Article

Wind turbine maintenance cost reduction by deep learning aided drone inspection analysis

ASM Shihavuddin[1]* , **Xiao Chen[2]**, **Vladimir Fedorov[3]**, **Anders Nymark Christensen[1]**, **Nicolai Andre Brogaard Riis[1]**, **Kim Branner[2]**, **Anders Bjorholm Dahl[1]** and **Rasmus Reinhold Paulsen[1]**

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Lyngby, Denmark

[2] Department of Wind Energy, Technical University of Denmark (DTU), Roskilde, Denmark

[3] EasyInspect ApS, Brondbyvester, Denmark

* Correspondence: shihav@dtu.dk; Tel.: +4571508691

Abstract: Timely detection of surface damages on wind turbine blades is imperative for minimising downtime and avoiding possible catastrophic structural failures. With recent advances in drone technology, a large number of high-resolution images of wind turbines are routinely acquired and subsequently analysed by experts to identify imminent damages. Automated analysis of these inspection images with the help of machine learning algorithms can reduce the inspection cost, thereby reducing the overall maintenance cost arising from the manual labour involved. In this work, we develop a deep learning based automated damage suggestion system for subsequent analysis of drone inspection images. Experimental results demonstrate that the proposed approach could achieve almost human level precision in terms of suggested damage location and types on wind turbine blades. We further demonstrate that for relatively small training sets advanced data augmentation during deep learning training can better generalise the trained model providing a significant gain in precision.

Dataset: doi:10.17632/hd96prn3nc.1

Dataset License: CC BY NC 3.0

Keywords: Wind Energy; Wind Turbine; Drone Inspection; Damage Detection; Deep Learning; Convolutional Neural Network (CNN)

---

1. Introduction

Reducing the Levelized Cost of Energy (LCoE) remains the overall driver for development of the wind energy sector [1,2]. Typically, the Operation and Maintenance (O&M) costs account for 20 - 25% of the total LCoE for both onshore and offshore wind in comparison to 15% for coal, 10% for gas and 5% for nuclear [3]. Over the years, great efforts have been made to reduce O&M cost of wind energy using emerging technologies, such as automation [4], data analytics [5,6], smart sensors [7], and artificial intelligence (AI) [8]. The aim of these technologies is to achieve more effective operation, inspection and maintenance with minimal human interference.

One of similar emerging technologies is to use drone-based inspection of the wind turbine blades. Drone inspections are currently being successfully used for structural health monitoring of diverse infrastructures such as bridges, towers, power plants, dams [9] and so forth. The drone-based approach enables low-cost and frequent inspections, thereby allowing predictive maintenance at lower costs. Wind turbine surface damages exhibit recognisable visual traits that can be imaged by drones with optical cameras. These damages for example leading edge erosion, cracks, damaged lightning receptors, damaged vortex generators and so forth are externally visible even in their early stages of development. Moreover, some of these damages even indicate severe internal structural damages [10].

Extracting damage information from a large number of high-resolution inspection images requires a significant manual effort, which is one of the reasons for the overall inspection cost still remaining at a high level. In addition, manual image inspection is tedious and therefore error-prone. By automatically providing

suggestions to experts on highly probable damage locations, we can significantly reduce the required man hours and simultaneously enhance human's detection performance, as a result minimising the labour cost involved with the analysis of inspection data. With regular cost efficient and accurate drone inspection, the scheduled maintenance of wind turbines can be performed less frequently, bringing down the overall maintenance cost contributing to the lower LCoE.

Only very few research works have addressed the machine learning based approaches for surface damage detection of wind turbine blades from drone images. One example, however, is Wang *et al.* [11] who used drone inspection images for crack detection. To automatically extract damage information, they used Haar-like features [12,13] and ensemble classifiers selected from a set of base models including logitBoost [14], decision trees [15] and support vector machines [16]. Their work was limited to detecting crack and relied on classical machine learning methods.

Recently deep learning technology became efficient and popular, providing ground breaking performances in detection systems for the last four years [17]. In this work, we addressed the problem of damage detection by deploying a deep learning object detection framework to aid human annotation. The main advantages of deep learning over other classical object detection methods are: it automatically finds the most discriminate features for the identification of objects and it goes through an optimisation process to minimise the errors.

Large size variations of different surface damages of wind turbine blades in general is a challenge for machine learning algorithms. In this study, we overcame this challenge with the help of advanced image augmentation methods. Image augmentation is the process of creating extra training images by altering images in the training sets. With the help of augmentation, different versions of the same image are created encapsulating different possible variations during drone acquisition [18].

The main contributions of this work are threefold:

1. **Automated suggestion system for damage detection in drone inspection images**: Implementation of a deep learning framework for automated suggestions of surface damages on wind turbine blades captured by drone inspection images. We show that deep learning technology is capable of giving reliable suggestions with almost human level accuracy to aid manual annotation of damages.
2. **Higher precision in suggestion model achieved through advanced data augmentation**: The advanced augmentation step called 'Multi-scale pyramid and patching scheme' enables the network to achieve a better learning. This scheme significantly improves the precision of suggestions specially for high-resolution images and for object classes that are very rare and difficult to learn about.
3. **Publication of wind turbine inspection dataset**: This work produced a publicly available drone inspection image of 'Nordtank' turbine over the years of 2017 and 2018. The dataset is hosted at doi:10.17632/hd96prn3nc.1 within the Mendeley public dataset repository [19].

Surface damage suggestion system is trained using faster R-CNN [20] which is a state of the art deep learning object detection framework. Convolutional neural network (CNN) is used as the backbone architecture in that framework for extracting feature descriptors with a high discriminative power. The suggestion model is trained on drone inspection images of different wind turbines. We also employed advanced augmentation methods (as described in details in method section) to generalise the learned model. The more generalised model helps to perform better for challenging test images during inference. The inference is the process of applying the trained model to an input image to receive the detected or in our case the suggested object in return.

Figure 1 illustrates the training flowchart of the proposed method. Initially, damages from the drone-based inspection images are annotated in terms of bounding boxes by experts. In the second stage, a deep learning suggestion model is trained with the faster R-CNN object detection framework using both original and augmented images.
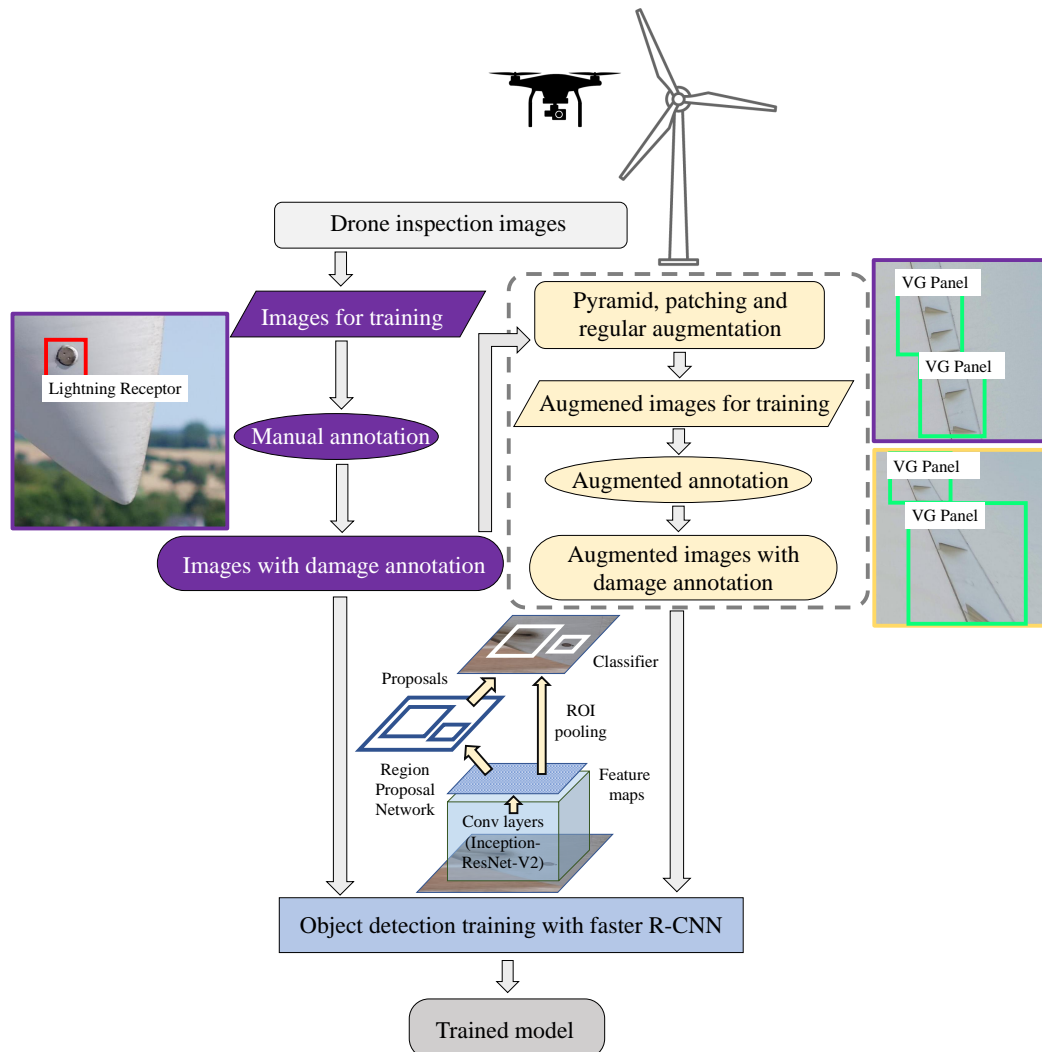
Figure 1. Flowchart of the proposed automated damage suggestion system. To begin with, damages on wind turbine blades that are imaged using drone inspections are annotated in terms of bounding boxes by field experts. Annotated imaged are also augmented with the proposed advanced augmentation schemes to increase the number of training samples. Faster-RCNN Deep learning object detection framework is applied to train from these annotated and augmented annotated images. After the training the deep learning framework produces a detection model that can be applied for new inspection image analysis.

2. Materials and Methods

2.1. Manual annotation.

For this work, four classes are defined; leading edge erosion (LE erosion), vortex generator panel (VG panel), VG panel with missing teeth and lightning receptor. Examples of each of these classes are illustrated in Figure 2. A VG panel and a lightning receptor are not any specific damage types, rather external components on wind turbine blade that often has to be visually checked during inspections. Both damaged and non-damaged lightning receptors are considered as the same class in this work, however they are typically considered separately. These two classes are specifically considered to evaluate the ability of the developed system to identify varied object classes related to wind turbines. These classes serve as passive indicators to the health condition of the wind turbine blades. Table 1 summarises the number of annotations for each class, that are annotated by experts and considered as ground truths. For inference, we used 2 human participants after training to detect damages, and considered their results with and without suggestions as human precision.
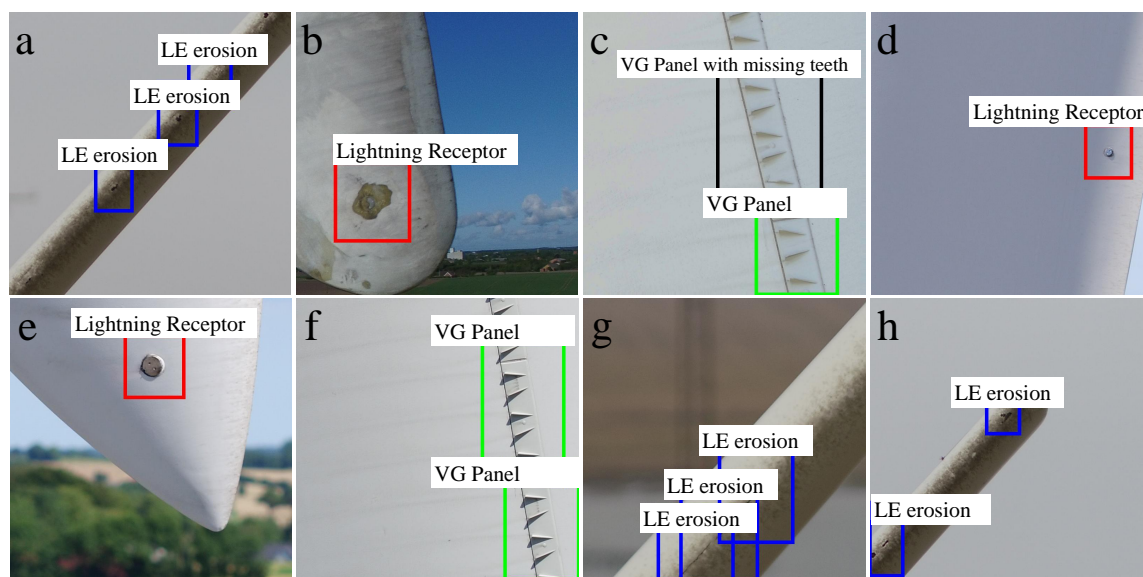

Figure 2. Examples of manually annotated damages related to wind turbines blades. a, g and h illustrate the examples of leading edge erosion annotated by experts using bounding boxes. b, d, e illustrate the lightning receptors. c shows the example of a Vortex generator (VG) panel with missing teeth. c and f show the examples of well functioning vortex generator (VG) panels.

Table 1. Details of the training and testing sets related to EasyInspect Dataset. From the pool of available images, 60% were used for training and 40% was used for testing. Annotations in the training set comprises of the annotations from full resolution images that were randomly selected. Annotations in the testing set comprises of the annotations from the remaining raw images together with the patched images.

List of classes		
Classes	Annotations - Training	Annotations - Testing
LE erosion	135	67
VG panel	126	65
VG with missing teeth	27	14
Lightning receptor	17	7

2.2. Image augmentations

Some types of damages on wind turbine blades are scarce and it is hard to collect representative samples during inspections. For deep learning, it is necessary to have thousands of training samples that are representative of the depicted patterns in order to obtain a good detection model [21]. The general approach is to use example

images from the training set and augment them in ways that represent probable variations in appearances maintaining the core properties of object classes.

2.3. Regular augmentations

Regular augmentations are defined as conventional ways of augmenting images for increasing the number of training samples. There are many different types of commonly used augmentation methods that could be selected based on the knowledge about object classes and the possible occurrences of variances during acquisition. Taking drone inspection and wind turbine properties into considerations, four types of regular augmentation are chosen to be used in this work which are listed as below.

- Perspective transformation for camera angle variation simulation.
- Left to right flip or top to bottom flip to simulate blade orientation: e.g. pointing up or down.
- Contrast normalisation for variations in lighting conditions.
- Gaussian blur simulating out of focus images.

Figure 3 illustrates the examples of regular augmentations of the wind turbine inspection images containing damage examples.
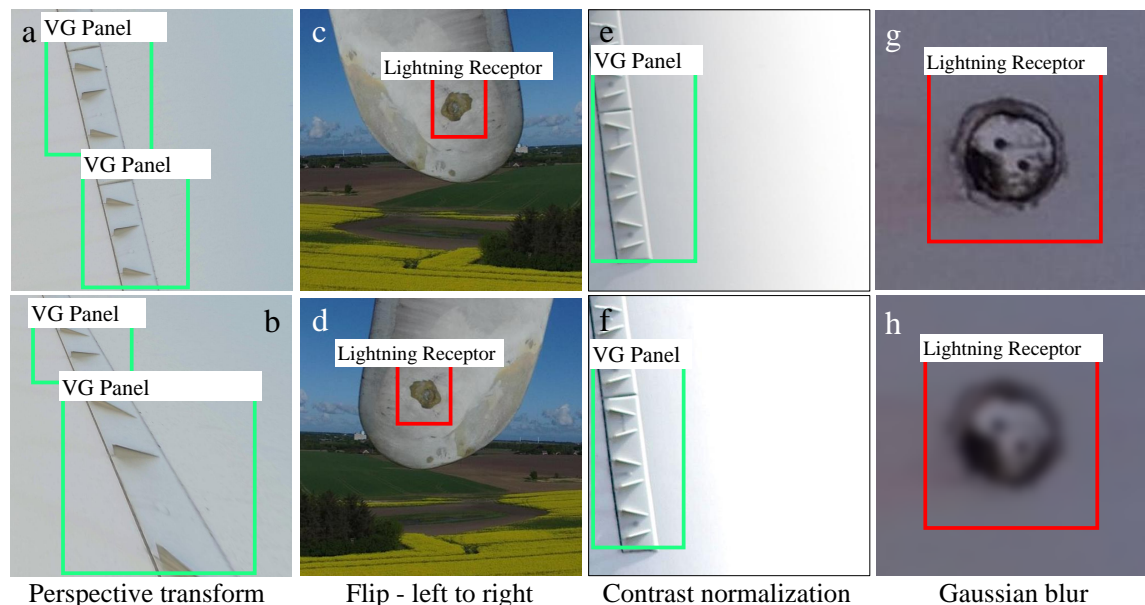

Perspective transform	Flip - left to right	Contrast normalization	Gaussian blur

Figure 3. Illustrations of the regular augmentation methods applied to drone inspection images. b is the perspective transform of a where both images illustrating the same VG panels; d is left to right flipping of the image patch c containing a lightning receptor. f is the contrast normalised version of e. h is the augmented image of the lightning receptor in g simulating de-focusing of the camera using approximate Gaussian blur.

2.4. Pyramid and patching augmentation

Drones deployed to acquire the dataset typically are equipped with very high-resolution cameras. High-resolution cameras allow drones to capture essential details being at a further and safer distance from the turbines. These high resolution images allow the flexibility of training from rare types of damages and wide variety of backgrounds at a different resolution using the same image.

Deep learning object detection framework such as Faster R-CNN framework have predefined input image dimension that typically allows for a maximum input image dimension (either height or width) to be 1000 pixels. It is maintained through re-sizing higher resolution images, while keeping the ratio between width and height in situ. For high-resolution images, this limitation creates new challenges due to damages being minimal in pixel sizes compared to full image size. For example, in Figure 4 top right, the acquired full resolution image is of size 4000×3000 pixels, where the lightning receptor only occupies around 100×100 pixels. When fed to CNN

during training in full resolution, it would be resized to the pre-defined network input size, where the lightning receptor would be occupying tiny portion of 33×33 pixels. Hence, it is rather complicated to acquire enough recognisable visual traits of the lightning receptor.

Using the multi-scale pyramid and patching scheme on the acquired high resolution training images, scale varied views of the same object is generated and fed to the neural network. In this scheme, the main full resolution image is scaled to multi-resolution images ($1.00\times$, $0.67\times$, $0.33\times$) and on each of these images, patches containing objects are selected with the help of a sliding window with 10% overlap. The selected patches are always of size 1000×1000 pixels.

The flow chart of this multi-scale pyramid and patching scheme are shown in Figure 4. This scheme helps to represent object capture at different camera distances allowing the detection model to be efficiently trained on both low and high-resolution images.
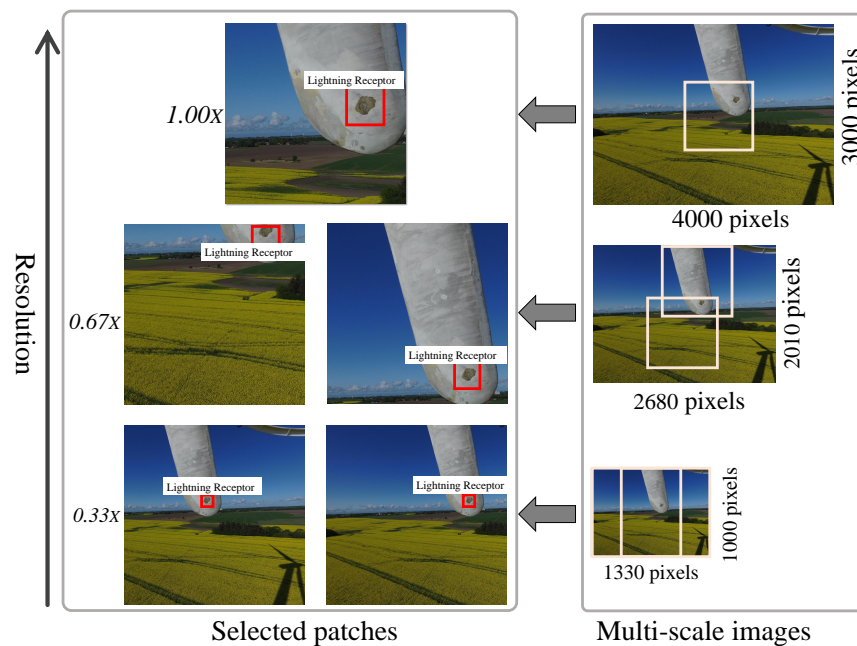

Figure 4. Proposed multi-scale pyramid and patching scheme for image augmentation. On the right side, there is the pyramid scheme and on the left side, is the patching scheme. The bottom level of pyramid scheme is defined to the image size where either height or width is of 1000 pixels. In the pyramid, from top to bottom, images are scaled from $1.00\times$ to $0.33\times$ simulating from the highest to the lowest resolutions. Sliding windows with 10% overlap are scanned over the images at each resolution to extract patches containing at least one object. Resolution conversions are performed through linear interpolation method [22].

2.5. Damage detection framework

With recent advances in deep learning for object detection, new architectures are frequently proposed establishing groundbreaking performances. Currently, different stable meta architectures are made publicly available and have already been successfully deployed for many challenging applications. Among deep learning object detection frameworks, one of the best-performing methods is the faster R-CNN [23]. In our work, for the surface damage detection and classification using drone inspection images are performed using faster R-CNN [20].

Faster R-CNN uses a region proposal network (RPN) trained on feature descriptors extracted by CNN to predict bounding boxes for objects of interest. CNN architecture automatically learns features such as texture, spatial arrangement, class size, shape and so forth from training examples. These automatically learned features are more appropriate than hand-crafted features.

Convolutional layers in CNN summarises information based on the previous layer's content. The first layer usually learns edges; the second finds patterns in edges encoding shapes of higher complexity and so

forth. The last layer contains feature map of much smaller spatial dimensions than the original image. The last layer feature map summarises information about the original image. We experimented with both lighter CNN architectures such as InceptionV2 [24] and ResNet50 [25] and heavier CNN architectures such as ResNet101 [25] and Inception-ResNet-V2 [26].

Inception-ResNet-V2 [26] is a very deep CNN network containing a combination of Inception and ResNet modules. This particular CNN network is used as the backbone architecture in our final model within faster R-CNN framework for extracting highly discrimination feature descriptors.

2.6. Performance measure - Mean Average Precision (MAP)

All the reported performances in terms of Mean Average Precision (MAP) are measured during inference on the test images, where the inference is the process of applying the trained model to an input image to receive the detected and classified object in return. In this work, we termed it as suggestions if the trained model from deep learning is being used.

Mean average precision (MAP) is commonly used in computer vision to evaluate object detection performance during inference. An object proposal is considered accurate only if it overlaps with the ground truth with more than a certain threshold. Intersection over Union (IoU) is used to measure the overlap of a prediction and the ground truth [1].

$$\text{IoU} = \frac{P \cap GT}{P \cup GT} \tag{1}$$

The IoU value corresponds to the ratio of the common area over the sum of the proposed detection and ground truth areas (as shown in Equation 1, where P and GT are the predicted and ground truth bounding boxes respectively). If the value is more than 0.5, the prediction or in our case the suggestion is considered as a true positive. This 0.5 value is relatively conservative as it makes sure that the ground truth and the detected object has very similar bounding box location, size and shape. For addressing human perception diversities, we used 0.3 IOU threshold for considering a detection as true positive.

$$\text{P}_C = \frac{N(\text{True Positives})_C}{N(\text{Total Objects})_C} \tag{2}$$

$$\text{AP}_C = \frac{\sum \text{P}_C}{N(\text{Total Images})_C} \tag{3}$$

$$\text{MAP} = \frac{\sum \text{AP}_C}{N(\text{Classes})} \tag{4}$$

Per class precision for each image, P_C is calculated using Equation 2. For each class, average precision, AP_C is measured over all the images in the dataset using Equation 3. Finally, mean average precision (MAP), is measured as the mean of average precision for each class over all the classes in the dataset (see Equation 4). Through out this work, the MAP is reported in percentage.

2.7. Software and codes

We used Tensorflow [27] deep learning API for experimenting with the Faster R-CNN object detection method with different CNN architectures. These architectures are compared with each other under proposed augmentation schemes. For implementing the regular augmentations, we used the 'imgaug' [2] package, and for the pyramid and patching augmentation, we developed an in-house python library. Inception-ResNet-V2 and other CNN weights are initialized from the pre-trained weight on Common Objects in COntext (COCO) dataset [28]. COCO dataset consists of 80 categories of regular objects and commonly used for bench-marking deep learning object detection performance.

---

[1] Ground truth refers to the original damages identified and annotated by experts in the field
[2] https://github.com/aleju/imgaug

2.8. Dataset

2.8.1. EasyInspect Dataset

The EasyInspect dataset is a non-public inspection data set provided by Easyinspect ApS [3] which contains images (4000 × 3000 pixels in size) of different types of damages in wind turbines from different manufacturers. The four classes are LE erosion, VG panel, VG panel with missing teeth and lightning receptor.

2.8.2. DTU Drone Inspection Data set

In this work, we produced a new public dataset entitled 'DTU - Drone inspection images of the wind turbine'. It is to our knowledge, the only public wind turbine drone inspection image dataset containing a total of 701 high-resolution images. This dataset has temporal inspection images of 2017 and 2018 covering 'Nordtank' wind turbine located at DTU Wind Energy's test site at Roskilde, Denmark. The dataset comes with the examples of damages or mounted objects as VG panel, VG panel with missing teeth, LE erosion, cracks, lightning receptor, damaged lightning receptor, missing surface material, and others. It is hosted at doi:10.17632/hd96prn3nc.1 [19].

3. Results

3.1. Augmentation of training images provides significant gain in performance

Comparing different augmentation types shows that a combination of the regular, pyramid and patching augmentations produces the more accurate suggestion model, specially for the deeper CNN architectures as backbone for faster R-CNN framework. Using CNN architecture ResNet-101 for example (as shown in Table 2 in column 'all'), without any augmentation the mean average precision (MAP) (detailed in Materials and Methods section) of damage suggestion is very low with the value of 25.9%. With help of the patching augmentation, the precision improves significantly (as for this case, the MAP raised to 35.6%).

---

[3]   https://www.easyinspect.io/

Table 2. Experimental results for different CNN architectures and data augmentation methods. All the experimental results are reported in terms of MAP in this table. E, VG, VGMT and LR represents Erosion, VG panel, VG with missing teeth and Lightning receptor respectively. 'All' is the overall MAP comprising all the four classes.

CNN	Augmentation	MAP (%)				
		E	VG	VGMT	LR	All
Inception-V2	Without	20.34	13.15	4.21	41.39	19.77
	Patching	37.98	53.86	32.74	22.12	36.67
	Patching + regular	36.90	52.58	36.08	23.54	37.28
	Pyramid + patching	88.11	**86.61**	58.28	**70.18**	**75.80**
	Pyramid + patching + regular	**89.94**	84.15	**71.85**	40.76	71.67
ResNet-50	Without	37.51	17.54	7.02	40.88	25.74
	Patching	34.48	54.92	32.17	34.96	39.13
	Patching + regular	35.06	47.87	24.88	34.96	35.69
	Pyramid + patching	90.19	85.15	**61.43**	**73.85**	**77.66**
	Pyramid + patching + regular	**90.47**	**88.84**	35.27	73.15	71.93
ResNet-101	Without	29.61	20.29	4.21	49.33	25.86
	Patching	34.79	47.35	25.35	34.96	35.61
	Patching + regular	36.06	51.41	32.16	33.46	38.27
	Pyramid + patching	**88.86**	**87.46**	41.53	64.19	70.51
	Pyramid + patching + regular	86.77	86.15	**41.53**	**77.00**	**72.86**
Inception-ResNet-V2	Without	31.54	18.35	5.96	54.14	27.50
	Patching	36.30	44.43	33.37	42.12	39.06
	Patching + regular	32.73	47.59	19.58	34.96	33.72
	Pyramid + patching	90.22	**91.66**	69.85	69.56	80.32
	Pyramid + patching + regular	**90.62**	89.08	**73.11**	**71.58**	**81.10**

Together with the patching and the regular augmentations, the MAP increased slightly to 38.3%. However, the pyramid scheme dramatically improves the performance of the trained model up to 70.5%. The best performing configuration is last one with the combination of pyramid, patching and regular augmentation schemes generating MAP of 72.9%.

As shown in Figure 5 (a-d), the proposed combination of all the proposed augmentation methods significantly and consistently improves the performance of the model and lifts it to above 70% for all four CNN architecture. Note that regular augmentation on top of the multi-scale pyramid and patching scheme adds on average 2% gain in MAP. The results also demonstrate that for the small dataset (where some types of damages are extremely rare), it is essential to generate as many augmented images as possible.

3.2. CNN architecture performs better as it goes deeper

When comparing the four selected CNN backbone architectures for faster R-CNN framework, the Inception-ResNet-V2 which is the deepest performs the best among all. If we fix the augmentation to the combination of pyramid, patching and regular, the MAP of the Inception-V2, ResNet-50, ResNet-101, Inception-ResNet-V2 are 71.67%, 71.93%, 72.86% and 81.10% respectively (as shown in Figure 5 and in Table 2). The number of layers in each of these networks representing the depth of the network could be arranged in the same order as well (where inception-V2 is the lightest and Inception-ResNet-V2 is the deepest). It demonstrates that the performance regarding MAP increases as the network goes deeper. The gain in performance of deeper networks comes with the cost of longer training time and higher requirements on the hardware.

3.3. Faster damage detection with suggestive system in inspection analysis

Time required for automatically suggesting damage locations for new images using the trained model depends on the size of input image and depth of the CNN architecture used. In our case, average time required for inferring a high resolution image using Inception-V2, ResNet-50, ResNet-101 and Inception-ResNet-V2 networks (after leading the model) are respectively 1.36 seconds, 1.69 seconds, 1.87 seconds and 2.11 seconds. Whereas, for human based analysis without suggestions, it can take around 20 seconds to 3 minutes per image
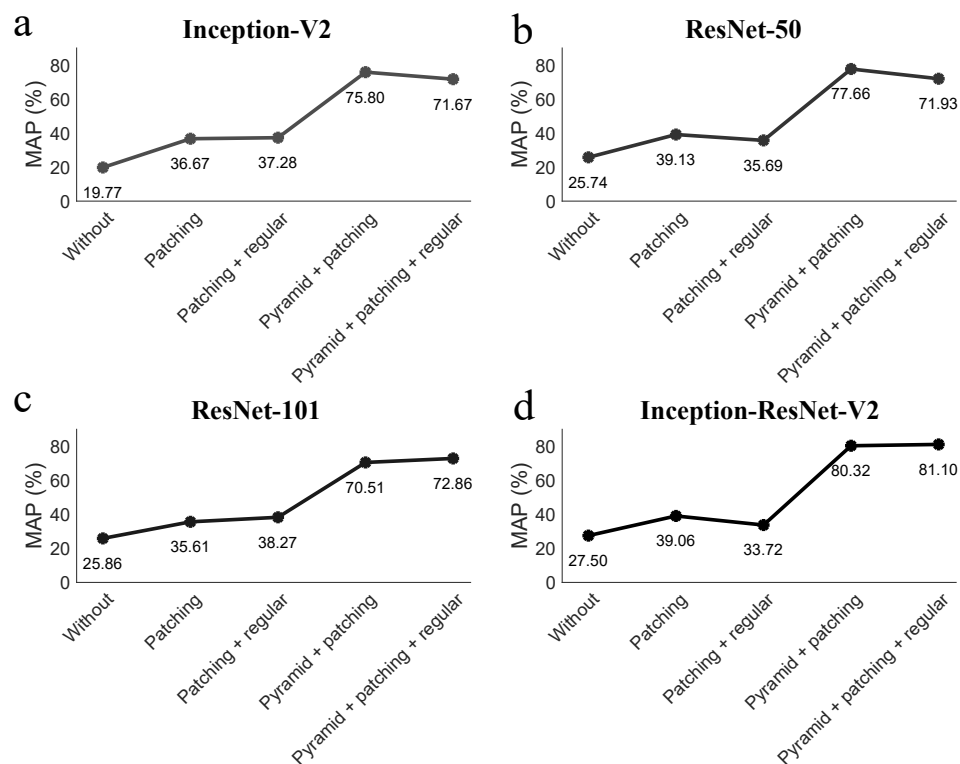
Figure 5. Comparison of the precision of trained models using various network architectures and augmentation types. a-d represents sequentially lighter to deeper CNN backbone architectures used for deep learning feature extraction. CNN networks explored in this work are: Inception-V2, ResNet-50, ResNet-101 and Inception-ResNet-V2. In each individual figure (a-d): the y-axis represents mean average precision (MAP) of the suggestion on test set and are reported in percentage. As a summery, it illustrates that with the pyramid, patching and regular augmentation the MAP goes higher than 70% across all the tested CNN architectures. For any specific type of augmentation, MAPs in general are higher for the deeper networks than for the lighter ones.

depending on the difficulty level for identification. With the deep learning aided suggestion system for human, the analysis time goes significantly lower (almost to two-third) and also produces better accuracy (see Table 3).

Table 3. Summary of the experimental results.

	Mean Average Precision (MAP) in percentage		
Classes	Suggestions	Human	Suggestions aiding human
LE erosion	90.62	68.00	86.60
VG panel	89.08	85.91	81.57
VG with missing teeth	73.11	80.17	91.37
Lightning receptor	71.58	98.74	81.71
Overall	81.10	83.20	85.31

All the experiments reported in this work are performed on a GPU cluster machine with 11 GB GeForce GTX 1080 Graphics Cards within a Linux operating system. The initial time required for training (275 epochs [4]) using Inception-V2, ResNet-50, ResNet-101 and Inception-ResNet-V2 networks are on average 6.3 hours, 10.5 hours, 16.1 hours and 27.9 hours respectively.

4. Discussion

With deep learning based automated damage suggestion system, the best performing model in our proposed method produces 81.10% precision, which is within 2.1% range of average human precision of 83.20%. In the case of deep learning aided suggestion for human, the MAP improves significantly to 85.31%, and the required processing time became two-third in average for each image. This result suggests that, humans can get benefited by suggestions on where to look for damages in images, specially for difficult cases like VG panel with missing teeth.

The experimental results show that for smaller image dataset of wind turbine inspection, the performance is more sensitive to the quality of image augmentation than the selection of CNN architecture. One of the main reasons is that most damage types can have considerably large variance in appearances that makes the deployed network dependent on the larger number of examples to learn from.

The combination of ResNet and inception modules in Inception-ResNet-V2 learns difficult damage types such as missing teeth in a vortex generator with more reliability than by the other CNN architectures. Figure 6 illustrates some of the suggestion results on inspection images for testing. The suggestion model developed in this study performed well for challenging images providing almost human level precision. When there is no valid class present in the image, the trained model finds only the background class and presents 'No detection' as label for that image (an example is shown in Figure 3g).

This automated damage suggestion system has a significant cost advantage over the current manual one. Drone inspections nowadays typically can cover up to 10 or 12 wind turbines per day. Damage detection, however, is much less efficient as it involves considerable data interpretation for damage identification, annotation, classification, etc., which have to be conducted by skilled personnel. This process would incur significant labour cost considering the huge amount of images taken by the drones from the field. Using the deep learning framework for suggestion to aid manual damage detection, the entire inspection and analysis process can be partially automated to minimise human intervention and reduce the overall O&M cost of wind turbines.

With suggested damages and subsequent corrections by human experts, along the time, the number of annotated training examples would be increased and fed to the developed system for updating the trained suggestion model. This continues way of learning through the help of human experts can increase the accuracy of deep learning model (expected to provide 2%-5% gain in MAP) and also reduce the required time for human corrections.

Relevant information about the damages, i.e., their size and location on blade can be used for further analysis in estimating wind turbine structural and aerodynamic conditions. The highest standalone MAP achieved

---

[4] one epoch is defined as when all the images in the training set had been used at least once for optimisation the detection model
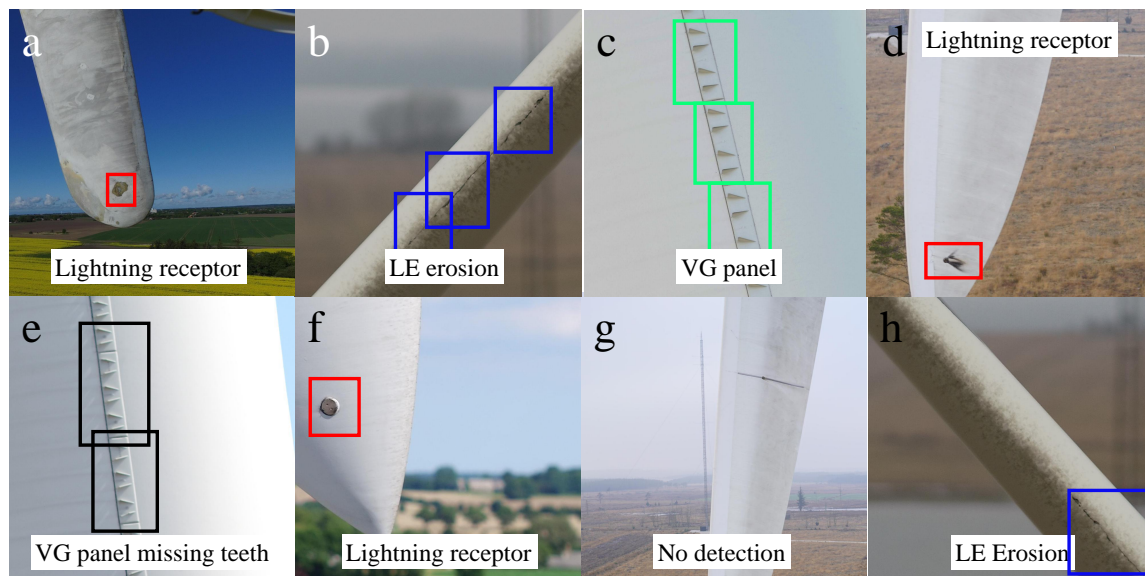
Figure 6. Suggestion results on the test images for the trained model using Inception-ResNet-V2 together with the proposed augmentation schemes. a, d and f illustrates suggested lighting receptors; b and h show LE erosion suggestion; c and e illustrate the suggestion of VG panels with intact teeth and those with missing teeth respectively. The latter exemplifies one of the very challenging tasks for automated damage suggestion method. g shows the example of when no damage is detected.

with our proposed method is 81.10%, which is almost within human level precision given the complexity of the problem and the conservative nature of the performance indicator. The developed automated detection system at its current state can safely work as suggestion system for experts to finalise the damage locations from inspection data.

5. Conclusion

The cost of wind turbine maintenance can be significantly reduced using automated damage suggestion based on regular drone inspection images. This work presented a method to minimise human intervention in wind turbine damage detection providing accurate suggestion of damages on drone inspection images. Using the Inception-ResNet-v2 architecture inside Faster R-CNN, 81.10% MAP was achieved on four different types of damages; LE erosion, vortex generator panel, vortex generator panel with missing teeth, and lightning receptor. We adopted a multi-scale pyramid and patching scheme that significantly improved the precision by 35% in average across tested CNN architectures. The experimental results demonstrate that deep learning with augmentation can successfully overcome the challenge of scarce availability of damage samples for learning.

In this work, a new image dataset of wind turbine drone inspection was published for the research community [19].

Author Contributions: R.R.P., K.B. and A.B.D. designed and supervised the research. A.S., V.F. designed experiments, implemented the system and performed the research. X.C., A.N.C. and N.A.B.R. analysed the data. A.S. wrote the manuscript with inputs from all authors.

References

1. Aabo, C.; Scharling Holm, J.; Jensen, P.; Andersson, M.; Lulu Neilsen, E. MEGAVIND Wind power annual research and innovation agenda, 2018.

2. Ziegler, L.; Gonzalez, E.; Rubert, T.; Smolka, U.; Melero, J.J. Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK. *Renewable and Sustainable Energy Reviews* **2018**, *82*, 1261–1271.

3. Verbruggen, S. Onshore Wind Power Operations and Maintenance to 2018, 2016.

4. Stock-Williams, C.; Swamy, S.K. Automated daily maintenance planning for offshore wind farms. *Renewable Energy* **2018**. doi:https://doi.org/10.1016/j.renene.2018.08.112.

5. Canizo, M.; Onieva, E.; Conde, A.; Charramendieta, S.; Trujillo, S. Real-time predictive maintenance for wind turbines using Big Data frameworks. 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), 2017, pp. 70–77. doi:10.1109/ICPHM.2017.7998308.

6. Zhou, K.; Fu, C.; Yang, S. Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews* **2016**, *56*, 215 – 225. doi:https://doi.org/10.1016/j.rser.2015.11.050.

7. Wymore, M.L.; Dam, J.E.V.; Ceylan, H.; Qiao, D. A survey of health monitoring systems for wind turbines. *Renewable and Sustainable Energy Reviews* **2015**, *52*, 976 – 990. doi:https://doi.org/10.1016/j.rser.2015.07.110.

8. Jha, S.K.; Bilalovic, J.; Jha, A.; Patel, N.; Zhang, H. Renewable energy: Present research and future scope of Artificial Intelligence. *Renewable and Sustainable Energy Reviews* **2017**, *77*, 297 – 317. doi:https://doi.org/10.1016/j.rser.2017.04.018.

9. Morgenthal, G.; Hallermann, N. Quality assessment of unmanned aerial vehicle (UAV) based visual inspection of structures. *Advances in Structural Engineering* **2014**, *17*, 289–302.

10. Chen, X. Fracture of wind turbine blades in operation—Part I: A comprehensive forensic investigation. *Wind Energy* **2018**, p. 06 June. doi:https://doi.org/10.1002/we.2212.

11. Wang, L.; Zhang, Z. Automatic detection of wind turbine blade surface cracks based on uav-taken images. *IEEE Transactions on Industrial Electronics* **2017**, *64*, 7293–7303.

12. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, Vol. 1, pp. I–I.

13. Shihavuddin, A.; Arefin, M.M.N.; Ambia, M.N.; Haque, S.A.; Ahammad, T. Development of real time Face detection system using Haar like features and Adaboost algorithm. *International Journal of Computer Science and Network Security (IJCSNS) 2010 pp 171-178* **2010**, *10*.

14. Friedman, J.; Hastie, T.; Tibshirani, R.; others. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* **2000**, *28*, 337–407.

15. Liaw, A.; Wiener, M.; others. Classification and regression by randomForest. *R news* **2002**, *2*, 18–22.

16. Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* **1995**, *20*, 273–297.

17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *nature* **2015**, *521*, 436.

18. Hauberg, S.; Freifeld, O.; Larsen, A.B.L.; Fisher, J.; Hansen, L. Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. Artificial Intelligence and Statistics, 2016, pp. 342–350.

19. Shihavuddin, A.; Chen, X. DTU - Drone inspection images of wind turbine, 2018. This dataset set has temporal inspection images for the years of 2017 and 2018 of the 'Nordtank' wind turbine at DTU Wind Energy's test site in in Roskilde, Denmark.

20. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015, pp. 91–99.

21. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621* **2017**.

22. Blu, T.; Thévenaz, P.; Unser, M. Linear interpolation revitalized. *IEEE Transactions on Image Processing* **2004**, *13*, 710–719.

23. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; others. Speed/accuracy trade-offs for modern convolutional object detectors. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

24. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. International Conference on Machine Learning, 2015, pp. 448–456.

25.   He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

26.   Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. AAAI, 2017, Vol. 4, p. 12.

27.   Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Software available from tensorflow.org.

28.   Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in COntext. European Conference on Computer Vision (ECCV). Springer, 2014, pp. 740–755.