Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Data Science Lab Fall 24

ETH AI CENTER

Few-Shot Anomaly Detection in a Real-World Aero-Engine Blade Inspection Use Case

Berke Arda [1], Katarína Osvaldová[1], Florian Scheidegger [2], Julia Vogt [3]
[1]ETH Zurich Student, D-INFK [2]IBM Research Europe [3]ETH AI Center, ETH Zurich

1. Introduction

The aim of this project is the investigation of the performance of state-of-the-art few-shot anomaly detection methods on a real-world aero-engine blade dataset, AeBAD.

Anomaly Detection: Anomaly detection focuses on identifying patterns in data that do not conform to expected behaviour. It can mean sample level classification or pixel level image segmentation. This task is crucial for quality and safety control and its automatisation poses financial benefits.

Few-Shot Learning: Few-shot learning addresses scenarios where only a limited amount of training data is available. In many instances, as collecting diverse a dataset is often infeasible. Few-shot methods leverage the ability to generalize from minimal examples to improve performance.

2. Related Work

The field of anomaly detection is constantly advancing with state-of-the-art approaches quickly evolving. Classical methods, such as PatchCore and ReverseDistillation, rely on reconstruction-based or feature-matching techniques to identify anomalies. While these methods perform well under standard conditions, they struggle with domain shifts, such as changes in lighting or background. Recently introduced methods like DRAEM and MMR aim to enhance anomaly detection through data-efficient architectures. However, these approaches require large training datasets or extensive domain-specific fine-tuning. Our work focuses on a few-shot method, WinCLIP published by Jeong et al. [1], that incorporates language-guided features, and evaluates its robustness under domain shifts of the AeBAD dataset.

3. Data

The AeBAD dataset is a real-world benchmark specifically designed for anomaly detection in aero-engine blade inspections. It was introduced by Zhang et al. [2], and it comprises diverse images of aero-engine blades captured subject to varying domain shifts, such as changes in background, illumination, and viewpoint, pictured in Figure 1. This diversity reflects the operational challenges of anomaly detection in real-world industrial scenarios.

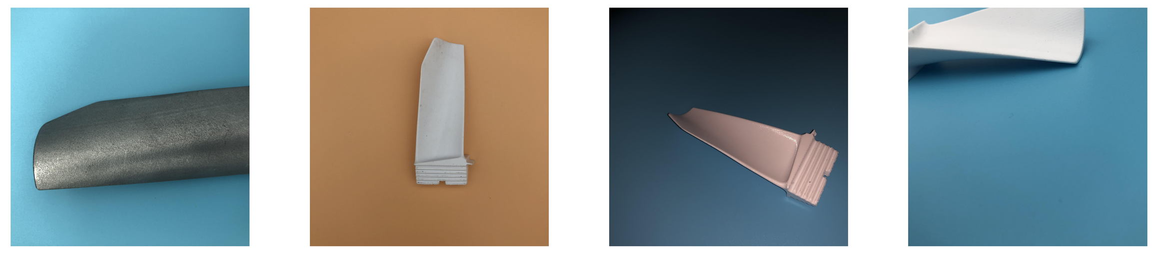

Figure 1: Example non-anomalous images with domain shift, from left to right: none, background, illumination, view.

4. Method

We conducted a comprehensive evaluation of anomaly detection methods using the AeBAD dataset, focusing on understanding the performance of WinCLIP under zero-shot and few-shot conditions. Curation of a large training dataset is very often infeasible. Thus, methods trained on hundreds of normal-shots, often, have different use case than few-shot methods. To facilitate a fair comparison, we designed two experiments. On one hand, we investigated WinCLIP+ trained with high number of normal-shots. On the other hand, we explored the decay of performance of the highest performing competition method.

5. Results

	Method	Same	Background	Illumination	View	Mean
MMR[2]	PatchCore	75.2 ± 0.3	74.1 ± 0.3	74.6 ± 0.4	60.1 ± 0.4	71.0
	ReverseDistillation	82.4 ± 0.6	84.3 ± 0.9	85.5 ± 0.9	71.9 ± 0.8	81.0
	DRAEM	64.0 ± 0.4	62.1 ± 6.1	61.6 ± 2.7	62.3 ± 0.9	62.5
Our	MMR	**85.7 ± 0.3**	**84.4 ± 0.7**	**88.7 ± 0.6**	79.5 ± 0.5	**84.6**
	WinCLIP (0-Shot)	80.3 ± 0.2	82.9 ± 0.5	67.0 ± 0.3	82.0 ± 0.3	78.0
	WinCLIP+ (1-Shot)	80.7 ± 0.5	83.1 ± 0.5	67.4 ± 0.6	**82.1 ± 0.4**	78.3
	WinCLIP+ (4-Shot)	80.9 ± 0.2	83.7 ± 0.4	67.7 ± 0.4	81.9 ± 0.3	78.6

Table 1: Sample-Level AUROC% on AeBAD dataset. Best values are highlighted.

Table 1 shows the sample level AUROC% of the investigated methods.

WinCLIP (Zero-Shot): This approach demonstrated strong generalization capabilities for anomaly classification and segmentation without any fine-tuning. However, the lack of pixel-level alignment limited segmentation precision in certain cases.

WinCLIP+ (Few-Shot): Incorporating a small number of normal samples significantly enhances performance. Between 1 to 4 shots, WinCLIP+ reduced false positives and improved anomaly localization, particularly in scenarios with domain shifts.

WinCLIP+ in a many-shot scenario: Table 2 presents the performance of WinCLIP trained with increased number of normal-shots. The performance improves only slightly, highlighting the robustness of WinCLIP+ in low-data regimes.

Method	Same	Background	Illumination	View	Mean
WinCLIP+ (10-Shot)	81.1 ± 0.3	83.7 ± 0.4	67.7 ± 0.4	81.8 ± 0.6	78.6
WinCLIP+ (50-Shot)	81.2 ± 0.1	83.8 ± 0.3	67.4 ± 0.5	81.9 ± 0.6	78.6
WinCLIP+ (75-Shot)	81.2 ± 0.2	84.0 ± 0.2	67.5 ± 0.3	81.9 ± 0.6	78.6

Table 2: Sample-Level AUROC% of WinCLIP+ in many-shot settings on AeBAD dataset.

MMR in a few-shot scenario: The best performing benchmark, MMR, a reconstruction-based method, was originally trained on the full dataset. Our investigation presented in Figure 2, shows a substantial performance decay with decreasing training dataset size. With the exception of the illumination domain shift, WinClip+ (1-shot) performs on par with MMR trained on at least 100 normal shots.

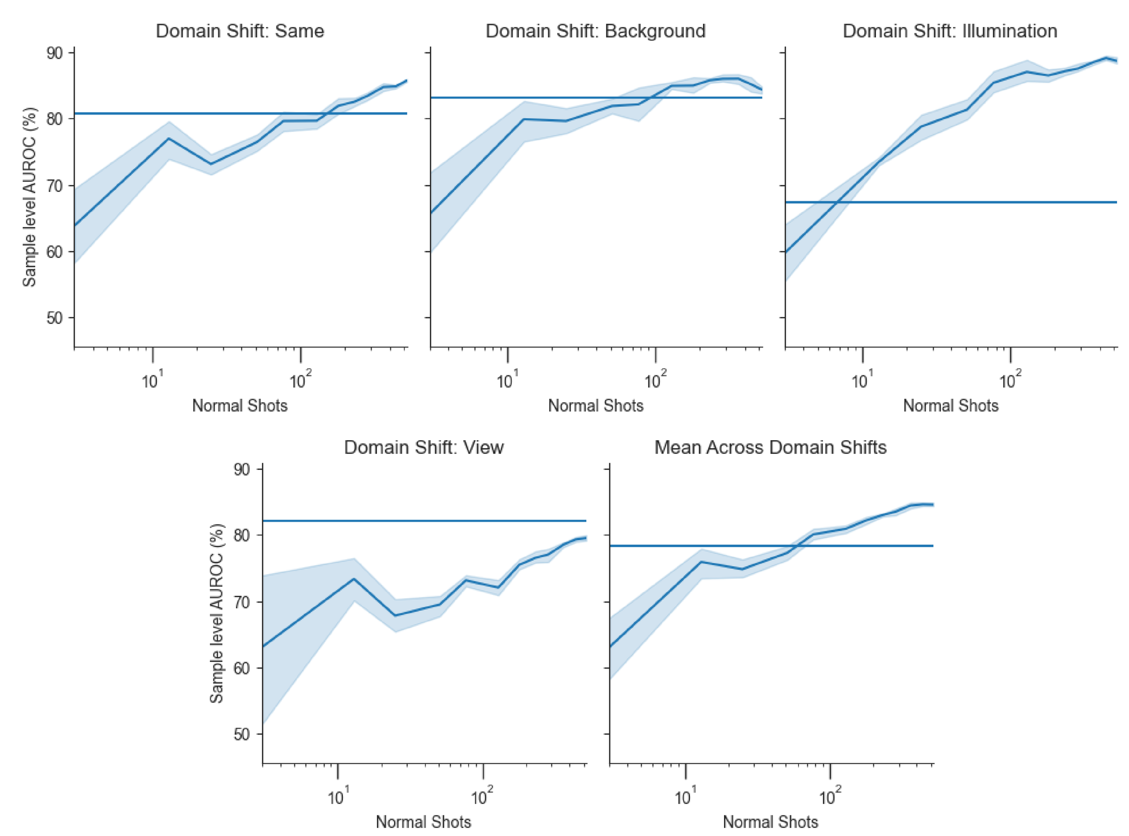

Figure 2: Performance of MMR at different training dataset sizes with the 95% confidence interval. The performance of WinCLIP (1-shot) is represented with a horizontal line.

6. Conclusions

In our work, we successfully tested WinCLIP's anomaly classification capability in a real-world use case, the AeBAD dataset. We also compared its performance with the current state-of-the-art anomaly detection method, MMR, and showed WinCLIP+'s superiority in a few-shot scenario.

7. Future Work

- So far, our work did not entail experiments aimed at improving WinCLIP+'s performance. We are hoping to explore the possibility of the improvement by aiming the focus of the method with purposefully designed training shot set containing multiple domain shifts, not necessarily the ones present in the testing dataset.
- In our work, we investigated the performance of a non-few-shot method MMR in a few-shot scenario. MMR is not the only well-performing method. Investigating possible performance decay of methods based on different techniques might yield novel insight.

References

[1] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19606–19616, 2023.

[2] Zilong Zhang, Zhibin Zhao, Xingwu Zhang, Chuang Sun, and Xuefeng Chen. Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction. *Computers in Industry*, 151:103990, 2023.

Contributions

Berke Arda: Methodology, Software, Investigation, Writing - Original Draft
Katarína Osvaldová: Methodology, Software, Visualization, Investigation, Writing - Original Draft
Florian Scheidegger: Conceptualization, Supervision
Julia Vogt: Supervision, Project Administration