Towards Automated Image-Based Detection of Non-Structural Elements.

Mauricio Guamán[1], Marco Todaro[1], Volkan Ozsarac[1], Davit Shahnazaryan[1], Daniele Sivori[1], Gerard J. O'Reilly[1],

1- Centre for Training and Research on Reduction of Seismic Risk (ROSE Centre), Scuola Universitaria Superiore IUSS di Pavia, Pavia 27100, Italy

mauricio.guaman@iusspavia.it

Abstract

Italy is highly exposed to seismic hazards, and past earthquakes have repeatedly shown that a significant share of losses and functional disruption stems from damage to non-structural elements (NSEs). Their vulnerability not only threatens occupant safety but also prolongs downtime and complicates post-event recovery, making rapid and reliable assessment essential. However, current practice relies heavily on manual inspection by engineers, a process that is subjective, slow, and resource intensive. In this context, image-based AI recognition provides a powerful opportunity to automate the damage characterisation of NSEs, reducing human workload and enabling faster, more consistent evaluations. The SONATA project directly addresses this need for next-generation seismic risk-mitigation tools in Italy. Within SONATA, vision-based damage detection is conceived as a sequential AI pipeline that first identifies the element being observed (Stage 1: object detection) and subsequently evaluates its damage state (Stage 2: damage characterisation). As an initial step, this paper introduces the proposed workflow based on a YOLO-based instance segmentation model, which is employed to detect and localise NSEs, including partitions and suspended ceilings. Due to the limited data available, the model is trained on a curated and manually annotated subset of the IDEA dataset, augmented to improve robustness. The results of the first stage demonstrate that the proposed model can reliably detect and segment NSEs in complex indoor scenes, achieving high precision and consistent convergence during training. These findings confirm the feasibility of the proposed detection pipeline and establish a basis for the subsequent development of automated damage detection. This study therefore represents an initial step toward a fully integrated, AI-enabled workflow for rapid seismic damage characterisation of NSEs within SONATA's broader prototype system.

Keywords: Non-structural elements, Image recognition, Deep learning, Damage detection

1- INTRODUCTION

Object damage detection has gained increasing attention due to recent advances in artificial intelligence (AI) technologies in civil engineering, particularly deep learning algorithms. Convolutional Neural Networks (CNNs) are the primary models used for object detection and segmentation, and have been successfully applied across many fields, including earthquake engineering. For example, Azimi et al. [1] provided a comprehensive review of deep learning-based damage detection and its integration with unmanned aerial systems (UASs) to enable autonomous assessments with minimal or no human intervention. This methodology is relevant for post-earthquake evaluation because it offers a viable solution for reducing inspection costs and overcoming difficulties in accessing hard-to-reach locations, where UASs play a crucial role in data acquisition. In parallel, advances in deep learning have also increased interest in computer vision applications within civil engineering, with algorithms adapted to address a wide range of structural assessment problems. Recent studies have demonstrated the use of computer vision for multi-class damage detection [2], corrosion detection [3], concrete crack identification [4], damage recognition of structural and non-structural components [5], among others. As a result, these approaches can complement or partially replace traditional inspection methods, greatly improving time efficiency and possibly the accuracy and objectivity in performance assessment procedures.

Quick post-earthquake evaluation is a critical approach for determining the level of damage sustained by buildings after a seismic event, enabling structures to be promptly tagged as safe for immediate reoccupation or restricted until repair. However, conventional post-earthquake assessments are influenced by several limiting factors, including accessibility to affected areas, the risk posed by entering damaged structures, the availability of qualified experts, and the inherent subjectivity of evaluators. Fukutomi et al. [6] highlighted how the business continuity and functional recovery of buildings were impacted by the 2011 Tohoku earthquake and emphasised the utility of structural health monitoring (SHM) systems for seismic performance evaluation. Japan's rapid post-event response system, Q-NAVI [7], uses sensor-based engineering demand parameters (EDPs) to rapidly estimate both structural and non-structural damage and assign safety tags (i.e., green, yellow, or red). While this approach significantly reduces assessment time, ambiguity remains in the interpretation of the yellow tag, which can represent either moderate but potentially life-threatening damage or minor reparable damage. This uncertainty underscores the importance of more accurate damage discrimination, especially for non-structural components, which often account for the largest proportion of economic losses and can suffer significant damage even under low levels of ground shaking [8].

Considering the importance of non-structural elements (NSEs) in seismic risk and economic loss assessment, this study focuses on common NSEs such as partitions (e.g., masonry, gypsum, glass), and suspended ceiling. This selection is motivated by three main reasons. First, the availability of images databases to train the AI models. Second, compatibility with a representative office-building scenario considered in this initial phase of the project. Third, the fact that office buildings concentrate a large share of economic activity, meaning that NSEs damage can lead to significant downtime costs and operational disruption, making mitigation strategies particularly relevant for stakeholders.

To address these challenges, engineers are increasingly leveraging AI-based methodologies due to their rapid response capabilities and reduced subjectivity in post-earthquake evaluations. Accordingly, this paper provides an overview of object detection techniques applied to NSEs, considering their critical role in economic losses and building downtime following seismic events. In addition, this work develops a two-stage vision-based pipeline (i.e., object detection and damage classification) tailored for NSEs within the SONATA project framework. CNN-based models are explored to develop algorithms capable of predicting damage states of non-structural components, thereby enhancing the accuracy and efficiency of current assessment practices. It is relevant to note that the model performance depends not only on the algorithm but also on the quality of the dataset used for training, which typically requires extensive and time-consuming annotation effort as well as a resource-intensive dataset construction process.

This work addresses the limited research on automated damage detection of NSEs in indoor post-earthquake environments and their integration into seismic risk assessment frameworks through fragility curve updating. Within this context, a two-stage vision-based pipeline is proposed, consisting of instance segmentation for NSE localisation and a subsequent damage classification stage. The first stage is implemented using a YOLO (You Only Look Once) based architecture [9] trained on a curated and manually annotated subset of the IDEA, Image Database for Earthquake damage Annotation [10], adapted to indoor non-structural components. The proposed framework is developed as part of the SONATA project, a multidisciplinary initiative aimed at establishing an integrated workflow for rapid post-earthquake damage assessment. The results demonstrate the feasibility of reliable NSE detection in complex indoor scenes, providing a foundational step toward fully automated element-level damage characterisation and seismic risk evaluation.

2- METHODOLOGY

This study proposes a two-stage computer vision framework for the automated damage recognition of NSEs in indoor post-earthquake environments. The system is designed as a sequential pipeline that processes raw visual data into element-level damage information to support rapid and consistent post-event assessment. The workflow consists of two main stages: 1) object detection and instance segmentation of NSEs; and 2) damage classification at element level. In the first stage, a YOLO based instance segmentation model [11] is employed to detect and localise NSEs such as partitions and suspended ceilings in indoor scenes, producing pixel-level masks and corresponding bounding boxes for individual components. In the second stage, the detected regions are extracted and analysed individually using a convolutional neural network (CNN) specifically designed for damage classification. Although this stage is not implemented in the present study, it is defined as a conceptual extension within the SONATA project.

This modular architecture enables decomposition of complex indoor scenes into structured element-level representations, facilitating scalable and automated post-earthquake damage assessment. A schematic representation of the workflow is presented in Figure 1.

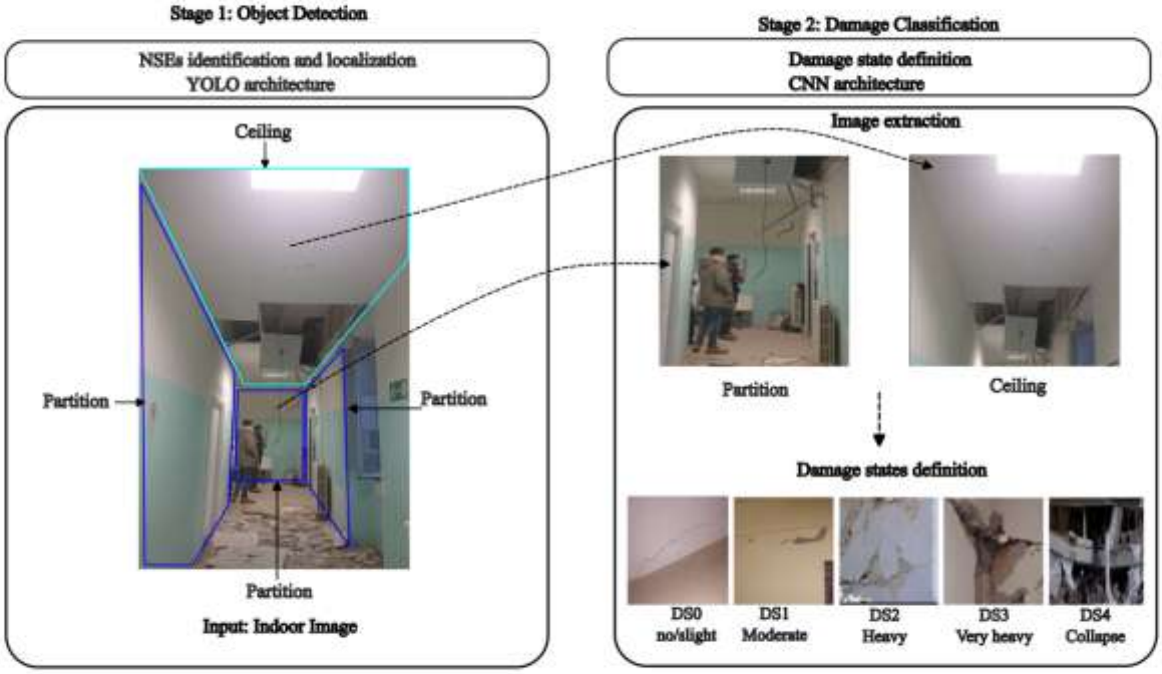

Figure 1: Proposed two-stage vision-based pipeline for automated NSE damage assessment: Stage 1 performs instance segmentation and localisation of NSEs (partitions and suspended ceilings); Stage 2 classifies the extracted regions into discrete damage states ranging from DS0 (no damage) to DS4 (collapse) (adapted from [2],[10]).

2-1- IMAGE DATASET DESCRIPTION

Deep learning-based damage detection models require large and well-annotated image datasets to achieve robust performance. However, one of the major limitations of datasets available in the literature is that they are typically developed for specific objectives, reflecting the outcomes targeted by their creators. Consequently, these datasets may not be directly applicable to alternative applications such as non-structural damage assessment. Therefore, in the context of the SONATA project, where such datasets remain limited, it is necessary to adapt existing databases by reorienting them toward NSEs for the present application and to populate them with new images. This strategy aims to enable the accurate identification of the damage state of an element following a seismic event without the need for expert intervention, thereby supporting the updating of fragility curves and the assessment of overall building risk [12].

Accordingly, in this study, the image dataset was derived from IDEA, originally developed to support image-based damage detection for post-earthquake structural assessment [10]. The complete IDEA dataset contains approximately 5,324 high-quality RGB images collected from post-earthquake inspections in Italy, including events such as L'Aquila (2009), Emilia (2012), and Central Italy (2016–2017), as well as periodic inspection activities. The images are annotated according to the Pascal VOC standard [13], enabling their use in object detection and segmentation tasks. Since the original annotations primarily target structural components, NSEs are not explicitly emphasised. Nevertheless, a subset of images containing relevant indoor environments was identified and extracted for the purposes of this study. Specifically, 288 images were selected, highlighting the limited availability of suitable data for non-structural damage assessment in the current literature. These images were manually annotated using the Computer Vision Annotation Tool [14], focusing on key NSEs such as partitions and suspended ceilings. The resulting annotations were formatted to support instance segmentation and used as input for the first stage of the proposed framework.

However, it must be noted that this dataset is not finalised and is still under development, with additions from multiple sources from the time of writing to the time of presentation. In particular, for the second stage, which focuses on damage state classification, the dataset will be expanded. These images will be obtained from experimental testing of NSEs at the EUCENTRE 9DLAB [15] within the SONATA project, as well as from supplementary images collected from the literature and web-based sources. Additionally, the classification task is defined in terms of discrete damage states, including DS0 (no or slight damage), DS1 (moderate damage), DS2 (heavy damage), DS3 (very heavy damage), and DS4 (collapse). These labelled data are intended to support the training of a CNN for element-level damage assessment and to contribute to the updating of fragility curves for seismic risk assessment.

2-2- STAGE 1: OBJECT DETECTION

Object detection is a fundamental component of the SONATA project, as it enables the localisation of NSEs in real post-earthquake scenarios, where these elements are typically found indoors. This capability is essential for improving data acquisition during inspections, allowing images or videos to be processed efficiently for rapid damage assessment. Several convolutional neural network architectures have been developed for object detection, including R-CNN, Faster R-CNN, RetinaNet, and YOLO. Among these, the YOLO architecture is selected due to its computational efficiency and real-time detection capability. Additionally, its availability as a pre-trained model makes it particularly suitable for applications with limited datasets.

In this work, YOLO26 [16], a recent extension of the YOLO family, is employed as a preliminary implementation to validate the feasibility of the proposed framework. At this stage, the objective is not to identify the optimal detection model, but to demonstrate that the pipeline is capable of localising NSEs in indoor post-earthquake scenarios. Previous post-disaster studies have successfully applied YOLO-based approaches for damage identification, reporting accurate results in detecting structural damage such as cracks or spalling in beams, columns, and walls, as well as in classifying building damage states (e.g., collapse and non-collapse). These findings support the use of YOLO as a suitable baseline model for the object detection module. The model was trained using a transfer learning strategy based on pre-trained weights and dataset described previously, using 80/20 training-validation split. Before the training, the input images were resized to 640 × 640 pixels and to mitigate the limited dataset size, data augmentation techniques were applied to increase dataset variability and improve generalisation. This included photometric transformations in the HSV colour space (i.e., hue, saturation, and value adjustments) and mild geometric transformations such as small rotations, translations, and scaling. The exact augmentation configuration used during training was as follows: hsv_h=0.015, hsv_s=0.7, hsv_v=0.4, degrees=5, translate=0.1, scale=0.3, shear=1.0, fliplr=0.5, flipud=0.0, mosaic=0.8, mixup=0.1, copy_paste=0.1, with mosaic disabled after 10 epochs to improve stability. These augmentations were selected to preserve segmentation consistency while increasing data variability in indoor environments. Further details on the definition and implementation of these hyperparameters are available in the Ultralytics documentation [17].

2-3- STAGE 2: DAMAGE CLASSIFICATION

Once the NSEs are localised using bounding boxes generated by YOLO, the corresponding regions are cropped from the original images to form individual samples for further analysis. This step is necessary because multiple NSEs coexist within a single scenario, and earthquake events often cause significant disorganisation of these components. Therefore, processing them as individual images is more appropriate, since each element may exhibit different damage types and damage states. Thus, isolating each element reduces background noise and enables a more focused analysis of its visual characteristics. After extraction, the resulting images are intended to be processed using CNN-based classification models, such as ResNet [18], AlexNet [19], VGG [20], Inception [21] and other deep learning architectures. The objective is to identify the most suitable model in terms of accuracy, computational cost, and efficiency, capable of correctly classifying different damage states. This classification framework is intended to support the update of fragility curves and improve post-earthquake damage assessment procedures.

The proposed approach aims to facilitate post-earthquake inspections by enabling rapid and consistent analysis of indoor images. Further development of this stage will rely on the availability of a dedicated dataset with consistent damage state definitions for each non-structural element. Such a dataset is currently under development within the SONATA project and will integrate experimental data from shaking table tests together with additional sources from the literature and web-based repositories.

3- PRELIMINARY RESULTS

The proposed YOLO-based instance segmentation model was evaluated for the detection and instance segmentation of NSEs elements in indoor post-earthquake environments. The model was trained for 80 epochs, and its performance was assessed using standard object detection metrics, including precision, recall, and mean

Average Precision (mAP) at different Intersection over Union (IoU) thresholds. The overall training behaviour is illustrated in Figure 2, which shows the evolution of training and validation losses. A consistent decreasing trend is observed across all loss components, including bounding box regression, classification and segmentation losses. The most significant improvements occur during the first 20-50 epochs, where the model rapidly learns discriminative features from the dataset. Beyond this phase, the curves progressively stabilise, indicating convergence. Furthermore, the close agreement between training and validation losses suggests good generalisation capability, with no significant evidence of overfitting.

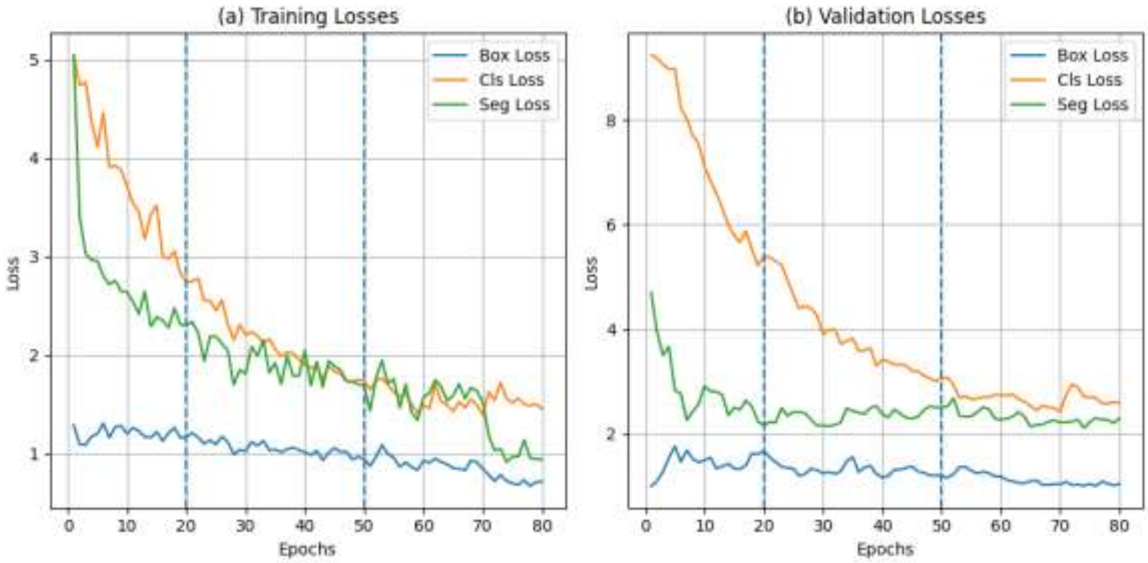

Figure 2: Training and validation loss curves for the YOLO-based instance segmentation model.

The evaluation metrics are presented in Figure 3, which summarises precision, recall, and mAP metrics for both detection and segmentation tasks. For the detection branch, precision consistently remains high, stabilising around 0.86, indicating a low false-positive rate. In contrast, recall stabilises at approximately 0.38, suggesting that a portion of the objects is not detected, particularly in complex or partially occluded scenarios. This behaviour reflects a conservative detection strategy, in which the model prioritises prediction accuracy over completeness. The quantitative results are summarised in Table 1. The model achieves a mAP@0.5 of approximately 0.44 and a mAP@0.5:0.95 of 0.31, indicating reliable detection under moderate IoU thresholds, while highlighting limitations in precise localisation under stricter evaluation criteria.

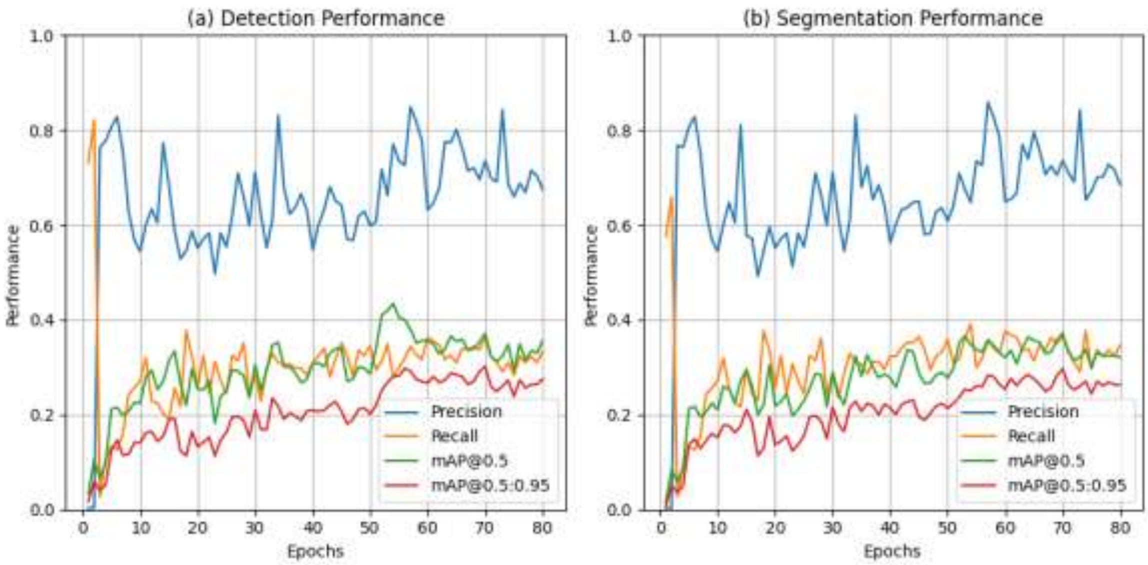

Figure 3: Performance metrics of the proposed model for detection and instance segmentation tasks, including precision, recall, and mAP.

A similar trend is observed in the instance segmentation branch (Figure 3). Precision is slightly higher, reaching approximately 0.87, while recall remains comparable at around 0.39. This indicates that segmentation benefits from spatial refinement but still inherits limitations from the detection stage. In terms of performance, segmentation achieves a mAP@0.5 of approximately 0.38 and a mAP@0.5:0.95 of 0.30. Overall, these results demonstrate consistent behaviour between detection and segmentation tasks, with slightly increased variability due to the higher complexity of pixel-level prediction. A summary of the result of both tasks is provided in the Table 1, highlighting their overall performance consistency.

Table 1: Evaluation metrics for object detection and instance segmentation models

Task	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Detection (B)	0.86	0.38	0.44	0.31
Segmentation (M)	0.87	0.39	0.38	0.30

A qualitative example of the model predictions is presented in Figure 4, where both bounding box detections and instance segmentation masks are visualised for a representative indoor scene. The results demonstrate the ability of the model to accurately localise and delineate NSEs under realistic conditions. Confidence scores associated with each detection provides additional insight into prediction reliability. High-confidence detections are generally associated with clearly visible and well-defined objects, whereas lower-confidence predictions are observed in cases involving partial occlusion, small object size, or complex backgrounds. These observations are consistent with the quantitative metrics reported earlier, particularly the high precision and moderate recall.

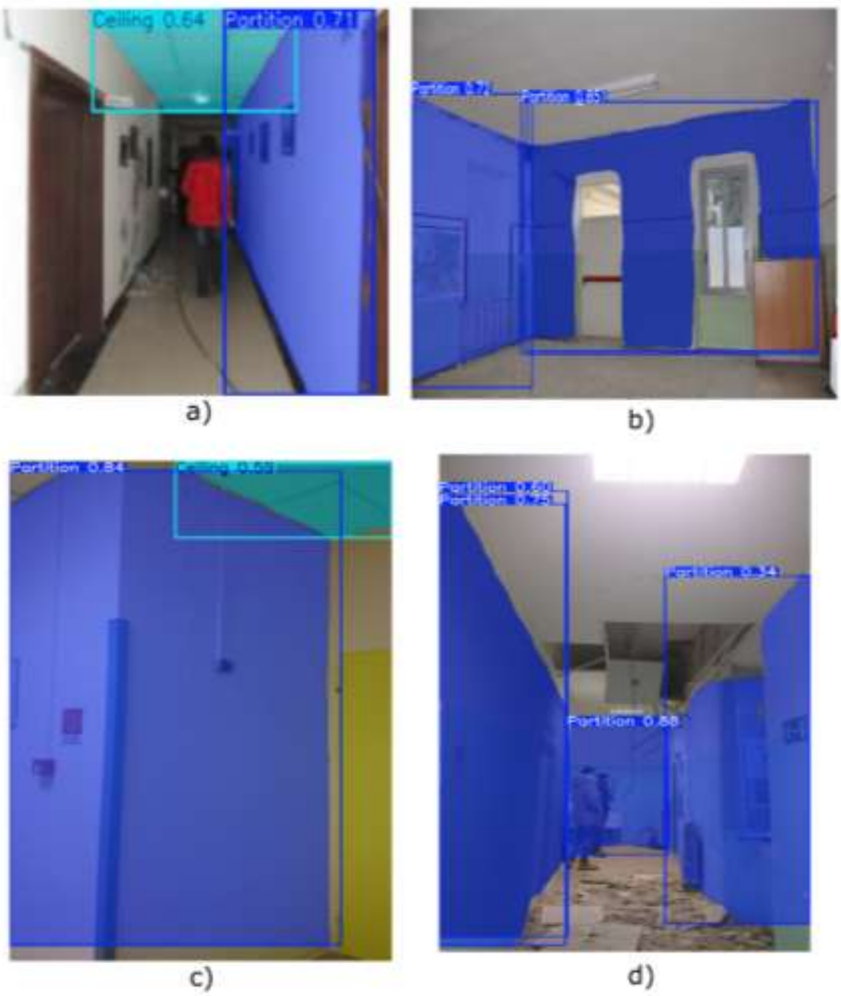

Figure 4: Qualitative results of the YOLO-based model showing bounding box detections and instance segmentation masks in a representative indoor scene.

Despite the limited size of the dataset, the proposed framework demonstrates the feasibility of applying YOLO-based models for the detection and segmentation of NSEs in indoor post-earthquake environments. At this stage, the system should be interpreted as a prototype implementation, aimed at validating the effectiveness of the first stage of the pipeline described previously in Section 2. The results confirm that the model is capable of reliably localising relevant elements, providing a solid foundation for further development.

4- CONCLUSIONS

This paper presented a preliminary framework for automated image-based detection of NSEs in indoor post-earthquake environments as part of the SONATA project. The proposed approach follows a two-stage pipeline, consisting of instance segmentation for element localisation and a subsequent damage classification stage defined as a conceptual extension. The results of the first stage demonstrate that a YOLO-based model is capable of reliably detecting and segmenting NSEs under realistic and complex indoor conditions. Although recall remains limited, the model achieves high precision and stable convergence, confirming the feasibility of the proposed detection strategy as an initial implementation.

It is important to emphasise that this study represents an early-stage showcase of the workflow rather than a fully optimised solution. The current dataset is limited in size, and future work will focus on expanding dataset to improve generalisation performance; in particular, to enhance recall performance in complex scenarios and detection completeness. A comprehensive comparison with alternative object detection and segmentation models, including both CNN-based (e.g., R-CNN, Faster R-CNN, RetinaNet) and transformer-based architectures (e.g., SAM 3.1, Mask2Former), will be conducted to evaluate their performance in terms of accuracy and computational cost, and to identify the most suitable approach given the available data and application constraints. The second stage, currently under development within the SONATA project, will extend this framework to element-level damage classification. This will enable a complete automated pipeline for post-earthquake assessment of NSEs, contributing to faster and more consistent seismic risk evaluation.

5- ACKNOWLEDGMENT

The work presented here has been developed within the framework of the “Sensor-driven statistical tools to evaluate risks and manage safety in the built environment” (SONATA) project, which is a Starting Grant awarded by the Italian Science Fund (Fondo Italiano per la Scienza) at IUSS Pavia under grant number FIS2023-03215.

6- REFERENCES

[1] Azimi, M.; Eslamlou, A.D.; Pekcan, G.; Data-driven structural health monitoring and damage detection through deep learning: State-ofthe- art review, Sensors (Switzerland), Volume 20, 2020, ISSN 14248220, https://doi.org/10.3390/s20102778

[2] Bhatta, S.; Dang, J.; Multiclass seismic damage detection of buildings using quantum convolutional neural network, Computer-Aided Civil and Infrastructure Engineering, Volume 39, 2024, 406–423, ISSN 14678667, https://doi.org/10.1111/mice.13084

[3] Nash, W.; Zheng, L.; Birbilis, N.; Deep learning corrosion detection with confidence, npj Materials Degradation, Volume 6, 2022, 1–13, ISSN 23972106, https://doi.org/10.1038/s41529-022-00232-6

[4] Dais, D.; Bal, İ.E.; Smyrou, E.; Sarhosis, V.; Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning, Automation in Construction, Volume 125, 2021, ISSN 09265805, https://doi.org/10.1016/j.autcon.2021.103606

[5] Wen, W.; Xu, T.; Hu, J.; Ji, D.; Yue, Y.; Zhai, C.; Seismic damage recognition of structural and non-structural components based on convolutional neural networks, Journal of Building Engineering, Volume 102, 2025, ISSN 23527102, https://doi.org/10.1016/j.jobe.2025.112012

[6] Fukutomi, Y.; O'Reilly, G.; Nakashima, M.; Kanda, K.; Kusaka, A.; Ogasawara, S.; Three benefits of “Caution (Yellow Tag)” in SHM-driven condition assessment of buildings: Eight years experience with market-based SHM, Earthquake Spectra, Volume 41, 2025, 979–994, ISSN 19448201, https://doi.org/10.1177/87552930241296377

[7] Kanda, K.; Nakashima, M.; Suzuki, Y.; Ogasawara, S.; “q-NAVI”: A case of market-based implementation of structural health monitoring in Japan, Earthquake Spectra, Volume 37, 2021, 160–179, ISSN 19448201, https://doi.org/10.1177/8755293020935884

[8] Perrone, D.; Calvi, P.M.; Nascimbene, R.; Fischer, E.C.; Magliulo, G.; Seismic performance of non-structural elements during the 2016 Central Italy earthquake, Bulletin of Earthquake Engineering, Volume 17, 2019, 5655–5677, ISSN 15731456, https://doi.org/10.1007/s10518-018-0361-5

[9] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A.; You Only Look Once: Unified, Real-Time Object Detection, 2016

[10] Senaldi, I.; Casarotti, C.; Mandirola, M.; Cantoni, A.; IDEA: Image database for earthquake damage annotation, Data in Brief, Volume 61, 2025, ISSN 23523409, https://doi.org/10.1016/j.dib.2025.111733

[11] Sapkota, R.; Cheppally, R.H.; Sharda, A.; Karkee, M.; YOLO26: Key Architectural Enhancements and Performance Benchmarking for Real-Time Object Detection, 2026

[12] Dondi, P.; Gullotti, A.; Inchingolo, M.; Senaldi, I.; Casarotti, C.; Lombardi, L.; Piastra, M.; Post-earthquake structural damage detection with tunable semi-synthetic image generation, Engineering Applications of Artificial Intelligence, Volume 147, 2025, ISSN 09521976, https://doi.org/10.1016/j.engappai.2025.110302

[13] Everingham, M.; Eslami, S.M.A.; Gool, L. Van; Williams, C.K.I.; Winn, J.; Zisserman, A.; The Pascal Visual Object Classes Challenge: A Retrospective, International Journal of Computer Vision, Volume 111, 2015, 98–136, ISSN 15731405, https://doi.org/10.1007/s11263-014-0733-5

[14] CVAT.ai Corporation; Computer Vision Annotation Tool (CVAT), github.com, 2026

[15] EUCENTRE; 9D System | 2 shake-tables

[16] Hidayatullah, P.; Tubagus, R.; YOLO26: A Comprehensive Architecture Overview and Key Improvements A PREPRINT, 2026

[17] Ultralytics; Model Training with Ultralytics YOLO, Model Training with Ultralytics YOLO, 2026

[18] He, K.; Zhang, X.; Ren, S.; Sun, J.; Deep Residual Learning for Image Recognition, 2015

[19] Krizhevsky, A.; Sutskever, I.; Hinton, G.E.; ImageNet Classification with Deep Convolutional Neural Networks

[20] Simonyan, K.; Zisserman, A.; Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015

[21] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z.; Rethinking the Inception Architecture for Computer Vision, 2015