Towards Small Anomaly Detection

Thomas Messerer
Fraunhofer Institute for Cognitive Systems (IKS), Hansastraße 32, 80686 Munich, Germany

Keywords: FOD, Foreign Object Debris, Small, Anomaly, Detection, Airport, Ramp, Apron, ML, AI.

Abstract: In this position paper, we describe the design of a camera-based FOD (Foreign Object Debris) detection system intended for use in the parking position at the airport. FOD detection, especially the detection of small objects, requires a great deal of human attention. The transfer of ML (machine learning) from the laboratory to the field calls for adjustments, especially in testing the model. Automated detection requires not only high detection performance and low false alarm rate, but also good generalization to unknown objects. There is not much data available for this use case, so in addition to ML methods, the creation of training and test data is also considered.

1 INTRODUCTION

Loose objects that are sucked into turbines can cause tremendous damage to aircraft. These objects are called FOD (Foreign Object Debris), they need to be removed from the vicinity of an aircraft, so we want to detect these anomalies with a camera-based system. [1]

There are different systems for the detection of FOD on runways available, they often use radar, sometimes in combination with vision systems and ML (Machine Learning). Before starting and after landing, the aircraft is at the parking position (ramp / apron). At this parking position there are usually a lot of working groups, loading stuff into and out of the aircraft. During this processes objects can break, or parts could fall from one of the vehicles. Due to the heavy weight of the aircraft, fragments can break off the surface, they are also dangerous. In general the foreign objects can have any size, shape, color, texture and sometimes they are flexible. Another difficulty is changing weather and lighting conditions; in addition, there are various ground markings at the parking position.

The article (Yuan et al., 2020) provides an overview of the FOD problem and detection systems for the runway. They want to detect the material the FOD consists of, which we neglect because we consider every FOD to be dangerous. In the article (Dai et al., 2020) they use a deep learning approach to detect foreign objects in metro doors, they mainly take into account complete objects which are typically clamped there. Their use case differs from ours, especially in terms of distance; as theirs, like ours, is not covered by standard data sets, they have created their own data set. A FOD detection on runways by drones is described in (Papadopoulos and Gonzalez, 2021), they want to detect different classes of objects, and compare different models in their paper. To capture the images for their data set, the integrated cameras of different drone models were used.

To our knowledge, there is no work on ML-based FOD detection at the parking position, where FOD searches are usually performed manually. Therefore we wanted to create a transportable, camera based system, which can detect FOD at the ramp at a low cost, supporting the ramp manager in locating FOD. Our camera perspective is planned to be sloping downwards, because we wanted to start with a ground based solution. The system should watch the safety area around the aircraft from the outside, so as not to disrupt workflows within this area. This results in large distances from the camera to the edge of the monitored area, so the objects in the images can become very small. The great variety of FOD is a challenge for image processing technologies and machine learning, especially when it comes to small objects. So we wanted to explore possible approaches and ML methods to tackle the problem of FOD detection at the ramp, without using reference images.

Following contributions are made in this paper: Description of our own data set, applicable ML methods and their generalization, detection of small ob-

[1] In this paper, FOD stands for Foreign Object Debris. Generally, it can also mean Foreign Object Damage, depending on the context.

Messerer, T.
Towards Small Anomaly Detection.
DOI: 10.5220/0012459800003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 860-865
ISBN: 978-989-758-684-2; ISSN: 2184-4313

Figure 1: Example images with FOD.

jects or anomalies, data augmentations and image synthesis, open questions and future directions.

2 DATA SET

Selecting an appropriate data set is the starting point for any image analysis project. There is a FOD data set, FOD-A (Travis Munyer, 2022), with 31 classes of objects. However, we wanted images with greater distances, a different perspective and including ground markings, so we took our own pictures of exemplary objects over a large area. To get some variance in the environmental conditions of the data, the images were taken on 2 different surfaces and under different lighting conditions. A digital single-lens reflex camera with a resolution of 5184 x 3456 pixels was used to take the pictures.

To reduce the input size of the models, the images were broken down into tiles. Because different experiments were carried out in the project, different data sets with different image sizes were created and used in the course of the project. In order to simplify the image content and thus the analysis by the model, the perspective was chosen so that no surroundings or horizon are visible in the images.

As can be seen in the Figures 1, in addition to the FOD, there may also be ground markings and stains, as well as moving objects such as leaves and foil. All FOD objects can be distributed at any position in the image, it cannot be assumed that they are in the central position.

As the system normally only analyzes images without FOD, the test data set must also contain im-

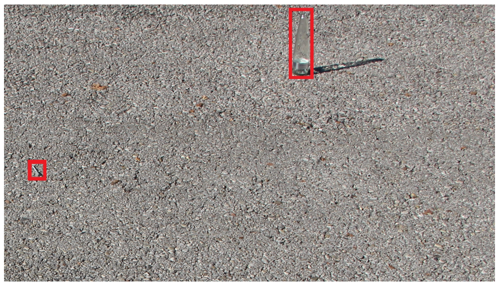

Figure 2: Example image with BoundingBoxes.

ages of the empty background to prevent false positives and false alarms. The data sets declared as ”unknown” are images with unknown objects that are not contained in the training and test data.

We decided not to label ground markings and stains, the model has to learn to ignore them. In our object detection annotations we used only one class, named ”object”, because we do not care about the actual type of the FOD. It is also difficult to have a class assignment for fragments. The same applies to the detection of materials. A black object could be made of metal or plastic, or be a broken piece of luggage. Figure 2 shows an image with 2 annotated FOD.

Especially the labeling for object detection is cumbersome, therefore we selected only some hundred images for our training data set. So we had, in addition to the small object detection, to handle the problem of the small training data pool.

3 EXPERIMENTAL METHODOLOGIES

This section describes the experiments conducted so far, dealing with the problem of little data, and methods to detect FOD.

3.1 Classification

The FOD detection is in principle an anomaly detection problem, whereby we have relatively small anomalies here in our images. The planned solution should work without any reference images of the observed area, and the model has to ignore ground markings, stains and similar.

The FOD detection can be treated as a classification problem, to distinguish between normal or anomalous scenes, this means without or with FOD. This makes annotating training data very easy, as the images only need to be sorted into two different folders. We used a ResNet50 (Kaiming He, 2015) for classification and 800 images for each class, the training to test ratio was 70/30.

After the first successful experiments, we tested several other architectures and their ability to generalize. The Table 1 also contains the number of parameters of the models, as an indication of their size and execution time.

Table 1: Accuracy [%] for different models and generalization to unknown objects.

Model	Param. count	Test (30 %)	Unknown objects
ResNet50	25.6M	98.8	95.7
ResNet18	11.7M	98.8	96.3
Inception-V3	27.2M	97.5	88.2
DenseNet121	8.0M	99.2	96.5
DenseNet161	28.7M	99.2	96.5
GoogLeNet	6.6M	99.0	93.9
ViT-Base-patch16	82.3M	98.7	97.7

Most of the results in Table 1 are in the same range, only Inception-V3 is behind. GoogLeNet is a small model with good results for test data, but in this use case the generalization is bad.

We used transfer learning for our experiments, to make use of pre-trained model parameters. To do this, we replaced the last fully connected layer in the model with a new one with the desired number of classes, here 2, and retrained the model. We tested training a network from scratch, but with transfer learning training converged faster to better accuracy values.

Our classification data set had a lot of images with big objects, so the classification for this images was easy. Classification becomes more difficult with smaller objects because the area of the object becomes (very) small in relation to the whole image. The proportion of foreign objects in the image can drop to 1% or less. For this reason, further tests were carried out with object detection.

3.2 Object Detection

With the object detection approach it should be easier to detect small anomalies, compared to classification. For the ground markings, visible aircraft parts, stains and other anomalies on the underground one can decide if these objects should be annotated or if the model should learn to ignore them. We have decided not to annotate them, so we perform a one-class object detection task with the single class ”object”. For FOD detection it is important that the model generalizes well to unknown objects, because the diversity of FOD is huge.

We used a Faster R-CNN model (Shaoqing Ren and Sun, 2015) with a ResNet-50-FPN backbone, the data set consisted of 402 images, randomly split into 80% training and 20% test data. We again used transfer learning by adjusting the number of detectable classes in the predictor, and retrained the model.

Table 2 shows the mAP (mean Average Precision) for the Pascal VOC metric (IoU = 0.5) and the COCO metric (IoU = 0.5 ... 0.95), tested with the 20% test data and three different data sets of unknown objects. The unknown datasets contain 100, 10 and 12 images.

Table 2: mAP for Pascal VOC and COCO metric.

Metric	Test (20%)	Unkn. 1	Unkn. 2	Unkn. 3
Pascal VOC	94.6	83.4	90.1	91.1
COCO	76.4	61.7	76.2	70.0

Due to the stricter constraints the numbers for COCO are lower than for the Pascal VOC metric. The difference in the results of the three unknown datasets is further evaluated.

We tested also the DINO (Mathilde Caron, 2021) vision transformer, but only the mAP value of the Pascal VOC metric improved a little bit to 95.5%, the 72.2% mAP for COCO is in the same range as the values for Faster R-CNN in Table 2.

The mAP values for the accuracy of the detection are the common starting points for a first rating of the detection performance. But in our use case, it's also important that the object detector works correctly when there are no foreign objects in the captured image; the system must not generate false alarms. If the position accuracy is not taken into account, as a simplified approach, these four cases can occur:

- True Positive (TP): There is FOD and the system has detected it.
- True Negative (TN): No FOD, everything is fine.
- False Negative (FN): The model missed FOD.
- False Positive (FP): The model has incorrectly classified an object as FOD when there is none.

We evaluated this on a dataset of 1019 images, consisting of 459 anomalous and 560 normal images, and obtained the following detection results:

- True Positive (TP): 454
- True Negative (TN): 559
- False Negative (FN): 5
- False Positive (FP): 1

This is a True Positive Rate (TPR, sensitivity) of 98.9% and a True Negative Rate (TNR, specificity) of 99.8%. In the FP image, a ground marking was detected as FOD. This is a problem that can perhaps be circumvented by explicitly annotating ground markings - if one chooses to do so. The FN cases were small and thin objects that were not detected, a challenge to be addressed in the future.

3.3 Data Augmentation

With tool-based image augmentation, it is possible to add more variance to the training data set. We tested random changes in the brightness and contrast of the images, this transformation can be integrated into the data loader. The results of different augmentation parameters in the classification are shown in Table 3. The data set contained 800 images in each of the 2 classes (normal and anomalous) with a 70/30 % training/test split; the unknown objects data set contained 125 images.

Table 3: Parameters for data augmentations, and the resulting classification accuracy [%] on test data and unknown objects.

Data augmentation	Test (30%)	Unknown objects
No augment.	99.0	95.7
brightness(±0.5) + contrast(±0.3)	98.8	97.1
brightness(±0.7)	99.0	97.1
brightness(±0.7) + contrast(±0.7)	99.2	97.3

Compared to the test accuracy without augmentation, the classification accuracy for unknown objects improved in all three tests, see Table 3.

The image augmentation was also tested with the object detection model. We used random ColorJitter with the following parameters: brightness=0.5, contrast=0.5, saturation=0.1 and hue=0.1. The training data set consisted of 248 images and a random 80/20 split between training and test data. The test data set 1 is from the same photo session as the training data, the test data set 2 is from a second session with different environmental conditions. No image augmentation was used in the baseline.

Table 4: mAP [%] at IoU = 0.5 with augmentation.

Augmentation of training data	Test (20 %)	Test data set 1	Test data set 2
None (baseline)	96.8	95.5	90.9
ColorJitter	97.9	95.5	94.7

The results in the Table 4 show that with ColorJitter data augmentation, the model generalizes better for the test data set 2, whose environmental conditions differ from the training data set.

To increase the number of images in the data set, one can crop certain regions from the images. But then one has to pay attention to the annotations so that the objects and bounding boxes do not get lost. The small amount of training material is problematic on the one hand, but on the other hand it seems to make the model easier to generalise. One must therefore observe the training and test accuracy curves, to avoid overfitting the model.

3.4 Synthetic Data

To increase the number of training images, we synthesized images by merging objects onto the underground. Because the objects are unknown, it is possible to control the size, position and shape of the objects at will. So it is possible to create data with selected properties, to improve detection capabilities as needed. This also has the advantage that the labeling data can be written automatically while creating the synthetic images.

With image generation, it is possible to add images with specific characteristics (object size and position) to the data set to improve training and thus detection performance. In the example image Figure 3 we inserted a screw into the image of the ground.

Figure 3: Synthesized image with screw.

As the objects in the synthetic images sometimes look a bit artificial, the synthetic images had only a small share (10 - 50%) in the training data sets to avoid a large distribution offset in the data. The same baseline as before (section 3.3) was used to evaluate the influence of synthetic data on training. In addition to the 198 basic training images, 200 synthetic images were added in this experiment. The tests with the test data sets were carried out solely with real images, they showed improvements in the results, compared to the baseline, see Table 5.

The combination of ColorJitter enhancement and synthetic data did not bring any improvement; on the contrary, the performance for test data set 2 did not increase very much, reaching only 91.1%.

Table 5: mAP [%] at IoU = 0.5 with synthetic images.

Augmentation of training data	Test (20 %)	Test data set 1	Test data set 2
None (baseline)	96.8	95.5	90.9
With 200 synth. images	98.0	96.4	92.3

3.5 Minimum Object Size

The most important number for FOD systems is, how small the recognizable objects can be. This is an ideal use case for synthetic images, as they can be created en masse and there is no distribution shift between training and test data.

We created images with FOD sizes starting at 5x5, up to 25x25 pixels, on background images from 300x300 to 1280x960 pixels. The tests showed that due to internal transformations in the model, the minimum detectable object size depends on the image size. This internal transformation resizes the images, typical input sizes of the first layer of our Faster R-CNN model had been 800x800 and 1088x800 features. It is important to be aware of this behaviour, if one want to analyse detection capabilities.

As a criterion we defined an accuracy level of 95% in detecting FOD, in this case a screw on the ground. The detectable object sizes, depending on the image size, are listed in Table 6.

Table 6: Recognizable object size of a screw, depending on the image size, threshold is 95% accuracy.

Image size	Bounding Box size
300x300	15x15
320x240	15x15
600x600	25x25
640x480	25x25
800x600	not detected
1280x960	not detected

The probability of recognition also depends on the contrast; a golden ball was recognized starting at a size of 7x7 pixels on a 300x300 pixel image, that corresponds to an area of 0.05% of the image. The correct identification of empty images as normal case was at a high level for all image sizes, above 95% for the image sizes where the model recognized the screw.

The minimum recognizable FOD sizes in our experiments are in pixels, which does not help the people in charge, they want to know concrete dimensions. The results from Table 6, in pixels, can be converted to absolute object sizes, using the focal length of the optics, the distance to the object and the pixel pitch of the camera sensor.

4 OUTLOOK AND FUTURE WORK

Despite the good results, there are some open questions and possible directions for improvements. There are promising approaches for improving performance, especially in the algorithms.

4.1 Unsupervised or Self-Supervised Learning

To avoid the burden of annotating data, it would be nice to work without annotated data, or to largely avoid the effort of annotating data. This is where unsupervised or self-supervised learning algorithms come in, together with algorithms that use self-attention to find anomalies. Experiments will test whether these approaches can detect FOD in the input images without drawing attention to markings or stains.

4.2 Improved Architectures or More Training Data

Detecting Small Objects, and in parallel to differentiate them from stains is difficult. Do we need other architectures for improved detection of small objects, at the cost of more computing effort - or should we just throw more training data on the problem?

One can also use more training data, with more small objects, in order to make the model more sensitive to small objects. Then the higher efforts only apply during the training phase. When building a real-world system, all of these design choices must be balanced.

One architecture that will be tested is CNN (Convolutional Neural Network), it should be able to learn local structures, for example the shape of ground markings. In this way the CNN should ignore ground markings in the image, in connection with a high FOD detection accuracy.

4.3 Detection and System Reliability

The detection of FOD is a safety critical application, and machine learning systems can support humans in this task. When the systems improve, the users rely more and more on their results. How to handle the probabilistic nature of ML systems in safety critical applications? Do we need to certify the reliability and generalizability of these systems and models?

In our images we had small objects down to 5x5 pixels. Sometimes the detection of this objects worked, but there is always the danger that the model is to sensitive, treating stains on the ground as FOD. The regulations request finding objects down to a size of a few centimeters, with at the same time a low false alarm rate. How can these two goals be balanced? Experiments should be conducted in real-world scenarios to gain more knowledge and experience.

5 CONCLUSIONS

The implementation of the FOD detection system is an engineering task, with some design decisions with advantages and disadvantages. We plan to do some real-world testing with live detection. And we continue to work on improving FOD detection on the ML side, especially with unsupervised and self-supervised algorithms.

ACKNOWLEDGEMENTS

This work was funded by the Bavarian Ministry for Economic Affairs, Regional Development and Energy as part of a project to support the thematic development of the Institute for Cognitive Systems (IKS).

REFERENCES

Dai, Y., Liu, W., Li, H., and Liu, L. (2020). Efficient foreign object detection between psds and metro doors via deep neural networks. *IEEE Access*, 8:46723–46734.

Kaiming He, Xiangyu Zhang, S. R. J. S. (2015). Deep residual learning for image recognition. https://arxiv.org/pdf/1512.03385.pdf , retrieved 2023-11-13.

Mathilde Caron, Hugo Touvron, I. M. H. J. J. M. P. B. A. J. (2021). Emerging properties in self-supervised vision transformers.

Papadopoulos, E. and Gonzalez, L. F. (2021). Uav and ai application for runway foreign object debris (fod) detection. In *Proceedings of the 2021 IEEE Aerospace Conference*, IEEE Aerospace Conference Proceedings. Institute of Electrical and Electronics Engineers Inc., United States of America.

Shaoqing Ren, Kaiming He, R. G. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. https://arxiv.org/pdf/1506.01497.pdf , retrieved 2023-11-13.

Travis Munyer, Pei-Chi Huang, C. H. X. Z. (2022). Fod-a: A dataset for foreign object debris in airports. https://arxiv.org/pdf/2110.03072.pdf , retrieved 2023-11-20. Download: https://github.com/FOD-UNOmaha/FOD-data.

Yuan, Z.-D., Li, J.-Q., Qiu, Z.-N., and Zhang, Y. (2020). Research on fod detection system of airport runway based on artificial intelligence. *Journal of Physics: Conference Series*, 1635(1):012065.