
- Guo, Z., Kumar, S. (2025). AI-Driven Recognition and Sustainable Preservation of Ancient Murals: the DKR-YOLO Framework. In *Heritage*, 8(10), 402. <https://doi.org/10.3390/heritage8100402>.
- Hu, G. X., Hu, B. L., Yang, Z., Huang, L., Li, P. (2021). Pavement Crack Detection Method Based on Deep Learning Models. In *Wireless Communication and Mobile Computing*. <https://doi.org/10.1155/2021/5573590>.
- Nguyen, S. D., Tran, T. S., Tran, V. P., Lee, H. J., Piran, M. J., Le, V. P. (2023). Deep Learning-Based Crack Detection: a Survey. In *International Journal of Pavement Research and Technology*, n. 16, pp. 943-967. <https://doi.org/10.1007/s42947-022-00172-z>.
- Sacco, G. L. S., Ferrero, C., Battini, C., Calderini, C. (2023). Combined Use of Deformation and Structural Analysis for the Structural Damage Assessment of Heritage Buildings: A Case Study in the Liguria Region (Italy). In *Engineering Failure Analysis*, n. 147 <https://doi.org/10.1016/j.engfailanal.2023.107154>.
- Sapkota, R., Flores-Calero, M., Qureshi, R., Badgajar, C., Nepal, U., Poullose, A., Zeno, P., Vaddevolu, U. B. P., Khan, S., Shoman, M., Yan, H., Karkee, M. (2025). YOLO Advances to Its Genesis: A Decadal and Comprehensive Review of the You Only Look Once (YOLO) Series. In *Artificial Intelligence Review*, vol. 58, n. 274. <https://doi.org/10.1007/s10462-025-11253-3>.
- Tran, T. S., Nguyen, S. D., Lee, H. J., Tran, V. P. (2023). Advanced Crack Detection and Segmentation on Bridge Decks Using Deep Learning. In *Construction and Building Materials*, n. 400. <https://doi.org/10.1016/j.conbuildmat.2023.132839>.
- Valentini, M., Battini, C., Vecchiattini, R. (2023). HBIM to Support the Executive Design of a Restoration. Critical Issues Related to Geometric and Semantic Modeling. In *SCIRES-IT - SCientific REsearch and Information Technology*, vol. 13, n. 2, pp. 125-136. <https://doi.org/10.2423/i22394303v13n2p125>.
- Wang, P., Fan, X., Yang, Q., Tian, S., Yu, L. (2025). Object Detection of Mural Images Based on Improved YOLOv8. In *Multimedia Systems*, vol. 31, n. 93. <https://doi.org/10.1007/s00530-025-01687-8>.
- Wu, Y., Li, S., Li, J., Yu, Y., Li, J., Li, Y. (2025). Deep Learning in Crack Detection: A Comprehensive Scientometric Review. In *Journal of Infrastructure Intelligence and Resilience*, vol. 4, n. 3. <https://doi.org/10.1016/j.jintel.2025.100144>.
- Yu, G., Zhou, X. (2023). An Improved YOLOv5 Crack Detection Method Combined with a Bottleneck Transformer. In *Mathematics*, n. 11. <https://doi.org/10.3390/math11102377>.
- Zhang, K., Tekin, A., Mea, C., Younan, S., Chalhoub, M., Fassi, F. (2026). From Survey to Action: AI-Driven Severe Damage Mapping. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII, pp. 519-526. <https://doi.org/10.5194/isprs-archives-XLVIII-2-W12-2026-519-2026>.

Authors

Carlo Battini, University of Genoa, carlo.battini@unige.it
Matilde Ridella, University of Genoa, matilde.ridella@edu.unige.it

To cite this chapter: Carlo Battini, Matilde Ridella (2026). Utilizzo di reti neurali convoluzionali nella tecnica Structure from Motion/The Use of Convolutional Neural Networks in the Structure from Motion Technique. In M. Balzani, M. Incerti, F. Maietti, L. Rossato, F. Raco (Eds.) *Disegno furioso. Atti del 47° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione/Furious Drawing. Proceedings of the 47th International Conference of Representation Disciplines Teachers*. Roma: UID Press, pp. 1867-1888.