

## ABSTRACT

---

Anomaly detection (AD) is crucial for industrial defect detection, as identifying defects early can prevent costly issues. Traditional methods rely on classical image processing techniques like filtering and signal processing, which don't require data but struggle with changing conditions, such as lighting. Modern approaches, based on deep learning, are more robust and have largely replaced these older methods. Supervised learning algorithms perform well but require extensive labeled data, while unsupervised models solely relies on normal data to detect defects. To reduce the data dependency, zero-shot and few-shot models have emerged. They offer a promising alternative, requiring little to no training data, making them incredibly flexible.

This thesis investigates the capabilities of select zero- and few-shot models based on the vision-language model CLIP for defect detection and localization on textile surfaces. These models are evaluated at the image and pixel level on two distinct fabric datasets. In addition, few-shot models with domain specific knowledge are repurposed for zero-shot AD by evaluating them in a cross-dataset setting to test their generalization ability.

The experimental findings reveal that the zero-shot approach was unable to detect defects effectively, while the few-shot models achieved accurate results on both datasets. One of the investigated models, termed SDP+, which leverages a pre-trained CLIP model, stands out with impressive defect localisation capabilities (i.e. with a dice scores of up to 0.50). Based on these capabilities, this technique was improved by computing an anomaly score (for image classification) from the anomaly maps, leading to a near-perfect image-level f1-score of 0.997. Moreover, the few-shot models demonstrated their potential as zero-shot models with domain-specific knowledge, achieving a cross-dataset image-level f1-score of up to 0.92, significantly outperforming the traditional zero-shot approach. These results have important implications for industrial applications. The models' ability to transfer knowledge between datasets can reduce the need for extensive data labeling, making them cost-effective tools for automating defect detection in new production environments. Additionally, the thesis serves as a starting point for future research, particularly in leveraging pre-trained vision-language models for industrial tasks. The findings demonstrate how few-shot models can be adapted to create flexible, scalable solutions for defect detection, helping to streamline quality control processes.

**Keywords:** Anomaly Detection and Localisation, Fabric Defect Detection, Zero-shot Learning, Few-shot Learning, CLIP, Vision-Language Models, Win-CLIP, AnomalyCLIP, CLIP-AD, Knowledge Transfer.

## ZUSAMMENFASSUNG

---

Anomaliedetektion (AD) ist entscheidend für die Erkennung von Fehlern, da das frühzeitige Erkennen von Defekten kostspielige Probleme verhindern kann. Traditionelle Methoden basieren auf klassischen Bildverarbeitungstechniken wie Filterung und Signalverarbeitung, die keine Daten erfordern, jedoch bei wechselnden Bedingungen, wie Lichtverhältnissen, Schwierigkeiten haben. Moderne Ansätze, die auf Deep Learning basieren, sind robuster und haben diese älteren Methoden weitgehend ersetzt. Überwachtes Lernen liefert gute Ergebnisse, benötigt jedoch viele annotierte Daten, während unüberwachte Modelle nur mit Normaldaten Anomalien erkennen. Um die Datenabhängigkeit zu verringern, wurden Zero-Shot- und Few-Shot-Modelle entwickelt. Sie bieten eine vielversprechende Alternative, die wenig bis gar keine Trainingsdaten benötigt und äußerst flexibel ist.

Diese Arbeit untersucht die Leistungsfähigkeit ausgewählter Zero- und Few-Shot-Modelle, die auf dem vision-language Modell CLIP basieren, zur Fehlererkennung und -lokalisierung auf Textiloberflächen. Diese Modelle werden sowohl auf Bild- als auch auf Pixelebene auf zwei unterschiedlichen Stoffdatensätzen evaluiert. Zudem werden Few-Shot-Modelle mit domänenspezifischem Wissen zu Zero-Shot-AD umfunktioniert, indem sie in einem Cross-Dataset-Setting getestet werden, um ihre Generalisierungsfähigkeit zu überprüfen.

Die Experimente zeigen, dass der Zero-Shot-Ansatz Defekte nicht effektiv erkennen konnte, während die Few-Shot-Modelle auf beiden Datensätzen präzise Ergebnisse erzielten. Eines der untersuchten Modelle, genannt SDP+, das auf einem vortrainierten CLIP-Modell basiert, hebt sich durch beeindruckende Fähigkeiten zur Defektlokalisierung hervor (z.B. mit Dice scores von bis zu 0,50). Diese Methode wurde durch die Berechnung eines Anomalie-Scores (zur Bildklassifizierung) aus den Anomalie-Maps verbessert, was zu einem nahezu perfekten Bild-F1-Score von 0,997 führte. Darüber hinaus zeigten die Few-Shot-Modelle ihr Potenzial als Zero-Shot-Modelle mit domänenspezifischem Wissen und erreichten in einem Kreuz-Datensatz-Setting einen Bild-F1-Score von bis zu 0,92, womit sie den Zero-Shot-Ansatz deutlich übertrafen. Diese Ergebnisse sind für industrielle Anwendungen von großer Bedeutung. Die Fähigkeit der Modelle, Wissen zwischen Datensätzen zu übertragen, kann den Bedarf an umfangreicher Datenannotierung reduzieren und sie zu kostengünstigen Werkzeugen für die Automatisierung der Fehlererkennung in neuen Produktionsumgebungen machen. Darüber hinaus dient die Arbeit als Ausgangspunkt für zukünftige Forschung, insbesondere bei der Nutzung vortrainierter vision-language Modelle für industrielle Aufgaben. Die Ergebnisse zeigen, wie Few-Shot-Modelle angepasst werden

können, um flexible, skalierbare Lösungen zur Fehlererkennung zu schaffen, die den Qualitätskontrollprozess effizienter gestalten.

**Keywords:** Anomaly Detection and Localisation, Fabric Defect Detection, Zero-shot Learning, Few-shot Learning, CLIP, Vision-Language Models, Win-CLIP, AnomalyCLIP, CLIP-AD, Transfer Learning.