

# Intelligent Image-Based Detection of Surface Anomalies in Aircraft Structures

*Oumaima Mesbahi*<sup>1,2\*</sup>, *Souhila Chabane*<sup>1,3</sup>, *Nuno Pereira Santos*<sup>1,2</sup>, *Adriano Del Pino Lino*<sup>3</sup>, *Mouhaydine Tlemçani*<sup>1,2</sup> and *José Manuel Mota Lourenço da Saúde*<sup>3</sup>

<sup>1</sup>Instrumentation and Control Laboratory- Center for Sci-Tech Research in Earth System and Energy – CREATE, Universidade de Évora, Portugal.

<sup>2</sup>Cátedra CEiiA de Ciência e Tecnologia Aeroespacial, Universidade de Évora, Portugal.

<sup>3</sup>CTI Aeroespacial, Cátedra CEiiA de Ciência e Tecnologia Aeroespacial, Évora, Portugal.

**Abstract.** In the aviation industry, the structural integrity of aircraft is paramount to ensure safety and reliability. The process of regular inspections and maintenance is integral to identifying abnormalities on the surface of the aircraft, which could cause catastrophic failures if left unchecked. Usually, the traditional process of inspections involves the visual assessment of the aircraft surface by a human, who may take a long time, is prone to inaccuracies and requires abundant knowledge. This work presents a discussion on the use of the YOLO image processing technique for the automated detection of abnormalities that exist on the surface of aircraft, including cracks, corrosion, dents, and delamination. The use of image processing for the detection of aircraft surface abnormalities stands to revolutionise the process of maintenance associated with aircraft. A second stage was added to the process, which focuses on including a human in the loop verification, where the inspector will be able to evaluate and validate the results. The technology possesses the potential for ease of scalability for the assessment of numerous aircraft in the fleet. The process leads to beneficial outcomes, including safety, cost reduction, and improved reliability.

## 1 Introduction

Maintaining aircraft structural integrity remains a core focus in the aviation industry, leading aircraft operators, whether private or military, to undertake stringent inspection procedures to fulfil all safety requirements and regulations. However, over time, aircraft external surfaces tend to have defects such as cracking, corrosion, abrasion, and paint degradation. International regulatory bodies such as ISO, EASA, and the FAA [1,2] have established strict guidelines prescribing just how such inspections should be carried out and documented to ensure consistent, traceable, and repeatable evaluation procedures.

Despite these regulations, conventional inspection practices rely predominantly on manual visual inspection by trained personnel. Despite its importance, this process, although time-consuming, calls for sustained attention to detail, particularly when dealing with large aircraft and areas of high defect density. Accordingly, this process could be subject to fatigue and several other training and environmental conditions. Increased availability of high-resolution imagery and recent advances in computer vision offer a path to augment inspectors' capabilities with automated tools that can speed detection while enhancing consistency, without necessarily eliminating expert oversight.

Recently, studies on different deep learning techniques for anomaly detection on the exterior of the

aircraft were conducted [3]. Various studies utilised convolutional neural networks, generative models, object detection models, etc., for defect detection on aircraft surfaces, and amongst these, YOLO methods were found to be promising due to a good balance between speed and accuracy [4-8]. The family of YOLO variants has been evaluated across a range of defect classes, and some newer variants exhibit higher performance on small, irregular, or low-contrast features, which are typical of aircraft-surface defects [5,6,9-11].

In this work, we evaluate the performance of a YOLO-based defect detection system trained on a custom dataset of annotated aircraft exterior images. The contribution presented here focuses on the major structural sections of an aircraft and evaluates YOLO's performance in defect detection under varying visual conditions. We therefore integrate a dedicated human-in-the-loop validation interface into our proposed approach, enabling the certified inspector to review and refine automated predictions. This integrated framework aims to improve the speed and reliability of surface inspections while expert judgment remains central to decision-making.

## 2 Aircraft surface anomalies detection methodology

\* Corresponding author: [omesbahi@uevora.pt](mailto:omesbahi@uevora.pt)

## 2.1 Dataset preparation

The first step of the proposed methodology was to create a high-quality dataset specifically for aircraft surface defect detection. A custom dataset was collected comprising approximately 7,000 images obtained during routine aircraft inspections. These included all types of surface anomalies, such as corrosion, erosion, cracks, dents, scratches, fretting damage, and missing paint. All images were annotated in Roboflow using bounding boxes and defect class labels, leveraging the expertise of certified aircraft inspectors to ensure annotations were accurate and domain-relevant [12].

This involved preprocessing steps of normalising the contrast and resizing images in a standard way to increase defect visibility and maintain uniformity across different inspection conditions. Subsequently, data augmentation was performed using the following transformations: horizontal and vertical flipping, random brightness and contrast adjustments, and a slight rotational offset. The goal of such transformations is to simulate real changes in natural lighting conditions, orientation, and distance from the point of maintenance inspection. The dataset was split into 83% training, 8% testing, and 9% validation sets. For each split, the same class distribution was maintained.

## 2.2 YOLO method for detection

After preparing the dataset, the detection architecture was selected. In this work, the You Only Look Once (YOLO) family of object detectors was adopted because it enables real-time inference while maintaining high detection accuracy, including under conditions with significant object-size variation. YOLO is a one-stage detector that treats object detection as a single regression problem, predicting bounding box coordinates and class probabilities directly from the image in a single forward pass.

YOLO was first proposed by Redmon et al. [13], who introduced a unified detection framework that outperforms traditional region-proposal methods in speed while achieving very competitive accuracy.

Unlike two-stage detectors such as Faster R-CNN [14], YOLO is a one-shot detector, enabling near-real-time inference. Where two-stage models have generally achieved much better performance on very small or heavily occluded objects, modern YOLO architectures have closed the gap by introducing multi-scale detection capabilities and attention-enhanced feature refinement modules. In the context of aircraft inspection, this ensures the detection of diverse defect types without compromising inference speed, an essential property for possible use in UAV-based or automated inspection pipelines.

The YOLO architecture used in this work had three major components:

### 1. Backbone:

In this work, the proposed convolutional network, based on CSP(Cross Stage Partial), was used for extracting the spatial and semantic features of input images. CSP modules alleviate computational overhead but preserve high representational capacity by partitioning the

gradient flow paths and improving the diversity in features.

### 2. Neck:

A PANet/FPN hybrid neck was added to combine low-level spatial detail with high-level semantic information. This multi-scale feature fusion is critical to reliably detecting small cracks or corrosion clusters that could be overshadowed by the large details of the structure.

### 3. Head:

The detection head of the network is designed to yield bounding boxes, confidence, and class probabilities, considering multiple scales. An anchor-free/adaptive anchor technique is incorporated into the updated versions of YOLO to enable the flexible localisation of irregular defect shapes, which are commonly encountered on aircraft skins.

## 2.3 Implementation and Model Configuration

### 2.3.1 Model Setup and Dataset Handling

The YOLOv8-s model was initialised with pre-trained weights on MS COCO, followed by fine-tuning for two target classes: Defect and Background. The configuration files were adapted to match the characteristics of the aircraft inspection dataset. Training was performed on a 24 GB NVIDIA RTX 3090 GPU with 24 GB of VRAM.

All the images are of size 640 x 640 pixels, which is the default size for training the YOLO model. Similarly, annotations of the same size are converted to the YOLO format, which contains bounding-box coordinates and class identifiers.

### 2.3.2 Anchor Boxes and Detection Strategy

In anchor-based variants, YOLO relies on pre-defined bounding box anchors across several scales. As a preprocessing step for training, the authors ran an anchor-clustering operation using k-means on the dataset to determine optimal anchor dimensions that best fit the shapes of defects on aircraft surfaces. This should ensure better localisation accuracy for elongated cracks, circular corrosion pits, or paint defects of arbitrary shapes.

For anchor-free variants in YOLOs, the regression for bounding boxes was based on dynamic point-based/centre-based prediction, making it less sensitive to the selection of the anchor configuration, especially when dealing with small geometries in defect structures.

### 2.3.3 Loss Functions and Optimisation Strategy

Training used the default YOLO loss function, which combines three ingredients:

**Bounding Box Loss:** Proposed based on Complete IoU, taking into account overlap, distance between centres, and consistency of aspect ratios.

**Objectness Loss:** a binary cross-entropy penalty which decides whether a predicted region contains a defect.

Classification loss: Also based on binary cross-entropy, optimised for distinguishing defect vs. background.

We used an SGD optimiser with a learning rate of 0.01, momentum 0.937, and weight decay 0.0005 to train the network utilising standard YOLO training heuristics. Training proceeded for numerous epochs with early stopping enabled in order to avoid overfitting.

### 2.3.4 Inference and Post-Processing

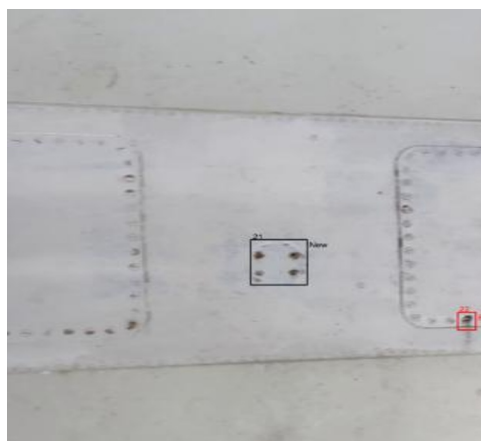
During inference, the model generated bounding boxes with associated confidence and class scores. These are suppressed using a confidence threshold of 0.5 for low-certainty detections. Non-maximum suppression is also integrated natively into YOLO to eliminate redundant bounding boxes using an IoU threshold of 0.45. This post-processing step is critical for detecting closely spaced surface anomalies, such as a cluster of corrosion or a network of cracks. By filtering overlapping predictions, the system produces a cleaner, more accurate representation of the actual surface defects.

## 2.4 Results of anomalies detection

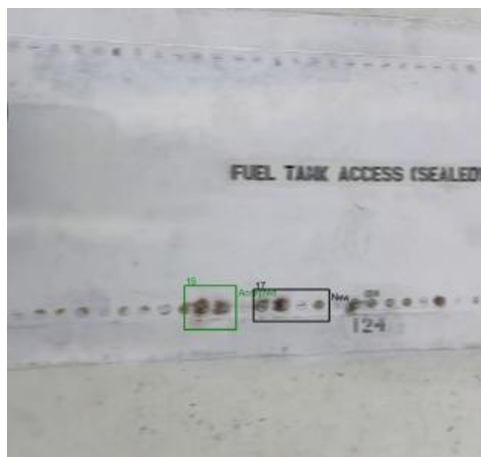
The precision of 0.5497 after training the model for 40 epochs indicates that the model correctly detected just over half of its total detections and was not overly plagued by false positives. On the other hand, with a recall value of 0.5842, the model was able to identify a bit more than half of all the actual defects present in the images. This shows the model's moderate ability to capture true defects, but it may need further improvement for detecting small or subtle anomalies. Finally, an F1-score of 0.5669, which is the balance of both precision and recall, reflects an overall moderate detection performance and points out the trade-off between missed defects and false alarms of non-defective regions. This interpretation is further corroborated by the model's accuracy of 0.55669, where more than half of all of its predictions, both for defect and background classes, were correct. Altogether, these results suggest that, even though the model has learned meaningful features of the defects, additional training data, improved annotation quality, or further hyperparameter tuning may be needed to achieve higher reliability when using this model in an operational inspection context.

Figures 1 to 4 show the detection results, indicating that the model can detect visible surface anomalies with high accuracy across different aircraft surface images. In all the images presented, the YOLO detector locates the regions of interest of the defects, showing stable bounding-box placement and consistent confidence scoring across changes in lighting conditions, textures, and background complexity. Qualitative examples such as these help demonstrate the model's ability to generalise to new inspection images, providing clear visual cues useful to inspectors during an assessment. Of course, some fine-scale defects may still require expert verification, but the overall outputs demonstrate a

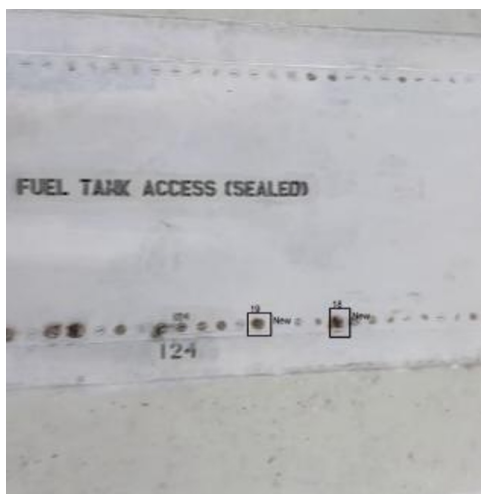
reliable first-pass detection capability suitable for integration into an inspection workflow.



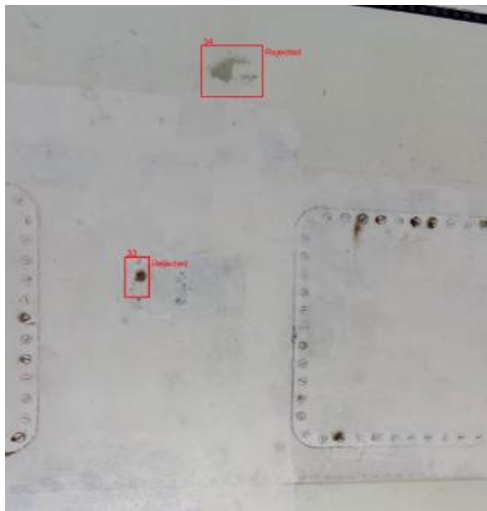
**Fig. 1.** Illustration 1 of the results of the trained YOLO model.



**Fig. 2.** Illustration 2 of the results of the trained YOLO model.



**Fig. 3.** Illustration 3 of the results of the trained YOLO model.



**Fig. 4.** Illustration 4 of the results of the trained YOLO model.

### 3 Human in the loop interface implementation

The Inspector Validation Interface was a custom Python and Tkinter implementation that incorporated a critical human-in-the-loop review stage in the automation process. In this process, certified inspectors can analyse, edit, and approve or reject the YOLO model's output. The inspectors will have the privilege to edit the bounding box, remove false readings, create a defect, and apply a set of standard defects by using a user-friendly editor. Validation data are saved in XML format, which is further enriched with metadata such as defect category, inspector notes, and validation status. Additional spatial context is provided by a built-in 3D aircraft visualisation panel, which allows users to view defects relative to specific aircraft zones.

This human-in-the-loop methodology greatly improves the system's reliability and regulatory compliance by ensuring that expert judgment remains the focus of any defect assessment. Verified annotations can be used for high-quality ground truth documentation and as continuous feedback to improve the model by retraining it. The risk of both false positives and missed anomalies decreases, while the overall accuracy, accountability, and reliability of the automated inspection system steadily increase.

### 4 Conclusion

In this work, a YOLO-based method for detecting defects on aircraft surfaces has been proposed, along with a dataset and validation process. Based on the results, the proposed model shows stable detection performance across different defect types. The inspection results show that YOLO can deliver a reliable first-pass assessment of aircraft surfaces. While performance metrics indicate room for improvement,

the system effectively identifies the most visible anomalies and supports inspection.

This work adds a human-in-the-loop validation interface to ensure that expert oversight remains central to the process, increasing trust, enabling model error correction, and enabling iterative dataset refinement. Together, the automated detector and validation tool provide a practical solution that enhances inspection efficiency while maintaining the safety and quality standards expected in aviation maintenance.

### Acknowledgment

The authors would like to thank Aero. Next, Portugal - ILAN VR (C645727867-00000066) for supporting this work.

### References

1. International Organization for Standardization, "ISO 23665:2021 — Artificial intelligence — Guidance on the use of AI in medical imaging," ISO, Geneva, Switzerland, 2021. [Online]. Available: <https://www.iso.org/standard/80124.html>
2. European Union Aviation Safety Agency (EASA), "EASA Airworthiness Directives (ADs)," [Online]. Available: <https://ad.easa.europa.eu>
3. Cumbajin, E., Rodrigues, N., Costa, P., Miragaia, R., Frazão, L., Costa, N., & Pereira, A. (2023). A systematic review on deep learning with CNNs applied to surface defect detection. *Journal of Imaging*, 9(10), 193.
4. Huang, B., Ding, Y., Liu, G., Tian, G., & Wang, S. (2024). ASD-YOLO: An aircraft surface defects detection method using deformable convolution and attention mechanism. *Measurement*, 238, 115300.
5. Li, X., Wang, C., Ju, H., & Li, Z. (2022). Surface defect detection model for aero-engine components based on improved YOLOv5. *Applied Sciences*, 12(14), 7235.
6. Qu, Y., Wang, C., Xiao, Y., Yu, J., Chen, X., & Kong, Y. (2023). Optimization algorithm for surface defect detection of aircraft engine components based on YOLOv5. *Applied Sciences*, 13(20), 11344.
7. Liao, K. C., Lau, J., & Hidayat, M. (2024). Aircraft skin damage visual testing system using lightweight devices with yolo: An automated real-time material evaluation system. *AI*, 5(4), 1793-1815.
8. Huang, J., Wei, M., Li, Z., Qu, H., Zhao, W., & Bai, X. (2025). J-DDL: Surface Damage Detection and Localization System for Fighter Aircraft. *arXiv preprint arXiv:2506.10505*.
9. Connolly, L., Garland, J., O'Gorman, D., & Tobin, E. F. (2024). Deep-Learning-Based Defect Detection for Light Aircraft With Unmanned Aircraft Systems. *IEEE Access*, 12, 83876-83886.

10. Avdelidis, N. P., Tsourdos, A., Lafiosca, P., Plaster, R., Plaster, A., & Droznika, M. (2022). Defects recognition algorithm development from visual UAV inspections. *Sensors*, 22(13), 4682.
11. Suvittawat, N., Kurniawan, C., Datephanyawat, J., Tay, J., Liu, Z., Soh, D. W., & Ribeiro, N. A. (2025). Advances in Aircraft Skin Defect Detection Using Computer Vision: A Survey and Comparison of YOLOv9 and RT-DETR Performance. *Aerospace*, 12(4), 356.
12. Chabane, S., Mesbahi, O., Santos, N. P., Tlemçani, M., & de Saude, J. L. (2025, October). Detection and Classification of Aircraft Structural Defects for Database Creation and Findings Identification. In 2025 13th International Conference on Systems and Control (ICSC) (pp. 144-149). IEEE.
13. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
14. Mesbahi, O., Chabane, S., Santos, N. P., Lino, A. D. P., Tlemçani, M., & Da Saúde, J. M. L. (2025, October). Automated Detection of Aircraft Surface Defects Using Deep Learning with Integrated Human Validation. In 2025 13th International Conference on Systems and Control (ICSC) (pp. 13-18). IEEE.