Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Automated Defect Detection in Aircraft using YOLOv8 and Vision Transformer

P. Subalakshmi,
ME Student,
Department of Computer Science and Engineering,
Government College of Technology,
Coimbatore, India

Dr. R. Muthuram
Associate Professor,
Department of Computer Science and Engineering,
Government College of Technology,
Coimbatore, India

***Abstract* - Visual inspection of aircraft surfaces is essential for ensuring safety. The traditional manual inspections are time-consuming, labour-intensive, and prone to human error. To overcome these challenges, this project proposes an automated deep learning based defect detection system that combines YOLOv8 for real-time defect localization with a Vision Transformer (ViT) for classification. High-resolution aircraft skin images collected under various environmental conditions are preprocessed and analyzed to detect defects such as cracks, dents, scratches, paint-off areas, corrosion, and missing rivet heads. YOLOv8 provides fast and accurate bounding-box predictions, while ViT enhances semantic understanding through attention-based feature extraction, enabling precise classification and severity analysis. The combined YOLOv8–ViT framework attains a detection accuracy of 94.3% and a classification accuracy of 95%, making it suitable for automated aircraft maintenance. This approach reduces human error, speeds up inspection time, and provides a safer and more efficient inspection process.**

Keywords*— *Aircraft defect detection, YOLOv8, Vision Transformer, deep learning.

I. INTRODUCTION

Visual inspection commonly referred to as a General Visual Inspection (GVI)[7], is one of the most important and commonly done maintenance tasks in the aviation industry. GVIs are done several times a day on every commercial aircraft to check for any damage or surface problems before they become serious safety issues. These regular checks make sure the aircraft stays safe to fly and follows all rules. It is thought that nearly 80% of aircraft maintenance activities involve visual inspection, very much depends on humans [20]. If an aircraft fails a GVI, it is considered unsafe to fly and cannot operate until the problem is fixed.

Traditional visual inspections depend on trained engineers, pilots, and ground handling staff who checks the aircraft exterior by eye for dents, cracks, corrosion, missing rivet heads, and other surface defects [8]. These checks are usually done before and after flights, as well as during regular maintenance. However, manual inspections take a lot of time and prone to human error, especially when working under poor lighting conditions on large aircraft surfaces.

Advancements in computer vision and deep learning have made it easier to automatically inspect aircraft. Technologies such as Unmanned Aerial Vehicles (UAVs)[1] provided with high-resolution cameras and AI-powered detection systems can significantly reduce inspection time while improving reliability. In this work, we propose a deep-learning model that uses YOLOv8 to quickly find defects and Vision Transformer (ViT) to classify them and understand their context. Images from the Aircraft Skin Defects Dataset were used for training, validating, and testing the system to check how accurate, reliable, and practical it is. The proposed hybrid method quickly detects defects, reduces human involvement, and supports maintenance strategies for safer, smarter, and more efficient aircraft operations.

II. RELATED WORK

Deep learning is now widely used in aircraft maintenance to automate inspections. Research shows that tools like convolutional networks, lightweight detectors, and transformer models help find structural defects more quickly and accurately than manual checks. Ding et al. [2] used a Faster R-CNN system to detect building damage like cracks, rust, and layer separation, achieving an mAP of 89.7%. Their study showed that region-based detectors work well, but larger datasets and using drones for data collection could make the system more reliable.

Donecle, a French company that uses drones to inspect airplanes, studied how to detect defects in aircraft bodies even when there were very few examples of some defects [4]. They used deep learning to get important features and combined it with Prototypical Networks to detect rare defects more effectively, achieving a mean average precision (mAP) of 79%. This study showed that using advanced learning methods with deep detectors helps detect rare but important aircraft issues more easily. Fotouhi et al. [5] applied transfer learning with a pre-trained AlexNet to classify damage in laminated composite materials, reaching a validation accuracy of 96.15%. While their approach worked well, it could not easily identify the level of damage, showing that CNNs can struggle with detecting small differences in defects.

Deep learning has been used for checking vehicle damage [10], helping to automatically find and analyze defects in vehicles. [6] Van Ruitenbeek et al and colleagues created a dataset of 3,797 damaged vehicle images and tested models such as YOLOv3 [14] and FSSD [15], which achieved the best mean average precision (mAP) of around 33%. Their tests also found that bad lighting and reflections can lead to incorrect detections in real-life situations. Another study on automating insurance claims used CNNs, transfer learning, and combined models, achieving 89.5% accuracy. However, low recall indicates that these models still need to understand features better and must handle different image conditions more effectively.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Vision Transformers (ViTs) are now commonly used in inspection systems because they can understand the overall structure of an image. Unlike CNNs that look at small local areas of an image, ViTs split the image into patches and use self-attention to understand the relationships between all parts of the image [11]. This makes them very good at telling apart similar-looking defects like scratches, tiny cracks, and areas where paint is missing, which is very important for airplane maintenance.

The evolution of YOLO models has greatly improved the detection of aircraft faults. Earlier versions (v1–v3) focused on detecting objects quickly in real time. Later versions like YOLOv5, YOLOv7, and YOLOv8 [13] added new features like attention mechanisms, anchor-free detection, and better loss functions to make detection more accurate. YOLOv8 provides a good balance of speed and accuracy, making it useful for inspecting high-resolution images of aircraft. Early research on detecting aircraft defects was done by Malekzadeh et al., [17] who created an automatic system using deep neural networks and achieved 96.37% accuracy. However, their model used small 65×65 image patches, which made it hard to see the full picture and difficult for maintenance crews to understand. Also, they did not test the system in real-world conditions with different lighting, dirt, or background noise.

Overall, previous studies show that a lot of progress has been made in automatically detecting defects in structures, cars, and airplanes. However, problems like small datasets, poor lighting, and difficulty spotting rare defects still exist. To overcome these issues, this work uses YOLOv8 for fast and accurate defect detection, combined with a Vision Transformer for detailed classification. This approach provides a scalable, precise, and drone friendly solution for inspecting modern aircraft.

III. PROPOSED METHODOLOGY

To automatically detect, classify, and identify aircraft surface defects, a hybrid deep-learning framework uses YOLOv8 to find defects, a Vision Transformer (ViT) to classify them, and a severity analysis module to check how serious each defect is (see Fig 1). The system first uses YOLOv8 to look at the aircraft image and find any defect areas by drawing boxes around them. [16] These boxed regions are then sent to the Region Cropping step, which cuts out only the defect parts to reduce background interference during analysis.

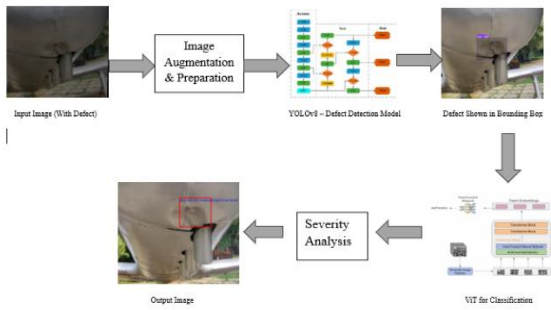

Fig. 1. Overall process of the proposed system.

Each cropped image is then given to the Vision Transformer (ViT) model, which identifies what type of defect it is. After this, the system performs Severity and Confidence Analysis, where it checks how serious the defect is based on the model's confidence. Finally, all results like the bounding box, defect name, and severity are shown in the Visualization and Reporting section. This makes it easy for the maintenance team to understand the defects and decide what to do next. This step-by-step process ensures defects are detected, clearly identified, and given the right priority for aircraft safety checks.

A. Dataset collection

The dataset used in this research is the Aircraft Skin Defect Dataset, which contains clear, good-quality images of different types of defects found on real aircraft surfaces. These images include parts like the body, wings, rivet lines, and panel joints. The dataset include many defect types such as cracks, dents, scratches, corrosion, paint peeling, and damaged or missing rivets defects that are important for checking the aircraft's safety. The images were taken in different lighting conditions, angles, and backgrounds to look like real inspection condition, including shadows, reflections, and small obstacles. This variety helps the model learn better and perform well in real life. All images are in JPG format with the same size, so they are easy to use for deep-learning models. Overall, the dataset is good and realistic for training and testing the YOLOv8–ViT model.

B. Image Augmentation

Image augmentation is used in this project to create more training images because the original aircraft skin dataset has very few defect samples. Using the Albumentation library, several new versions of each defect image are created. These include flips, rotations, brightness and contrast changes, and small size adjustments, which help copy real inspection conditions on aircraft surfaces. After augmentation, the total number of training and validation images becomes much larger. This helps the model learn better, prevents overfitting, and helps it better detect defects even when the lighting, angles, or aircraft materials.

C. Data Preparation & Annotation

Image annotation was done using the LabelImg tool. Each defect on the aircraft surface was marked with a bounding box and given a label based on its type. The dataset included five defect types:

- crack
- dent
- scratch
- paint peel
- missing rivet

For each labeled image, LabelImg created a YOLO-format .txt file that includes the class ID and the bounding box coordinates (x, y, width, height) in a normalized form. To save time and reduce manual work, a semi-automated labeling method was used. First, a small set of images was labeled by hand and then used to train an initial YOLOv8 model. This model then predicted labels for the remaining images, which were checked and corrected for accuracy. This process made labeling faster and keep the data accurate. Finally, the dataset was then divided into 80% for training and 20% for validation, making sure all defect types were included and no data was repeated.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

D. YOLOv8 Detection Module

The YOLOv8 detection module is the first step in the aircraft defect identification system. It finds and identifies defects directly from the input images. YOLOv8 uses a backbone called C2f and a FPN–PAN neck, which helps it detect defects of different sizes, like cracks, dents, corrosion, and missing rivets. During training, the model learns to recognize visual patterns for each type of defect and creates bounding boxes with confidence scores for the detected areas. Lightweight versions like YOLOv8n and YOLOv8s were chosen to keep detection fast and suitable for real-time. After training, YOLOv8 can quickly look at each image and give the defect locations and their predicted types. These detected areas are then cropped and sent to the Vision Transformer for more detailed classification. This way YOLOv8 focus on finding defects, while the classifier works on clear images of just the defects to give better results (see Fig 2).

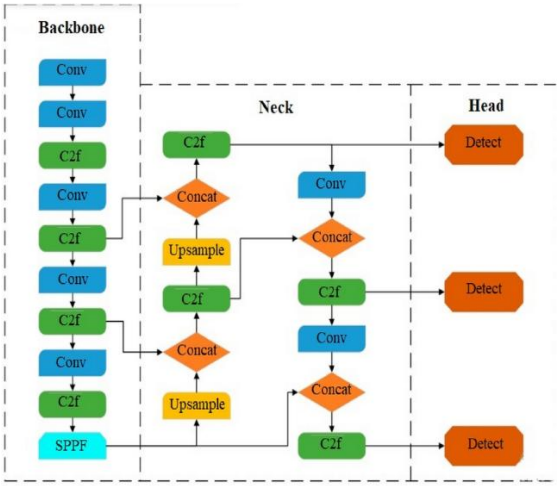

Fig. 2. YOLOv8 architecture

E. Vision Transformer (ViT) Classification Module

The Vision Transformer (ViT) classification module is the second part of the proposed system. It checks each cropped defect image from YOLOv8 and identifies the type of defect. Unlike normal CNN models that look at small part of the images, ViT splits the cropped image into small patches and uses self-attention to understand the entire image. This helps the model understand long-range patterns, making it easier to tell apart defects that look almost the same, like cracks and scratches or corrosion and paint peel. The ViT model used in this project was first trained on large image datasets and then trained again using aircraft defect images, so it can learn features specific to aircraft defects. During training, the model learns to look at the shape, colour, edges, and texture of each defects, which helps it classify them very accurate. With its attention mechanism, ViT understands the important features of every defect type and gives more reliable results. Finally, it shows what type of defect it found and how sure the model is, and these results are then used to check severity of the defect (see Fig 3).

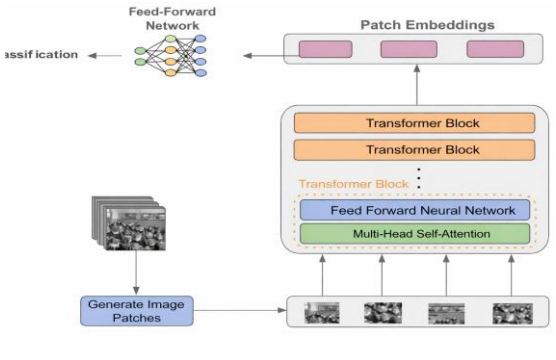

Fig. 3. ViT classifier

F. Severity Analysis Module

After the classification stage, each detected defect is assigned a severity level using a rule-based and confidence scores. The severity depends on the type of defect and the confidence scores from both the YOLOv8 detector and the Vision Transformer classifier.

TABLE 1

RULE-BASED SEVERITY ANALYSIS

Defect type	Confidence range	Severity level
Crack	> 0.85	High
Corrosion	0.70–0.85	Medium
Missing rivets	> 0.90	High
Scratch	< 0.60	Low

IV. EVALUATION METRICS

YOLOv8 Detection Accuracy (mAP):

The mean Average Precision (mAP) tells how correctly YOLOv8 detects and marks defects like cracks, dents, scratches, corrosion, missing rivets, and paint peel.

- mAP50 means accuracy when the overlap (IoU) is at least 50%.
- mAP50–95 means the average accuracy from IoU 50% to 95%.

Higher mAP values indicate better defect localization and detection performance.

A. ViT Classification Accuracy:

This measures how accurately the Vision Transformer identifies the defect type for each detected region.

$$\text{Classification Accuracy} = \frac{\text{Number of correctly classified defect regions}}{\text{Total number of detected defect regions}} \times 100$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

B. Quantitative Results

We tested YOLOv8 and ViT together, and its detection accuracy, classification accuracy were measured. The results are summarized in the table below:

TABLE II

PERFORMANCE METRICS

Model	YOLOv8 Detection (mAP50)	YOLOv8 Detection (mAP50–95)	ViT Classification Accuracy
YOLOv8 + ViT	0.94	0.82	95%

The combined YOLOv8 + ViT model gives the best results. It has the highest detection accuracy (mAP50–95 = 0.82) and the highest classification accuracy (95%). This shows that the model can clearly identify even small and difficult defects like corrosion, fine cracks, and partly missing rivets.

V. RESULTS AND DISCUSSION

A. Visualization of Model Interpretability

Figure 4 shows the output from the YOLOv8 detection module, where the model draws boxes around the defects found on the aircraft surface. These images clearly show how the system identifies problem areas like cracks, dents, corrosion, paint peel, and missing rivets. Each box marks the exact area the model detects as a defect, helping maintenance engineers understand the system easily. The results also show that the model mainly focuses on important parts of the aircraft body, such as rivet lines, panel joints, and areas that usually face high stress and are more likely to get damaged.

Fig. 4. YOLOv8 Bounding-Box

Figure 5 shows how the Vision Transformer (ViT) explains its classification using attention heatmaps. These heatmaps highlight the parts of the defect patch the model looks at most. Red and yellow areas show where the model is strongly focusing like crack edges, corrosion patterns, or missing rivet marks while blue and green areas show regions that matter less. This helps users easily understand why the model selected a particular defect type.

Fig. 5. ViT Attention Heatmaps for Defect Classification

Figure 6 shows how the system assigns a severity level Low, Medium, or High to each detected defect based on the model's confidence, the type of defect, and the size of the bounding box. The colour markings help engineers easily spot serious issues, like deep cracks or missing rivets, which are shown in red. This gives a clear and understandable process from detection to severity rating, making the system transparent and reliable for aircraft inspection.

Fig. 6. Severity-Level Visualization for Aircraft Defects

B. Discussion

The proposed YOLOv8–ViT system works well for automatically checking aircraft defects. YOLOv8 accurately finds the defect locations on different aircraft surfaces, and the ViT model helps correctly identify small or detailed issues like corrosion, tiny cracks, or missing rivets. Using both together gives better results than detection alone. The severity module adds more value by marking which defects are most serious, so maintenance teams can focus on urgent problems first. Overall, the system is fast, easy to understand, and suitable for real aircraft inspections, helping reduce manual work and improve safety.

VI. CONCLUSION AND FUTURE WORK

This research introduces an automated aircraft defect detection system that uses YOLOv8 to quickly find defects and a Vision Transformer (ViT) to correctly classify them. The system can identify many types of defects such as cracks, dents, scratches, corrosion, paint peeling, and missing rivets, even when images are taken in different lighting or working conditions. By combining YOLOv8's fast detection with the ViT model's strong ability to understand detailed features, the system can accurately analyse and classify defects on aircraft surfaces.

The proposed model performs more effectively than traditional CNN-based approaches. It gives high detection and classification accuracy while still running efficiently. The explainable

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

outputs, like bounding boxes and ViT heatmaps, clearly match the real defect areas, making the system more interpretable, reliable, and practical for maintenance engineers.

Because it is fast, light, and accurate, this system is useful for real aircraft maintenance. The system can support early defect identification and improve maintenance decision-making and make inspections more reliable, reduce mistakes made by humans, and highlight which parts need repair first using its severity analysis module. Overall, the YOLOv8–ViT model offers a practical, accurate, and easy-to-understand solution for aircraft inspection, helping improve safety and efficiency in aviation maintenance.

A. Future Work

The proposed YOLOv8–ViT defect detection system works well, but there are many ways to make it even better for real use. In the future, the dataset can be improved by adding more aircraft images from different places, aircraft types, and real working conditions so the model can understand all kinds of situations. Using video-based inspection or drone-captured images can help in continuous real-time monitoring of aircraft surfaces. Adding other types of data like thermal images, infrared scans, or ultrasonic sensor readings can help detect hidden defects that cannot be seen in normal photos.

The severity analysis part can be made stronger by creating a model that learns from past maintenance records and how defects grow over time. Using more advanced transformer-based models may also improve accuracy when the defects are complicated. Finally, running the system on edge devices or maintenance drones can support the development of a fully automated aircraft inspection platform.

REFERENCES

[1] Luke Connolly, James Garland(Member, IEEE), Diarmuid O'gorman , Edmond F. Tobin "Deep-Learning-Based Defect Detection for Light Aircraft With Unmanned Aircraft Systems,"May2024.https://doi.org/10.1109/ACCESS.2024.3412204.

[2] M. Ding, B. Wu, J. Xu, A. N. Kasule, and H. Zuo, ''Visual inspection of aircraft skin: Automated pixel-level defect detection by instance segmentation,'' Chin. J. Aeronaut., vol. 35, no. 10, pp. 254–264, Oct. 2022. https://doi.org/10.1016/j.cja.2022.05.002

[3] Y. Li, Z. Han, H. Xu, L. Liu, X. Li, and K. Zhang, ''YOLOv3-lite: A lightweight crack detection network for aircraft structure based on depthwise separable convolutions,'' Appl. Sci., vol. 9, no. 18, p. 3781, Sep. 2019. https://doi.org/10.3390/app9183781

[4] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, ''Albumentations: Fast and flexible image augmentations,'' Information, vol. 11, no. 2, p. 125, Feb. 2020. https://doi.org/10.3390/info11020125

[5] S. Fotouhi, F. Pashmforoush, M. Bodaghi, and M. Fotouhi, ''Autonomous damage recognition in visual inspection of laminated composite structures using deep learning,'' Compos. Struct., vol. 268, Jul. 2021, Art. no. 113960. https://doi.org/10.1016/j.compstruct.2021.113960

[6] R. E. van Ruitenbeek and S. Bhulai, ''Convolutional neural networks for vehicle damage detection,'' Mach. Learn. Appl., vol. 9, Sep. 2022, Art. no. 100332. https://doi.org/10.1016/j.mlwa.2022.100332

[7] Y.D.V.Yasuda,F.A.M. Cappabianco, L.E.G.Martins, and J.A.B.Gripp, ''Aircraft visual inspection: A systematic literature review,'' *Comput. Ind.*, vol. 141, Oct. 2022, Art. no. 103695. https://doi.org/10.1016/j.compind.2022.103695

[8] R. Girshick, ''Fast R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017. https://doi.org/10.1145/3065386

[10] K. Patil, M. Kulkarni, A. Sriraman, and S. Karande, ''Deep learning based car damage classification,'' in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 50–54. https://doi.org/10.1109/ICMLA.2017.0-179

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ''MobileNetV2: Inverted residuals and linear bottlenecks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[12] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLOv8*. [Online]. ,juhttps://github.com/ultralytics/ultralytics

[13] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, ''A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS,'' *Mach. Learn. Knowl. Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023. https://doi.org/10.3390/make5040083

[14] Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767

[15] Z. Li, L. Yang, and F. Zhou, ''FSSD: Feature fusion single shot multibox detector,'' 2017, arXiv:1712.00960. https://doi.org/10.48550/arXiv.1712.00960

[16] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

[17] T. Malekzadeh, M. Abdollahzadeh, H. Nejati, and N.-M. Cheung, ''Aircraft fuselage defect detection using deep neural networks,'' 2017, arXiv:1712.09213. https://doi.org/10.48550/arXiv.1712.09213

[18] J. Miranda, J. Veith, S. Larnier, A. Herbulot, and M. Devy, ''Machine learning approaches for defect classification on aircraft fuselage images aquired by an UAV,'' Proc. SPIE, vol. 11172, pp. 49–56, Jul. 2019. https://doi.org/10.1117/12.2520567

[19] Tzutalin. (2015).Labelimg. [Online]. https://github.com/tzutalin/labelImg

[20] C. G. Drury and J. Watson, ''Good practices in visual inspection,'' Hum. Factors Aviation Maintenance-Phase Nine, Prog. Rep., FAA/Hum. Factors Aviation Maintenance, USA, Tech. Rep. 1, 2002. https://doi.org/10.1109/TIP.2019.2910414

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016,pp.779-788. https://doi.org/10.48550/arXiv.1506.02640