CEAS - AIDAA Conference 2025 Materials Research Forum LLC
Materials Research Proceedings 69 (2026) 37-41 https://doi.org/10.21741/9781644904251-7

Deep-Learning-based Dent Detection of Aircraft Surfaces using Synthetic Data

Aditi MHATRE[1,a *], Salvatore MEROLA[2,b], Ann-Kathrin KOSCHLIK[1,c], Rebecca RODECK[1,d] and Gerko WENDE[1,e]

[1]German Aerospace Center (DLR) – Institute of Maintenance, Repair, and Overhaul, Hamburg, Germany

[2]University of Naples Federico II – Dept. of Industrial Engineering, Naples, Italy

[a]aditi.mhatre@dlr.de, [b]salvatore.merola@unina.it, [c]ann-kathrin.koschlik@dlr.de, [d]rebecca.rodeck@dlr.de, [e]gerko.wende@dlr.de

Keywords: Aircraft Surface Inspection, Dent Detection, Deep Learning, Aircraft Maintenance, Visual Check, Generative AI

Abstract. Detecting and localizing dents on aircraft surfaces is crucial for maintaining their structural integrity. However, this task can be challenging for humans as dents are not very prominent to the naked eye and require the assistance of light reflections to reveal the damages across the surface. Latest state-of-the-art technologies such as lasers or cameras digitize this step, however the workload is shifted to identifying the dents in the virtual image. The integration of deep-learning methodologies can help automate dent detection. This study compares two object detection architectures: You Look Only Once (YOLOv11) and RealTime DETection TRansformer (RT-DETR) for dent detection. A high-quality dent dataset is prepared, consisting of real and synthetic images of common long- and mid-range aircraft fuselages, to train and test the models. The results indicate that YOLOv11 marginally outperforms RT-DETR in detecting dents with a mean accuracy precision (mAP50) score of 0.66 against the mAP50 value of 0.57 for RT-DETR.

Introduction

Dents are one of the most common damages found in aircraft surfaces that are caused by external forces and affect the structural integrity. These dents need to be classified based on their size, shape, and location to assess their severity and guide maintenance decisions [1,2]. This inevitable inspection process is performed manually by technicians, and requires a significant number of work hours. Unlike other types of defects like scratches or corrosion, dents are more difficult to detect visually as there is usually no prominent change in colour [3]. Inspectors must rely on close observation and additional lighting, which can lead to eye fatigue and missed detections. Introducing deep-learning methodologies offers a promising solution by automatically detecting dents on large aircraft surfaces with high accuracy, reducing human errors and inspection time. However, most existing studies focus on larger dents and are constrained by limited training datasets for dents, which hinders the models' ability to learn diverse dent features [3,4]. This study aims to identify the most suitable approach for detecting dents with the help of light reflections, serving as the Region of Interest (ROI) for further damage assessment and decision-making. This case study focuses on dent detection on aircraft structures using visual images. In addition to the accuracy of the models, it is required that the algorithms are fast and efficient, enabling them to operate in real-time. To address this, our study implements and compares the object detection architectures, YOLOv11[1] and the RT-DETR [5]. One of the key contributions of our work is the generation of an enriched dent dataset, consisting of both real images of dents and synthetic images generated with the help of generative AI models, improving the overall results

[1] YOLOv11: github.com/ultralytics/ultralytics

Methodology

The YOLOv11 is a versatile foundational model that is adept at object detection and instance segmentation, and suitable for real-time applications. It is available in different model sizes allowing it to cater to diverse computing environments. Similarly, RT-DETR is another state-of-the-art model in this domain, which is an end-to-end Transformer-based detector. These models are the current state-of-the-art in real-time object detection, which provide a good trade-off between accuracy and speed, making them suitable for aircraft inspection. To train the models, we prepared a high-quality dent dataset, which contains 1166 annotated images of dents across various aircraft parts including wing sections of B737 and A380. The dataset is prepared by combining the 614 real images and 552 synthetic images, which are generated with the help of a Diffusion model [6, 7], an image-to-image model. The real images capture artificial impacts created on a B737 wing section [2], and various dents caused due to hailstorms on a slat section of an A380. These images are captured under varying light conditions to replicate the diverse illumination found in aircraft visual inspections, including natural light and flashlights as well as photos from different angles. This helps capture the effect of specular light reflections across the surface to highlight dents on reflective materials and help recognize dents more accurately.

Various image acquisition methods are employed to collect the images, namely a smartphone camera, an Intel RealSense camera mounted on an inspection robot, and a digital single-lens reflex (DSLR) camera. To further augment the dataset for training, we generate synthetic images based on 276 out of 614 real images, with two synthetic images being generated for every real image, leading to a total of 552 synthetic images. More details can be found in [7]. The annotated labels on the dataset were “dent” and “no_dent”. The “no_dent” is introduced to label other types of defects such as corrosion or peeling paint to differentiate it from dents. Some sample images of the dataset are showcased in Table 1. The dataset is split into train and test datasets in a ratio of 80:20, with all the synthetic images being assigned to the training subset and only real images being used for the test subset. In total, the training dataset consisted of 933 images and the test dataset contained 233 images. To optimize training, we leveraged pre-trained YOLOv11 and RT-DETR models, which were initially trained on the popular COCO (Common Objects in Context)[2] dataset. The NVIDIA GeForce RTX 4070 SUPER GPU is utilized to finetune these pre-trained models for our specific task. The training process involved 200 epochs with a learning rate of 0.001.

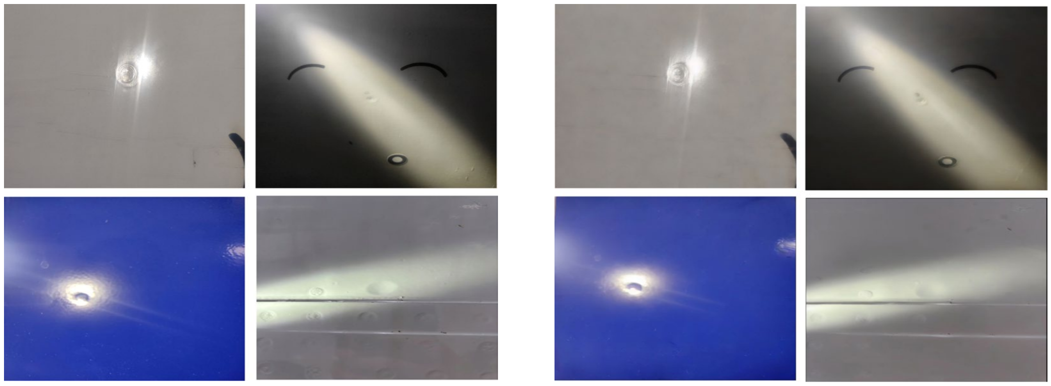
Results and Discussion

After training the models for 200 epochs on the training dataset, they are evaluated on the test dataset in two ways: quantitatively and qualitatively. For the quantitative method, the mean accuracy precision of the model predictions was compared with the ground truth labels to evaluate how accurately the model localized the dents. There are two variations of mAP used to evaluate the results: mAP50 and mAP50-95, which are both valued in the range of 0–1, where a higher value indicates a better model accuracy. The mAP50 is calculated at an intersection over union (IoU) threshold of 0.5. In the mAP50-95 metric, the model’s average performance is taken across IoU thresholds from 0.5 to 0.95 in steps of 0.05. Due to this difference in the calculation, the scores of mAP50 are generally higher than mAP50-95. The quantitative results for the models are summarized in Table 2. The YOLOv11 has a mAP50 score of 0.64 and mAP50-95 of 0.27. After training the RT-DETR with the same learning rate and epochs as the YOLOv11, it achieved a mAP50 score of 0.57 and a corresponding mAP50-95 value of 0.23. Compared to other existing studies [4], where the performance for dent identification was around 0.5, our approach achieved a higher score of 0.64, indicating an improvement in dent detection accuracy.

[2] https://cocodataset.org

Table 1: Overview of dent dataset showing real and synthetic images under different lighting.

(a) Real Images (left box) vs. Synthetic Images (right box).



(b) Effect of light reflections across dented surface in real images.

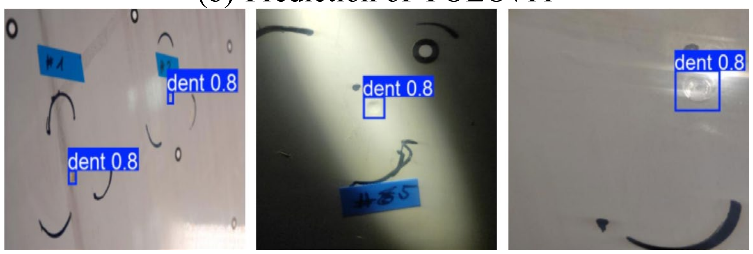
Table 2: Quantitative results for the accuracy precision of the models.

Model	mAP50	mAP50-95
YOLOv11	0.64	0.27
RT-DETR	0.57	0.23

The qualitative method involves observing the bounding box predictions of the models, and comparing them against the annotated labels of the test dataset. The models predict bounding boxes along with a prediction label and a corresponding confidence score for each detected dent. The confidence score is in the range of 0–1 to indicate the probability of the classification being correct. A comparison of the ground truth labels against the dent predictions of YOLOv11 and RT-DETR is shown in Fig.1.

(a) Ground Truth of Test Dataset

(b) Prediction of YOLOv11



(c) Prediction of RT-DETR

Figure. 1: Comparison of (a) ground truth labels against dent detection predictions of (b) YOLOv11 and (c) RT-DETR.

Based on the tests conducted, both YOLOv11 and RT-DETR demonstrated a high degree of accuracy in detecting dents across aircraft surfaces, with the YOLOv11 achieving marginally better results compared to RT-DETR. It was also observed that the introduction of synthetic images for training helped improve the accuracy of both the models as compared to when they were only trained on real images. As real images are used to generate more training images, the damages produced in the synthetic images are very similar to the dents found in a real scenario with pixel-level deviations. This introduces the required variation needed in a training dataset to enable better training of the models. The unique illumination modes employed during image capture enabled the models to comprehend the diverse ways dents can appear under different environmental conditions. This is observed in Table 1(b), where the left image reveals dents due to the gap in the line of light as it deviates around the dent, creating a stark contrast. The right image showcases shadows around the curve and partial illumination of the dent, resulting in a more nuanced representation. This approach helps develop a deeper understanding of how environmental factors influence dent appearance as well as recognize the variations in lighting conditions during inspections.

Conclusion

In this study, two deep-learning models, YOLOv11 and RT-DETR, are trained on a novel dent dataset to develop a dent detection system for aircraft damage assessment. A diverse dataset comprising real and synthetic images was used, captured under varying light conditions to account for the impact of specular light reflections on dent visibility. Comparative evaluation of YOLOv11 and RT-DETR models showed that YOLOv11 outperformed RT-DETR, albeit with a modest margin. The findings contribute to the existing literature on dent detection using deep-learning methods and highlight the effectiveness of synthetic images in datasets for more comprehensive understanding of lighting effects on surface features. Future work can build upon this work by expanding the dataset to include more synthetic images for training deep-learning architectures and exploring the use of light reflections to reveal damages.

References

[1] R. Arora, V. Bhatia and S. Sidharth. Military Aircraft Surface Defect Detection Using Machine Learning (2024). International Conference on Computing, Sciences and Communications (ICCSC), pp. 1-6. https://doi.org/10.1109/ICCSC62048.2024.10830372

[2] A.K. Koschlik, M. J. Scott, T. Keser, F. Rauscher, H. Meyer, W. Verhagen, P. Marzocca, F. Raddatz, and G. Wende. Mixed reality for aircraft structural dent and buckle non-destructive evaluation (2024). In 34th Congress of the International Council of the Aeronautical Sciences, ICAS 2024. Congress of the International Council of the Aeronautical Sciences.

[3] A. Plastropoulos, K. Bardis, G. Yazigi, N.P. Avdelidis, and M. Droznika. Aircraft Skin Machine Learning-Based Defect Detection and Size Estimation in Visual Inspections. Technologies 12, no. 9 (2024): 158. https://doi.org/10.3390/technologies12090158.

[4] P. Shailaja and S. Padmanabhan. A Survey on Autonomous Damage Detection on Aircraft Surfaces using Deep Learning Models (2022). 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India. pp. 1135-1140, doi:10.1109/ICCMC53470.2022.9753755.

[5] Y. Zhao, W. Lv, S.L. Xu, J.M. Wei, G.Z. Wang, Q.Q. Dang, Y. Liu, and J. Chen. DETRs Beat YOLOs on Real-time Object Detection (2024). 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16965-16974. https://doi.org/10.1109/CVPR52733.2024.01605.

[6] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. 21st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10674-10685. https://doi.org/10.1109/CVPR52688.2022.01042.

[7] S. Merola, A. Mhatre, A.K. Koschlik, M. Guida, and F. Marulo. Comparative Analysis of Generative Data Augmentation Techniques for Aircraft Damage Detection Algorithms: A Case Study (2025). Proceedings of 28th AIDAA International Congress & 10th CEAS Aerospace Europe Conference.