2022-06-26

Drone Footage Wind Turbine Surface Damage Detection

Foster, A

10.1109/ivmsp54334.2022.9816220
2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)
IEEE

2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP) | 978-1-6654-7822-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IVMSP54334.2022.9816220

Drone Footage Wind Turbine Surface Damage Detection

Ashley Foster, Oscar Best, Mario Gianni, Asiya Khan, Keri Collins, Sanjay Sharma
School of Engineering, Computing and Mathematics (SeCAM)
University of Plymouth, UK
Email: ajifoster3@gmail.com, {oscar.best, mario.gianni, asyia.khan, kery.collins, sanjay.sharma}@plymouth.ac.uk

***Abstract*—In this work a new publicly available dataset of wind turbine surface damage images is presented. Moreover, a comparison between ResNet-101 Faster R-CNN and YOLOv5 for Wind Turbine Surface Damage Detection is analysed and performance of these models on drone footage with active turbines is also discussed. Results show that YOLOv5 outperforms ResNet-101 Faster R-CNN in predicting the bounding box coordinates of the damaged surfaces of the wind turbines. However, unlike YOLOv5, ResNet-101 Faster R-CNN estimates an entire instance of damage as a single prediction.**

I. Introduction

Offshore Wind Turbine inspection is a high risk and expensive operation and alternative approaches to reduce both the human exposure to dangerous situations [1], [2], [3], [4], [5] and the life-cycle cost of maintenance of these assets [6] are vital to success of this sector growth [7]. One approach to leverage these issues is the use of computer vision techniques, specifically Convolutional Neural Networks (CNNs), to identify defects and to avoid greater future damage [8], [9], [10].

CNNs for examining wind power infrastructure imagery have been initially used by Moreno *et al* [11]. However, although the proposed approach provided an accuracy of 81.25%, it was limited by the small size of the dataset as it included only 78 images.

A larger dataset has been presented in [9], providing a significant number of publicly available high resolution, unannotated, wind turbine drone images. In this work the authors also provided a comparison of a number of models performances on a labelled version of this dataset. The comparison highlighted that Inception-ResNetv2 outperforms other ResNet models, providing a Mean Average Precision (mAP) of 81.1%.

Sakar *et al*. [12] extended the work in [9] by enhancing and labelling the dataset. The authors also tested the performance of YOLOv3 on this dataset reaching an mAP of 96%. However, this dataset is not publicly available.

In this work we present a publicly available Wind Turbine Surface Damage Detection dataset [13]. This dataset has been derived from an unlabelled publicly available dataset [14] and provides over 13000 images, with 3000 labelled instances of damaged and dirt wind turbines. We also present a comparison between the performance of ResNet-101 Faster R-CNN [15] and YOLOv5 [16] on this dataset. Finally, we discuss the results obtained by these models in detecting damage in videos of an active wind turbine obtained from a moving drone.

The paper is organized as follows. In the next section we present the state-of-the-art on Wind Turbine Surface Damage Detection using CNNs. Section III describes the method used to build our new dataset and the CNNs architectures used for benchmarking. In Section IV we present the comparison between these models on the proposed dataset and our attempt on detecting damages on videos from a moving drone of an active wind turbine. Section V concludes the work.

II. State of the art

In [11] the authors proposed a CNN architecture for detecting certain damages in the face of a wind turbine blade, i.e. by impact of rays, wearing and fractures. To train the network they also constructed a dataset of 78 150x150 pixel images containing the types of damages of real wind turbines from the public Internet. Moreover, they built a mock-up of a wind turbine to test the CNN on a mechatronics system. However, although the proposed approach achieved an accuracy of 81.25%, the CNN was trained with few number of images. Moreover, the sample images were real photographs of blades and the experiments were done over a mock-up. Nevertheless, this work demonstrated that a large number of samples should be considered and transfer learning was suggested for better classification.

A model based on Mask R-CNN for Wind Turbine Surface Damage Detection from photos taken by an Unmanned Aerial Vehicle (UAV) was proposed in [8]. In this work the authors also proposed an approach for automatic annotation of the images. This approach relied on a combination of image classification based on Resnet-18 CNN and thresholding to label the pixels of the images. Resnet-18 was chosen to reduce the spatial distortion in the class activation mapping. However, the results highlighted that with a mAP@0.30 of 0.2, the proposed method was 92% likely to cover the damage in full, providing a high recall with a low precision.

In Shihavuddin *et al.* [9] four Faster R-CNN models were trained to identify two classes of damage and two classes of common wind turbine features using images taken by a UAV: Inception-V2 [17], ResNet-50, ResNet-101 [18], and Inception-ResNet-V2 [19]. Images from a non-public dataset provided commercially by EasyInspect ApS as well as from the DTU-Drone inspection images of the wind turbine dataset [14] were used. The images were annotated to detect leading edge erosions, lighting receptors and vortex generator

Fig. 1: Sample instances of labelled images in the damage class

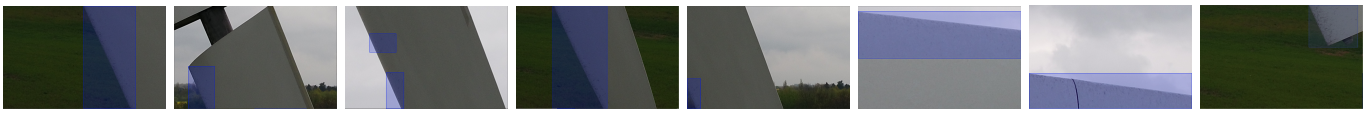


Fig. 2: Sample instances of labelled images in the dirt class

panels with or without missing teeth. Due to the limited number of instances in some classes, multiple augmentation methods were used including mirroring the images both vertically and horizontally, changing perspective, varying contrast to simulate different light conditions, applying Gaussian blur to simulate the image being out of focus and pyramid and patching methods to simulate different camera distances. All the models were pre-trained on the COCO dataset [20]. This work showed that there exist a direct correlation between the depth of the networks, the mAP value, and the time taken to infer. Moreover, the use of pyramid and patching augmentation was found to almost double performance when compared to patching alone. Finally the comparison showed that Inception-ResNet-V2 outperformed the other models with a mAP of 81.1%.

The work in [21] demonstrated that the application of techniques for image enhancement (e.g. white balance and histogram equalisation) to normalise the images such that they have the same temperature, rather than to increase the size of the training dataset or optimise the network model, indirectly increase the precision of a CNN model inference. In this work the authors also proposed a novel architecture called Image Enhanced Mask R-CNN which combined these image enhancement techniques with image augmentation and a tuned Mask R-CNN model to improve the detection performance of the standard Mask R-CNN model, resulting in a mAP at an Intersection over Union (IoU) threshold of 0.5 increase of 1.64%.

An approach based on YOLO to surface damage detection was presented by Sakar *et al.* [12]. Here the authors compared the performance of YOLOv3, YOLOv2 and Faster R-CNN models on the Nordtank WT dataset [14], along with 300 additional images collected from the Internet. The resulting mAP values were very impressive with the Faster-RCNN reaching 87% and YOLOv3 reaching 96%. These mAP values were all calculated at an IoU threshold of 0.65.

III. Method

Figure 1 and Figure 2 show samples of instances of the labelled images in the proposed dataset [13]. This dataset has been built as follows. The 5280x2970 pixel images from the DTU-Drone inspection images of the wind turbine unlabelled dataset [14] have been split into 72 smaller 586x371 images. This resolution has been chosen taking into account the YOLOv5 default input resolution, i.e. 640x640. Images without wind turbine surfaces were disregarded from the dataset to make the labelling process more manageable. The effects of a lower ratio of background to labelled images has also been taken into account as a trade off for recall over precision. Two classes have been considered: dirt and damage. While dirt may have energy production consequences, damage may have more significant safety consequences. Images labelled in the dirt class represents a dark shading seen on the surface of the wind turbine blades. The shading can be staining due to dirt and dust, or a buildup of flying insects, which is an issue that can have a significant impact on energy producing performance [22]. Images labelled in the damage class represent markings on the tower and nacelle as well as the blade (e.g. large bare patches in the outer coating on the leading edges [23], dark burn marks on the blades which can cause damage ranging from removal of the laminating layer to full blade tip removal [24] and blistering and cracking of the coating on the tower and nacelle leading to exposure of the steel underneath which can be identified from inconsistencies in the smooth surface, and later a red rusting). The labelling has been done by hand using the free online tool makesense.ai [25] which also provides tools for exporting the labels directly into the format required by the proposed CNN architectures. As a result of this process, we labelled 9351 images with 8770 instances of damage (see Figure 1) and 581 instances of dirt (see Figure 2). The reason why the data labels turned out to be heavily uneven toward the damage class is that dirt tended to cover a large continuous area, specifically the leading edge of the blade, being labelled with a large bounding box. Conversely, damage forms more discrete markings, lending itself towards separate bounding boxes for each mark.

Four CNNs architectures have been used to benchmark the proposed dataset: ResNet-101 Faster R-CNN, YOLOv5 Small(S), YOLOv5 Medium(M) and YOLOv5 Large(L). The ResNet-101 classifier is composed of 101-layers with 33 3-layered residual blocks as described by He *et al.* [26]. The Faster R-CNN has been used with the Detectron2 library [27]. ResNet-101 Faster R-CNN has been initialised with the weights pretrained on the ImageNet dataset [28] as provided by Detectron2. YOLOv5S, YOLOv5M and YOLOv5L differ from modifying the width and the height of the BottleneckCSP

feature extraction modules of the Cross Stage Partial Network inspired backbone. These networks have been initialised with the weights pretrained on the COCO dataset [20].

IV. Experimental Results

To ensure a fair comparison, the CNNs architectures mentioned in Section III have been initially fine tuned on the NEU-DET dataset [29]. NEU-DET is a database of 1,800 gray scale images of six different kinds of typical surface defects used for training and validating models for metallic surface defect detection.

To mitigate the problem of the unequal distribution of classes within our dataset, data augmentation techniques have been used as well. These include HSV augmentation, translation, scaling, flipping and the mosaic technique introduced in [30]. The mosaic augmentation loads four images in a 2x2 mosaic, reportedly allowing for detection outside normal context. HSV augmentation applies a random gain in the HSV values of a given image within a bound given as a hyper-parameter, and applies uniformly across a mosaic when used. 0.015, 0.7 and 0.4 have been used as values for the hue, the saturation and the gain hyper-parameters. The scale, translation and horizontal flip augmentation are then randomly applied within the supplied bound of the mosaic. More precisely, a 50% chance of horizontal flipping, a random scaling between 0.5x-1.5x of the original image and a random translation in the x and y of 10% of the total image resolution.

The dataset was split 70%-30% into a training and test set. For the YOLOv5 models, training consisted of 1000 epochs with a batch size of 16. This duration was chosen as it allowed the convergence of the loss calculated on the validation set. The batch size was chosen to support the memory requirements of the YOLOv5L. ResNet-101 Faster R-CNN was run for 100,000 epochs with a batch size per image of 512.

The Stochastic Gradient Descent (SGD) has been used to train the parameters of all the proposed models and a greedy k-fold cross validation approach [31] has been used to estimate the hyper-parameters of the optimiser (see Table I).

	Learning Rate	Weight Decay	Momentum
YOLOv5 (S, M, L)	0.01	0.0005	0.937
Faster R-CNN	0.0025	0.0001	0.9

TABLE I: Hyper-parameters of SGD.

For ResNet-101 Faster R-CNN the weighted sum of four loss values across both the Region Proposal Network (RPN) output and the final output has been used as loss function. This function includes the Binary Cross Entropy loss for the RPN, the L1 Localisation loss for the RPN, the Softmax Cross Entropy Classification loss for the final output and the L1 Localisation loss for the final output. The combination of the Means Square Error of the predicted bounding boxes, the Binary Cross Entropy of the predicted object confidences and the Binary Cross Entropy of the predicted classes have been used as compound loss function for all the YOLOv5 architectures.

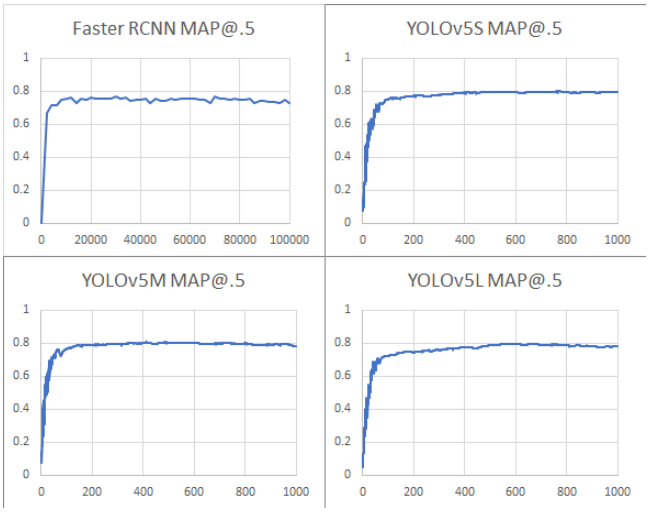


Fig. 3: mAP values at 0.5 IoU

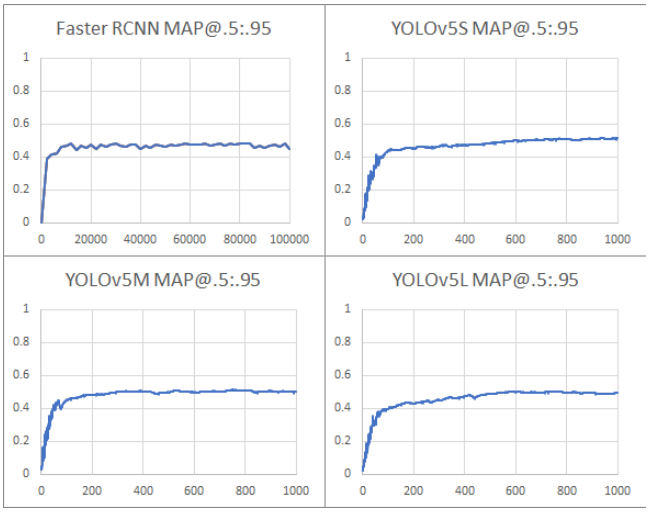


Fig. 4: mAP values at 0.5 through 0.95 IoU

The Mean Average Precision (mAP) metric has been used to compare the performance of the aforementioned CNNs architectures. This metric relies on the Intersection over Union (IoU) threshold for the prediction. A threshold of 0.5 IoU as well as an average over 0.5-0.95 at 0.5 intervals were decided [32].

Figure 3 and Figure 4 show the mAP@0.5 and mAP@0.5:0.95 values, respectively, for all the proposed CNNs architectures over the training phase and the final results are summarised in Table II.

	mAP@.5:.95	mAP@0.5	Recall	Precision
Faster R-CNN	0.4743	0.7539	0.73123	0.76422
YOLOv5s	**0.5121**	**0.7937**	0.79045	**0.82386**
YOLOv5m	0.4989	0.7837	0.79152	0.80792
YOLOv5L	0.4931	0.7836	**0.82118**	0.78386

TABLE II: mAP values, precision and recall at highest F1-score obtained in the testing phase

.

As it can be seen from Table II, YOLOv5S outperforms all the other architectures. However, YOLOv5L shows a better ability to estimate the samples labelled as damaged

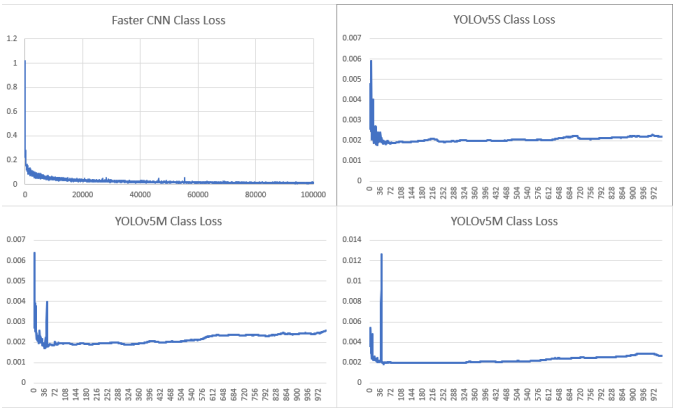


Fig. 5: Class Loss Values

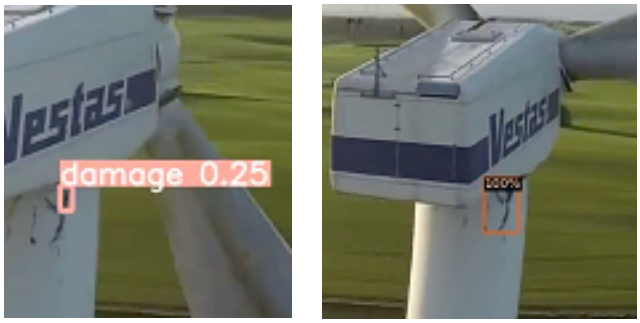
However, the trends of the loss on the validation set depicted in Figure 5 show that unlike ResNet-101 Faster R-CNN, the YOLOv5 networks suffer from over-fitting. Moreover, on images containing densely packed bounding boxes ResNet-101 Faster R-CNN demonstrated better performance on bounding box matching with respect to the ground truth than YOLOv5 (S, M, L).

Finally, an attempt of using both ResNet-101 Faster R-CNN and YOLOv5 (S, M, L) for detecting the surface damages on moving wind turbine blades using videos of drone footage from YouTube has also been made [33], [34], [35]. Table III reports the performance of each model with respect to the inference speed.

	YOLOv5S	YOLOv5M	YOLOv5L	Faster R-CNN
Speed	**17.6** FPS	12.2 FPS	8.5 FPS	1.7 FPS

TABLE III: Inference rate of ResNet-101 Faster R-CNN and YOLOv5 (S, M, L). .

As expected all the YOLOv5 CNNs significantly outperformed ResNet-101 Faster R-CNN in terms of inference rate. This is due to the design of the architecture of YOLOv5 which minimises inference time through the single neural network [36]. However, an interesting result in the estimation of the bounding boxes on the produced YouTube videos has been observed [34], [35]. While ResNet-101 Faster R-CNN estimates an entire instance of damage as a single prediction, YOLOv5S tends to identify only portions of the damage (see Figure 6. Besides this, YOLOv5 generally seemed to perform better, and while detected many false positives on the background, it was a lot less discriminative with its predictions on the blades too lending itself towards human assisted damage detection.



(a) (b)

Fig. 6: Surface Damage Detection on moving wind turbines from images captured by a drone [33]: (a) YOLOv5S Video detection [34]; (b) ResNet-101 Faster R-CNN Video detection [35]

V. Conclusion

This work presented a fully annotated publicly available dataset for Wind Turbine Surface Damage Detection, and provides some benchmark performance values using Faster R-CNN with a ResNet-101-C4 classifer and a number of YOLOv5 implementations. This work further uses the trained models to perform object detection on drone footage of a moving drone on an active wind turbine, proving the applicability of the used technologies towards AI-assisted or fully autonomous inspections on actively running wind turbines, potentially saving significant costs due to downtime. An interesting avenue of future research is the use of segmentation techniques on video footage to provide more specific descriptions of the wind turbine damage, and the annotating of wind turbine damage for segmentation using class activation mapping representations as described in [8]. The proposed dataset could easily be expanded given additional footage to allow for higher quality predictions and, in the authors opinion, doing so would be of considerable value.

Acknowledgment

This research is in part supported by the UK EPSRC funded Programme Grant Project Supergen ORE hub 2018 - grant reference number EP/S000747/1, and by the UK EPSRC funded Programme Grant Project National Centre for Nuclear Robotics (NCNR) - grant reference number EP/R02572X/1.

References

[1] "Big spring man accidentally falls off wind turbine; dies," 2021, https://www.cbs7.com/2020/09/23/big-spring-man-accidentally-falls-off-wind-turbine-dies. [Online]. Available: https://www.cbs7.com/2020/09/23/

big-spring-man-accidentally-falls-off-wind-turbine-dies/#:~:text=TOM%20GREEN%20COUNTY%2C%20TX%20(KOSA,11000%20block%20of%20Allen%20Road.

[2] "Finnish worker dies in sirdal accident," 2021, https://norwaytoday.info/news/finnish-worker-dies-in-sirdal-accident. [Online]. Available: https://norwaytoday.info/news/finnish-worker-dies-in-sirdal-accident/

[3] "Calexico man identified as victim of fatal 100-foot fall inside whitewater turbine," 2021, https://eu.desertsun.com/story/news/2021/01/04/calexico-man-idd-victim-fatal-fall-inside-whitewater-turbine/4130827001/. [Online]. Available: https://eu.desertsun.com/story/news/2021/01/04/calexico-man-idd-victim-fatal-fall-inside-whitewater-turbine/4130827001/

[4] M. Gianni, "Towards expendable robot teaming in extreme environments," *International Journal of Mechanical Engineering and Robotics Research*, vol. 8, no. 6, pp. 830–838, 2019.

[5] P. A. Bogdan, J. Wheadon, F. B. Klein, and M. Gianni, "Magnetic tracked robot for internal pipe inspection," in *2021 European Conference on Mobile Robots (ECMR)*, 2021, pp. 1–6.

[6] B. Steffen, M. Beuse, P. Tautorat, and T. S. Schmidt, "Experience curves for operations and maintenance costs of renewable energy technologies," *Joule*, vol. 4, no. 2, pp. 359–375, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542435119305793

[7] HM Government, "Net Zero Strategy: Build Back Greener," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033990/net-zero-strategy-beis.pdf, 2021, online; accessed 28 March 2022.

[8] M. Crous, B. O. K. Intelligentie, and A. Visser, "Combining weakly and strongly supervised segmentation methods for wind turbine damage annotation," in *Computer Science*, 2018.

[9] A. Shihavuddin, X. Chen, V. Fedorov, A. Nymark Christensen, N. Andre Brogaard Riis, K. Branner, A. Bjorholm Dahl, and R. Reinhold Paulsen, "Wind turbine surface damage detection by deep learning aided drone inspection analysis," *Energies*, vol. 12, no. 4, p. 676, 2019.

[10] 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, 2019, 52nd CIRP Conference on Manufacturing Systems (CMS), Ljubljana, Slovenia, June 12-14, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827119305918

[11] S. Moreno, M. Peña, A. Toledo, R. Treviño, and H. Ponce, "A new vision-based method using deep learning for damage inspection in wind turbine blades," in *2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*. IEEE, 2018, pp. 1–5.

[12] D. Sarkar and S. K. Gunturi, "Wind turbine blade structural state evaluation by hybrid object detector relying on deep learning models," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8535–8548, 2021.

[13] Ashley Foster and Oscar Best and Mario Gianni and Asiya Khan and Keri Collins and Sanjay Sharma, "YOLO Annotated Wind Turbine Surface Damage," https://www.kaggle.com/datasets/ajifoster3/yolo-annotated-wind-turbines-586x371, 2021, available online.

[14] A. Shihavuddin and X. Chen, "Dtu - drone inspection images of wind turbine," 2018.

[15] Q. Guo, L. Liu, W. Xu, Y. Gong, X. Zhang, and W. Jing, "An improved faster r-cnn for high-speed railway dropper detection," *IEEE Access*, vol. 8, pp. 105 622–105 633, 2020.

[16] D. Liao, Z. Cui, X. Zhang, J. Li, W. Li, Z. Zhu, and N. Wu, "Surface defect detection and classification of si3n4 turbine blades based on convolutional neural network and yolov5," *Advances in Mechanical Engineering*, vol. 14, no. 2, p. 16878132221081580, 2022. [Online]. Available: https://doi.org/10.1177/16878132221081580

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[19] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[20] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2015.

[21] J. Zhang, G. Cosma, and J. Watkins, "Image enhanced mask r-cnn: A deep learning pipeline with new evaluation measures for wind turbine blade defect detection and classification," *Journal of Imaging*, vol. 7, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2313-433X/7/3/46

[22] G. Corten and H. Veldkamp, "Aerodynamics - insects can halve wind-turbine power," *Nature*, vol. 412, pp. 41–2, 08 2001.

[23] R. Herring, K. Dyer, F. Martin, and C. Ward, "The increasing importance of leading edge erosion and a review of existing protection solutions," *Renewable and Sustainable Energy Reviews*, vol. 115, p. 109382, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032119305908

[24] A. Garolera, S. Madsen, M. Nissim, J. Myers, and J. Holboell, "Lightning damage to wind turbine blades from wind farms in u.s." *IEEE Transactions on Power Delivery*, vol. 31, pp. 1–1, 01 2014.

[25] P. Skalski, "Make Sense," https://github.com/SkalskiP/make-sense/, 2019.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[27] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[28] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[29] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," May 2019. [Online]. Available: https://hdl.handle.net/2134/12249215.v1

[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.

[31] D. S. Soper, "Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation," *Electronics*, vol. 10, no. 16, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/16/1973

[32] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014. [Online]. Available: https://arxiv.org/abs/1405.0312

[33] T. Nøhr, "Wind turbine, close up with drone. watch in 1080p!!!" 2015. [Online]. Available: https://www.youtube.com/watch?v=JzRNRy3ArJA

[34] A. Foster, "Yolov5s wind turbine surface damage object detection," 2021. [Online]. Available: https://www.youtube.com/watch?v=3SXwnXqh-AY&ab_channel=ajifoster3

[35] ——, "Faster r-cnn resnet-101-c4 wind turbine surface damage detection," 2021. [Online]. Available: https://www.youtube.com/watch?v=DLjP1QVptws&ab_channel=ajifoster3

[36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.