MEKATRONIKA
JOURNAL OF MECHATRONICS AND INTELLIGENT MANUFACTURING
E-ISSN: 2637-0883
VOLUME 7, ISSUE 2, JULY 2025, 108 - 123
DOI: https://doi.org/10.15282/mekatronika.v7i2.11787

RESEARCH ARTICLE

Automated Aircraft Structural Defect Detection Using Deep Learning and Computer Vision

Rexcharles Enyinna Donatus[1]*, Osichinaka Chiedu Ubadike[1], Mathias Usman Bonet[1], Samuel David Iyaghiba[1], Ifeyinwa Happiness Donatus[2] and Ndubuise Isaac Mbada[3]

[1]Department of Aerospace Engineering, Air Force Institute of Technology, Kaduna, Nigeria
[2]Department of Computer Science, Kaduna State University, Kaduna, Nigeria
[3]Metallurgical and Material Engineering Department, Air Force Institute of Technology, Kaduna, Nigeria

ABSTRACT - Manual aircraft inspections are labor-intensive and susceptible to human error, potentially compromising safety and accuracy. This study presents an automated defect detection framework based on the Mask R-CNN instance segmentation model for identifying cracks and dents in aircraft structures. A dataset of 2,000 annotated images was generated using augmentation techniques and used to train a ResNet-101-based Mask R-CNN model. The system achieved high detection performance, with crack detection reaching a precision of 92.8%, recall of 88.7%, and F1 score of 90.75%; dent detection achieved 91.2% precision, 88.1% recall, and an F1 score of 89.62%. Evaluation using mean Intersection over Union (IoU) and Average Precision (AP@[IoU=0.50:0.95]) confirmed accurate defect localization and segmentation. These findings demonstrate the model's potential to improve inspection reliability and operational efficiency, contributing to safer, more consistent aircraft maintenance practices.

ARTICLE HISTORY
Received : 27th Dec. 2024
Revised : 6th March 2025
Accepted : 20th June 2025
Published : 30th July 2025

KEYWORDS
Computer vision
Deep learnings
Defect detection
Instance segmentation
Mask R-CNN
Object Detection

1. INTRODUCTION

Air travel is among the safest transportation methods, effectively covering vast distances while accommodating numerous passengers and significant cargo [1], [2].It is noteworthy that throughout the aircraft's life cycle, structural inspection emerges as a pivotal factor in ensuring safety and is indispensable for ensuring safe air transportation [3], [4]. The crash of China Airlines Flight 611 resulted from undetected cracks that progressively deteriorated, eventually causing the aircraft to break apart [5]. Failure to conduct thorough inspections during airline maintenance poses significant risks to flight safety. Undetected issues or defects in aircraft components can compromise their integrity, potentially leading to in-flight malfunctions or accidents. As a result, thorough structural inspections are crucial for maintaining the airworthiness and dependability of aircraft, ensuring the safety of both passengers and crew. With aircraft undergoing continuous cycles of loading and unloading during flights, landings, taxiing, and cabin pressurization, many components are susceptible to developing fatigue cracks over time. Additionally, external factors such as lightning strikes pose another risk for cracking. To address these issues, airlines implement pre- and post-flight inspections[6], [7]. These inspections follow checklists detailing various components, which, if improperly maintained, could lead to accidents [3], [4].

Routine visual checks before takeoff and after landing ensure the optimal condition of the aircraft. Aircraft fuselage exteriors typically consist of metallic alloys or composites. Maintenance personnel perform many procedures manually or visually, relying on their expertise [8], [9].Globally, most aerospace inspections rely on manual human visual assessment of structural parts, which is often unreliable, time-consuming, labor-intensive, and risky for inspectors, depending heavily on their expertise[6], [10].Researchers have explored automated inspection using image processing methods, UAVs, traditional machine learning techniques, and deep learning for cost-effective, efficient, and safer aircraft maintenance [5], [11], [12]. Edge detection automates crack detection but is sensitive to noise [13], [14]. Percolation-based processing and image binarization have limitations with small or blurry cracks[6], [15], [16]. Deep learning, especially CNNs, improves feature extraction and flaw identification with greater precision [12], [17], [18].

Due to inherent weaknesses in traditional aircraft surface crack detection methods, primarily reliant on manual, visual inspections conducted by highly trained personnel, this approach is subject to limitations. These include inspector fatigue, lapses in concentration, and subjective judgments, potentially leading to missed or misdiagnosed defects[19]. Deep learning has recently demonstrated remarkable strides across various visual tasks, including object classification and detection [20], [21]. This development has transformed the way traditionally difficult visual tasks are approached, resulting in a wide range of successful applications. As a result, incorporating deep learning into structural defect detection has become a primary focus of this study.

***CORRESPONDING AUTHOR | R. E. Donatus** | ✉ charlly4eyims@yahoo.com

Traditional vision-based techniques face several challenges, such as heavy reliance on human input, time consumption, and susceptibility to error. Technicians must remain highly focused, increasing the risk of missing or misidentifying defects, which can lead to safety issues and inefficiencies. With the rapid growth of the airline industry, these methods struggle to meet rising demand and accuracy standards. Furthermore, technician fatigue can heighten the risk of oversight, emphasizing the urgent need for an automated, reliable, and efficient system to accurately analyze defect data from images or videos. This necessity drives the exploration of advanced methods to enhance defect detection and address these challenges.

Prior research has established a foundation for understanding defect detection; We emphasize the key aspects of these studies, with a focus on their methodologies and findings. Studies conducted by Wang *et al.,* [22], proposed a method utilizing different-source sensors and support vector machines for improved detection of aircraft skin cracks. The technique involved fusing data from image sensors and ultrasonic sensors to enhance the accuracy and distinguishing degree of crack detection. Experimental results indicated a higher recognition degree compared to other methods, with the Genetic Algorithm-Support Vector Machine (GA-SVM) method achieving an accuracy of 93.3% and SVM method achieving 88.9%. The result showed the effective utilization of information fusion from multiple sensors, leading to enhanced inspection performance and overcoming limitations of traditional visual inspection methods. While the proposed GA-SVM method achieved higher accuracy in aircraft skin crack detection compared to traditional SVM, the computational cost associated with the GA optimization process may raise concerns. This increased computational complexity could potentially limit real-time application scenarios and scalability.

Research undertaken by George, (2016) [23], conducted a crack analysis of the slat and flap sections of an aircraft using non-destructive testing (NDT) techniques. The primary objective was to identify any cracks present in the slat and flap sections of a Boeing 757-200 aircraft wing. The techniques used for this inspection included ultrasonic bond testing and eddy current testing. The results of the inspection revealed disbonding of the skin to core wedge in the slat section and cracks in the support fitting assembly of the inboard flaps. However, the study did not provide specific percentages for the detected defects; by incorporating state-of-the-art methods the accuracy and efficiency of the crack detection process can be further enhanced.

In 2017 Malekzadeh *et al.,* [24], proposed a novel approach utilizing Deep Neural Networks (DNNs) and transfer learning to automatically detect defects in aircraft structures. The technique involved selecting regions of interest using the Speeded up Robust Features (SURF) interest point extractor, followed by pre- and post-processing techniques to handle washed and unwashed fuselage conditions. The results of the study demonstrated impressive performance, with an accuracy of approximately 96.37% and a sensitivity of about 96.48% in detecting defects in aircraft fuselage images. The proposed algorithm significantly reduced the workload of manual inspection, with a 6x speed-up in defect detection by evaluating only selected patches of regions of interest. While the proposed approach in aircraft fuselage defect detection showed high accuracy and sensitivity, there may be limitations in handling complex defect patterns or diverse fuselage conditions due to the reliance on the SURF interest point detector. This could lead to missed detections or false positives, potentially impacting the algorithm's adaptability and robustness in real-world scenarios.

Findings from another researcher [25] , they developed an automated UAV-based system for concrete spalling and crack inspection, utilizing deep convolutional neural networks (CNN) for detection and labeling. The system achieved a detection accuracy of 93.36% with the CSSC database and over 70% accuracy in field tests, showcasing the effectiveness of deep learning approaches in concrete defect detection. These findings validate the feasibility of locating cracks in images using a sliding window approach. However, a key challenge lies in determining the optimal window size for defects of varying scales, alongside the high computational overhead incurred due to the repeated application of the DCNN classifier. To enhance the efficiency of detecting and pinpointing objects like cracks, further exploration of advanced object detection technology is warranted.

Another study conducted in 2018 by Gopalakrishnan *et al.,* [26] , worked on enhancing crack damage detection in unmanned aerial vehicle (UAV) images of civil infrastructure using pre-trained deep learning models with transfer learning. The technique involved leveraging deep learning models trained on large image datasets and transferring this knowledge to automatically detect cracks in complex UAV images of civil infrastructure systems. The results demonstrated that the proposed method achieved up to 90% accuracy in identifying cracks in realistic scenarios without the need for additional preprocessing. These findings are significant because it has the potential to streamline and improve the efficiency of civil infrastructure inspection processes by utilizing UAVs and deep learning algorithms. By automating crack detection with high accuracy, this approach can enhance safety, cost-effectiveness, and overall infrastructure maintenance. However, there may be concerns regarding the generalizability and robustness of the model.

A separate study conducted by Shen *et al.,* (2019) [17], proposed a deep learning framework utilizing Fully Convolutional Networks (FCN) for automatic damage detection in aircraft engine borescope images, achieving high prediction accuracy for identifying cracks and burns. The model demonstrated pixel-wise classification accuracy (PA and mA) scores of 0.9803 and 0.9726, as well as region overlap scores (mIU and fwIU) of 0.6789 and 0.6150, respectively. While the results showcased the framework's efficiency and accuracy in damage detection, FCN's performance in complex image scenarios is hindered by its inability to achieve detailed object localization and segmentation at the pixel level.

In [6], the authors Spencer *et al.,* (2019) developed deep learning methods, specifically convolutional neural networks (CNNs), to automate the inspection and monitoring of civil infrastructure. By utilizing CNNs for image analysis and damage detection, the research achieved notable progress in identifying visual defects such as cracks and spalling. The results indicated a significant enhancement in accuracy, with a detection rate of 85% for structural damage. This advancement holds promise for more reliable and efficient civil infrastructure assessment, particularly through the use of physics-based models for training data. While the study showcased significant advancements in accuracy, there is a potential trade-off between performance metrics. Specifically, the focus on enhancing detection accuracy through deep learning techniques may come at the cost of model generalization and adaptability to diverse real-world scenarios. This emphasis on accuracy could lead to overfitting on the training data, limiting the model's ability to effectively handle new or unseen instances of structural damage. Another work By Li et al., (2019) [5], worked on lightweight crack detection in aircraft structures using YOLOv3-Lite. In their work, they revealed significant advancements in detection speed and accuracy compared to existing methods. While YOLOv3-Lite outperformed other models in detection efficiency, achieving a 50% faster speed than YOLOv3, the results obtained from the experiments demonstrate the effectiveness of YOLOv3-Lite in detecting cracks in aircraft structures. However, there were concerns regarding the generalization performance due to limited dataset diversity. Enhancements in hyper parameter tuning and data augmentation strategies may further improve the robustness and accuracy of YOLOv3-Lite in challenging detection scenarios.

A recent study published by Bouarfa *et al.,* (2020) [27] **,** proposed the use of Mask R-CNN, a deep learning technique, to automate the visual inspection of aircraft for detecting dents. The experiments conducted focused on evaluating the performance of the proposed approach by considering metrics such as precision, recall, F1 score, and F2 score. The results indicated that augmentation techniques improved prediction performance, with Experiment 3 achieving a precision of 50.4% and a recall of 5.6%. Additionally, the study found that increasing the number of epochs enhanced overall performance. The significance of these results lies in the potential for improving aircraft maintenance efficiency and accuracy through automated defect detection. By successfully detecting dents on aircraft surfaces, this technology could help maintenance engineers identify and address issues promptly, leading to enhanced safety and operational reliability. However, the limited dataset size is used for training the model, which may impact the generalizability of the results. Hence there is need for larger and more diverse datasets to train the model effectively and ensure its applicability across various aircraft types and maintenance scenarios.

In [28], Doğru *et al.,* (2020) building upon their previous work (Bouarfa *et al.,* (2020)) , by applying Mask R-CNN, a deep learning technique, to enhance aircraft maintenance inspections by detecting defects such as dents in images. By addressing the challenge of limited datasets, the research employs data augmentation techniques (flipping, rotating, and blurring) and a pre-classifier to improve model performance. The results indicate a highest F1 score of 67.50% and a recall of 66.29%, showcasing the model's effectiveness in identifying defects, although precision was lower at 21.56%. These findings highlight the potential of deep learning to improve inspection efficiency and accuracy in the aviation industry. Despite its contributions, the study faces limitations, including reliance on a small dataset, which may affect the model's generalizability and lead to a high number of false positives.

Another study by Medak *et al.,*(2021) [12] , presented automated defect detection from ultrasonic images using deep learning, the authors proposed a novel approach for defect detection in ultrasonic images. They utilized the EfficientNet object detection algorithm, specifically the EfficientDet-D0 model, to analyze ultrasonic images for defects. The study demonstrated that by calculating aspect ratios and scales as proposed in their work, the mean average precision improved by almost 6%. Additionally, they found that using a smaller input image resolution decreased the model's performance, attributing this to information loss due to early down sampling in the EfficientNet architecture. Comparison with other models like YOLOv3 and RetinaNet showed that even the smallest baseline model, EfficientDet-D0, outperformed the best version of RetinaNet by more than 4%. Their result showed the potential for more accurate and efficient defect detection in ultrasonic images, which is crucial for various industries such as manufacturing and infrastructure maintenance. Moreover, the improved performance of the EfficientNet model suggests that deep learning techniques can enhance defect detection processes, leading to more reliable inspection outcomes. However, there may be limitations in terms of fine-grained defect segmentation. The EfficientNet model focuses on object detection rather than instance segmentation, potentially leading to challenges in precise delineating defect boundaries. This limitation could impact the model's ability to provide detailed defect information necessary for certain applications, such as identifying subtle defects or measuring defect areas accurately.

In 2022 a study released by Huang *et al.,* [29] , proposed a deep learning-based method using Mask R-CNN with a morphological closing operation for crack instance segmentation in shield tunnel lining images. By incorporating these advanced techniques, the model efficiently detected cracks and generated high-quality segmentation masks for each crack in the dataset, containing 1171 labeled crack instances in 761 images. The integrated model achieved a balanced accuracy of 81.94%, a F1 score of 68.68%, and an intersection over union (IoU) of 52.72% on 76 test images. The study excels in integrating Mask R-CNN with a morphological closing operation, resulting in superior accuracy in crack detection. However, additional validation is necessary to evaluate the method's applicability to diverse aircraft structural defect detection systems, as it currently exhibits limited adaptability across various structural materials and defect types. Improvements to the F1 score could be achieved through optimization and exploration of alternative backbone architecture or post-processing techniques. Another recent work by Vorobev *et al.,* (2023) [30] presented an approach that used U-Net and DeepLabv3 neural networks for semantic segmentation of structural defects in CT images of CFRP

specimens. The U-Net models achieved Dice coefficients of 0.83 and 0.68 for dataset 1 and 2, while the DeepLabv3 models achieved 0.63 and 0.44, respectively. These results highlighted the effectiveness of simpler models like U-Net for defect segmentation in CT images, emphasizing the critical role of dataset size and model complexity in achieving accurate segmentation. While the study demonstrated the effectiveness of simpler models like U-Net for defect segmentation in CT images, the study's limited exploration of hyper parameters may have missed opportunities to optimize model performance for defect segmentation accuracy. Additionally, its evaluation on specific CFRP datasets may not fully capture real-world defect variability, potentially limiting the model's robustness across diverse composite materials and defect types in practical applications.

A separate work conducted in 2023 by Denhof *et al.,* [31] , developed non-destructive methods based on vibration diagnostics of fatigue damage in aircraft gas turbine engine blades using the higher harmonics and damping characteristic techniques. The experiments focused on detecting small cracks by testing blades at low stress amplitudes, revealing that the second harmonic exhibited high sensitivity to crack presence, surpassing principal resonance sensitivity by up to two orders of magnitude. The acceleration vibration response showed significantly higher non-linear distortions than the strain response, indicating its effectiveness for damage diagnostics. The results underscored the potential of these methods to enhance the reliability of early-stage damage detection in blades, contributing to improved safety and performance in aircraft engines.

To address the limitations of traditional inspection methods, which often rely on subjective visual assessments, and may suffer from issues such as limited accuracy, time-consuming manual processes, and potential oversight of critical defects by human inspectors leading to inconsistencies. While conventional object detection models aim to encompass regions of interest (ROI) within bounding boxes, the precision required to delineate defect shape, and location necessitates the utilization of instance semantic segmentation methods or pixel-level classification that offers the advantage of not only identifying object categories and their boundaries but also distinguishing between individual instances of objects within the same category. This research focuses on utilizing Mask R-CNN, a deep learning technique, to enhance the precision and efficiency of defect detection. By automating the identification of critical defects like cracks and dents, this approach streamlines the pre-flight inspection process, significantly reducing inspection time while improving accuracy.

The proposed approach in this research, which leverages advanced techniques, Mask R-CNN can detect multiple defects simultaneously while generating segmentation masks for each instance, thereby enhancing the accuracy and efficiency of detection of defect. Moreover, Mask R-CNN does semantic segmentation, which extends beyond mere identification to include path delineation.

1.1 Aim and Objectives

This research work aims to develop a deep learning-based computer vision system using Mask R-CNN to improve the efficiency and accuracy of detecting aircraft structural defects. The following objectives are set to be achieved in this research work.

i. To develop an aircraft defect detection system utilizing the Mask R-CNN architecture, trained on visual inspection images.

ii. To validate the effectiveness of the developed model and evaluate its performance through key metrics such as precision, accuracy and recall enhancing overall safety in aircraft operations.

1.2 Object Detection

Object detection is a branch of computer technology within computer vision and image processing. It focuses on identifying and locating instances of semantic objects from specific categories, such as building , humans, or cars, in digital images and videos [32]. Object detection has garnered significant attention recently due to its broad applications and technological advancements. Object detection is a crucial task in autonomous driving within contemporary traffic systems. More recent deep learning models, along with various AI sensors, are vital for effectively managing and controlling these systems. [33].This area is heavily researched in academia and practical fields such as drone scene analysis, autonomous driving, security monitoring, and transportation surveillance, detection of defects in civil and aerospace engineering and robotic vision. The rapid advancement of object detection methodologies is as a result of the development of enhanced GPU computing power and deep convolutional neural networks. Currently, deep learning models are extensively used in computer vision for both domain-specific and general object detection, employing these networks for feature extraction, classification, and localization in images and videos [34].

In contrast to image classification, object detection from images presents a more complex challenge necessitating the use of more complex methodologies. Through the ongoing exploration by researchers, numerous exemplary models of target detection algorithms have emerged. Pre-existing domain-specific image object detectors are typically classified into two categories: two-stage detectors and single-stage detectors. Some of the two-stage algorithms includes Fast R-CNN, Faster RCNN, R-FCN, SPPNet, Mask R-CNN and others[35], [36]. These algorithms are characterized by high localization and higher detection accuracy but slightly slower speed, posing challenges for real-time detection tasks. On the other hand, single-stage detection algorithms like the YOLO series, SSD, DSSD, RetinaNet, and are notable for their high inference speed [21], [37].

However, they tend to exhibit slightly lower detection accuracy when compared to two-stage detection algorithms. The distinguishing feature of two-stage detectors is the RoI (Region of Interest) pooling layer that separates the two stages [34]. The two phases of two-stage detectors are divided by the RoI (Region of Interest) pooling layer. In Faster R-CNN, for instance, the first phase, called the Region Proposal Network (RPN), proposes potential object bounding boxes. In the second phase, the RoI Pooling layer extracts features from each candidate box for tasks related to classification and bounding box refinement [38]. Figure 1 shows the fundamental structure of two-stage detectors.

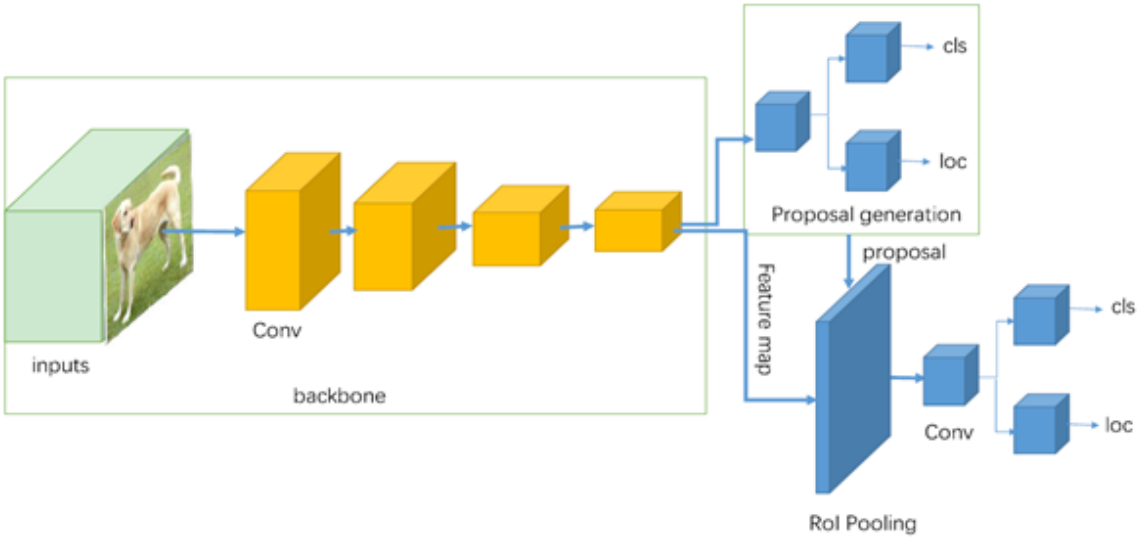

Figure 1. Fundamental Structure of Two-Stage Detectors [34]

Figure 1 depicts the core architecture of two-stage detectors, highlighting a region proposal network that directs region proposals into both a classifier and a regressor. Yellow cubes represent consecutive convolutional layers, grouped into blocks with consistent resolution in the backbone network. Due to down sampling operations after each block, subsequent cubes gradually reduce in size. Thick blue cubes denote convolutional layers, which may encompass multiple layers. The flat blue cube signifies the RoI pooling layer, responsible for producing feature maps of uniform size for detected objects [34].

1.3 Instance Segmentation

Instance segmentation involves delineating objects in their precise shapes, rather than merely enclosing them within bounding boxes that approximate their spatial location. It integrates classification, localization, and segmentation into a single framework, producing pixel-level masks for each identified object. Popular models for instance segmentation include U-Net, DeepLab, and Mask R-CNN. Compared to traditional object detection and image classification methods, instance segmentation offers improved accuracy in detecting fine-grained or irregularly shaped objects, such as surface defects [32].

Mask R-CNN excels at detecting structural defects such as cracks and dents, which often have irregular shapes and non-uniform orientations. While many existing studies focus on object detection tasks involving targets with well-defined features and consistent aspect ratios, these models often struggle to detect cracks due to their variability in shape, scale, and texture. Cracks may appear thin, elongated, or partially occluded, posing challenges for traditional detection frameworks. In contrast, Mask R-CNN performs instance segmentation with precise mask generation, making it well-suited for identifying such complex and variable defects.

Additionally, the model incorporates instance segmentation loss during training, which enhances its performance in spatially detailed tasks. Its multi-branch architecture and use of RoI Align allow for accurate pixel-level segmentation and effective feature extraction. Mask R-CNN is widely recognized as a state-of-the-art approach for instance segmentation and has been successfully applied in various domains, including road surface defect detection, industrial inspection, aerospace component analysis, bolt fastener inspection, and leather surface segmentation [29].

1.4 Mask Region-Based Convolutional Neural Network

The development of the Mask Region-Based Convolutional Neural Network (Mask R-CNN) model by the Facebook AI team in 2017 represents a significant advancement building upon the Faster R-CNN model. While Faster R-CNN exhibits high efficiency in target identification and classification, they face limitations in distinguishing between different individuals within the same target class. Similarly, despite the effectiveness of these two-stage detectors, the quest for high-speed detection has led to the emergence of masks region-based Convolutional Neural Network (Mask R-CNN). The Mask R-CNN, expands the capabilities of Faster R-CNN by incorporating functionalities for semantic segmentation, object localization, and instance segmentation of natural image alongside the existing regression and classification branches [32]. These enhancements are illustrated in Figure 2.

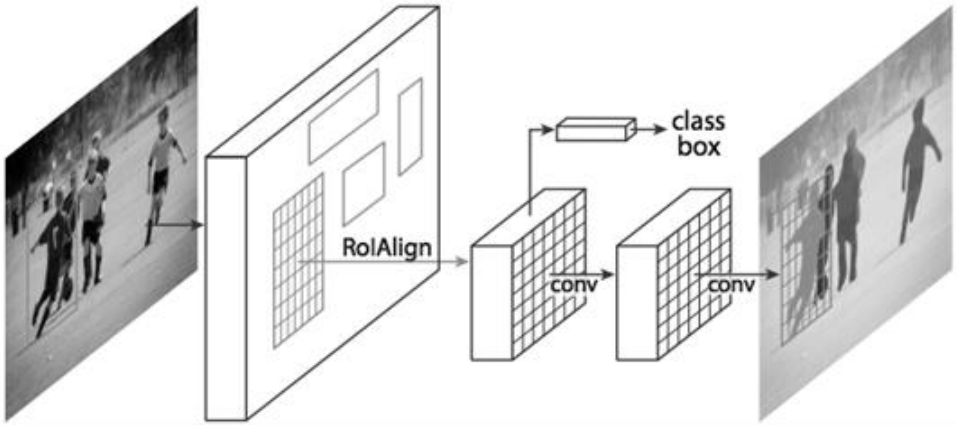

Figure 2. Mask R-CNN [38]

Mask R-CNN offers improved object detection through segmentation. It introduces a sub module known as the Mask Branch onto Faster R-CNN, enabling the learning of segmentation mask [38]. This sub module simultaneously predicts segmentation masks for each Region of Interest (RoI) by using convolutional arrays for classification and bounding box regression. The mask module employs a compact fully convolutional network (FCN) for each RoI, producing a segmentation mask for every pixel within it [32], [33].

The key modifications to the Faster R-CNN architecture. Is that it replaces the conventional RoI Pooling operation with an improved RoI Align operation, enabling precise generation of instance segmentation masks. Secondly, Mask R-CNN integrates a dedicated network head, which functions as a compact fully convolutional neural network, facilitating the production of desired instance segmentations. Thirdly, it decouples mask and class predictions, with the mask network head autonomously predicting masks independently from the network head responsible for class prediction. Figure 3 depicts the structure of Mask R-CNN, featuring several key components: a head architecture, backbone architecture, and a region proposal network (RPN), [29].

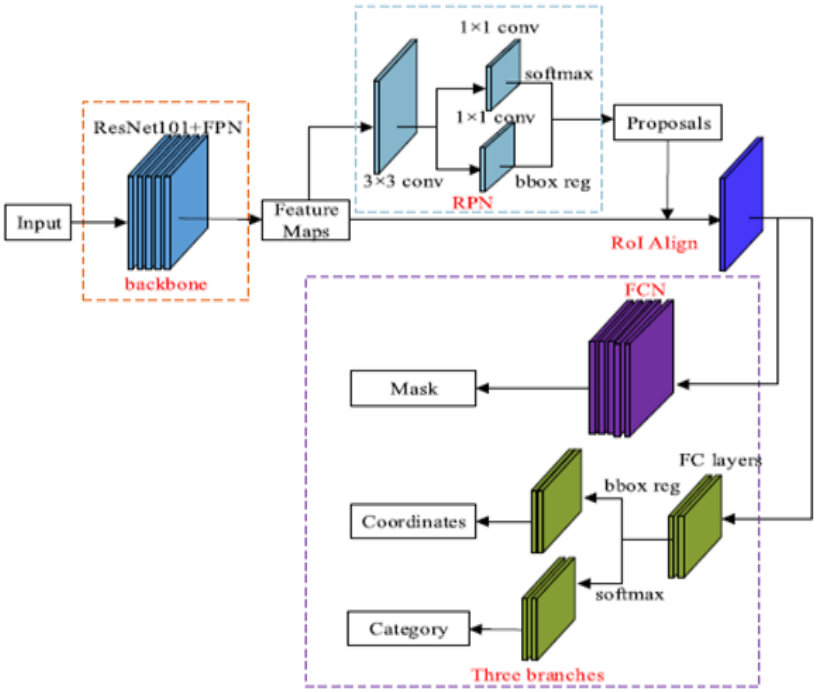

Figure 3. Architecture of the Mask R-CNN model used for defect detection [39]

Although Mask R-CNN enhances the performance of these earlier models, it imposes a higher computational cost. For instance, YOLO, a widely used object detection algorithm, is significantly faster when only bounding box predictions are required [39]. However, detecting cracks and dents often necessitates precise segmentation rather than just bounding boxes, making Mask R-CNN a more suitable choice despite its computational demands. Another limitation of Mask R-CNN is the annotation process for segmentation masks. Labeling data for mask generation is a labor-intensive task, as annotators must manually outline a polygon around each crack or dent within an image, making the process time-consuming and challenging.

2. METHODS AND MATERIAL

The system architecture adopted in this research utilizes a robust deep learning framework tailored for precise defect identification in aircraft structures. Instance segmentation was chosen over traditional object detection due to its ability to localize objects with pixel-level accuracy. Rectangular bounding boxes are often inadequate for detecting cracks, which are typically thin, irregular, and non-axis-aligned. Bounding boxes tend to enclose excessive background, reducing localization precision and introducing noise that may degrade model performance. As a result, traditional object detection frameworks struggle to accurately identify such defects, whereas instance segmentation particularly through Mask R-CNN provides a more effective solution [32]. The system incorporates a preprocessing pipeline that includes data augmentation and normalization to enhance dataset variability and improve generalization during training. At its core, the architecture employs the Mask R-CNN model, which combines object detection and instance segmentation to accurately locate and classify defects in high-resolution images. This design supports efficient processing and high detection accuracy, making it well-suited for real-time or near-real-time pre-flight inspection scenarios.The sequential flow of the system is illustrated in the architectural block diagram (Figure 4), which outlines the stages from input preprocessing through to final defect prediction and segmentation output.

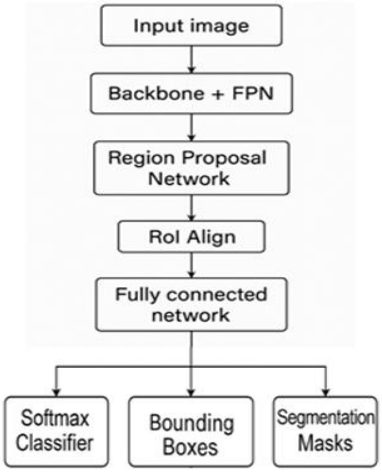

Figure 4. Sequential Pipeline of the Mask R-CNN Architecture

Once the input image is processed by the ResNet-101 backbone, feature maps are extracted at multiple levels. This deep residual network captures hierarchical image features, enhancing representation learning through skip connections. To improve detection across multiple spatial scales, a Feature Pyramid Network (FPN) is applied atop ResNet-101, integrating low- and high-level feature maps. This fusion enhances the model's ability to detect cracks and dents of varying sizes and complexities. The enhanced feature maps are passed to the Region Proposal Network (RPN), which generates region proposals by predicting anchor boxes and classifying regions as either foreground (potential defects) or background (non-defective areas). The RPN performs binary classification only it does not distinguish between cracks and dents—but its outputs bounding box deltas to refine anchor placements, improving alignment with candidate defect regions. Refined proposals are then processed by the Region of Interest Align (ROIAlign) layer, which extracts fixed-sized feature maps, preserving spatial alignment for accurate localization. These features are subsequently passed to the Mask R-CNN detection head, which produces three outputs: (1) a classification label with confidence score (e.g., Crack, Dent, or No Defect); (2) bounding box regression for precise object localization; and (3) a pixel-wise segmentation mask. The segmentation branch is particularly valuable for handling irregular defect geometries that are not adequately captured by bounding boxes alone.

To address the variable shapes and scales of structural defects in aircraft surfaces, the Matterport Mask R-CNN model was adapted with a carefully selected anchor box configuration. Anchors were generated at each feature map level using aspect ratios of 1:1, 1:2, and 2:1, along with scales of 32, 64, 128, 256, and 512 pixels, enabling the RPN to detect both narrow linear cracks and wider dent regions effectively. The anchor stride was maintained at 16 pixels to balance proposal density and computational efficiency. This configuration, in combination with the ResNet-101 backbone and FPN, enabled the generation of high-quality region proposals across multiple resolutions, ensuring robust and accurate instance segmentation for defect detection.

The overall workflow of the proposed deep learning-based framework for structural defect detection is presented in Figure 5, outlining the sequential stages from image preprocessing to defect localization and segmentation. Figure 5 presents a high-level workflow of the aircraft defect detection pipeline developed in this study, showing the flow from image preprocessing and augmentation to model training and evaluation. The process starts with image acquisition and preprocessing, which includes resizing, normalization, and data augmentation to standardize the input and improve model generalization. These images are then passed through a deep learning-based detection system that extracts features, proposes candidate regions, and performs classification and segmentation to identify structural defects such as cracks and dents. Each stage in the pipeline from data preparation to prediction output is designed to support accurate and robust

detection, especially for irregular and small-scale defects. The figure provides an end-to-end overview of how the proposed system transforms raw image data into actionable detection results.

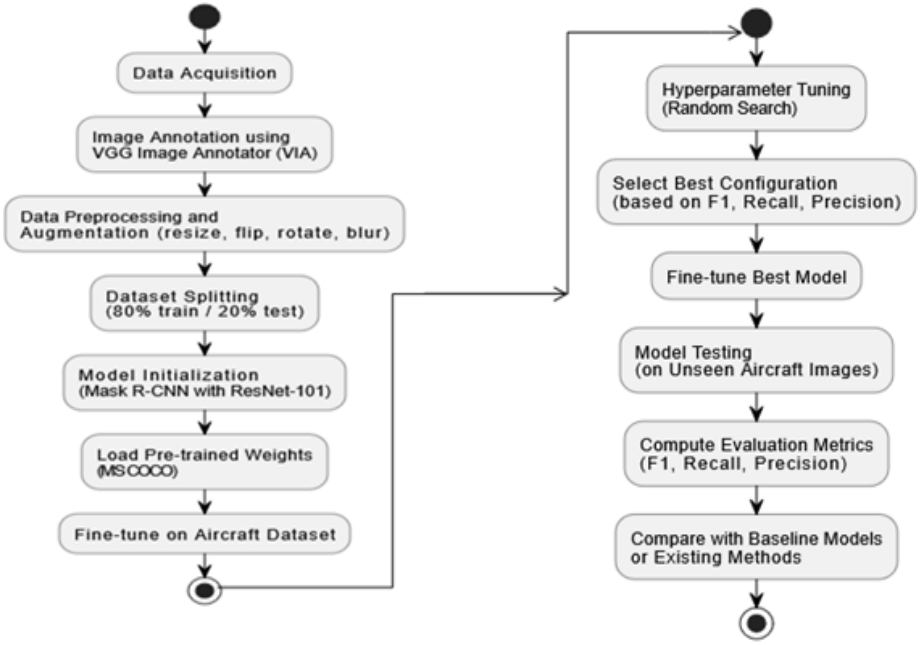

Figure 5. Training pipeline of the proposed Mask R-CNN model for aircraft defect detection

2.1 Dataset

In developing a deep learning model using Mask Regional Convolutional Neural Network (Mask R-CNN) architecture for automated defect detection annotated dataset of aircraft visual inspection images was used. Cracks and dents from various structural parts are collected. The dataset, obtained from Roboflow's repository in 2024, was accessed through Universe at https://universe.roboflow.com/. The platform is comprehensive, tailored to streamline dataset preparation for machine learning endeavors, encompassing tasks like object detection, image segmentation, and classification, satisfying all my criteria for data selection. The selected dataset was augmented to 2000 static photos; it includes information about cracks from various aircraft surfaces and has been hand annotated. Figure 6 displays sample images from the dataset folder, illustrating aircraft surface features.

Figure 6. Images sampled from the dataset folder depicting aircraft surface

2.2 Data Annotation

This research aims to train a computer system to detect cracks and dents in aircraft, which falls under supervised learning. Therefore, data labeling is crucial. In computer vision object detection, this labeling is known as annotation. In our case, annotation involves accurately identifying the damaged areas in images and outlining the boundaries around the aircraft dents. To perform this annotation, we utilize the VGG Image Annotator [27] , depicted in Figure 7 (a) and (b), Instance of data annotation carried out with VGG image annotator. The VGG Image Annotator is an open-source tool developed by the Visual Geometry Group at Oxford University. With this tool, we upload all images and carefully draw polygon masks around the dents in each image. The images are annotated to ensure precise labeling of cracks and dents.

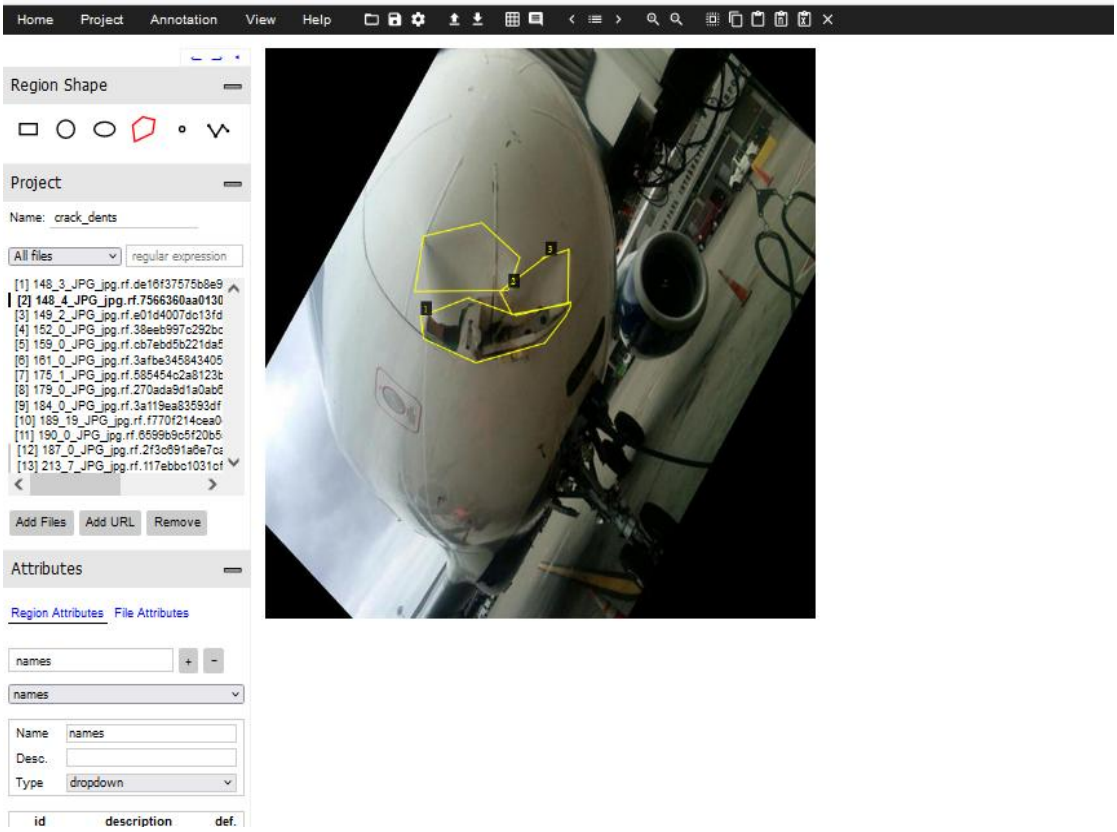

Figure 7. Instance of data annotation carried out with VGG image annotator

2.3 Environmental Setup

Establishing an appropriate computational environment is crucial for optimizing model performance and ensuring development efficiency. In this study, Python was employed due to its extensive ecosystem of open-source libraries, which facilitated efficient code development and rapid prototyping. The implementation of the Mask R-CNN model, along with other benchmark models for aircraft defect detection, was conducted using Python. Key libraries and framework include TensorFlow/Keras, OpenCV, Numpy, Scikit-learn, and Matplotlib, are selected for their respective capabilities in deep learning, image processing, data analysis, and visualization. A GPU-enabled workstation was used to accelerate training and experimentation. All training and evaluation procedures were executed on a system equipped with an Intel Core i7-12700K CPU and an NVIDIA RTX 3090 GPU (24 GB VRAM), running Windows 10 (64-bit). This configuration ensured enough computational resources to handle the complexity of the model and the size of the dataset.

2.4 Model Configuration and Training

In this study, a pre-trained Mask R-CNN model, based on the Matterport implementation, was employed. Mask R-CNN is a two-stage object detection architecture known for its high accuracy, attributed to its advanced design that supports robust feature extraction and precise instance-level segmentation. To adapt the model to the specific task of detecting structural defects in aircraft components, the Matterport framework was fine-tuned using a curated dataset of annotated defects. The training process was structured around three main elements: model configuration, training configuration, and evaluation metrics. The dataset consisted of annotated images of cracks and dents on aircraft surfaces. It was split into training and testing subsets using an 80:20 ratio, yielding 1,600 training images and 400 testing images. This split ensured a reliable evaluation of the model's generalization performance on unseen data.

Fine-tuning of the Mask R-CNN model was conducted over 50 epochs, with each epoch comprising 50 training steps and 4 validation steps [40]. Empirical observations indicated optimal model performance between epochs 37 and 40. Training was conducted using the Adam optimizer, with a learning rate of 1e-4 and a batch size of 4. These parameters were selected to balance memory efficiency and model stability. To enhance generalization and prevent overfitting,

dropout layers with a dropout rate of 0.3 and batch normalization were incorporated into the custom classification and mask heads of the network. During the training process, a multi-task loss function was employed to jointly optimize the model for classification, localization, and segmentation tasks.

$$L = L_{cls} + L_{bpx} + L_{\text{mask}} \tag{1}$$

where:

L_{cls}: Classification loss – which measures the model's ability to correctly classify objects (e.g., crack vs. dent).

L_{bpx}: Bounding-box regression loss – which evaluates the accuracy of the predicted object localization.

.L_{mask} : Mask loss – which quantifies the pixel-level accuracy of the predicted segmentation masks.

This formulation follows the standard multi-task loss proposed by [36] and subsequently refined in practical implementations such as [29]. The convergence of the total loss L is commonly used as an indicator for training completion; a stabilized or plateauing loss value typically signifies that the model has reached optimal performance under current hyperparameters.

2.5 Preprocessing

Preprocessing was a critical step in this study to enhance model performance, ensure input consistency, and improve defective localization accuracy. All images were resized to a standardized resolution of 512×512 pixels to maintain uniform input dimensions, which is essential for stable learning and computational efficiency [28]. Following resizing, pixel values were normalized to a [0,1] range to minimize variation due to lighting and contrast differences, facilitating faster convergence during training [34]. To improve model generalization and robustness, data augmentation techniques including random rotations, horizontal flipping, and Gaussian noise addition were applied to the training images. These augmentations expanded the dataset's diversity and helped the model adapt to various defect appearances and orientations. The combined preprocessing pipeline not only stabilized training but also improved the model's ability to detect cracks and dents under varying real-world conditions. As such, preprocessing significantly contributed to the overall effectiveness of the Mask R-CNN model in structural defect detection.

2.6 Validating Model Effectiveness Using F1 Score, Recall, and Precision

For decision-makers using this decision-support system, identifying the presence of dents and/or cracks is prioritized over precisely measuring their area. Consequently, this study emphasizes the accurate detection of cracks and dents and evaluates performance based on the quality of these predictions. To achieve this, established metrics such as precision, recall, and F1 scores are employed. Hence, the following step is taken.

i. Test the optimized Mask R-CNN models on a separate test dataset containing unseen aircraft inspection images.

ii. Compute evaluation metrics, including F1 score, recall, and precision, to quantitatively assess the performance of the models.

2.7 Evaluation of the Developed Model using Precision, Recall, and F1 Score

To determine the precision, recall and F1 score of the developed model, equations 2, 3and 4 were utilized.

i. Recall: Recall involves adding True Positives (TP) to False Negatives (FN) and then dividing by TP, yielding the fraction of positive predictions that were accurate as depicted in equation (2) [26] [26].

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

ii. Precision: Precision, on the other hand, adds TP to False Positives (FP) and divides by TP, providing insight into how many positive predictions were correct, as illustrated in equation (3) [41].

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

iii. F1-Scores: The F1-Score, as denoted in equation (4), incorporates both recall and precision. It is imperative to consider F1-Score, as it integrates precision and recall, offering a more comprehensive evaluation [42].

$$F1\ Score = 2 \times \left(\frac{Precision\ \times Recall}{Precision + Recall}\right) \tag{4}$$

It is imperative to consider a multitude of performance metrics when assessing the proposed approach to accurately describe the comprehensive performance of the detection and identification system. In this research work, we have highlighted the most pertinent evaluation metrics that should be considered during the evaluation of the developed model.

3. RESULTS AND DISCUSSION

This section presents and discusses the research results, including a performance comparison between existing techniques and the developed model.

3.1 Evaluation of Training Loss

Presents the training loss versus epoch graph, which provides key insights into the learning dynamics of the Mask R-CNN model during the training phase. The graph illustrates the progressive reduction in loss as the model learns to localize and segment structural defects more accurately. A steady decline in the loss curve indicates effective optimization and model convergence. Notably, the loss stabilizes between epochs 35 and 40, suggesting that the model reaches its optimal training state within this range. This convergence implies that the model has successfully minimized prediction errors and is effectively learning to segment cracks and dents in aircraft surface images. The loss curve thus serves as a critical diagnostic tool for validating training performance and determining appropriate stopping criteria.

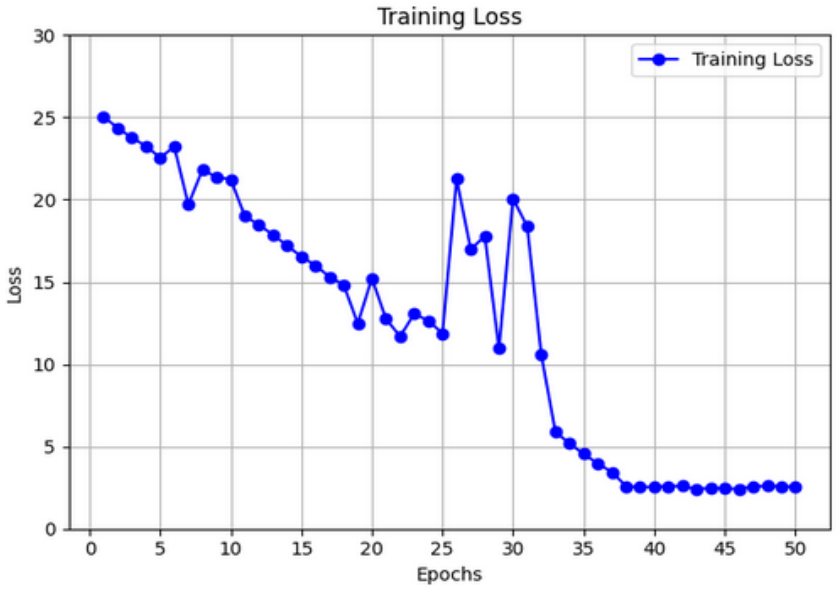

Figure 8. Training loss over 50 epochs with specified fluctuations

3.2 Evaluating Performance of the Developed Model

To evaluate the performance of our developed deep learning model, metrics such as the confusion matrix, precision, recall, and F1 score are used. The confusion matrix visually represents the performance of the object detection model, particularly in the context of detecting cracks and dents. It displays the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This helps in understanding how well the model is distinguishing between the presence and absence of defects. In Figure 10 and Figure 11 Each cell in the matrix represents instances of actual versus predicted class labels, which is crucial for evaluating the model's performance, identifying improvement areas, and understanding error types.

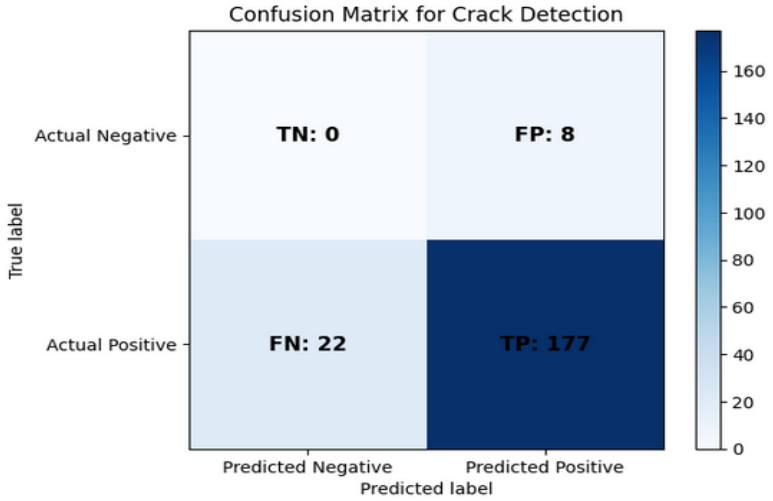

Figure 9. Confusion Matrix evaluating the performance of the developed model in detecting cracks

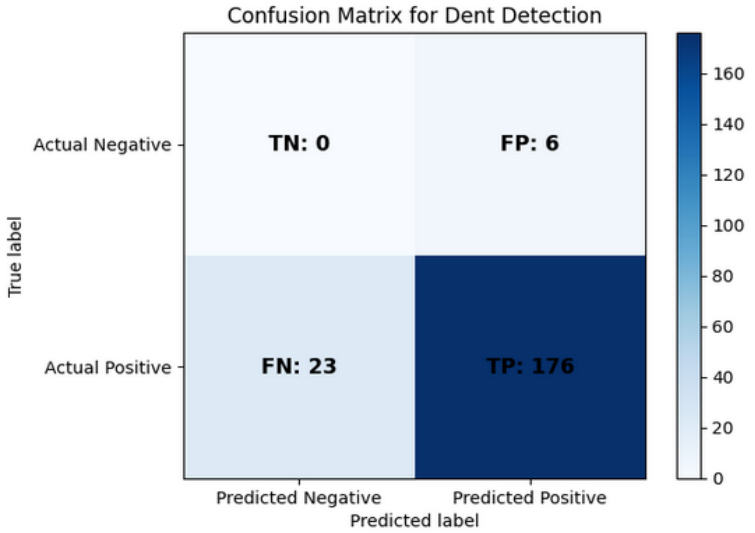

Figure 10. Confusion Matrix evaluating the performance of the developed model in detecting dents

True Positive (TP) refers to instances where the model correctly identifies the presence of a structural defect, such as a crack or dent. False Positive (FP) occurs when the model incorrectly predicts a defect in a region where none exists. False Negative (FN) represents cases where the model fails to detect an actual defect present in the image. True Negative (TN) is less commonly used in object detection tasks because the model does not explicitly label background regions as "non-defect." Since object detectors like Mask R-CNN focus only on positive detections, true negatives are typically considered to be zero or undefined. Evaluating these quantities allows for the computation of key performance metrics such as precision, recall, and F1-score, which are essential for assessing the model's effectiveness in identifying structural defects. These metrics play a critical role in validating the reliability of the defect detection system for real-world aircraft inspection and maintenance applications.

3.3 Performance Evaluation of the Developed Model vs. Baseline Models

The developed Mask R-CNN model achieved strong performance in detecting structural defects on aircraft surfaces. For crack detection, the model attained a precision of 92.8%**,** recall of 88.7%**,** and F1-score of 90.75%**.** Dent detection yielded a precision of 91.2%**,** recall of 88.1%**,** and F1-score of 89.62%**,** with an average precision (AP@0.50:0.95) of 0.889 for cracks and 0.861 for dents. The mean IoU across all test samples was 0.83 for cracks and 0.79 for dents, indicating accurate mask localization**.** Table 1 shows the detailed evaluation results of the developed Mask R-CNN model for crack and dent detection using key metrics such as mean Average Precision (mAP) at different IoU thresholds, as well as Intersection over Union (IoU), precision, F1 scores and recall values. These results reflect the model's effectiveness in accurately localizing and classifying structural defects on aircraft surfaces.

Table 1. Table caption (no period at the end of the caption)

Metric	Crack Detection	Dent Detection	Bouarfa et al. (2020)	Doğru et al. (2020)
AP at IoU=0.50:0.95	0.885	0.855	-	-
AP at IoU=0.50	0.953	0.932	-	-
AP at IoU=0.75	0.81	0.78	-	-
Mean IoU	0.83	0.79	-	-
Precision	0.928	0.912	0.909	0.731
Recall	0.887	0.881	0.667	0.6408
F1 Score	0.9075	0.8962	0.7694	0.6769

From Table 1, the results clearly show that the crack detection model outperforms the dent detection model by a margin of 2.75% in mAP (AP@IoU=0.50:0.95), indicating better overall localization and classification performance. At an IoU threshold of 0.5, both models maintain high precision levels, with cracks reaching 95.5% and dents 92.5%. Even at the stricter IoU threshold of 0.75, crack detection sustains an AP of 80.5%, compared to 77.5% for dents. These results, combined with strong IoU, precision, recall, and F1 scores, confirm the model's effectiveness and reliability in segmenting structural defects under realistic conditions, likely due to more distinct spatial features, a larger training dataset, or improved learning generalization. In addition, these results demonstrate the model's stronger capability in identifying and precisely segmenting cracks, while still maintaining robust performance in dent detection. Recall values confirm that the model effectively detects a majority of true defects in both categories. Compared to Bouarfa et al. (2020) and Doğru et al. (2020), the developed model achieves notably higher precision, recall, and F1 scores, demonstrating superior effectiveness in defect detection tasks.

Figure 12 (a) and (b) presents a bar chart comparing the precision, recall, and F1 score of both the developed and existing models. The developed model for crack detection exhibited higher precision (92.8%) and recall (88.7%) compared to the existing models, which achieved lower precision (71.31%) and recall (64.08%) [28]. As shown in the chart, the F1 score of the developed model (90.75%) surpassed that of the existing models (67.69%), indicating superior overall performance in crack detection tasks. Similarly, the dent detection models achieved higher performance metrics, with F1 score, precision, and recall of 89.62, 91.20, and 88.10, respectively, surpassing the baseline models this improvement highlights the effectiveness of incorporating a larger dataset (1000 images) and combining augmented data with advanced techniques like Mask R-CNN, leading to enhanced robustness and accuracy.

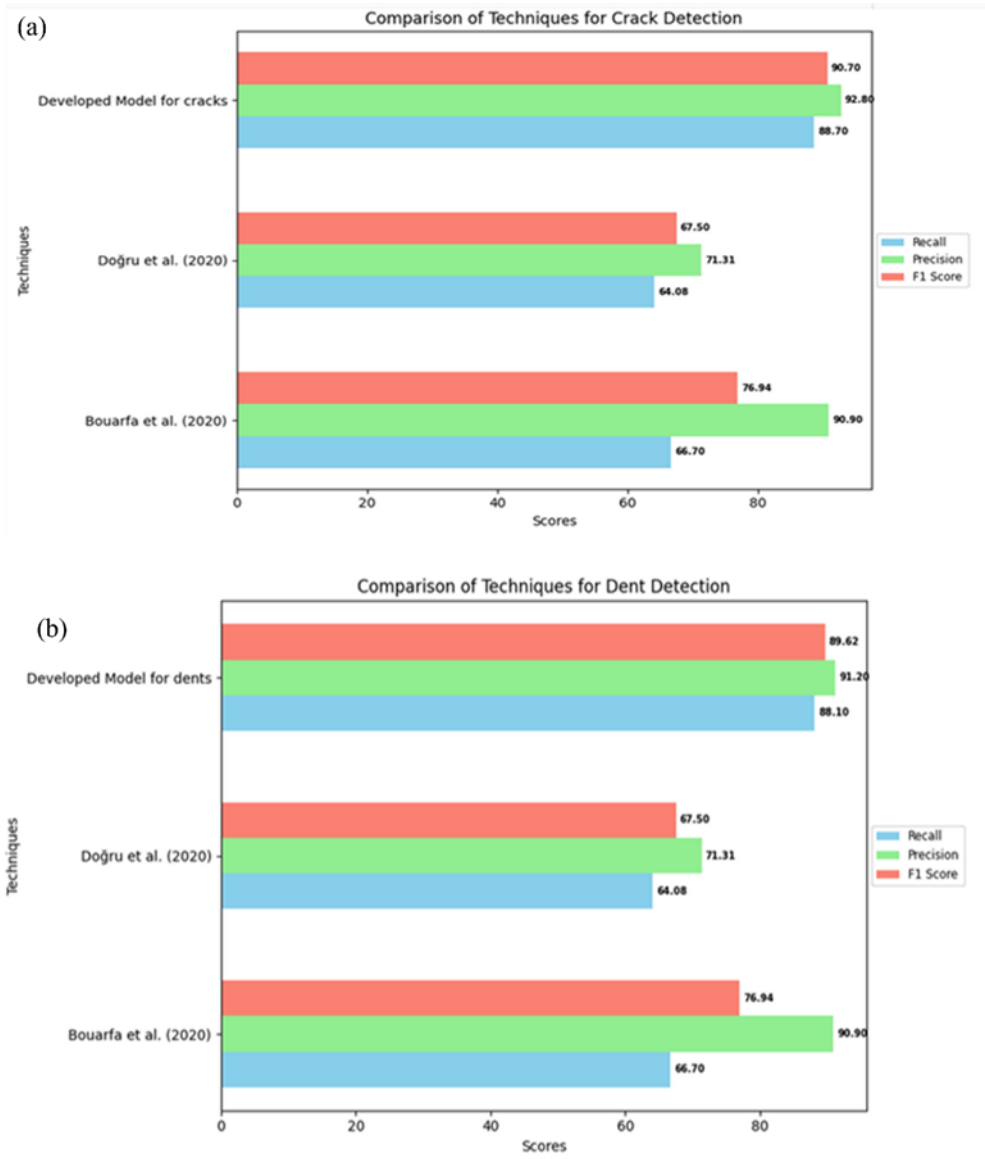

Figure 11. Precision, recall, and F1 score comparison: developed vs. existing models

The high precision achieved by our proposed model indicates fewer false positives, which reduces unnecessary resource allocation. Additionally, the high recall signifies fewer false negatives, ensuring that most objects are detected. Overall, the high accuracy demonstrates the model's reliable performance. The Mask R-CNN model was chosen for its superior ability to perform instance segmentation**,** making it more suitable for pixel-level defect detection compared to bounding-box-only detectors like YOLOv5. Its Feature Pyramid Network (FPN) enhances performance across scales, which is crucial for detecting both fine cracks and broader dents. Furthermore, RoI Align ensures precise localization by eliminating quantization errors found in RoI pooling used in older R-CNN variants. These architectural strengths contribute significantly to the higher mean IoU and AP scores observed. Despite strong performance, the model exhibited limitations in specific cases. Misclassifications were observed when cracks appeared in low contrast regions or had irregular shapes, and when dents overlapped with structural textures, leading to false positives. High false negatives occurred in underrepresented classes, possibly due to dataset imbalance. Addressing these may involve expanding the dataset, using contrast enhancement preprocessing, or integrating focal loss to better penalize hard-to-classify examples.

Although this study focuses solely on the Mask R-CNN architecture, it is important to contextualize its performance against other state-of-the-art object detection models commonly applied in defect detection tasks. For instance, YOLOv5

has been used in surface inspection tasks and typically achieves precision scores around 85–88%, as reported by Jiao et al. (2021) in automated weld defect classification [43]. Similarly, Faster R-CNN remains a strong baseline, with precision levels of approximately 86% in metallic surface flaw detection tasks [44]. Transformer-based models like DETR have shown promising results in recent studies, achieving comparable accuracy but often requiring larger datasets and longer training times [45]. In comparison, the Mask R-CNN model developed in this work demonstrates competitive or superior performance, particularly in its ability to perform pixel-level segmentation, which is crucial for accurately localizing irregularly shaped defects like cracks and dents.

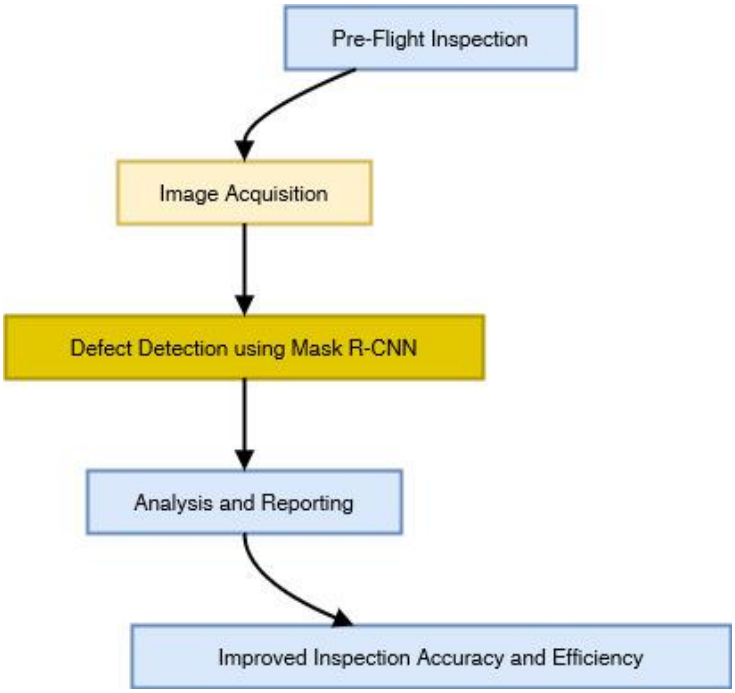

Figure 12. Conceptual integration of the mask R-CNN-based defect detection system into real-world aircraft maintenance workflows

Stage: Pre-Flight Inspection

i. Current Workflow Stage: During the pre-flight inspection, technicians manually examine aircraft surfaces for visible defects such as cracks and dents. This process involves visual inspection and may use traditional methods like dye penetrants or ultrasonic testing.

ii. Integration of the Deep Learning System:
   a) Image Acquisition: High-resolution images of aircraft surfaces are captured using cameras or imaging devices.
   b) Defect Detection: The deep learning system, using Mask R-CNN, processes these images to detect and classify cracks and dents automatically.
   c) Analysis and Reporting: The system provides real-time analysis, highlighting detected defects and generating reports or alerts for the maintenance team.

iii. Outcome: This integration enhances pre-flight inspection by providing accurate, automated defect detection. Technicians receive detailed information about defects quickly, allowing them to address issues more efficiently and improve overall inspection accuracy. This reduces manual inspection time and increases the reliability of the maintenance process.

Equation should be presented by using the Microsoft equation of Cambria Math (or MathType) font 10 in justify format.

$$\boldsymbol{X_{best}}(t) = \min_{i \in 1,\ldots,n} fit_i(\boldsymbol{X}(t)) \tag{5}$$

4. CONCLUSION

This study presents a deep learning-based framework for automated detection of structural defects in aircraft using the Mask R-CNN architecture. Unlike traditional manual inspections that are time-consuming and prone to human error, the proposed system enables precise instance segmentation of cracks and dents from aircraft surface images. The integration of Feature Pyramid Networks and RoI Align within Mask R-CNN ensures high accuracy across multiple defect scales and shapes, making it suitable for practical deployment in aviation safety protocols. Beyond performance metrics, the system's ability to produce pixel-level defect masks offers substantial value for automated documentation, quantification, and prioritization of maintenance tasks. Looking forward, future work will explore the adoption of

Transformer-based vision architectures such as the Swin Transformer, which may offer enhanced global context awareness and better generalization across varied aircraft types. Additionally, real-world testing across multiple aircraft platforms and structural materials will be essential to ensure robustness. Another promising direction is the integration of edge computing capabilities to enable real-time inference onboard UAVs, paving the way for autonomous in-field inspections in both military and commercial aviation settings.

ACKNOWLEDGEMENT

Author gratefully acknowledge Air Force Institute of Technology and Kaduna State University, Kaduna, Nigeria for providing the necessary resources and facilities.

REFERENCES

[1] C. V. Oster, J. S. Strong, and C. K. Zorn, "Analyzing aviation safety: Problems, challenges, opportunities," *Research in Transportation Economics*, vol. 43, no. 1, pp. 148–164, 2013.

[2] I. Savage, "Comparing the fatality risks in United States transportation across modes and over time," *Research in Transportation Economics*, vol. 43, no. 1, pp. 9–22, 2013.

[3] A. Barnett, "Aviation safety: A whole new world?" *Transportation Science*, vol. 54, no. 1, pp. 84–96, 2020.

[4] N. E. Andenyangtso, R. D. Enyinna, and M. U. Bonet, "Effective Maintenance Of Aircraft Antiskid Brake System," *Mekatronika: Journal of Intelligent Manufacturing and Mechatronics*, vol. 6, no. 1, pp. 19–29, 2024.

[5] Y. Li, Z. Han, H. Xu, L. Liu, X. Li, and K. Zhang, "Applied sciences YOLOv3-Lite : A lightweight crack detection network for aircraft structure based on depthwise separable convolutions," Applied Sciences, vol. 9, no. 18, p. 3781, 2019.

[6] B. F. Spencer, V. Hoskere, and Y. Narazaki, "Advances in computer vision-based civil infrastructure inspection and monitoring," *Engineering*, vol. 5, no. 2, pp. 199–222, 2019.

[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," *Computers in Industry*, vol. 141, p. 103695, 2022.

[8] W. Chen and S. Huang, "Human reliability analysis for visual inspection in aviation maintenance by a Bayesian network approach," *Transportation Research Record*, vol. 2449, pp. 105–113, 2014.

[9] C. G. Drury and A. K. Gramopadhye, "Human factors in aviation maintenance: how we got to where we are," *International Journal of Industrial Ergonomics*, vol. 26, no. 2, pp. 125–151, 2000.

[10] W. Wang and C. Su, "Semi-supervised semantic segmentation network for surface crack detection," *Automation in Construction*, vol. 128, p. 103786, 2021.

[11] S. Teng, Z. Liu, G. Chen, and L. Cheng, "Concrete crack detection based on well-known feature extractor model and the YOLO_v2 network," *Applied Sciences*, vol. 11, no. 2, pp. 1–13, 2021.

[12] D. Medak, L. Posilovic, M. Subasic, M. Budimir, and S. Loncaric, "Automated defect detection from ultrasonic images using deep learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 10, pp. 3126–3134, 2021.

[13] I. Abdel-Qader, O. Abudayyeh, and M. E. Kelly, "Analysis of edge-detection techniques for crack identification in bridges," *Journal of Computing in Civil Engineering*, vol. 17, no. 4, pp. 255–263, 2003.

[14] D. Lattanzi and G. R. Miller, "Robust automated concrete damage detection algorithms for field applications," *J. Journal of Computing in Civil Engineering*, vol. 28, no. 2, pp. 253–262, 2014.

[15] Y. Tomoyuki and H. Shuji, "Improved percolation-based method for crack detection in concrete surface images," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, 2008.

[16] H. Kim, E. Ahn, S. Cho, M. Shin, and S. H. Sim, "Comparative analysis of image binarization methods for crack identification in concrete structures," *Cement and Concrete Research*, vol. 99, pp. 53–61, 2017.

[17] Z. Shen, X. Wan, F. Ye, X. Guan, and S. Liu, "Deep learning based framework for automatic damage detection in aircraft engine borescope inspection," in *2019 International Conference on Computing, Networking and Communications*, pp. 1005–1010, 2019.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research.*, vol. 15, pp. 1929–1958, 2014.

[19] S. Wu and J. I. E. Fang, "Sample and structure-guided network for road crack detection," *IEEE Access*, vol. 7, pp. 130032–130043, 2019.

[20] X. Tang, "DeepID3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[22] C. Wang, X. Wang, X. Zhou, and Z. Li, "The aircraft skin crack inspection based on different-source sensors and support vector machines," *Journal of Nondestructive Evaluation*, vol. 35, no. 3, p. 46, 2016.

[23] T. George, "Crack analysis of aircraft slat and flap sections using NDT techniques," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3, no. 5, pp. 48–51, 2016.

[24] T. Malekzadeh, M. Abdollahzadeh, H. Nejati, and N.-M. Cheung, "Aircraft fuselage defect detection using deep neural networks," *arXiv preprint arXiv:1712.09213*, 2017.

[25] L. Yang, B. Li, W. Li, Z. Liu, G. Yang, and J. Xiao, "Deep concrete inspection using unmanned aerial vehicle towards CSSC database," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 24-28, 2017.

[26] K. Gopalakrishnan, H. Gholami, A. Vidyadharan, and A. Agrawal, "Crack damage detection in unmanned aerial vehicle images of civil infrastructure using pre-trained deep learning model," *International Journal for Traffic and Transport Engineering*, vol. 8, no. 1, pp. 1–14, 2018.

[27] S. Bouarfa, A. Doğru, R. Arizar, R. Aydoğan, and J. Serafico, "Towards automated aircraft maintenance inspection. A use case of detecting aircraft dents using mask r-cnn," *AIAA Scitech 2020 Forum*, p. 0389, 2020.

[28] A. Doğru, S. Bouarfa, R. Arizar, and R. Aydoğan, "Using convolutional neural networks to automate aircraft maintenance visual inspection," *Aerospace*, vol. 7, no. 12, pp. 1–22, 2020.

[29] H. Huang, S. Zhao, D. Zhang, and J. Chen, "Deep learning-based instance segmentation of cracks from shield tunnel lining images," *Structure and Infrastructure Engineering*, vol. 18, no. 2, pp. 183–196, 2022.

[30] R. Vorobev, I. Vasilev, and I. Kremnev, "Tomography of materials and structures segmentation of structural defects in polymer composite computed tomography images with deep learning models," *Tomography of Materials and Structures*, vol. 3, p. 100014, 2023.

[31] D. Denhof, B. Staar, M. Lütjen, and M. Freitag, "Automatic optical surface inspection of wind turbine rotor blades using convolutional neural networks," *Procedia CIRP*, vol. 81, pp. 1166–1170, 2023.

[32] X. Xu, M. Zhao, P. Shi, R. Ren, X. He, X. Wei, et al., "Crack detection and comparison study based on faster R-CNN and mask R-CNN," *Sensors*, vol. 22, no. 3, p. 1215, 2022.

[33] H. Fujita, M. Itagaki, K. Ichikawa, Y. K. Hooi, K. Kawano, and R. Yamamoto, "Fine-tuned pre-trained mask R-CNN models for surface object detection," *arXiv preprint arXiv:2010.11464*, 2020.

[34] L. Jiao *et al.*, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.

[35] K. Zhang, H. D. Cheng, B. Zhang, and D. Ph, "Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning," *Journal of Computing in Civil Engineering*, vol. 32, no. 2012, pp. 1–12, 2018.

[36] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[37] A. N. Azhar and M. L. Khodra, "Fine-tuning pretrained multilingual BERT model for Indonesian aspect-based sentiment analysis," in *2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications*, pp. 2980–2988, 2020.

[38] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2020.

[39] Z. Zhou, M. Zhang, J. Chen, and X. Wu, "Detection and classification of multi-magnetic targets using mask-RCNN," *IEEE Access*, vol. 8, pp. 187202–187207, 2020.

[40] L. Zhang, Z. Wang, L. Wang, Z. Zhang, X. Chen, and L. Meng, "Machine learning-based real-time visible fatigue crack growth detection," *Digital Communications and Networks*, vol. 7, no. 4, pp. 551–558, 2021.

[41] M. M. Rosso, G. Marasco, S. Aiello, A. Aloisio, B. Chiaia, and G. C. Marano, "Convolutional networks and transformers for intelligent road tunnel investigations," *Computers & Structures*, vol. 275, p. 106918, 2023.

[42] O. Yaman, F. Yol, and A. Altinors, "A fault detection method based on embedded feature extraction and SVM classification for UAV motors," *Microprocessors and Microsystems*, vol. 94, p. 104683, 2022.

[43] N. Hütten, M. A. Gomes, T. Meisen, F. Hölken, K. Andricevic, and R. Meyes, "Deep learning for automated visual inspection in manufacturing and maintenance: A survey of open-access papers," *Applied System Innovation*, vol. 7, pp. 1–38, 2024.

[44] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN : Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 1–14, 2016.

[45] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," *European Conference on Computer Vision*, pp. 213-229, 2020.