XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Visual detection of surface defects in wind turbine blades using two step image segmentation

Stevan H. R. Góes[2], Carlos G. Carvalho Jr[2], André R. F. Oliveira[1], Jean L. Silva[1], Rafael Tasinato[1], Hermes Senger[2], Edson S. Gomi[1]

[1]*University of São Paulo*
Av. Prof. Luciano Gualberto, nº380, 05508-010, São Paulo/SP, Brazil
gomi@usp.br
[2]*Federal University of São Carlos*
Rod. Washington Luís, 235-A, 13565-905, São Carlos/SP, Brazil
hermes@ufscar.br

Abstract. Automated visual inspection of wind turbine blades (WTBs) using drone imagery is a crucial yet complex task, primarily due to the large scale of the structures, varying lighting conditions, and the wide variety of potential surface defects. These challenges demand solutions that are both precise and scalable. To address this, we propose a robust two-stage segmentation pipeline designed to efficiently localize and analyze defects in real-world drone images of WTBs. Our approach leverages two cascaded deep learning models operating at different spatial resolutions. In the first stage, a global segmentation model identifies and isolates the blade from its background, terrain, or turbine tower by generating a binary mask. This step significantly reduces the search space, focusing the analysis on the most relevant region and minimizing computational overhead. In the second stage, the extracted blade area is subdivided into overlapping high-resolution patches. These patches retain detailed texture and structural information necessary for accurate defect recognition. A specialized segmentation network is then applied to these patches, producing multi-channel binary masks where each channel represents a specific defect class, such as adhered material, erosion, mechanical impact and mechanical failure. This hierarchical architecture offers several key advantages: enhanced computational efficiency by concentrating fine-grained analysis on smaller regions of interest; modularity that allows each model to be optimized independently; and support for multi-class segmentation, enabling simultaneous detection of multiple defect types—an essential feature in real-world inspections, where composite damage is common. Moreover, the spatially coherent masks produced by both stages allow for geometric post-processing. This makes it possible to reconstruct the location of defects across the entire blade, enabling the measurement of geometric attributes such as defect area, relative position, and clustering. Such spatial insights are vital for guiding maintenance operations and supporting predictive maintenance strategies. By decomposing the task into global blade localization and local defect segmentation, our pipeline balances accuracy, efficiency, and flexibility. In summary, the proposed method offers a scalable and extensible framework for reliable WTB inspection, with strong potential for real-world deployment and further research exploration.

Keywords: Segmentation, visual inspection, wind turbine blades, neural networks and deep learning.

1 Introduction

Wind energy is critical to the global transition to sustainable energy. Wind turbine blades (WTBs) play a crucial role in energy conversion, but they are vulnerable to a variety of defects—such as erosion, cracks, impacts, lightning strikes, and materials adhering to the blade surface—that affect performance and safety. Inspecting WTBs reliably is essential but challenging due to their size, complex shapes, and harsh environments. Traditional manual inspections using rope access or platforms are slow, risky, and subjective, often missing subtle defects [1].

To address this, the industry increasingly uses drones (UAVs) with high-resolution cameras for fast, contactless inspections. However, analyzing the large volumes of images remains difficult, highlighting the need for automated systems that accurately detect and classify multiple defect types [2].

Deep learning, especially convolutional neural networks (CNNs) and U-Net architectures, has advanced semantic segmentation in many fields, including defect detection. Improvements such as multi-stage segmentation, attention mechanisms, and domain adaptation have further boosted results [2]. Building on this, we propose a two-

stage deep learning pipeline designed to handle the large spatial scale of WTBs and perform detailed multi-class defect segmentation [3].

The first stage performs global segmentation to separate the blade from complex backgrounds, producing a binary mask that reduces the search area and computational cost. The second stage divides the segmented blade into overlapping high-resolution patches, which a specialized U-Net analyzes pixel-wise to classify defects like erosion, adhered materials, and mechanical damage [4].

This hierarchical design balances high-resolution analysis with efficiency by focusing only on relevant regions. Its modularity allows separate training and optimization of each stage, supporting transfer learning and architectural tuning. Multi-class segmentation is vital since blades often have multiple defect types simultaneously [4]. The spatially coherent masks enable geometric post-processing to map defect locations and calculate metrics such as area and clustering, aiding maintenance planning and structural health monitoring [3].

Moreover, the pipeline integrates well with predictive maintenance and digital twin systems by converting raw images into structured data for trend analysis and risk assessment. Tested on real UAV data, it handles challenges like lighting variations, blade curvature, perspective distortion, and occlusion, suitable for both onshore and offshore farms.

In summary, we present a scalable, accurate framework for automated WTB inspection using UAV imagery. By combining global localization with local multi-class segmentation, our method improves efficiency and modularity, advancing autonomous, data-driven maintenance in wind energy to enhance safety, reduce downtime, and lower costs [5].

2 Related work

This section reviews recent methodologies for classifying and detecting defects in wind turbine blades (WTBs), with a focus on approaches that leverage image processing and deep learning for automation in renewable energy systems. Several studies emphasize preprocessing techniques to isolate regions of interest and improve analysis efficiency. For example, [5] proposes a regression crop technique to extract only the blade surface from UAV images, enhancing defect detection by removing irrelevant background data. In [6], an optimized U-Net is developed for infrared imagery, using hierarchical depthwise separable convolutions to reduce complexity while maintaining segmentation accuracy—ideal for real-time or resource-limited environments. The model in [7] extends U-Net with dual attention mechanisms—channel and spatial attention—to refine feature weighting and improve segmentation under varying lighting and surface conditions. Pixel U-Net [8] further enhances segmentation by replacing traditional pooling with pixel shuffle operations, preserving spatial resolution and enabling better detection of fine-grained features across diverse imaging scenarios.

For surface defect detection, [1] adapts YOLOv5 to improve sensitivity to small anomalies on turbine blades. The optimized architecture supports early detection of subtle damage, reducing downtime and maintenance costs. In image classification, AlexNet [9] remains influential due to its capacity for high-level feature extraction and compatibility with transfer learning. Its reuse for WTB inspection allows effective training on limited labeled data, a common challenge in industrial settings [3].

Finally, [10] presents a hybrid framework combining Otsu thresholding, transfer learning, and random forest classification. This integration outperforms traditional models like SVM, offering greater robustness and generalization for real-world defect detection.

3 Metodology

The proposed methodology uses artificial intelligence and data analytics to improve maintenance and increase the operational reliability of wind turbines. The process begins with the use of deep neural networks applied to high-resolution images to automatically detect WTBs. Once detected, the region is processed and fed to another deep neural network capable of segmenting different kinds of defects.

The pipeline developed in this work proposes a solution based on computer vision and deep neural networks for the detection and classification of defects in wind turbine blades (WTBs). The process begins with the capture of high-resolution images by drones, which record the blades in fine detail. These images then undergo an automatic segmentation step that isolates the blade from the background using a neural network specifically trained for this purpose (Fig. 1). With the blade segmented, the image is divided into smaller blocks. Each block then undergoes a secondary segmentation process to isolate the defects on the WTB.

Automated visual inspection of WTBs using drone imagery faces challenges due to blade size, varied surface defects, and the need for high-resolution analysis. To tackle these, we propose a double segmentation pipeline, consisting of two cascaded deep learning models. This enables efficient defect localization by first identifying the

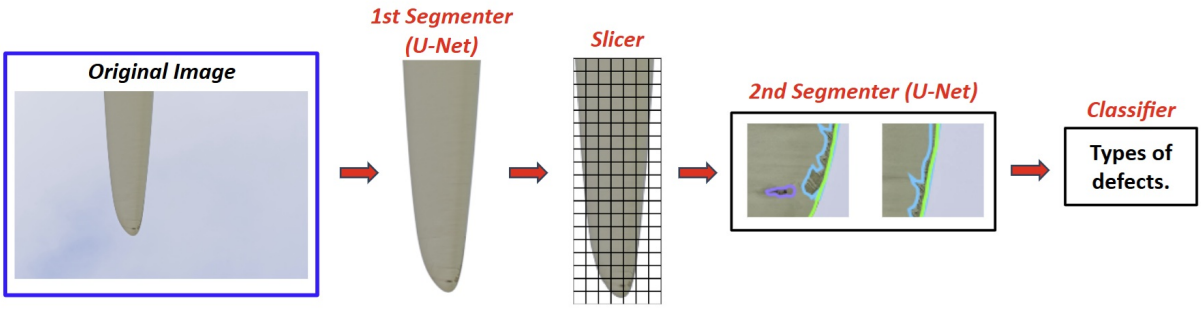


Figure 1. Workflow for defect classification.

blade globally, then detecting fine defects within localized patches (Fig. 1) .

The first stage performs global segmentation, distinguishing the blade from background elements (sky, terrain, turbine parts) in high-resolution images. This narrows the region of interest, producing a binary mask used to crop the blade area. Next, the second stage subdivides this area into overlapping high-resolution patches to detect structural defects like cracks, erosion or adhered material. Each patch is processed by a second segmentation network producing multi-channel binary masks, each channel representing a defect type.

This hierarchical method handles different spatial scales: global localization followed by detailed analysis. It improves computational efficiency by focusing on relevant subregions rather than the entire image, reducing inference time and enabling scalability for large datasets or real-time use. The modular design allows independent optimization and easier experimentation without affecting the entire pipeline.

Models based on two-stage segmentation usually require higher computational cost, since they involve additional network processing, leading to increased inference time and memory usage compared to single-stage approaches [11]. However, this extra cost can be justified by the accuracy gain, especially in small or difficult-to-segment regions [12]. In some cases, the overhead can be reduced when the second stage is applied only to regions of interest [13].

The pipeline supports multi-class segmentation, detecting multiple defect types simultaneously—critical in real scenarios where blades have several damages. Quantifying defect extent enables detailed maps for maintenance planning and repair prioritization.

3.1 Optimizing UNet-MiT Training

In our experiments, we used a UNet architecture with a MiT (Mix Transformer) encoder, trained for 40 epochs using the AdamW optimizer with a constant learning rate of 0.0001. This setup provided stable convergence and good initial performance in defect segmentation tasks.

In the latest experiments, we introduced two key improvements. First, we applied Exponential Moving Average (EMA) to the model weights with a decay factor of 0.99. This technique helped stabilize training and improved validation performance by reducing noise in weight updates.

Second, we enhanced the data augmentation strategy by combining RandomCrop and Resize. This approach increased data variability and encouraged the model to learn features at different scales and positions—important for detecting diverse defects on wind turbine blades.

These additions led to more consistent results across training runs, improved generalization on validation data, and reduced overfitting. EMA contributed to smoother loss curves and better model stability, while the new augmentation strategy helped the model focus on relevant local features.

Oversampling was applied to address imbalanced datasets, where one or more classes have significantly fewer examples than the others. This technique increases the frequency of examples from the rarer classes, causing them to appear more often during training. As a result, it helps balance the class distribution and enables the model to better learn the characteristics of the minority classes.

In this work, we applied *oversampling* to address the imbalance between defect classes. This technique artificially increases the number of samples from minority classes, leading to better balance and improved robustness of the model when detecting rare defects. Prior studies show that oversampling often outperforms other resampling methods without causing significant overfitting in convolutional neural networks, particularly under high imbalance scenarios [14, 15].

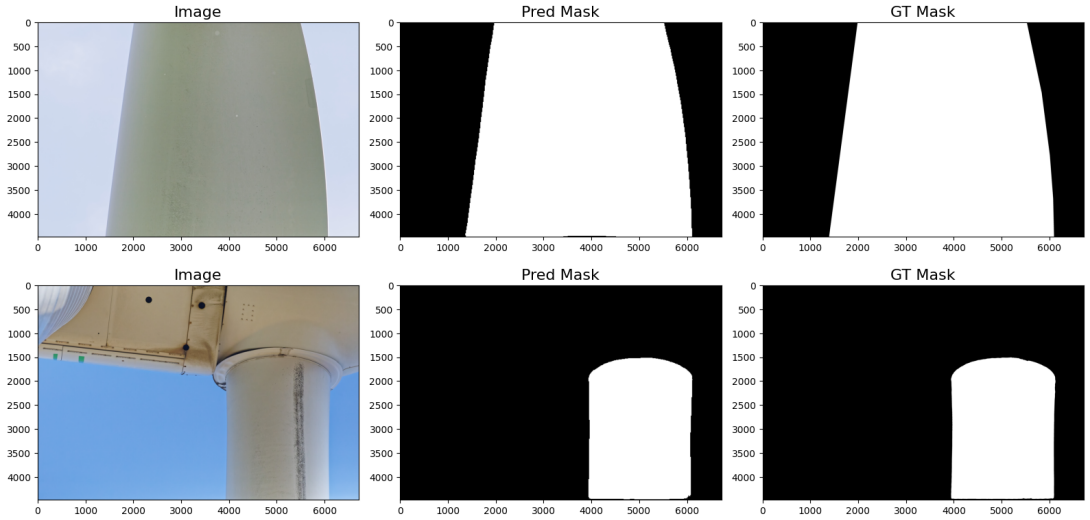


Figure 2. Results of the first segmentation using UNet.

3.2 Dataset Construction and Preprocessing

The success of a deep learning-based segmentation pipeline heavily depends on the quality and structure of the underlying dataset. The construction process leverages manual annotations from drone imagery of WTBs, systematically processed through custom scripts to generate task-specific datasets.

For the first segmentation task, which involves detecting the full blade within the drone image, we utilize annotations exported directly from CVAT (Computer Vision Annotation Tool). These annotations are provided as polygonal regions that encompass the entire blade. A preprocessing script converts these polygons into pixel-wise binary masks where the blade area is labeled as foreground and all other regions as background. These masks serve as the ground truth for training the first segmentation model, which operates on full-resolution images.

The accurate extraction of these masks ensures that the model learns to identify blades under varying environmental conditions, orientations, and lighting scenarios. Some results of this segmentation process are shown in Fig. 2.

The second segmentation task requires a more intricate preprocessing pipeline. After applying the first model to localize the blade, the script crops the blade region from the original image and slices it into smaller, square patches of fixed size.

These patches retain high resolution and focus exclusively on relevant areas. For each patch, defect-specific annotations from CVAT are retrieved and converted into multi-channel binary masks. Each channel corresponds to a distinct defect type, such as cracks, erosion, or adhered material. This format enables the training of multi-class segmentation models capable of distinguishing between various defect categories simultaneously.

For the construction of the dataset used in this study, 3482 images of wind turbine blades were collected and organized, containing various types of visual defects. These images were subsequently annotated using the CVAT, resulting in a total of 4267 annotations, indicating that some images contain more than one identified defect. The annotations were organized into defect categories based on the origin and nature of the damage. The classifications are adhered material, erosion, scratches and mechanical impacts, and mechanical failures. Adhered material includes defects like coating contamination caused by foreign materials on the blade surface. Erosion covers early and advanced erosion from natural agents such as wind, rain, or solid particles. Scratches and mechanical impacts range from light and moderate scratches to damage on the upper layer. Mechanical failures include lubricant oil leakage, vortex generator damage, protection tape damage, and mechanical actuator failures.

An important consideration in this stage is patch sampling. To ensure a balanced dataset and adequate representation of all defect types, we implemented a sampling strategy that biases the selection towards patches containing at least one defect. Non-defective patches are still included to preserve context and prevent overfitting, but their proportion is carefully regulated.

Furthermore, we store each patch and its corresponding masks in TIFF format to retain spatial precision and support multi-channel annotations without compression artifacts. Metadata, including original coordinates and defect labels, is saved alongside, enabling spatial reconstruction in later stages.

The Mean Intersection over Union (mIoU) and Dice Similarity Coefficient (DICE) metrics were employed to evaluate the quality of defect segmentation on wind turbine blades. Both metrics quantify the overlap between the predicted segmentation and the ground truth, yet they differ in their formulation and sensitivity.

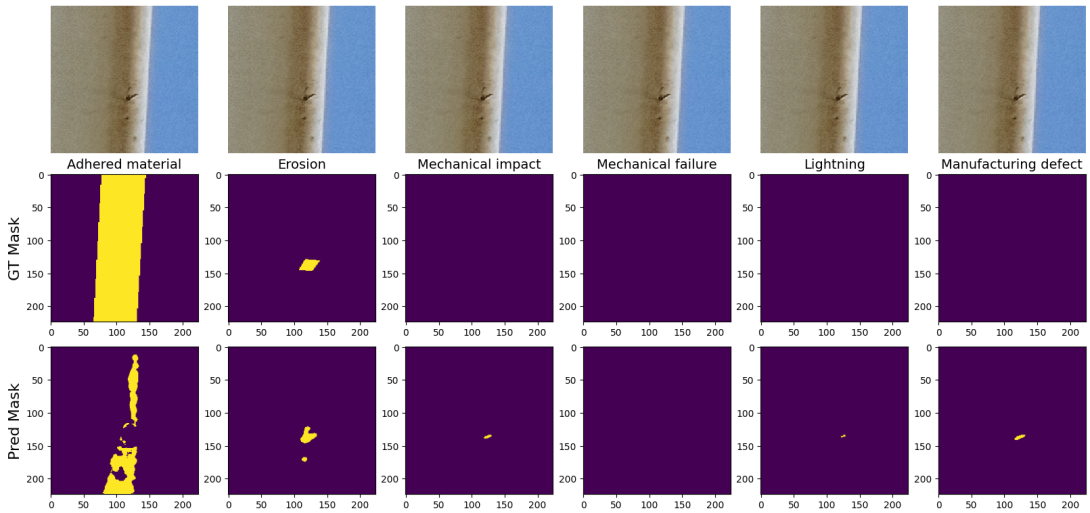


Figure 3. Results of the second segmentation using UNet with focal and dice loss.

4 Results and Discussions

The first U-Net segmentation, based on the classic encoder-decoder design, performs binary segmentation of the blade. Given the high contrast and smooth contours of blades in aerial images, this step converges quickly and achieves near-perfect results even with standard configurations. The second U-Net segmentation is more complex, aimed at multi-label segmentation to detect different types of defects in small patches (Fig. 4). Each output channel corresponds to a specific defect class. For this, we used deeper encoders such as ResNet34 and EfficientNet, leveraging pretrained weights to improve feature extraction and accelerate convergence.

Loss function design was essential due to the high class imbalance in the second task. While Cross-Entropy loss performed poorly, a combination of Dice and Focal Loss significantly improved defect segmentation, especially for small or sparse regions. After correcting a bug in the evaluation pipeline, improvements in IoU and Dice coefficients became clear (Fig. 5). Both networks were trained independently. The first U-Net used full images resized to 1024×1024, while the second used 256×256 patches with strong augmentation. Optimizers and hyperparameters were tuned separately to suit each task.

4.1 Validation of the Image Segmentation Process using mIoU and DICE Metric

The Adhered Material class showed the best performance overall, with mIoU = 0.5843 and DICE = 0.6187. This supports the notion that defects with well-defined edges and consistent textures tend to be segmented more accurately.

The Mechanically-Induced Failures class also achieved strong results, with mIoU = 0.5774 and DICE = 0.6023. This indicates that the segmentation model was able to identify this defect category with relatively high precision. Such performance can be attributed to the visual consistency of the defects, the clear definition of their contours in the images, and the higher representation of this class in the training dataset — all factors that contribute to more effective learning and detection.

In contrast, the Erosion class demonstrated lower performance, with mIoU = 0.4220 and DICE = 0.4658. This can be explained by the diffuse and subtle nature of erosion, which often involves smooth transitions between damaged and intact areas, making precise segmentation more challenging. Additionally, variations in lighting and surface texture may exacerbate this difficulty.

Erosion and scratch classes showed lower precision due to the subtle nature of these defects in RGB images. Integrating *multispectral or infrared* images can improve detection by providing additional physical and thermal information. Studies show that fusing RGB and thermal data increases accuracy in identifying surface and thermal faults in industrial components [16, 17].

The Scratch and Mechanical Impact class obtained the lowest performance across all metrics, with mIoU = 0.2818 and DICE = 0.3096. This may be due to the small scale and low visual salience of these defects, which can easily blend into the background or be interpreted as image noise. These results suggest that the current model still lacks the sensitivity required to detect fine or low-contrast features.

A consistent pattern observed across all classes is the higher DICE values compared to mIoU, which aligns

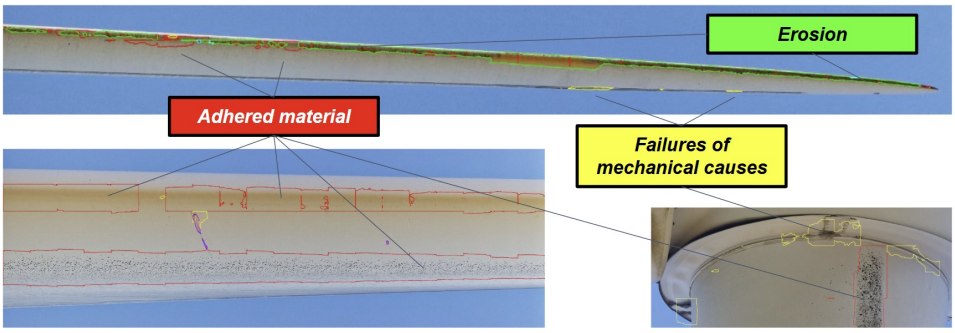


Figure 4. Exemple the Image Segmentation Process for Defects in WTB

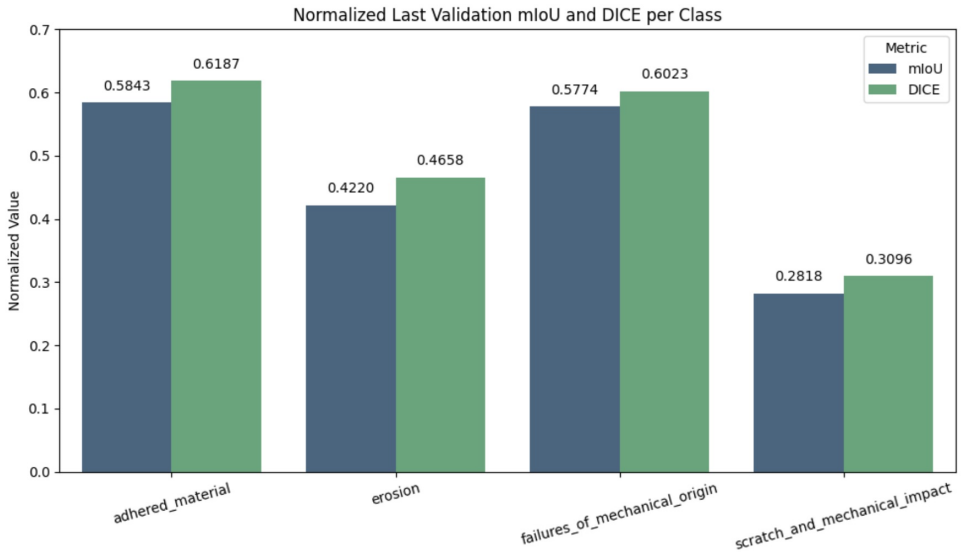


Figure 5. Validation of the Image Segmentation Process for Defects in WTB

with findings in the literature. The DICE metric tends to be more forgiving of small inaccuracies along object boundaries, emphasizing the overall overlap between predicted and ground truth regions. In contrast, mIoU is more stringent, heavily penalizing false positives and negatives. This difference is particularly relevant for classes with small or irregular shapes, where relative errors have a greater impact on metric scores.

5 Conclusions

This work presented a two-step segmentation pipeline for automated inspection of wind turbine blades (WTBs) using drone imagery. The proposed architecture proved effective by combining global segmentation for blade localization with detailed local segmentation for multi-class defect detection. Experimental results demonstrated the robustness of this approach, especially for defects with well-defined edges and consistent visual appearance. The use of two separately trained U-Net models brought key advantages in terms of modularity, computational efficiency, and adaptability. The initial segmentation stage accurately isolated the blade from the background, even under challenging lighting and perspective conditions. The second stage, focused on high-resolution patch-based segmentation, delivered promising results despite challenges related to underrepresented classes and subtle visual features.

Quantitative analysis using mIoU and DICE metrics highlighted important patterns, such as the superior performance of the DICE metric for small or irregularly shaped defects. Techniques such as oversampling, data augmentation, and the use of modern encoders (e.g., MiT and EfficientNet) contributed significantly to improved accuracy and generalization.

The model trained on onshore wind farm images is unlikely to generalize well to offshore wind farms due to harsher conditions, such as strong winds, fog, and water reflections; however the proposed pipeline should work properly if trained on offshore images.

Future work may explore integrating this pipeline with predictive maintenance and digital twin systems,

enabling automated, time-aware analysis of blade condition. Furthermore, adding modules for severity estimation, repair prioritization, and risk classification could greatly enhance the practical impact of this approach in the wind energy sector. In summary, this study provides a solid and scalable foundation for the advancement of autonomous visual inspection systems with real-world industrial applicability.

Acknowledgements. We gratefully acknowledge the support of the RCGI – Research Centre for Greenhouse Gas Innovation, hosted by the University of São Paulo (USP), sponsored by FAPESP – São Paulo Research Foundation, and sponsored by TotalEnergies, and the strategic importance of the support given by ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&DI levy regulation. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] R. Zhang and C. Wen. Sod-yolo: A small target defect detection algorithm for wind turbine blades based on improved yolov5. *Advanced Theory and Simulations*, vol. 5, n. 7, pp. 2100631, 2022.
[2] P. Rizk, F. Rizk, S. S. Karganroudi, A. Ilinca, R. Younes, and J. Khoder. Advanced wind turbine blade inspection with hyperspectral imaging and 3d convolutional neural networks for damage detection. *Energy and AI*, vol. 16, pp. 100366, 2024.
[3] M. Tan and Q. V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*, 2019.
[4] L. Peng and J. Liu. Detection and analysis of large-scale wt blade surface cracks based on uav-taken images. *IET Image Processing*, vol. 12, n. 11, pp. 2059–2064, 2018.
[5] W. Zhou, Z. Wang, M. Zhang, and L. Wang. Wind turbine actual defects detection based on visible and infrared image fusion. *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–8, 2023.
[6] J. Yu, Y. He, H. Liu, F. Zhang, J. Li, G. Sun, X. Zhang, R. Yang, P. Wang, and H. Wang. An improved u-net model for infrared image segmentation of wind turbine blade. *IEEE Sensors Journal*, vol. 23, n. 2, pp. 1318–1327, 2023.
[7] L. Wang, J. Yang, C. Huang, and X. Luo. An improved u-net model for segmenting wind turbines from uav-taken images. *IEEE Sensors Letters*, vol. 6, n. 7, pp. 1–4, 2022.
[8] S. Z. Rizvi, M. Jamil, and W. Huang. Enhanced defect detection on wind turbine blades using binary segmentation masks and yolo. *Computers and Electrical Engineering*, vol. 120, pp. 109615, 2024.
[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105, 2012.
[10] X. Yang, Y. Zhang, W. Lv, and D. Wang. Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. *Renewable Energy*, vol. 163, pp. 386–397, 2021.
[11] K. Wang and others. Organ at risk segmentation in head and neck ct images by using a two-stage segmentation framework based on 3d u-net. *International Journal of Radiation Oncology, Biology, Physics*, vol. 102, n. 2, pp. S694–S695, 2018.
[12] X. Liu and others. Dual knowledge-guided two-stage model for precise small organ segmentation in abdominal ct images. *IET Image Processing*, vol. 18, n. 7, pp. 1345–1356, 2024.
[13] H. Chen and others. Two-stage framework for faster semantic segmentation. *Sensors*, vol. 23, n. 7, pp. 3451, 2023.
[14] M. Buda, A. Maki, and M. A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, vol. 106, pp. 249–259, 2018.
[15] H. Li, Y. Zhang, X. Chen, and others. On-the-fly image-level oversampling for imbalanced datasets of manufacturing defects. *IEEE Transactions on Industrial Informatics*, vol. 20, n. 5, pp. 5123–5134, 2024.
[16] S. Svystun, P. Radiuk, O. Melnychenko, O. Savenko, and A. Sachenko. Yolo ensemble for uav-based multispectral defect detection in wind turbine components. *arXiv preprint arXiv:2509.04156*, 2025.
[17] X. Zhang and Y. Zhang. Progress in active infrared imaging for defect detection in photovoltaic panels. *Sensors*, vol. 23, n. 21, pp. 8780, 2023.