Informatics and Knowledge Management Graduate Program
Rua Vergueiro, 235/249, Liberdade Zip Code: 01504-001
São Paulo/SP, Brazil
Tel.: +55(11) 3385-9078

São Paulo, October 6, 2025.

Dear Organizers of **CILAMCE2025**,

We would like to express our sincere gratitude to you and the reviewers for the valuable feedback and insightful comments on our manuscript titled ***"AUTOMATIC INSPECTION OF BUILDING CRACKS FROM AERIAL IMAGES ACQUIRED BY DRONES"*** **– ID 611**.

We did our utmost to respond to the reviewers' feedback and improve the overall quality of the work. Please, find below our detailed responses to the comments of the reviewers.

Yours sincerely,



Sidnei Alves de Araújo, Ph.D.
Associate Professor at Informatics and Knowledge Management Graduate Program, UNINOVE,
Corresponding author, e-mail: saraujo@uni9.pro.br

Responses to the reviewers' comments

The manuscript adresses a relevant theme in vision-based structural health monitoring, presenting a method based on the pre-trained YOLO model for crack detection and measurement in civil constructions. The paper is clearly structured and well organized, however there are some aspects that could be further improved or clarified.

Authors: *Thank you very much for the constructive comments and careful evaluation of our manuscript*

0. Keywords: The authors used the acronym "RPAS". However, acronyms should be avoided as keywords, especially if they are not widely known.

Authors: *Thank you very much for your observation, we have changed the keyword "RPAS" to "drone".*

1. Introduction

1.1 It should be interesting adding a paragraph talking about structural health monitoring, their types (vibration and vision based) and why opt for the vision based approach.

Authors: *To address this question, we rewrote the begin of introduction Section with the following paragraph:*

Structural inspection of buildings and infrastructure is essential to ensure the safety, durability, and performance of civil constructions [1]. However, common approaches for anomaly inspection generally involve manual processes, require physical access to critical areas, and rely heavily on the inspector's experience [2]. In this context, Structural Health Monitoring (SHM) has emerged as an area dedicated to the continuous or periodic assessment of the integrity of buildings and infrastructures, aiming at the early detection of damage. Among the existing approaches, vibration-based methods and computer vision-based methods stand out. Vibration-based SHM relies on the analysis of dynamic parameters, such as natural frequencies and vibration modes, which undergo changes when cracks or loss of structural stiffness are present. Although this technique enables a global assessment and is already well consolidated, it has limitations regarding the precise localization of damage and its sensitivity to environmental factors. In contrast, methods based on computer vision (CV) have gained prominence by enabling the direct and localized identification of anomalies such as cracks, in addition to allowing more precise measurement of their dimensions. These characteristics have motivated the development of several CV-based methods, including the approach presented in this study.

1.2 Use of vague expressions such as good results. The authors should opt to a more suitable term, for instance, promising results, or describe it. Terms such as "traditional methods" and "classical approaches" also should be more detailed: what are the traditional methods / classical approaches?

Authors: *Thank you for pointing this out. We have revised the text to either remove or clarify these vague expressions.*

1.3 The authors should clarify the GAP found as well as the main contributions of this paper. How do this paper differ from previous works? As I understand it, the GAP and contribution consist of the issue of integrating damage detection and its quantification. But would not quantification already be a way of identifying damage?

Authors: *We rewrote the last paragraph of the introduction to clarify the GAP. The revised version now states: "Despite recent advances in computer vision-based crack detection methods, there remains a gap in the measurement of detected cracks, a task for which machine learning and deep learning methods alone are insufficient. In this context, this work contributes to the literature by proposing a computational method that integrates crack detection and length measurement, using images captured by drones and combining ML and CV techniques such as the YOLOv8 CNN, thresholding, segmentation, and feature extraction. This approach represents a step forward toward more efficient strategies for structural monitoring and preventive maintenance".*

2. Theoretical Background: The authors gave too much attention to the crack's width, but they did not use this information further. I suggest including this approach as a way of working in the future, even if it means classifying cracks according to the Brazilian standards mentioned (NBR 15575-2:2013 and NBR 6118:2023).

Authors: *We appreciate the reviewer's observation. Our initial intention was only to clarify why we used the term crack to refer to crack, fissure, or split. However, we agree with your comment and have modified the text in Section 2.1 accordingly to address this point.*

3. Material and Methods

3.1 The authors descibed both datasets used (IM_1 and IM_2) but there is a lack of information about them. It should be interesting to include more detail about them, for example, how many samples are available in each class (healthy and damaged)?

Authors: *The image datasets used for training contain cracks in all samples, with each occurrence visually annotated (segmented). Thus, the model was specifically trained to identify and segment the locations of cracks in the images, without performing a classification training between different classes (such as healthy and damaged). We have completed Section 3.1 with this information.*

3.2 In Stage 1: Automatic crack detection with YOLOv8, it ahould be interesting to include more information about this process and stablish a parallel with what is shown in Figure 2. For example, the proportion used in Training/Validation/Testing splitting was not clear. Including more information about this process would be useful to facilitate the reproducibility of the work.

Authors: *Thank you for your observation, the text of Section 3.1 has been revised to address this concern. For instance, in the conducted experiments, each dataset was divided into training, validation and testing sets in proportions of 70%, 20% and 10%, following a scheme commonly adopted in the literature.*

3.3 How many times the computatinal experiments were independently run? It should be mentioned.

Authors: *Throughout the computational experiments, several training sessions were performed with the aim of optimizing the hyperparameters of each model. Table 1 presents the best results obtained, as well as the associated hyperparameter configurations. The IM_1 dataset, which contains a smaller number of images (87), was subjected to a larger number of experiments. Although it did not yield good results, the tests conducted with IM_1 contributed to a better configuration of the YOLO model, which was later applied to the IM_2 dataset. This information has been included at the beginning of Section 4 (Results).*

4. Results

4.1 There is a lack of results assessment. Only IM_2 was evaluated. Although IM_1 struggled to achieve high performances, some interesting discussion can be rise, for example, why did YOLOv8n-seg achieve better results than YOLOv8s-seg, once the authors previously described the latter model as being more capable of offering better accuracy?

Authors: *We thank you for the observations. In the revised version of the text, the Results Section has been improved with additional explanations to address the issues raised.*

4.2 The results presented in Table 1 are average results? How about the standard deviation?

Authors: *Table 1 presents the hyperparameters and the best performance achieved by each of the YOLO models trained. This has been clarified in the revised version of the text.*

4.3 How the confidence value was measured?

Authors: *The confidence score is provided by the YOLO framework for each bounding box indicating a detection. It is computed as the product of the probability that an object exists in a region and the probability that it belongs to a specific class, representing how certain the model is about the detection. We have included this information at the end of Section 3.2 and added a reference for further details.*

5. Conclusions

5.1 Both in the conclusion and in the summary of the articles, the results presented represent only one of the scenarios evaluated by the authors (IM_2), without this condition being specified.

Authors: *We hope that the changes made in the Results and Conclusions sections were sufficient to clarify these issues.*

5.2 I also suggest including a brief discussion of the results obtained in IM_1, because when you state that the results achieved were satisfactory, it could be understood that the 34% achieved in IM_1 was also satisfactory. Was that the case?

Authors: *We hope that the changes made in the Results and Conclusions sections were sufficient to clarify these issues.*