**Abstract**

Wind turbines are at the heart of sustainable electricity generation from renewable sources. These blades, however, are constantly exposed to a myriad of environmental conditions, rendering them prone to wear and tear. This vulnerability underscores the importance of consistent monitoring not just to reduce risks, maintenance costs, and downtime but also to safeguard against structural damages that could compromise the constant supply of power. Considering this, our research delves into the intricate task of detection and segmentation of prevalent faults in turbine blades, specifically surface damage, and edge erosion. Furthermore, we've extended our investigation to two crucial components: the lightning receptor and the VG panel. To ensure the highest efficiency possible, we conducted a comparative analysis of three state-of-the-art models: Mask R-CNN, YOLOv7, and YOLOv8. We also researched the YOLOv8 model with three different Intersections of Unions, and they are IoU@0.6, IoU@0.7, and IoU@0.8.

**Introduction**

The primary driver of greenhouse gas emissions stems from fossil fuel-based energy sources. Renewable energy sources must be incorporated into our daily lives to reduce emissions and address environmental problems. The Paris Agreement, signed by 165 nations in 2015, marked a significant milestone in fortifying the global commitment to curbing carbon dioxide emissions and bolstering investments in renewable energy solutions [1]. Following the agreement, there was a significant increase in expenditure for green energy infrastructure. Notably, solar and wind energy emerged as front-runners in the competition for renewable energy production. Solar power provides 9% and wind production 24%, respectively, to the nation's energy supply, according to the United States Energy Information Administration [2]. Advancements in technology have revolutionized the design and production of wind turbines, resulting in a reduction in the cost of power generation compared to other renewable sources. The efficiency of a wind turbine is nearly double that of solar panels [3]. As of now, wind energy stands out as one of the most cost-effective renewable resources, with production costs averaging between one and two cents per kilowatt-hour [4]. Moreover, it is worth noting that harnessing one megawatt of wind energy can potentially offset approximately 2600 tons of carbon dioxide emissions [5].

The wind turbines are built on land (onshore) or in large bodies of water like oceans and lakes (offshore). Irrespective of the categories, the wind turbine architecture consists of many mechanical and electrical components such as a rotor, blades, generator, controller, and gearbox. Notably, turbine blades, typically measuring an impressive 100-140 meters in length, play a crucial role in electricity generation. Its efficiency in rotating determines the amount of electric power generated. These gigantic blades are subjected to aerodynamic and gravitational loads while operating under extreme climatic conditions, which causes vibration forces that result in structural damages such as cracks on the surface, erosion of the edge, pitch angle, and twisting blades [ 4] [ 5].

These forces can induce vibrations that manifest as structural damage, such as surface cracks, edge erosion, and even issues with pitch angles or twisted blades. Traditional inspection of these blades relies on time-based maintenance techniques. This approach is not only inefficient but also fraught with challenges. For instance, rope-based inspections pose significant risks to personnel, and telephotography methods can overlook microscopic structural damages due to human limitations. Furthermore, faults in wind turbines don't just compromise efficiency; they also pose environmental hazards. Traditional inspection methods, relying on visual or manual techniques, are often time-consuming and are becoming increasingly impractical with the growing size and complexity of wind farms and prone to human error [27].

As wind turbine technology advances, so does the need for robust fault detection and segmentation methods to ensure optimal performance and mitigate potential failures. It is necessary for the adoption of deep-learning approaches for efficient and precise fault detection and segmentation. Mask R-CNN, YOLOv7, and YOLOv8 are advanced object detection models capable of both bounding box detection and instance segmentation. So, this research addresses these challenges by proposing an autonomous detection system using Mask R CNN, YOLOv7, and YOLOv8 algorithms. The goal is to precisely identify faults, thereby reducing O&M costs while elevating safety standards. Our approach leans on cutting-edge object detection algorithms, namely Mask R-CNN, YOLOV7, and YOLOV8, to navigate the challenges of manual inspections. First of all, let's start with contributions and data acquisition.

**Contributions**

1. Data acquisition was the major contribution, where we gathered all the images of turbine blades using drones at different backgrounds.
2. Initially, The MaskRCNN algorithm was applied (single-stage detector) to wind turbine blades, and then YOLOv7 (multistage detector)
3. Finally, YOLOv8 (multi-stage detector) was applied with variable Intersection of Union thresholds to analyze model accuracy and efficiency.

**Data Acquisition**

In this research, we developed a dataset containing 2,127 augmented drone images. These images capture four distinct categories: edge erosion, surface damage, VG panel, and lighting receptor. These images encompass various backgrounds and fault variations. It's worth noting that while VG panels and lighting receptors aren't specific fault types, they are external components typically visible during wind turbine blade inspections [7]. We systematically divided the dataset into training, validation, and test subsets, consisting of 2,127, 181, and 41 images. Addressing concerns of underfitting and overfitting, we harnessed data augmentation strategies like flipping, rotating, shearing, blurring, and cropping. Each image underwent precise annotation, employing bounding boxes and labels to spotlight specific faults in wind turbine blades using the Roboflow Annotator tool, a publicly available resource. Though initially set to a pixel dimension of 856 x 856, images are resized to a 640 x 640 resolution for both the training and testing phases. Elevating the models' effectiveness, we incorporated transfer learning techniques, resulting in a notable enhancement in model performance.

**Transfer learning**

Transfer learning is used to leverage pre-trained models, reducing the need for extensive data and computational resources. Recognizing the challenge of amassing several labeled datasets, we adopted the transfer learning method. Specifically, we utilize a pre-trained model from the Microsoft Common Objects in Context (MS COCO) dataset [8]. Additionally, our study includes an analysis of the performance of transfer learning, with experiments involving the freezing of layers in the initial stages of the Mask R-CNN, YOLOv7, and YOLOv8 models. In the next section, we will delve into the algorithms utilized in this project.

**Algorithms and methodology:**

In this study, we leverage three cutting-edge object detection algorithms—Mask R-CNN, YOLOv7, and YOLOv8—to effectively detect and segment wind turbine faults. Our approach begins with the mAP-focused segmentation tasks handled by Mask R-CNN. Subsequently, we transition to the deployment of YOLOv7 and YOLOv8 for robust and real-time fault detection. The overarching goal is to achieve a harmonious balance between accuracy and efficiency in identifying defects in turbine blades. Additionally, our investigation delves into the YOLOv8 model, exploring its algorithms with different Intersection of Union parameters, further enhancing our understanding of its capabilities. First, we chose the Mask R-CNN algorithm because of its good performing capabilities.

**Mask R-CNN**

Mask R-CNN is a state-of-the-art instance segmentation model that builds on top of Faster R-CNN. It is a two-stage framework, where the first stage proposes regions of interest (RoI), and the second stage performs classification, bounding box regression, and instance segmentation on each RoI. The backbone of Mask R-CNN is Res Net, which handles the vanishing gradient problem in deep networks by introducing skip or residual connections, and it is integrated with a Feature Pyramid Network (FPN) to improve model detection accuracy and training time. Feature Pyramid Network is a top-down architecture with lateral connections developed to extract and build high-level semantic feature maps at different spatial resolutions by a bottom-up pathway, a top-down pathway, and lateral connections [19]. The feature maps produced by the Feature Pyramid Network (FPN) serve as input to the Region Proposal Network (RPN). Utilizing the concept of anchors, the RPN generates region proposals at various scales and aspect ratios for objects in the image. RPN processes all feature maps', and extracts RoI (Region of Interest) features from different sizes of the feature pyramids based on the size of the specific fault type. It operates like a sliding window and efficiently identifies areas containing objects in parallel due to its convolution operation. RoI Align is a critical layer in implementing the Mask R-CNN algorithm. It is responsible for extracting "M×M" feature maps from each RoI and unifying the output size of each RPN. Unlike RoI Pool, RoI Align eliminates aggressive quantization, significantly improving location accuracy. The RoI is divided into 2×2 sub-windows or bins, and bi-linear interpolation is used to interpolate values of generated features within each RoI bin. The layer then aggregates the results using the max operation [10]. The loss function in Mask R-CNN is a combination of RoI loss, classifier loss, and mask head loss. Classifier loss Lcls (Pi, Pi\*) = -lb [PiPi\* + (1- Pi) (1- Pi\*)] Where pi represents the predicted probability that anchor point i is the target; p i represents the predicted value of the corresponding real area label; lb represents the log loss function [23]. Lbox (ti,ti\*)= R(ti- ti\*), ti signifies the four parameterized coordinate vectors of the predicted frame; t i signifies the coordinate vector corresponding to the border of the real area [24]. RoI as Lall=Lcls + Lbox + Lmask, where Lcls, Lbox, and Lmask represent the classification loss, bounding box loss, and the average binary cross-entropy loss respectively [22]. The total architecture is shown below in Figure 1. Next, we chose YOLOv7 because of its speed and efficiency.

**YOLOv7**

YOLOv7 introduces several innovative features to enhance detection accuracy. One standout is its multi-head architecture, which incorporates a unique approach. Additionally, it implements the E-ELAN network, surpassing its predecessors in efficiency by focusing on layer aggregation. This network consists of residual blocks, each housing expands, shuffles, and merges cardinality operations. In the design of YOLOv7, a series of actions, including expansion, channel mixing, and merging, work in harmony to significantly enhance the network's learning capacity. The integration of the Bidirectional Feature Pyramid Network is particularly noteworthy. This design aims to optimize the flow of information throughout the network, significantly improving object detection accuracy. Building on this, the Path Aggregate Network enhances the structure by strengthening connections between feature pyramid levels. As a result, the features at each level of the feature pyramid combined to increase their degree of information [20]. It introduces a dual-head system for detection. The auxiliary head, being the first, is pivotal during training phases in intermediate layers, as it predicts coarse bounding boxes and class probabilities. Then, the lead head focuses on final detection outcomes, forecasting precise bounding boxes and class probabilities. Later in the process, refinement comes through a deliberate model re-parameterization strategy. Especially the identity connection is omitted in RepConv. This strategic design enhances gradient diversity for different feature maps, optimizing the network structure. Another notable aspect of YOLOv7 is its dynamic feature, which reviews model outcomes and ground truths, assigning soft labels to detected objects. These soft labels, derived from ground truth, contribute to fine-tuning predictions. Lastly, YOLOv7 collaborates with YOLACT, known for its rapid instance segmentation capabilities. While YOLACT autonomously handles instance segmentation tasks on feature maps without requiring bounding box regression, it's important to note that its accuracy slightly lags behind multi-stage algorithms like Mask R-CNN. The combination of YOLOv7 and YOLAC makes YOLOv7 the most precise and fastest instance segmentation algorithm for Wind turbine blade fault detection and instance segmentation. The loss function of Yolov7 is 𝐿𝑡𝑜𝑡𝑎𝑙\_𝑙𝑜𝑠𝑠=𝐿𝑜𝑏𝑗\_𝑙𝑜𝑠𝑠+𝐿𝑏𝑜𝑥\_𝑙𝑜𝑠𝑠+𝐿𝑐𝑙𝑠\_𝑙𝑜𝑠𝑠. BCE cross-entropy loss formula is represented as Ln=−wn[yn\*logσ(xn)+(1−yn)\*log(1−σ(xn))] and σ(xn)=1/(1+(e)^-x where σ(xn) represents the sigmoid function, wn signifies the average of the results, and yn signifies the real sample label. Intersection of Union (𝐼𝑜𝑈)=(|𝐴∩𝐵|) / (|𝐴∪𝐵|). YOLOV7 uses CIoU loss. In the below Figure 5 loss function formulae of YOLOv7 are mentioned. Finally, we chose YOLOv8 as the latest version of the YOLO family whose accuracy and speed are higher than previous versions.

**YOLOv8**

YOLOv8 is the latest version of the YOLO family. The architecture of YOLOv8 is state-of-the-art, capable of accommodating different resolutions, and even includes an instance segmentation model reminiscent of YOLACT. It is also anchor-free and a constituent of the Feature Pyramid Network (FPN) Path Aggregation Network (PAN), and SPPF modules. These enhance the ability to discern object shapes and textures across various scales and lead to improved accuracy. The backbone of the model is composed of four sections, each introduced by a single convolution and succeeded by an innovative C2f module [12]. This new addition to CSPDarknet53 involves a bottleneck segment featuring two 3x3 convolutions with residual connections. The model features a revamped detection head that separates classification and detection tasks. The loss computations have been refined, utilizing BCE Loss for classification and a combination of Complete Intersection of Union (CIOU) Loss + Distributed Focal Loss (DFL) for regression. The DFL approach models the box position as a general distribution, while VFL introduces an asymmetric weighting operation. These enhancements contribute to the robust performance of YOLOv8 in object detection tasks. [17][18]. Below Figure 3 is YOLOv8 architecture. Below Figure 4 are equations of DFL [26].

**Our research and results**

We applied advanced detection and segmentation algorithms—Mask R-CNN, YOLOv7, and YOLOv8—to wind turbine blades, achieving noteworthy results. Specifically, Mask R-CNN exhibited a detection mAP@IoU (0.5) of 86.30% and a segmentation mAP@mask (0.5) of 84.56%. YOLOv7 outperformed with a detection mAP@IoU (0.5) of 95.80% and a segmentation mAP@mask (0.5) of 96.30%. Subsequently, our research extended to the implementation of the YOLOv8 algorithm, introducing the Intersection of Union (IoU) tuning at thresholds of 0.6, 0.7, and 0.8. This comprehensive exploration showcases the nuanced performance of these algorithms in the realm of wind turbine fault detection and segmentation.

**Conclusion**

This research successfully employed deep learning models for automated detection and segmentation of wind turbine blade faults, paving the way for improved wind farm inspections. YOLOv8 surpassed Mask R-CNN and YOLOv7 in accuracy and efficiency, achieving an impressive mAP@IoU of 97.3% for detection and 97.6% for segmentation. YOLOv8 demonstrated robustness to different IoU thresholds, suggesting its adaptability to diverse detection requirements. Integrating these models into real-world systems can significantly reduce maintenance costs, downtime, and safety risks associated with manual inspections. Further research could explore incorporating environmental factors and real-time data streams to enhance model accuracy and robustness in various operating conditions. Overall, this study highlights the effectiveness of deep learning algorithms for automated wind turbine blade fault detection and segmentation, contributing to a more efficient and sustainable wind energy future. The proposed research potentially revolutionized wind turbine maintenance by enabling automated, accurate, and cost-effective fault identification.