**Detection and segmentation of wind turbine blades faults using Mask R-CNN, YOLOV7, And YOLOV8 with different Intersection of union**

**Pavan Sai Prasanth Sabnaveesu and Dr. Md Monirul Islam**

**Keywords**: Object detection, semantic segmentation, wind turbine blades faults, Mask R-CNN, YOLOV7, And YOLOV8

**Abstract**

Wind turbines play a critical role in the sustainable generation of electricity from renewable sources. The stability, safety, and reliability of blades face many challenges. However, these structures are exposed to various environmental conditions, and their components are susceptible to damage. To prevent structural damage, ensure the desired level of power demand, and reduce unscheduled downtime, risk, maintenance cost, and time, it is necessary to monitor the turbine blades consistently. This research helps to investigate and analyze the damage to wind turbine blades. Utilized drone images for data acquisition and two types of faults and two parts such as lighting receptor, VG panel, surface damage, and edge erosion are used in prediction. Literature review of Mask R-CNN, YOLOV7, and YOLOV8 Achieved a mAP score of 97.4% for detection and 94.7% with IoU@o.5 for segmentation with YOLOV8 algorithms.

**Introduction**

The primary cause of greenhouse gas emissions is fossil fuel-based energy sources. Sustainable integration of renewable sources is crucial for addressing environmental issues and reducing emissions. The Paris Agreement, which 165 nations signed in 2015, was a major step in strengthening the global response to reduce CO2 emissions and increase investment in renewable energy sources [1]. After the agreement, there was a substantial increase in the desire for investments in green energy infrastructure. Solar and wind energy are leading the charge in the competition for renewable energy production. The U.S. Energy Information Administration estimates that 9% of solar electricity and 24% of wind generation are used to meet the country's energy needs [2]. Technology advancements in the design and production of wind turbines have reduced the cost of power generation compared to other renewable sources. The efficiency of a wind turbine is almost two times greater than that of solar panels [3]. Wind energy, which costs around 1-2 cents per kilowatt-hour [4], is the most affordable renewable resource after production tax incentives. 1 Mega Watt of wind energy can reduce carbon dioxide emissions by about 2600 tons [5].

The wind turbine can be built on land (onshore) or in large bodies of water like oceans and lakes (offshore). Irrespective of the categories, the wind turbine architecture is a combination of many mechanical and electrical components such as a rotor, blades, generator, controller, and gearbox. The blade of the turbine is an integral part of the architecture of the wind turbine. The part is required to rotate at high speed and based on the effectiveness of the rotation, the generation of the electric power will vary. Usually, the length of the blades is around 100-140 meters, which is gigantic in size. 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].

Wind turbine blades are often inspected through a time-based maintenance technique, which is anomalous, costly, and dangerous for inspection. Rope-based inspection can be extremely dangerous for the maintenance staff, and using telephotography is ineffective because microscopic erosion and structural damage are sometimes difficult to notice by the human eye. The faults in a wind turbine can lead to hazards for the operating environment. As the number of maintenance requirements is quite high, the operation and maintenance (O&M) cost is significant. Up to 30% of the cost of producing wind electricity goes into operating and maintaining wind turbines[6]. Therefore, labor-intensive maintenance is unreliable, expensive, and risky. To make the operation safe, it is required to develop a method to monitor the turbine blades consistently with minimal human intervention. Therefore, the thesis aims to develop an autonomous wind turbine fault detection method to identify the type of fault and the position of the fault to lessen the O&M cost of wind turbines significantly while ensuring a safe monitoring process. To surmount risky and expensive challenges we used objection detection algorithms such as Mask R-CNN, YOLOV7, and YOLOV8[7]. First step explains about source of data.

**Data Acquisition**

The dataset used in this research includes 2,127 drone images (with augmentation) that can be classified into four different fault classes: edge erosion, surface damage, VG panel, and lighting receptor. The images are collected with different backgrounds as well as fault variations that describe the major four types of faults in wind turbine blades. VG panel and lighting receptor are not any specific fault type in wind turbine blades, but external components on wind turbine blades that are generally visualized during wind turbine blades inspection [7]. To train the models, the images are divided into training validation, and test data sets, which contain 2,127, 181, and 41 images, respectively. Applied data augmentation techniques such as Flipping, Rotating, Shear, blur, crop which prevents under-fitting and over-fitting. Image annotator is to draw bounding boxes and labels and it is done using the publicly available Roboflow Annotator tool for the fault of the wind turbine blades. The images are resized, and the pixel size of the images is 856 x 856. Finally, I used image size as 640\*640 resolution while training and testing. In our algorithms we further added transfer learning that enhanced our model performance.

**Transfer learning**

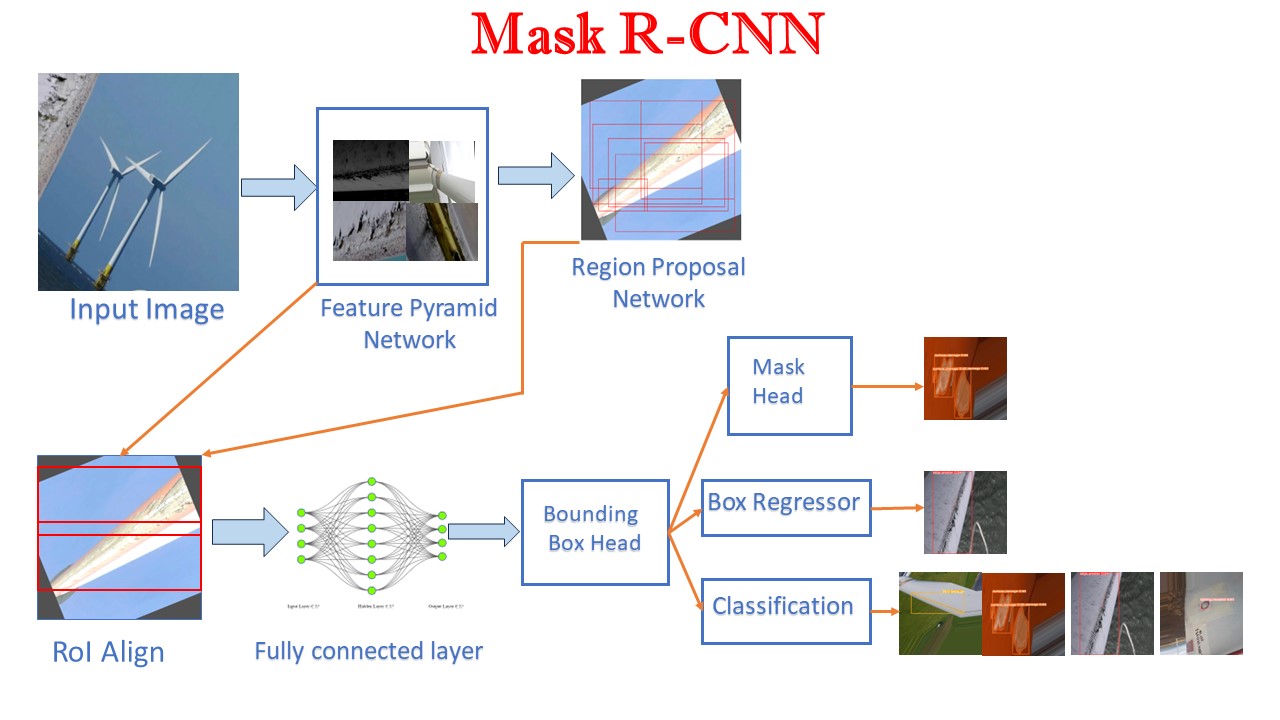
This research uses the transfer learning technique to train the model. To mitigate the precondition of a significant number of labeled data sets, the transfer learning method is applied, a pre-trained model of Microsoft Common Object in Context (MS COCO) dataset [8] is used. To analyze the performance of transfer learning, this research also conducted experiments to freeze the number of layers in the initial stage of the Mask R-CNN, YOLOV7 and YOLOV8. Next. We will dive algorithms used in project.

**Algorithms and methodology:**

The Mask R-CNN algorithm segments the image by precise features by different classes. Masked R-CNN involves a multi-step process. In the first phase, the images are given to the model, and they extract important features from the image and generate a proposal network. In the second stage, it filters and classifies the suggestions to generate bounding boxes and masks for defects [Figure 1].

**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. Backbone of Mask R-CNN is Res Net which handle the vanishing gradient problem in very deep networks by introducing skip or residual connections and it is integrated with a feature pyramid network, Feature Pyramid Network **(**FPN), to improve model detection accuracy and training time. Feature Pyramid Network is a top-down architecture with lateral connections is developed to extract and build high-level semantic feature maps at different spatial resolutions by 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 moving 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 consider interpolated values of generated features within each RoI bin. The layer then aggregates the results using the max operation [10]. The predicted bounding box and mask corresponding to the highest class score is the final prediction for each region. The total architecture is shown in the below figure 1

****

**Figure 1**

A diagram of a diagram

Description automatically generated

Figure 3

YOLOv7 (CSPDarknet53 with E-ELAN, CSP, PAN, BiFPN )

YOLOV7 introduced the multi-head concept in architectu­re that helps the model to improve detection accuracy. YOLOV7 offers an E-ELAN network for layer aggregation efficiently to the previous version.Each residual block in the E-ELAN architecture consists of a mix of expand, shuffle, and merge cardinality operations. Expanding the number of channels in the feature map, mixing the channels in the feature map, and merging the channels in the feature map according to their cardinality all boost the network's capacity for learning.

YOLOV7 is based on a planned model re-parameterization strategy. According to this strategy, RepConv should not contain an identity connection [11]. Usually, RepConv is constructed with a layer of residual layers or concatenation connections, followed by flattening and dense layers. RepConvN, a version of RepConv without identity connections, can be utilized [11]. Figure 4 illustrates the appropriate combination of Conv for the YOLOV7 model.

A diagram of a diagram

Description automatically generated with medium confidence

Figure 4

The lead head of the model oversees the final detection result whereas the auxiliary head helps the model during training time in the middle layers [11]. For the former, the author claims that accessing the cascade of ResNet [13] would give greater gradient variety for distinct characteristic graphs, thereby damaging the network structure since convolution and diverse networks, the identity connection in RepConv [14] is removed and built the intended reparametrized convolution [11]. YOLOv7 also proposed a dynamic that investigates model results and ground truths. It also assigns the soft label to the detected object. YOLOv7 architecture maps soft labels' that account for ground truth which helps to find better predictions.

YOLOv7 integrates with YOLACT [15] to perform instance segmentation. YOLACT performs instance segmentation tasks on feature maps independently of bounding box regression which makes it the fastest instance segmentation algorithm, though the detection accuracy of YOLACT is lower than multi-stage instance segmentation algorithms such as Mask R-CNN [15]. 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.

**YOLOv8 (**CSPDarknet53 with ShuffleNetV2 and ResBlockV2**,** C2f, SPPF**)**

YOLOv8 is the newest iteration of the YOLO object detection model. It is Anchor free. It combines both Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). This combination enhances its ability to recognize object shapes and textures across varied scales, leading to higher accuracy. The introduction of a state-of-the-art model accommodating different resolutions and even an instance segmentation model reminiscent of YOLACT. The backbone of YOLOv8 consists of four sections, each prefaced by a single convolution and followed by a c2f module, which is an innovative addition to CSPDarknet53. This module involves a bottleneck segment featuring two 3x3 convolutions with residual connections.

YOLOv8 has several key strengths:

1. A lightweight network architecture.
2. Efficient feature fusion techniques.
3. Enhanced detection accuracy by combining features from several real-time object detectors.

It still leverages the CSP idea from YOLOv5 and incorporates feature fusion (PAN-FPN) and SPPF modules. The main advancements in YOLOv8 include b) While retaining the c2f module concept, its design was influenced by the ELAN structure in YOLOv7. c) A revamped detection head, which separates classification and detection. d) Improved loss computations, utilizing BCE Loss for classification and a combination of CIOU Loss + DFL for regression. The DFL approach models the box position as a general distribution, with VFL introducing an asymmetric weighting operation [17][18].

A screenshot of a computer

Description automatically generated

**Our research and results**

Applied Mask R-CNN, YOLOV7 algorithms on wind turbine blades for detection and segmentation. Achieved mAP@IoU (0.5) 86.30% for detection and mAP@mask (0.5) for segmentation is 84.56% with Mask R-CNN. With mAP@IoU (0.5) YOLOv7 accomplished 95.80% for detection, 96.30% mAP@mask (0.5) for segmentation. Researched the project by changing Intersection of Union [IoU] to 0.4, 0.5, and 0.6 using YOLOV8 algorithm.

**Sample Images:**

A close-up of a white surface

Description automatically generated A person wearing a harness

Description automatically generated

Table1. Performance comparison between Mask R-CNN. YOLOv7, and YOLOV8

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Performance Metrics** | **Mask R-CNN** | **YOLOv7** | **YOLOV8** | **Improvement** |
| mAP@IoU(0.5) | 86.30% | 95.80% | 98.4% | +2.60% |
| mAP@mask(0.5) | 84.56% | 96.30% | 96.3% | 0% |

|  |  |  |
| --- | --- | --- |
| **Performance of different IoU’s** | **Detection** | **Segmentation** |
| mAP@IoU@default | 98.4% | 95.3% |
| mAP@mask(0.6) | 94%/94.9% | 95.3% |
| mAP@IoU(0.4) | 94.9% | 96.3% |

**Conclusion**

**References**

1. Seo, S. N. (2017). Beyond the Paris Agreement: Climate change policy negotiations and future directions. Regional Science Policy & Practice, 9(2), 121-140.
2. (June 10, 2022). U.S. Energy Information Administration - EIA - Independent Statistics and Analysis. U.S. energy facts explained - consumption and production - U.S. Energy Information Administration (EIA). https://www.eia.gov/energyexplained/us-energy-facts/
3. Wiser, R., Bolinger, M., Hoen, B., Millstein, D., Rand, J., Barbose, G., and Paulos, B. (2022). Land-based wind market report: 2022 edition. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).
4. Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., and Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. Renewable Energy, 133, 620-635.
5. Kusiak, A., and Li, W. (2011). The prediction and diagnosis of wind turbine faults. Renewable Energy, 36(1), 16-23.
6. Nielsen, J. J., and Sørensen, J. D. (2011). On risk-based operation and maintenance of offshore wind turbine components. Reliability Engineering & System Safety, 96(1), 218-229.
7. Shihavuddin, A. S. M., Xiao Chen, X., Fedorov, V., Nymark Christensen, A., Andre Brogaard Riis, N., Branner, K., and Reinhold Paulsen, R. (2019). Wind turbine surface damage detection by deep learning aided drone inspection analysis. Energies, 12(4), 676.
8. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., and Ramanan, D., and Zitnick, CL (2014, September). Microsoft COCO: Common objects in context. In European Conference on Computer Vision (pp. 740-755).
9. Roboflow Annotator. Roboflow. <https://roboflow.com/>.
10. Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick Facebook AI Research (FAIR) Mask R-CNN
11. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors Chien-Yao Wang1, Alexey Bochkovskiy, and Hong-Yuan Mark Liao1 1 Institute of Information Science, Academia Sinica, Taiwan
12. Real-Time Flying Object Detection with YOLOv8 Dillon Reis\*, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi Georgia Institute of Technology
13. Fu, Y., Wu, J., Hu, Y., Xing, M., and Xie, L. (2021, January). Desnet: A multi-channel network for simultaneous speech dereverberation, enhancement and separation. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 857-864). IEEE.
14. Soudy, M., Afify, Y., and Badr, N. (2022). RepConv: A novel architecture for image scene classification on Intel scenes dataset. *International Journal of Intelligent Computing and Information Sciences*, *22*(2), 63-73.
15. Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). Yolact: Real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9157-9166).
16. Jacob Solawetz and Francesco. What is yolov8? the ultimate guide., 2023. 04-30-2023. 1, 5, 8
17. DC-YOLOv8: Small size Object detection algorithm based on camera sensor Haitong Lou1, Xuehu Duan1, Junmei Guo1, Haiying Liu1 \*, Jason Gu2, Lingyun Bi1, Haonan Chen 1 7 April 2023 doi:10.20944/preprints202304. 0124.v1
18. DCF-Yolov8: An Improved Algorithm for Aggregating Low-Level Features to Detect Agricultural Pests and Diseases by Lijuan Zhang 1,2, Gongcheng Ding 1,2, Chaoran Li 3 and Dongming Li 1,\*ORCIDAgronomy 2023, 13(8), 2012; https://doi.org/10.3390/agronomy13082012.
19. Feature Pyramid Networks for Object Detection Tsung-Yi Lin, Piotr Dollar , Ross Girshick , Kaiming He, Bharath Hariharan, and Serge Belongie