A Computer Vision Deep Learning Tool for Automatic Recognition of Bearing Failure Modes

Stephan Baggeröhr[1], Sebastián Echeverri Restrepo [1,2], Mourad Chennaoui [1], Christine Matta [1] and Cees Taal[1],

[1] *SKF, Research and Technology Development Center, Houten, the Netherlands*
stephan.baggerohr@skf.com
cees.taal@skf.com
sebastian.echeverri.restrepo@skf.com
mourad.chennaoui@skf.com
christine.matta@skf.com

[2] *Department of Physics, King's College London, London, United Kingdom*

Abstract

We introduce an object detection model specifically designed to identify failure modes in images of bearing components, including the inner ring, outer ring, and rolling elements. The method effectively detects and pinpoints failure features, subsequently determining the associated failure mode within the image. With images sourced from real-world bearing applications, our model can recognize various ISO-failure modes such as surface-initiated fatigue, abrasive wear, adhesive wear, moisture corrosion, fretting corrosion, current leakage erosion, and indents from particles. The proposed model could be used in an assistive tool where failure modes give insights on how to prolong average future bearing life in an asset and therefore reduce related costs and environmental impacts.

1. Introduction

Bearings are extensively utilized in a wide range of rotating equipment and are essential for ensuring their proper function. Bearing failures can lead to unplanned downtime with unforeseen costs, or even result in hazardous situations. Sensor-based condition monitoring has been an important tool for the prediction of these undesired events and are a key ingredient for a predictive maintenance strategy (Randall & Antoni, 2011). In this paper, the focus is on a subsequent stage after sensor-based fault detection, that is, a visual inspection of the replaced disassembled bearing to further prolong the average future bearing life in an asset (SKF, 2017).

A visual inspection of the bearing provides additional information on how to prevent problems from reoccurring. This includes altering the bearing design, lubrication or operation and maintenance procedures. Another important application of bearing inspections is quantifying its damage severity, such as spall size. This information can be fed back to sensor-based condition monitoring systems enabling supervised machine learning for bearing diagnostics and prognostics. Inspections are also being used to determine whether a bearing qualifies for remanufacturing (Chiarot, Cooper Ordoñez, & Lahura, 2022). Remanufacturing is a process which enables re-using the bearing by means of polishing or grinding its components, potentially doubling its life. To summarize, visual bearing inspections can significantly prolong average future bearing life in an asset and therefore reduce related costs and environmental impacts, e.g., due to the manufacturing process of the bearing.

Unfortunately, visual postmortem analysis of bearings require an application engineer with many years of experience, which is something not always readily available. This limits its scalability to be applied to a large population of bearings used in an asset. In this work we propose an image-based deep learning algorithm, which can assist the technician replacing the bearing. For example, a picture can be taken of the bearing components with a smart-phone, where the software automatically provides insights on proposed maintenance actions, altering bearing designs, its remanufacturability and provide an automated connection in supervising condition monitoring algorithms.

Bearing failures can occur due to a wide variety of reasons (Liu & Zhang, 2020), where each failure category can lead to a unique footprint observable during visual inspection (SKF, 2017). The different categories of bearing failures have been standardized and well described in (ISO-15243-2017, 2017), also referred to as bearing failure modes, where, in total,

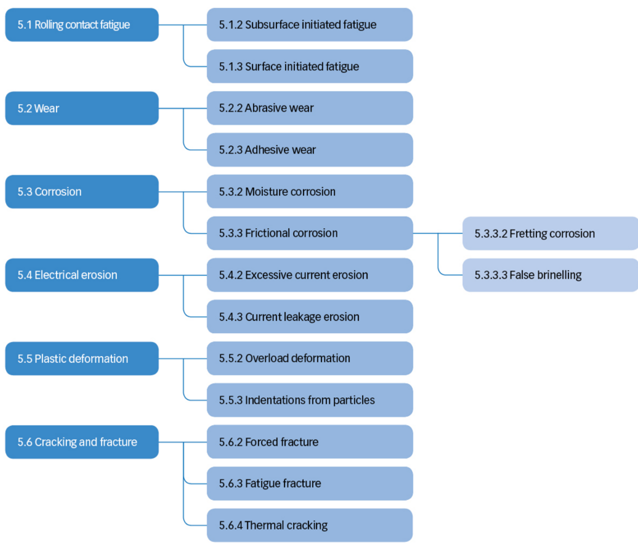

Figure 1. *ISO 15243-2017 failure mode classifications. Image taken from (SKF, 2017).*

seven main categories of failure modes are proposed. An overview of the different failure modes is shown in Figure 1. In Figure 2 an overview is shown on the most common failure modes based on collected statistics from bearing inspections (SKF, 2017).

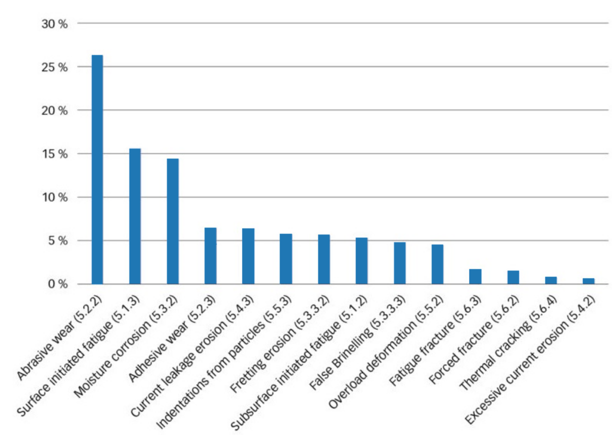

Figure 2. *An example of SKF's field failure statistics, detailing the frequency of various failure modes. Image taken from (SKF, 2022).*

Applying deep learning algorithms to automate visual inspections in PHM applications is not new. A significant amount of work has been done in the field of structure health monitoring. Examples include crack detection in concrete structures caused by, e.g., changing loading and corrosion (Azimi, Eslamlou, & Pekcan, 2020). More examples can be found from the steel industry, that is, detection and classification of steel surface defects (Fu et al., 2019; Wang, Xia, Ye, & Yang, 2021). However, to the authors knowledge there is no specific method to classify bearing failure modes.

In this work a framework of selecting a deep-learning based object detection model is introduced. The object detection model is specifically tasked to identify failure modes in images of bearing components, including the inner ring, outer ring, and rolling elements. This model effectively detects and pinpoints failure features, subsequently determining the associated failure mode within the image. As a first step, the selected model is trained for the top 7 most common failure modes, namely: abrasive wear, surface-initiated fatigue, moisture corrosion, adhesive wear, current leakage erosion, fretting corrosion, and indentations from particles (Figure 2).

2. Dataset

The foundation of bearing failure mode object detection model lies in the curated dataset. The dataset encompasses a broad spectrum of bearing images, taken from industrial centres across the globe and showcases various bearing types along with the one or more of the top seven primary failure modes identified for diagnosis. This breadth in dataset variety was crucial for the development of a model capable of accurately identifying and classifying a range of real-world bearing failures captured in their operational environment.

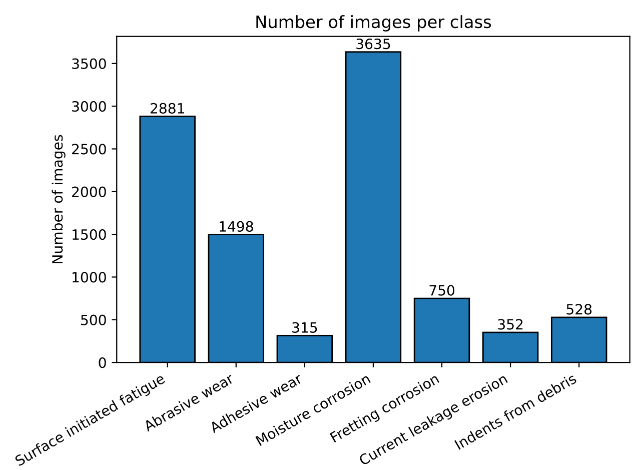

Figure 3. *Final number of images per failure class after selecting process and annotation done by expert.*

The precision in our dataset was ensured by an expert led data labelling team based on SKF employee's experience. Specialists in bearing maintenance meticulously labeled and annotated each image, drawing accurate bounding boxes around the designated failure modes. During the annotation process images were selected based on their representation of the failure mode, making sure the failure mode characteristics and features are within clear view according to the expert. Furthermore, an assessment was made on the quality of the image itself, filtering any blurry images. Images objects other than bearing components (maintenance tools, tables, etc) were also removed from the training set. In the end the dataset comprised of 11k images across the 7 chosen failure modes as shown in Figure 3. Images were normalized, padded and resized to 640x640 pixels. Additionally, augmentation techniques were applied to the images before ingesting into the model.

Here the bar graph illustrates a significant class imbalance within our object detection dataset, where certain classes are overly represented with a high number of images, while others have markedly fewer instances. This imbalance poses a challenge for effective model training, as it can lead to biases towards the more prevalent classes, potentially compromising the model's ability to accurately detect and classify less represented objects. Addressing this issue is crucial for enhancing the model's overall performance and ensuring a balanced sensitivity across all classes. To overcome this, as a first step, the shift-scale-rotate augmentation was applied with a rotation limit set to +/- 15 degrees. This method involves stochastic affine transformations that adjust the original images through shifting, scaling, and rotating. Such transformations significantly increase the dataset's variety without the need to collect new samples.

3. Proposed Methodology

The methodology employed in developing an object detection model aimed at detecting failure modes in images of bearings was twofold: firstly, leveraging out-of-the-box (pretrained) models, and secondly, fine-tuning these models on the earlier described dataset split into an 80%-20% training and test set, respectively.

3.1. Model Selection

To determine the optimal pre-trained model for our application, we conducted a comparative analysis of several state-of-the-art models. Each model was evaluated using its default parameters, with the only modifications being the image size and batch size. Specifically, all models were trained with images resized to 640x640 pixels and a batch size of 4. The models included in the study were as follows with their respective backbone (Zou, Chen, Shi, Guo, & Ye, 2023):

- EfficientDet (D0)
- EfficientDet (D4)
- Retinanet (Resnet - 101 - 2x)
- Retinanet (Resnet - 101 - 1x)
- Retinanet (Swin)
- Yolo-x (Yolo - Tiny)

The models were compared using the COCO metric. The COCO metric, used for evaluating object detection models, includes several key components: Average Precision (AP) and Average Recall (AR) across multiple IoU thresholds (0.50 to 0.95). The metric also evaluates performance across different object sizes (small, medium, large), providing a comprehensive and standardized assessment of a model's detection capabilities. This robust evaluation ensures accurate localization and detection across varied conditions (Lin et al., 2014).

3.2. Training the Model

The dataset, characterized by class imbalance among different failure modes, necessitated a tailored approach to model training. To mitigate the effects of class imbalance, focal loss was integrated into the model's loss function (Lin, Goyal, Girshick, He, & Dollár, 2017). This modification aimed to amplify the loss associated with misclassified examples, particularly those from underrepresented classes, thereby enhancing the model's sensitivity to such cases. The models were trained with a learning rate of 1e-4 for 20 epochs.

One of the paramount challenges encountered during training was the potential for overfitting. To counteract this, techniques such as early stopping, layer normalization, and weight decay were employed. Additionally, model performance was evaluated using the test set to ensure generalizability beyond the training data. Early stopping as a regularization technique was also used to prevent overfitting, by halting the training process before the model's performance on the test set starts to degrade. By terminating the training at this optimal point, early stopping ensures that the model retains its ability to generalize well to new, unseen data, thereby mitigating overfitting and improving the model's overall predictive performance.

4. Results

Figure 5 shows the results of the comparative study of different model architectures. The study revealed that EfficientDet and RetinaNet emerged as top candidates in terms of accuracy in contrast to the Yolo methods. The RetinaNet model, with its ResNet backbone, was ultimately selected based on its performance.

The implementation of early stopping mechanisms helped mitigate this risk by halting training once the test loss plateaued, as shown in Figure 6. This strategy proved invaluable in preserving the model's generalizability.

Example model predictions for the different failure modes are shown in Figure 4. Looking at the confusion matrix in Figure 7, the Retinanet model detection threshold was set in a way that left around 32% of the images without any predictions resulting in low recall. Among the images with predictions, there was a notable emphasis on precision, as evidenced by a significant number of predictions aligning along the matrix's

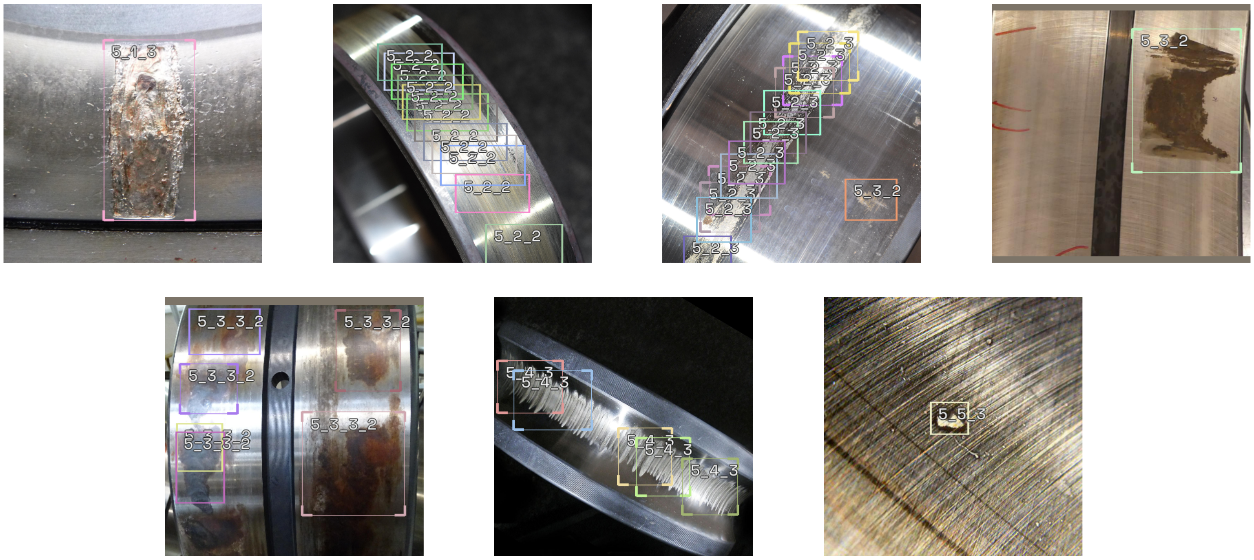

Figure 4. Example predictions for several failure modes. From left to right, top to bottom: Surface initiated fatigue, abrasive wear, adhesive wear, moisture corrosion, fretting corrosion, current leakage erosion and indentations from debris.

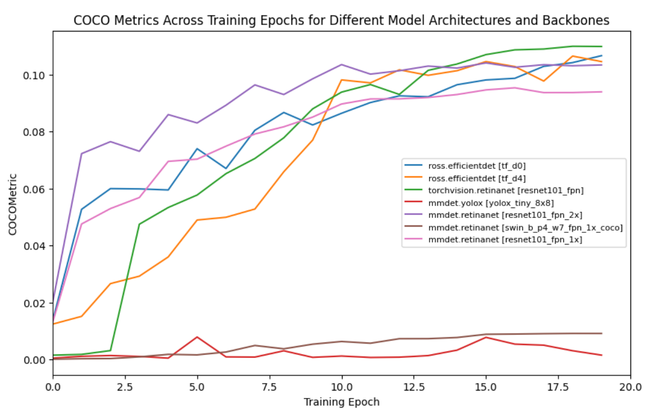

Figure 5. *Comparative analysis of investigated object detection models.*

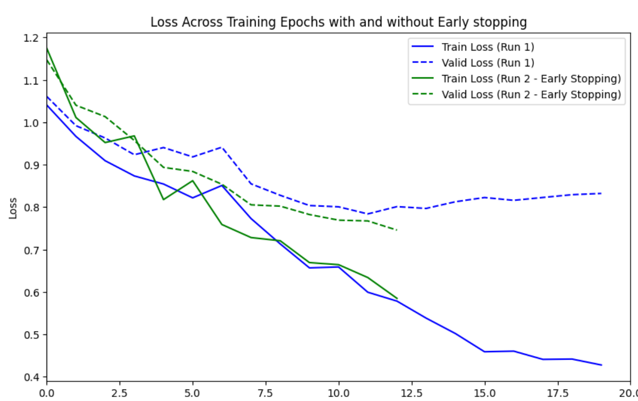

Figure 6. *Graph depicting model accuracy along epochs with and without early stopping indicated to prevent overfitting.*

diagonal.

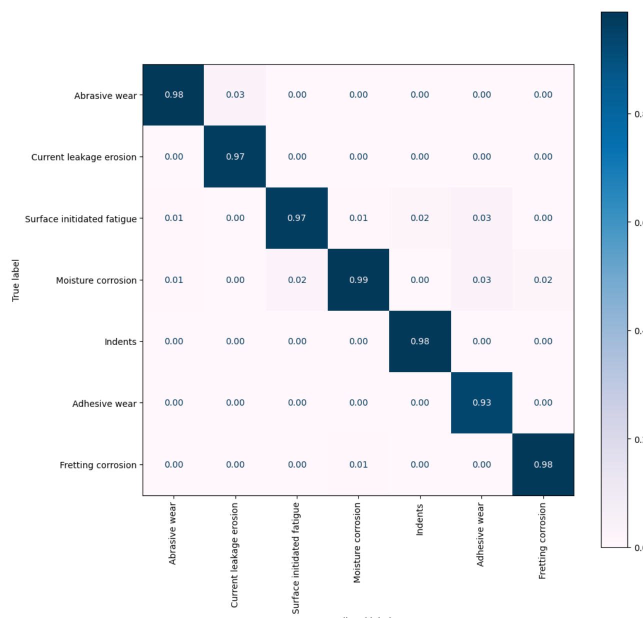

Figure 7. *Confusion matrix for the top performing model (RetinaNet - ResNet 101, fpn) model applied to the test set. Displayed results pertain exclusively to images with predictions.*

In evaluating the performance of the object detection models, we have observed a notable discrepancy between the model's precision and recall, as measured by the COCO metric system. Specifically, our model demonstrates high precision (as shown in Figure 7), indicating a strong ability to correctly identify and label objects when it decides to do so. However, this is adjacent to a significantly lower recall, suggesting that the model is more conservative in its detection, often missing objects that should have been detected. This characteristic leads to a lower overall COCO metric score, which incorporates both precision and recall into its evaluation. Despite this, the high precision of our model still presents substantial utility in specific applications where the cost of false positives is high, and accuracy in the detection of identified objects is paramount. In such scenarios, our model's ability to minimize incorrect detection — ensuring high confidence in the positive detection it makes — can be more valuable than detecting every possible object, underscoring the importance of considering application-specific requirements when evaluating model performance. Therefore, while the overall COCO metric may be lower, the high precision of our model affirms its applicability and effectiveness in contexts where precision is critically valued over recall.

5. Conclusion

The model, selected through bench-marking various neural network architectures, was trained to detect seven primary bearing failure modes, addressing challenges such as class imbalance and image rotation inconsistencies. Key to the success was the meticulous collection and preparation of images. A dataset comprising 11k images of bearings with annotated failure modes was curated to train the model. Through thorough data gathering, precise annotation, and strategic data augmentation, we created a robust dataset that improved the accuracy of the model and real-world applicability. RetinaNet, with its ResNet 101 - fpn backbone, was chosen for its performance. This work shows the feasibility of such a model to be used in an assistive tool where failure modes give insights on how to prolong average future bearing life in an asset and therefore reduce related costs and environmental impacts.

References

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, *20*(10).

Chiarot, C., Cooper Ordoñez, R. E., & Lahura, C. (2022). Evaluation of the applicability of the circular economy and the product-service system model in a bearing supplier company. *Sustainability*, *14*(19).

Fu, G., Sun, P., Zhu, W., Yang, J., Cao, Y., Yang, M. Y., & Cao, Y. (2019). A deep-learning-based approach for fast and robust steel surface defects classification. *Optics and Lasers in Engineering*, *121*, 397-405.

ISO-15243-2017. (2017). Rolling bearings—damage and failures—terms, characteristics and causes. *BSI Standards Publication*.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the ieee international conference on computer vision* (pp. 2980–2988).

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer vision–eccv 2014: 13th european conference, zurich, switzerland, september 6-12, 2014, proceedings, part v 13* (pp. 740–755).

Liu, Z., & Zhang, L. (2020). A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement*, *149*, 107002.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485 - 520.

SKF. (2017). Bearing damage and failure analysis. *SKF Group*.

SKF. (2022). *Bearing damage analysis: iso 15243 is here to help you.* Retrieved 2023-03-01, from `https://evolution.skf.com/bearing-damage-analysis-iso-15243-is-here-to-help-you/`

Wang, S., Xia, X., Ye, L., & Yang, B. (2021). Automatic detection and classification of steel surface defect using deep convolutional neural networks. *Metals*, *11*(3).

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, *111*(3), 257–276.