Pós-Graduação em Ciência da Computação

Felipe de Melo Battisti

ADDRESSING REAL-WORLD VARIABILITY IN POWER LINE INSPECTIONS WITH UNSUPERVISED VISUAL ANOMALY DETECTION

M.Sc. Dissertation

Federal University of Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE

2025

Federal University of Pernambuco
Center for Informatics
Graduate in Computer Science

Felipe de Melo Battisti

ADDRESSING REAL-WORLD VARIABILITY IN POWER LINE INSPECTIONS WITH UNSUPERVISED VISUAL ANOMALY DETECTION

A M.Sc. Dissertation presented to the Center for Informatics of Federal University of Pernambuco in partial fulfillment of the requirements for the degree of Master of Science in Computer Science.

Advisor: Francisco Paulo Magalhães Simões
Co-Advisor: Veronica Teichrieb

RECIFE
2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Battisti, Felipe de Melo.
   Addressing Real-World variability in power line inspections with unsupervised visual anomaly detection / Felipe de Melo Battisti. - Recife, 2025.
   87f.: il.

   Dissertação (Mestrado)- Universidade Federal de Pernambuco, Centro de Informática,Programa de Pós-Graduação em Ciência da Computação, 2025.
   Orientação: Francisco Paulo Magalhães Simões.
   Coorientação: Veronica Teichrieb.

   1. Anomaly detection; 2. Unsupervised learning; 3. Industrial inspection; 4. Power transmission; 5. Deep learning. I. Simões, Francisco Paulo Magalhães. II. Teichrieb, Veronica. III. Título.

UFPE-Biblioteca Central

Dissertação de mestrado apresentada por **Felipe de Melo Battisti** ao Programa de Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título **Addressing Real-World Variability in Power Line Inspections with Unsupervised Visual Anomaly Detection** , orientada pelo **Prof. Francisco Paulo Magalhães Simões**, co-orientado pela Profa. Veronica Teichrieb e aprovada pela banca examinadora formada pelos professores:

Prof. Francisco Paulo Magalhães Simões
Centro de Informática/UFPE

Prof. Flávio Arthur Oliveira Santos
Centro de Informática/UFPE

Prof. Alexandre Cardoso
Universidade Federal de Uberlândia (UFU)

RECIFE
2025

To my mother, Maria Aparecida Rodrigues de Melo, my greatest source of strength, resilience, and unconditional support. Her constant encouragement and unwavering belief in my potential have been the foundation of every step in my academic journey.

To my friends at Voxar Labs, whose companionship, advice, and support were invaluable throughout this process. Sharing this journey with you made every challenge lighter and every achievement more meaningful.

To my advisors, Prof. Veronica Teichrieb and especially Prof. Francisco Paulo Magalhães Simões, whose guidance, patience, and trust were essential. Francisco, in particular, played a pivotal role in shaping this work and in helping me grow both as a researcher and as a professional.

This dissertation is dedicated to all of you — for being not only part of this achievement, but fundamental to making it possible.

Acknowledgements

À minha mãe, Maria Aparecida Rodrigues de Melo, minha maior fonte de força, resiliência e apoio incondicional. Seu encorajamento constante e sua inabalável crença no meu potencial foram a base de cada passo da minha trajetória acadêmica.

Aos meus amigos do Voxar Labs, cuja amizade, conselhos e apoio foram inestimáveis ao longo de todo esse processo. Compartilhar essa jornada com vocês tornou cada desafio mais leve e cada conquista muito mais significativa.

Aos meus orientadores, Profa. Veronica Teichrieb e o Prof. Francisco Paulo Magalhães Simões, cujo acompanhamento, paciência e confiança foram essenciais. Francisco, em particular, teve um papel fundamental na construção deste trabalho e no meu desenvolvimento, tanto como pesquisador quanto como profissional.

Dedico este trabalho a todos vocês — por não serem apenas parte desta conquista, mas por terem sido fundamentais para que ela se tornasse possível.

Then as it was, then again it will be. An' though the course may change
sometimes, Rivers always reach the sea.
—LED ZEPPELIN — TEN YEARS GONE

Resumo

A detecção visual de anomalias em infraestruturas de transmissão de energia elétrica é fundamental para prevenir falhas críticas e garantir a confiabilidade da rede. À medida que essas redes se expandem para regiões remotas e complexas, inspeções manuais tradicionais enfrentam desafios operacionais, como alto custo, riscos à segurança e baixa escalabilidade. Embora métodos recentes de inspeção visual automatizada baseados em deep learning apresentem avanços, a maioria é avaliada em cenários controlados, o que limita sua eficácia em aplicações reais.

O problema central abordado neste trabalho é a baixa capacidade de generalização de modelos não supervisionados de detecção de anomalias quando aplicados a imagens aéreas obtidas em condições reais de inspeção. Modelos supervisionados exigem grandes volumes de dados anotados com falhas raras e variadas—algo impraticável nesse domínio. Por isso, métodos não supervisionados, que aprendem apenas a partir de dados normais, são mais viáveis, mas enfrentam dificuldades ao lidar com ruído de fundo, variações de iluminação e anomalias sutis.

Este trabalho demonstra que a integração de mecanismos de atenção e técnicas de remoção de fundo pode aprimorar significativamente a detecção de anomalias em cenários não controlados. Para isso, foi desenvolvido o InsPLAD-seg, um novo conjunto de dados com anotações em nível de pixel de falhas reais. Além disso, foram aplicados módulos de atenção (SENet e CBAM) e segmentação por objetos utilizando YOLO, com o objetivo de refinar a extração de características e reduzir interferências visuais.

Os resultados mostram que a combinação de atenção e remoção de fundo melhora consistentemente os escores de AUROC em níveis de imagem e pixel, com destaque para a detecção de corrosões e pequenas falhas. Modelos como RD++ e DifferNet, aprimorados com atenção, alcançaram desempenho de ponta, enquanto a remoção de fundo aumentou a robustez frente à variabilidade visual.

Esses achados reforçam que mecanismos de atenção e pré-processamentos que simulam condições controladas são essenciais para aproximar os modelos de detecção do uso prático em campo. Ao modelar diretamente anomalias em inspeções reais, esta abordagem contribui para sistemas de monitoramento mais robustos e escaláveis.

Este trabalho viabiliza a aplicação prática de modelos de detecção de anomalias em ambientes ruidosos, nos quais precisão e escalabilidade são requisitos críticos.

Palavras-chave: detecção de anomalias, transmissão de energia, aprendizado não supervisionado, segmentação semântica, deep learning, mecanismos de atenção, inspeção industrial.

Abstract

Visual anomaly detection in power transmission infrastructure is essential for preventing critical failures and ensuring grid reliability. As electricity networks expand into remote and complex environments, traditional manual inspections face significant challenges, including labor intensity, safety hazards, and limited scalability. Recent progress in deep learning has enabled automated visual inspection, but most anomaly detection models are evaluated under controlled conditions, limiting their real-world applicability.

The key problem addressed is the poor generalization of unsupervised anomaly detectors when deployed on aerial images from real-world inspections. Supervised models require large, labeled datasets of rare and diverse defect types, something infeasible in power line inspection due to the scarcity and unpredictability of faults. Unsupervised methods, which learn only from normal data, are more suitable but struggle with cluttered backgrounds, variable lighting, and subtle anomaly patterns.

This work shows that the integration of attention mechanisms and background removal techniques enhances the detection of unsupervised anomalies in uncontrolled scenarios. We created a novel dataset, InsPLAD-seg, with pixel-level annotations of real-world defects and applied attention modules (SENet, CBAM) and YOLO-based object segmentation to refine feature extraction and reduce noise interference.

Our results demonstrate that combining attention mechanisms with background removal consistently improves AUROC scores at both image and pixel levels, especially in corrosion detection and small defect localization. RD++ and DifferNet models enhanced with attention modules achieved state-of-the-art results, while background segmentation further improved robustness to environmental variability.

These findings suggest that controlled-like preprocessing and feature-aware attention are critical to bridging the gap between laboratory benchmarks and real-world deployments in infrastructure inspection. By modeling in situ power line anomalies, this approach supports more resilient and scalable grid monitoring solutions.

This enables real-world deployment of anomaly detection models in noisy environments, where precision and scalability are critical.

Keywords: anomaly detection, power transmission, unsupervised learning, semantic segmentation, deep learning, attention mechanisms, industrial inspection.

List of Figures

List of Tables

List of Acronyms

Contents

1

Introduction

The task of visually inspecting components for defects is fundamental to ensuring the safety and reliability of industrial operations. In large-scale environments, particularly those involving complex and geographically distributed infrastructure, this process is not only essential but also operationally challenging. Manual inspection remains the predominant approach in many sectors; however, it is labor-intensive, time-consuming, and prone to human error—especially under conditions of repetitive analysis, limited visibility, or safety risks. These limitations have driven the adoption of automated visual inspection systems, which promise greater consistency, efficiency, and scalability MENDU; MBULI (2025).

Such systems support a range of applications, including quality assurance, maintenance planning, video surveillance, and security monitoring. These applications span various domains, notably manufacturing BERGMANN et al. (2019a); RUDOLPH; WANDT; ROSENHAHN (2021); YU et al. (2021), healthcare SCHLEGL et al. (2019); HAN et al. (2021), and power systems GE et al. (2021); LEKIDIS; ANASTASIADIS; VOKAS (2022). In particular, this work focuses on the inspection of power transmission infrastructure, where the detection of visual and structural anomalies is critical to maintaining operational continuity and preventing large-scale disruptions. High-voltage transmission lines extend across vast and often remote regions. A single defective component—such as a cracked insulator or a corroded connector—can initiate cascading failures, leading to blackouts that affect entire towns or even major cities. As electricity becomes increasingly central to everyday life and industrial production, the consequences of such failures extend beyond operational delays to include substantial economic and social impacts GUAN et al. (2021); HE et al. (2024).

To mitigate these risks, electric utilities conduct routine inspections of transmission towers and substations. While traditionally performed by field teams, these inspections involve significant logistical challenges and safety hazards, including high-altitude access and proximity to live circuits. In recent years, unmanned aerial vehicles (UAVs), commonly referred to as drones, have become an effective tool to address these limitations, enabling the capture of high-resolution images from difficult-to-access components LI et al. (2023); CHEN et al. (2018). These images are then reviewed by engineers and trained professionals to assess potential

Figure 1.1: Example of power line towers for inspection

damage or wear. However, the manual analysis of UAV imagery remains a labor-intensive task, particularly as inspection programs scale to cover large grid segments, as the example in figure 1.1. The high volume of images, combined with the repetitive nature of the task, increases the likelihood of human fatigue and oversight DUTTA; SONI; GUPTA (2021). These challenges have reinforced the need for intelligent automation that can assist or even replace human evaluators in the inspection workflow ZHONG et al. (2021).

Historically, automated defect detection has relied on supervised learning models trained on labeled examples of defective and non-defective components HUANG; TIAN; CHEN (2022). These approaches are widely used in industrial settings due to their strong discriminative performance when ample annotated data is available. Yet, in real-world scenarios—particularly in power systems—fault cases are scarce and diverse, limiting the feasibility of supervised approaches. The imbalance between normal and defective samples introduces significant generalization challenges, as supervised models struggle to adapt to novel or rare fault conditions JALIL et al. (2019).

To address these limitations, unsupervised anomaly detection methods have emerged as a promising alternative. Rather than learning from explicit examples of faults, these methods model the distribution of normal data and identify deviations as potential anomalies. This makes them suitable for inspection contexts where defective cases are infrequent and heterogeneous BASIT; MANZOOR; AKRAM (2024); ZHOU et al. (2022). In power transmission systems, unsupervised models can be effectively combined with UAV imagery to enable scalable and autonomous fault detection across large infrastructures. These developments point toward a new paradigm of intelligent, data-driven inspection practices that improve both the coverage and reliability of

critical infrastructure monitoring.

1.1 Problem Statement

Despite recent advances in anomaly detection techniques, their performance tends to decline significantly when deployed in uncontrolled environments. Most state-of-the-art methods have been evaluated under controlled conditions—such as uniform backgrounds, consistent lighting, fixed resolutions, and standardized viewpoints—which do not reflect the complexities encountered in real-world industrial applications RUDOLPH; WANDT; ROSENHAHN (2021); YU et al. (2021); DEFARD et al. (2021); TIEN; WANG; LEE (2023). In particular, tasks like power transmission line inspection involve substantial variability in camera perspective, object scale, orientation, illumination, background clutter, and image resolution. These factors introduce noise and distortions that impair the effectiveness of conventional detection models.

As a result, anomaly detection in uncontrolled environments remains an open and underexplored challenge. Developing robust methods capable of generalizing across such variability is essential for advancing automated visual inspection in realistic field conditions. Addressing this problem requires not only the investigation of novel techniques but also the construction of dedicated datasets that reflect these real-world complexities—since, as identified in our literature review, existing public datasets do not adequately represent the visual diversity and unpredictability of such environments. Once such representative data is available, the focus must shift toward designing or adapting anomaly detection approaches that remain reliable and effective under the inherent noise and variability of uncontrolled conditions.

1.2 Research Objectives

This dissertation explores the hypothesis that integrating attention mechanisms and leveraging background removal through object segmentation can enhance unsupervised anomaly detection at both image and pixel levels. Attention modules, such as CBAM WOO et al. (2018) and SENet HU; SHEN; SUN (2018), improve neural network performance by enhancing the network's focus on salient spatial and channel information. These mechanisms selectively highlight foreground features while suppressing irrelevant noise from the background, offering a robust solution to the challenges posed by uncontrolled environments. Additionally, background removal using object segmentation can make real-world datasets more comparable to controlled settings, potentially improving model performance by eliminating extraneous variability.

The objectives of this study are fivefold:

1. To investigate the strengths and limitations of state-of-the-art anomaly detection methods in uncontrolled environments.

2. To develop a pixel-level dataset that enables fine-grained evaluation of anomaly detection performance under real-world conditions.

3. To benchmark unsupervised anomaly detection methods at both image and pixel levels using a representative dataset.

4. To analyze the impact of incorporating attention mechanisms on the performance of anomaly detection models.

5. To evaluate the influence of background removal on anomaly detection effectiveness in visually complex and uncontrolled settings.

1.3 Contributions

This dissertation makes the following key contributions to the field of unsupervised anomaly detection for industrial inspection:

- Creation of a novel dataset, *InsPLAD-Seg*, by annotating defect masks in the InsPLAD-Fault dataset to enable pixel-level evaluation, addressing the lack of publicly available pixel-level datasets for industrial inspection in the wild.

- Comprehensive evaluation of state-of-the-art unsupervised anomaly detection methods at both image and pixel levels, providing insights into their strengths and limitations.

- Integration of attention mechanisms, such as CBAM and SENet, with unsupervised anomaly detection methods, demonstrating their impact on improving pixel-level detection accuracy.

- Development of background-removed versions of the InsPLAD-Seg dataset using object segmentation, enabling the study of anomaly detection performance in controlled-like conditions.

- Comparative analysis of the impact of attention mechanisms and background removal on the performance of anomaly detection methods, contributing to a deeper understanding of their roles in real-world applications.

These promising results motivated the exploration of further improvements in unsupervised anomaly detection by integrating attention mechanisms with other methods and addressing additional challenges such as pixel-level anomaly detection and noisy backgrounds

1.3.1 Scientific Contributions

This dissertation has led to the publication of the following peer-reviewed works, listed in chronological order to reflect the progressive development of the research:

- *Attention Modules Improve Modern Image-Level Anomaly Detection: A DifferNet Case Study* SILVA et al. (2023).
  André Luiz Vieira e Silva, **Felipe Battisti**, Francisco Simões, Danny Kowerko, Tobias Schlosser, Veronica Teichrieb.
  This extended abstract presents the first version of **AttentDifferNet**, introducing attention mechanisms (SENet and CBAM) into the DifferNet architecture. Initial results on three industrial anomaly detection datasets demonstrated promising improvements in AUROC, forming the conceptual foundation for subsequent studies.

- *Attention Modules Improve Image-Level Anomaly Detection for Industrial Inspection: A DifferNet Case Study* SILVA et al. (2024a).
  André Luiz Vieira e Silva, **Felipe Battisti**, Francisco Simões, Danny Kowerko, Tobias Schlosser, Veronica Teichrieb.
  Published in the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024.
  This full paper builds upon the previous abstract, expanding the evaluation of AttentDifferNet across controlled and in-the-wild datasets. It establishes state-of-the-art results on the InsPLAD-fault dataset and offers qualitative evidence through Grad-CAM analysis, highlighting how attention blocks enhance the network's focus on relevant foreground regions.

- *Attention Modules Improve Efficient Anomaly Localization for Industrial Inspection in the Wild* SILVA et al. (2024).
  André Luiz Vieira e Silva, **Felipe Battisti**, Veronica Teichrieb, Francisco Simões.
  Published in the LatinX in CV Workshop at CVPR, 2024.
  This extended abstract introduces **InsPLAD-seg**, the preliminary version of a real-world dataset for pixel-level anomaly detection in power line components. It evaluates the integration of attention modules into lightweight methods such as SimpleNet and EfficientAD, demonstrating improved performance in both image- and pixel-level tasks. The work also highlights the computational efficiency of the proposed models, reinforcing their applicability in real-time industrial inspection scenarios.

These publications reflect the progressive maturation of this research—from conceptual exploration to dataset creation and benchmarking. The extended experiments, improved methods, and additional findings presented throughout this dissertation go beyond what has been published so far. As such, the results are being consolidated and expanded for a future journal-oriented publication targeting a broader scientific audience.

2

Literature Review

2.1 Overview of Visual Anomaly Detection

Visual Anomaly detection is a crucial task in industrial inspection, particularly for detecting structural defects, equipment degradation, and foreign object interference in power line towers. Unlike controlled environments, power line inspections face additional challenges due to **uncontrolled backgrounds**, **varying lighting conditions**, and **occlusions** caused by vegetation and atmospheric distortions. The visual anomalies encountered in power line towers include **corrosion, broken parts, missing components, birds' nests, disconnected or damaged wires, dirt accumulation**, and other environmental influences that affect infrastructure integrity.

Traditional inspection methods relied on manual assessments or basic statistical techniques HODGE; AUSTIN (2004); MARKOU; SINGH (2003). However, with advancements in deep learning, modern approaches leverage **convolutional neural networks (CNNs)**, **vision transformers (ViTs)**, and **self-supervised learning techniques** to improve detection accuracy in complex industrial settings CHALAPATHY; CHAWLA (2019); PANG et al. (2021).

2.1.1 Definitions and Categories

Anomaly detection refers to the process of identifying deviations from normal patterns in data CHANDOLA; BANERJEE; KUMAR (2009); PIMENTEL et al. (2014). In power line tower inspections, anomalies are characterized by structural failures, contamination, or external interference. The nature of these anomalies is diverse and can be categorized as follows:

- **Point Anomalies**: Individual instances that significantly deviate from normal conditions, such as a single corroded joint or a loose bolt GOLDSTEIN; UCHIDA (2016).

- **Contextual Anomalies**: Objects or structures that may appear normal in one setting but anomalous in another, such as a wire that is intact but misaligned due to high wind stress AHMED; MAHMOOD; HU (2016).

- **Collective Anomalies**: A group of elements forming an anomaly, such as clusters of dirt buildup on insulators or large bird nests obstructing transmission lines GOLDSTEIN; UCHIDA (2016).

Anomaly detection methods can also be categorized based on the level of analysis:

- **Image-Level Anomaly Detection**: Determines whether an image as a whole contains an anomaly without pinpointing specific regions RUFF et al. (2018).
- **Pixel-Level Anomaly Detection**: Focuses on localizing and segmenting anomalies within an image BERGMANN et al. (2019b).

Furthermore, methods can be divided based on the learning paradigm:

- **Supervised Methods**: Require labeled datasets containing both normal and anomalous samples KIRAN; THOMAS; PARAKKAL (2018).
- **Unsupervised Methods**: Operate without labeled anomalies, learning the distribution of normal data and detecting deviations ZONG et al. (2018).

2.1.2 Image-Level vs. Pixel-Level Anomaly Detection

Anomaly detection methods can be broadly categorized into **image-level** and **pixel-level** approaches, depending on the granularity of the detection process. **Image-level anomaly detection** focuses on classifying an entire image as normal or anomalous, while **pixel-level anomaly detection** localizes specific regions or pixels that deviate from normal patterns. The choice between these methods depends on the application requirements, such as whether a binary classification is sufficient or if precise defect localization is necessary.

2.1.2.1 Image-Level Anomaly Detection

Image-level anomaly detection determines whether an image as a whole contains an anomaly, without identifying the specific location of the irregularity. This approach is commonly used in **automated inspection systems**, where rapid screening of large datasets is necessary to filter out defective or anomalous samples for further evaluation.

2.1.2.1.1 Principles of Image-Level Anomaly Detection At a high level, image-level anomaly detection encodes images into a **feature space** and compares them against known distributions of normal samples. The anomaly decision is typically based on:

- **Feature-Based Methods**: The image is processed through a feature extractor (e.g., a convolutional neural network, or CNN) to generate a **high-dimensional representation** that captures relevant visual patterns. Anomalies are identified by computing

distances between the extracted features and a reference distribution of normal features.

- **Classification-Based Methods**: A model is trained using labeled normal and abnormal images to perform direct classification. This approach is widely used in **supervised settings**, where anomaly labels are available during training.
- **Statistical and Probabilistic Methods**: Techniques such as **one-class classification**, **density estimation**, and **reconstruction-based scoring** assign anomaly scores based on likelihood or deviation from learned distributions.

2.1.2.1.2 Applications and Challenges Image-level anomaly detection is frequently used in scenarios where **global image context** is sufficient to determine abnormalities. Examples include:

- **Industrial Quality Control:** Determining if an entire product image contains a visible defect, such as a broken component in manufacturing.
- **Power Line Tower Inspections:** Detecting the presence of corrosion, missing parts, or structural damage from aerial images.
- **Medical Imaging:** Identifying whether an X-ray or MRI scan contains anomalies without precisely localizing them.

One of the main **limitations** of image-level anomaly detection is its inability to pinpoint **where** an anomaly is present. This restricts its usability in applications requiring **precise defect localization**, making it necessary to use pixel-level anomaly detection for more detailed analysis.

2.1.2.2 Pixel-Level Anomaly Detection

Pixel-level anomaly detection, also referred to as **anomaly segmentation**, focuses on identifying **specific regions** or **pixels** that differ from expected normal patterns. Unlike image-level detection, which provides a binary classification, pixel-level detection generates **anomaly heatmaps or segmentation masks** that indicate **the exact locations of defects**.

2.1.2.2.1 Principles of Pixel-Level Anomaly Detection Pixel-level anomaly detection operates by analyzing **local spatial structures** within an image. The key techniques used include:

- **Reconstruction-Based Methods**: Models such as autoencoders (AEs) and generative adversarial networks (GANs) are trained to reconstruct normal images. Anomalous regions exhibit **high reconstruction errors**, as they deviate from the training distribution.

- **Density Estimation Methods**: Approaches like **normalizing flows** and **Bayesian deep learning** estimate the probability distribution of normal samples. Low-likelihood regions correspond to anomalies.

- **Feature Memory Banks**: Methods such as PatchCore and SPADE store normal feature embeddings and compare new image patches against stored representations. Anomalies are detected when new patches deviate significantly from memory-stored patterns.

2.1.2.2.2 Applications and Challenges Pixel-level anomaly detection is used in applications where **precise localization** of defects is critical. Some examples include:

- **Industrial Visual Inspection:** Identifying localized defects such as cracks, scratches, or missing components in products.

- **Power Line Tower Inspection:** Detecting corrosion, broken wires, or dirt accumulation in specific regions of a transmission tower.

- **Medical Imaging:** Segmenting tumor regions or structural abnormalities in radiological scans.

A key **advantage** of pixel-level methods is their ability to provide **detailed, interpretable outputs**, making them highly useful in safety-critical applications. However, they often require **more computational resources** and may suffer from **higher false-positive rates** due to fine-grained noise in images.

2.1.2.3 Comparison Between Image-Level and Pixel-Level Detection

Table 2.1 summarizes the main differences between image-level and pixel-level anomaly detection.

Table 2.1: Comparison of Image-Level and Pixel-Level Anomaly Detection

Feature	Image-Level Detection	Pixel-Level Detection
Output Type	Binary Label (Normal/Anomaly)	Segmentation Map / Heatmap
Localization	No localization	Precise localization
Computational Cost	Low	High
Model Complexity	Simpler	More complex
Common Methods	Feature-Based, Classification, PCA	Reconstruction, Density Estimation, Memory Banks
Best for	Rapid screening, simple defects	Complex structures, high-precision tasks
Example Use Cases	Industrial sorting, power grid scans	Medical imaging, defect segmentation

2.2 Controlled Datasets for Industrial Anomaly Detection

Datasets play a fundamental role in developing and evaluating machine learning models for anomaly detection, particularly in industrial settings. Many early studies rely on controlled datasets, where objects appear under standardized conditions, with uniform lighting and minimal environmental noise. These datasets help establish baseline performances and refine foundational methods. However, they often fail to capture the complexities of real-world applications, where variations in illumination, occlusions, and environmental factors significantly impact model performance.

A common strategy in novelty detection research is to repurpose multiclass classification datasets by selecting a subset of classes as anomalies. Models are trained exclusively on inlier classes and later tested on the artificially designated outliers. While this approach provides an abundance of labeled training and evaluation data, it does not reflect real-world anomaly detection scenarios, where anomalies are subtle deviations rather than entirely different object categories SALEH et al. (2013); HENDRYCKS; GIMPEL (2018). Some of the most frequently used datasets in this approach include MNIST, CIFAR-10, and ImageNet.

2.2.1 MNIST, CIFAR-10, and ImageNet-based Datasets

One of the simplest ways to benchmark anomaly detection methods is to adapt existing classification datasets. MNIST LECUN et al. (1998), CIFAR-10 KRIZHEVSKY; HINTON (2009), and ImageNet DENG et al. (2009) are commonly used in this manner. The approach consists of designating certain classes as anomalies while training the model exclusively on the remaining ones. During testing, the objective is to determine whether a sample belongs to an inlier or outlier class.

While this setup is computationally convenient, it does not provide a realistic assessment of anomaly detection capabilities. The primary issue is that anomalies in these datasets are entire object categories rather than subtle defects or structural irregularities. For example, if a model is trained on images of dogs and later tested on images of airplanes, it is trivially easy to identify the outliers. In contrast, industrial anomaly detection typically involves detecting small defects, such as microscopic cracks or missing components, within objects that otherwise appear normal.

Another limitation is the lack of fine-grained feature variations. In industrial applications, anomalies often share most of their visual characteristics with normal instances, making them much harder to distinguish. Classification datasets fail to capture this subtlety, as the outlier classes are inherently disjoint from the inliers. Furthermore, these datasets do not support pixel-wise annotation, preventing the evaluation of methods that require precise localization of defects.

2.2.2 Pascal VOC with Outliers

Recognizing the limitations of standard classification datasets, Saleh et al. SALEH et al. (2013) introduced an alternative approach based on the Pascal VOC dataset EVERINGHAM et al. (2010). Their dataset consists of abnormally shaped objects, such as irregularly structured cars, boats, and airplanes, which are compared against normal instances from Pascal VOC. The objective is to distinguish objects that deviate in shape from their expected class distributions.

While this dataset presents a more structured novelty detection problem, it still primarily focuses on entire objects rather than local anomalies. Many industrial applications require detecting small-scale deviations—such as minor deformations, missing parts, or surface defects—rather than identifying entirely different shapes. Additionally, the dataset is constructed from internet searches, which introduces inconsistencies and does not simulate real-world environmental challenges such as occlusions, lighting variations, or background noise.

2.2.3 MVTec AD – A Step Forward for Industrial Anomaly Detection

A significant advancement in industrial anomaly detection benchmarking was introduced with the **MVTec Anomaly Detection** dataset (**MVTec AD**) BERGMANN et al. (2019b). It consists of **5,354 high-resolution images** spanning **15 different real-world product categories**, including metal nuts, circuit boards, textiles, and pharmaceutical capsules. The dataset provides both normal and defective samples, with **pixel-wise ground truth annotations** for anomalies. This feature allows for a more comprehensive evaluation of anomaly detection methods, particularly in terms of **localization accuracy**, which was previously overlooked in classification-based benchmarks.

Figure 2.1 presents examples of normal and defective samples from MVTec AD, with normal instances in the top row and anomalies in the bottom row. The dataset captures a range of defects, including **scratches, dents, misaligned components, missing material, and contamination**, primarily focusing on **surface-level defects** and structural inconsistencies. While widely used for benchmarking unsupervised anomaly detection methods, MVTec AD remains a **controlled dataset**, with all samples captured under **ideal lighting and stable imaging conditions**. This controlled setting limits its ability to reflect real-world industrial inspection challenges, where **variable lighting, occlusions, background clutter, and environmental interference** significantly impact anomaly detection performance. Additionally, the dataset lacks anomalies caused by **external contaminants**, such as dirt accumulation, dust, or foreign object interference, which are common in practical applications like power line inspection.

The introduction of **MVTec 3D AD** WEIMAR et al. (2022) sought to address the limitations of **2D image-based inspection** by incorporating three-dimensional defect information. This dataset contains **over 4,000 high-resolution 3D scans** of industrial products across **10 categories**, where each sample is represented as an **organized point cloud** along with an **RGB image**. The alignment between the **3D structure and visual appearance** allows for detecting

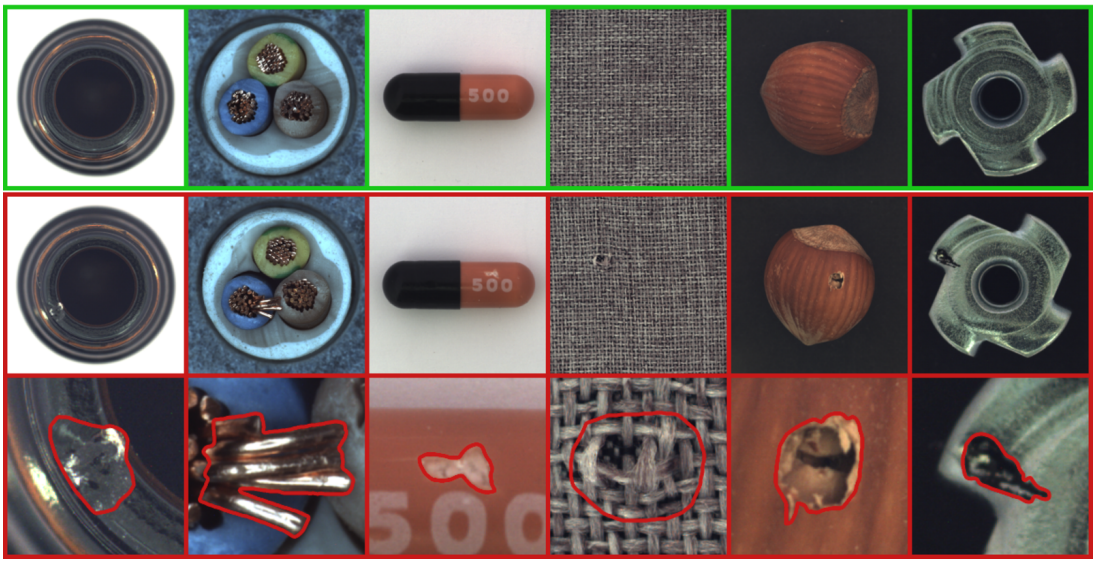


Figure 2.1: Example images from MVTec AD. The top row contains normal samples, while the bottom row shows defective samples with pixel-wise annotations of anomalies.

defects that are difficult to discern in 2D images, such as **subtle deformations, dents, and warping**. However, the dataset remains restricted to a **controlled acquisition environment**, where objects are scanned in an isolated setting without interference from external environmental factors.

Another extension, **MVTec LOCO AD (Logical Constraints Anomaly Detection)** MADER et al. (2022), introduces a different challenge by incorporating **logical anomalies**, in addition to conventional structural defects. The dataset consists of **3,644 images** across **five categories**, inspired by **real-world industrial inspection scenarios**. Logical anomalies arise from **misconfigurations, incorrect part placements, and assembly inconsistencies**, rather than material defects. While this dataset expands the definition of industrial anomalies, it still lacks **unstructured real-world conditions**, as it remains confined to **synthetic industrial environments** where logical inconsistencies are manually introduced rather than organically occurring.

Despite the advancements brought by these datasets, all MVTec benchmarks share common limitations. The controlled conditions under which images are captured fail to replicate real-world challenges such as **shadows, reflections, dirt, occlusions, and background distractions**, which can significantly impact model robustness. In industrial applications such as **power line inspection**, anomalies are not limited to **structural defects** but also include **foreign object interference**, such as **fallen branches, bird nests, and external debris**. These types of anomalies are entirely absent from the MVTec dataset family, making them insufficient for evaluating models intended for real-world deployment in outdoor industrial scenarios.

2.2.4 MIAD – A Synthetic Benchmark for Industrial Anomaly Detection

The MIAD dataset ZHANG et al. (2023) is a synthetic benchmark designed for anomaly detection in outdoor maintenance inspection scenarios. It consists of 105,000 images across seven industrial settings, including electrical insulators, metal welding, photovoltaic modules, wind turbines, catenary droppers, nuts and bolts, and witness marks. Unlike traditional manufacturing inspection datasets captured in controlled environments, MIAD introduces variability in surface conditions, backgrounds, and viewpoints to better simulate real-world challenges.

The dataset is generated using BlenderProc, a procedural rendering pipeline based on Blender, allowing precise control over lighting, textures, and object positioning while maintaining pixel-accurate annotations. However, despite its high-quality rendering, MIAD remains a synthetic dataset and lacks real-world imperfections such as sensor noise, motion blur, and natural lighting fluctuations. While it introduces environmental variability through procedural textures and randomized backgrounds, it does not capture the unpredictability of real-world industrial anomalies, such as foreign object interference or unexpected environmental conditions.

2.3 Real-World In-the-Wild Datasets for Power Line Inspection

While controlled and synthetic datasets provide valuable benchmarks, real-world datasets are essential for evaluating models under practical conditions. Power line inspection, in particular, demands datasets that account for environmental challenges such as uncontrolled lighting, occlusions, vegetation interference, and structural variations. Although several datasets have been proposed, many focus on detecting a **single power line component**, limiting their applicability to broader inspection tasks.

2.3.1 Single-Class Power Line Inspection Datasets

Several datasets focus on detecting and analyzing a single power line component, such as insulators, conductors, or transmission towers. While these datasets provide valuable data for specific tasks, they are inherently limited in scope, as they do not address the **complexity of full power line inspections**, where multiple components and their interactions must be considered.

The **Tomaszewski et al. Dataset** TOMASZEWSKI; RUSZCZAK; MICHALSKI (2018) contains **2,630 high-resolution images** (**5616×3744**) of power line insulators captured from nine video sequences. Despite offering numerous samples, it suffers from redundancy, as images are extracted from consecutive frames of the same recording. Additionally, lighting variations and insulator orientations are constrained by the original recording conditions, limiting dataset variability.

The **Chinese Power Line Insulator Dataset (CPLID)** TAO et al. (2020) comprises

600 real-world images of power line insulators and **248 augmented samples**, where defective insulators are synthetically placed against different backgrounds. It includes **1,569 annotations**, covering both intact insulators and missing cap defects. However, the use of synthetically generated defects raises concerns about domain shift when models are deployed in real-world scenarios.

For conductor segmentation, the **Powerline Dataset** LEE et al. (2017) provides **4,200 RGB images** and an equal number of infrared images, annotated for **semantic segmentation** of power lines. However, its relatively small image sizes (**128×128** and **512×512**) limit resolution, restricting fine-grained feature extraction.

Similarly, the **VEPL Dataset** CANO-SOLIS; BALLESTEROS; BRANCH-BEDOYA (2023) focuses solely on **vegetation encroachment** in power line corridors. While it provides vegetation masks to support automated maintenance, it does not address structural defects or failures in power line components.

The **Tower Dataset** BIAN et al. (2019) contains **1,300 images** of transmission towers collected from online sources and video footage. Although image resolutions vary and provide diverse tower settings, the dataset lacks detailed annotations and excludes other power line components.

The **XAI-Guided Insulator Anomaly Dataset** HOEFLER et al. (2024) introduces **explainable AI (XAI)-based anomaly detection** for power line insulators, specifically tackling **class imbalance**, a prevalent issue in industrial inspection datasets where defective samples are significantly underrepresented. By integrating **XAI techniques** into the anomaly detection pipeline, this dataset enhances interpretability, making it particularly relevant for real-world deployment. However, it remains limited to insulators and does not encompass broader power line infrastructure components or environmental interferences.

Despite their contributions, these datasets share a fundamental limitation: **they focus exclusively on a single power line component**, rather than addressing the broader challenge of comprehensive power line inspections. A real-world inspection scenario requires identifying **multiple interacting components**, detecting anomalies, and distinguishing between structural failures and external interferences.

2.3.2 Multi-Class Power Line Datasets

To overcome the limitations of single-class datasets, multi-class datasets have been introduced, covering various power line components within a single dataset.

The **Transmission Towers and Power Lines (TTPLA)** dataset consists of **1,100 aerial images** (**3840×2160 resolution**) with **pixel-wise instance segmentation annotations** for **transmission towers and power line conductors**. While providing a more detailed view of multiple power line components, it remains restricted to only two object classes.

A more diverse dataset is the **STN Power Line Asset Dataset (STN PLAD)**, which

contains **133 aerial images** annotated for object detection across five power line components: **transmission towers, insulators, tower ID plates, yokes, and Stockbridge dampers**. With high-resolution images, it provides dense annotations but lacks segmentation-level details.

The **PLT-AI Furnas Dataset** expands further by including **6,295 images** with **17,808 annotations** across five object classes: **insulators, balisers, bird nests, separators, and Stockbridge dampers**. In particular, it is one of the few datasets that includes **fault annotations**, offering valuable defect samples.

ImageNet also contains **1,290 images** of power lines and transmission towers with bounding box annotations. However, most images are ground-level shots, which makes them less applicable to aerial inspections.

Among the multiclass datasets designed for power line inspection, **InsPLAD-det** and **InsPLAD-fault** stand out due to their scale, diversity of component types, and inclusion of defective samples. As summarized in Table 2.2, InsPLAD-det provides object-level annotations across 17 distinct components under real-world UAV conditions, supporting large-scale object detection. InsPLAD-fault builds upon this structure by introducing image-level labels for five key component classes, enabling supervised and unsupervised anomaly detection. However, both versions are limited to coarse annotations, lacking the spatial detail required for fine-grained localization tasks.

Table 2.2: Comparison of publicly available datasets for industrial anomaly detection and power line inspection.

Dataset	Env.	Asset Classes	Annotation Type	Images	Defective Classes	Vision Tasks
MVTec AD	Controlled	15	Pixel-level	5,354	15	AD (Pixel), IC
MVTec LOCO AD	Controlled	5	Image-level + Logic	3,644	5	AD (Logical), IC
MVTec 3D AD	Controlled	10	3D + RGB	4,000	10	AD (3D), IC
MIAD	Synthetic	7	Pixel-level (synthetic)	105,000	7	AD, OD
TTPLA	Controlled	2	Instance segmentation	1,100	0	IS
STN PLAD	In-the-Wild	5	Bounding box	133	0	OD
PLT-AI Furnas	In-the-Wild	5	BBox + Faults	6,295	4	OD
CPLID	In-the-Wild	1	BBox + Synthetic	848	1	OD
VEPL	In-the-Wild	1	Vegetation masks	328	N/A	SS (Vegetation)
InsPLAD-det	In-the-Wild	17	Bounding box	10,607	—	OD
InsPLAD-fault	In-the-Wild	5	Image-level	10,378	5	IC, AD (Image-level)
InsPLAD-seg	In-the-Wild	5	Pixel-level	302	5	AD (Pixel), SS

2.3.3 Extending InsPLAD: From Fault Detection to Pixel-Level Segmentation

To address this limitation, our work introduces **InsPLAD-seg**, an extension of InsPLAD-fault that incorporates **pixel-level annotations of defects** for the same five component classes. With 302 manually segmented defect masks, this version enables **semantic segmentation and pixel-level anomaly detection**, supporting more accurate and interpretable inspection pipelines. Unlike existing datasets that focus on image-level classification, InsPLAD-seg facilitates the training and evaluation of models that require spatial precision, allowing for more robust and

detailed fault localization under in-the-wild inspection conditions.

2.4 Classical Approaches for Anomaly Detection

Traditional anomaly detection methods rely on statistical techniques, feature extraction, and shallow machine learning models. These approaches, while effective in controlled settings, often struggle with real-world complexities due to their reliance on predefined features and assumptions about data distribution.

One of the most widely used strategies in classical anomaly detection involves **dimensionality reduction techniques**, such as **Principal Component Analysis (PCA)**, **Linear Discriminant Analysis (LDA)**, and **Independent Component Analysis (ICA)**. These methods project high-dimensional data into lower-dimensional spaces, highlighting variations that may indicate anomalies. **PCA** is particularly effective in capturing dominant data trends, and it has been applied in industrial fault detection by identifying residual errors from PCA-reconstructed data DAGA (2019). **LDA**, on the other hand, maximizes class separability and has been successfully applied to gait anomaly detection SABOOR et al. (2020). **ICA** is particularly suited for mixed signal environments, such as EEG-based anomaly detection NEHA SARDANA; KANWADE; TEWARY (2021), where independent source separation aids in identifying unusual patterns. However, these techniques assume linear relationships in data and struggle with more complex anomalies.

Another category of classical approaches involves **frequency-domain transformations**, such as **Fourier Transform (FT)** and **Discrete Cosine Transform (DCT)**, which analyze data by decomposing it into frequency components. These methods have been widely used in industrial condition monitoring, where fault patterns manifest as changes in frequency spectra VITOR; GOEDTEL; CASTOLDI (2023). **DCT**, in particular, has been used in biomedical applications for detecting arrhythmias BOULIF; ANANOU; OULADSINE (2023). Despite their advantages in signal processing, these methods struggle with detecting local anomalies, as they focus on global frequency patterns rather than spatial irregularities.

Classical machine learning models, such as **Support Vector Machines (SVM)** and **Multilayer Perceptrons (MLP)**, have also been applied to anomaly detection. **SVMs** have been widely used due to their ability to handle high-dimensional feature spaces and small datasets SHARMA et al. (2023), but they require careful kernel selection and often perform poorly on complex, high-dimensional data. **MLPs**, while more flexible, lack spatial awareness, making them less effective for structured anomaly detection tasks such as defect localization CHANG'AN; MA; YANG (2022).

The emergence of **deep learning-based methods** has largely replaced classical approaches. **Convolutional Neural Networks (CNNs)** have demonstrated significant improvements in learning hierarchical feature representations for anomaly detection, outperforming traditional feature extraction-based methods SINGH et al. (2021). **Generative Adversarial**

Networks (GANs) have also become prominent for anomaly detection, leveraging their ability to model complex data distributions and generate synthetic anomalies for training LI et al. (2024).

Despite their historical significance, classical anomaly detection methods suffer from several limitations: they rely heavily on **manual feature engineering**, struggle to generalize to **complex, non-linear data**, and lack **localization capabilities** essential for real-world industrial anomaly detection. These constraints have led to the widespread adoption of deep learning approaches, which will be discussed in the following sections.

2.5 Modern Deep Learning Approaches for Anomaly Detection

With the rise of deep learning, anomaly detection has moved beyond traditional feature engineering, leveraging **object detection and semantic segmentation models** for improved accuracy and localization. Unlike classical methods, deep learning architectures extract hierarchical representations, making them highly effective for identifying **structural defects** and **out-of-distribution patterns**.

2.5.1 Object Detection and Segmentation for Anomaly Detection

Several datasets, including **InsPLAD-seg, MVTecAD, and STN PLAD**, provide object-level annotations for power line components, aligning with recent anomaly detection methods based on **object detection frameworks** such as **Single Shot Multibox Detector (SSD)**, **You Only Look Once (YOLO)**, and **Region-based CNN (RCNN)**. Additionally, segmentation-based models like **U-Net** have been widely adopted for **pixel-level anomaly localization**.

SSD and YOLO are particularly popular due to their **real-time performance**, enabling efficient detection of power line components **?**). While **RCNN-based models** provide **superior localization accuracy**, they come at the cost of higher computational complexity **?**). For finer-grained defect detection, **U-Net** has demonstrated strong **segmentation performance** by capturing both **global and local context ?**), laying the groundwork for more advanced anomaly detection architectures.

2.5.2 Advanced Architectures for Anomaly Detection

Beyond standard object detection models, several modern architectures have emerged, each introducing unique improvements to anomaly detection. Notable among them are **EfficientNet, SimpleNet, DifferNet, and RD++**, which offer different approaches to **feature extraction** and **representation learning**.

EfficientNet is a **high-performance model** originally designed for **image classification** but increasingly adopted for **anomaly detection** due to its **efficient scaling mechanisms**. SHETE

et al. (2024) proposed an **EfficientNet-based online-adaptive anomaly detection system**, leveraging **Mahalanobis distance-based feature extraction** for **defect identification in aircraft assembly**. Similarly, GULA; BERTOLDO (2023) demonstrated that **EfficientNet outperforms Principal Component Analysis (PCA)-based anomaly detection**, achieving **higher accuracy** with **fewer components**.

SimpleNet streamlines anomaly detection by employing a **pre-trained feature extractor**, a **shallow feature adapter**, and a **lightweight discriminator**.LIU et al. (2023) introduced a **SimpleNet model** that achieved **state-of-the-art performance** on the **MVTec AD benchmark**, attaining a detection **AUROC of 99.6%** while operating at **77 FPS**, making it highly suitable for **real-time applications**.

DifferNet is a **normalizing flow-based method** that models **image-level likelihoods** for **defect detection**. RUDOLPH; WANDT; ROSENHAHN (2020) first introduced **DifferNet** as a **semi-supervised anomaly detection model**, assigning **probability scores** to **image features**. More recently, **?**) improved upon **DifferNet** by incorporating **attention mechanisms**, significantly enhancing **anomaly localization** on the **InsPLAD-fault dataset**.

RD++ refines **representation disentanglement**, aiming for more robust **anomaly detection**. While recent studies have not directly applied **RD++** to industrial tasks, its **feature separation techniques** align with those used in **DifferNet**, making it a promising candidate for future applications in **power line anomaly detection**.

3

Methodology

In this chapter, we present the methodology employed to develop and evaluate an anomaly detection system for power line components in-the-wild. Our approach is evaluated with the InsPLAD-seg dataset, a new dataset designed for pixel-level semantic segmentation of anomalies derived from the InsPLAD-fault dataset SILVA et al. (2023). The newly introduced InsPLAD-seg dataset offers a comprehensive collection of labeled images that reflect real-world inspection conditions, serving as a robust foundation for training and benchmarking anomaly detection models.

To improve anomaly detection accuracy, we integrate attention mechanisms within convolutional neural networks (CNN), using state-of-the-art modules such as Squeeze and Excitation Networks (SENet) HU; SHEN; SUN (2018), Convolutional Block Attention Modules (CBAM) WOO et al. (2018) and Self-Supervised Predictive Convolutional Attentive Block (SSPCAB) RISTEA et al. (2022a). These mechanisms enhance feature extraction by refining spatial and channel-wise representations, thereby increasing model robustness against variations in lighting, occlusion, and background complexity. Furthermore, we examine multiple deep learning-based anomaly detection methods, including DifferNet, SimpleNet, EfficientAD, and Reverse Distillation++ (RD++), evaluating their performance under different conditions. A key component of our approach is background removal through object segmentation, where a YOLOv11-based segmentation model is trained to isolate components from noisy backgrounds, thereby reducing interference in feature learning.

To assess the effectiveness of our methodology, we employ AUROC (Area Under the Receiver Operating Characteristic Curve) as the primary evaluation metric, ensuring comparability with the existing literature. In addition, latency measurements are included to analyze computational efficiency.

The following sections detail the dataset, attention mechanisms, anomaly detection models with attention modules, segmentation-based background removal, and evaluation metrics, which form the basis for the experimental analysis in the next chapter.

3.1 InsPLAD-seg

The proposed InsPLAD-seg dataset is an evolution of the previous InsPLAD-fault dataset SILVA et al. (2023), focusing specifically on semantic segmentation to address the need for pixel-level detection of anomalies in the wild. The primary motivation for this adaptation was to enable detailed, precise identification of defects, facilitating better maintenance and monitoring of overhead power lines in real-world conditions.

InsPLAD-seg contains real-world inspection imagery of operating overhead power lines, encompassing both normal and anomalous components across different classes. This version includes segmentation masks for five critical components commonly encountered in power line inspections: Glass Insulators, Vari-grips, Lightning Rod Suspensions, Yoke Suspensions, and the Polymer Insulator Upper Shackle. Inclusion of detailed masks across multiple component types enables the development of robust, generalizable segmentation models capable of localizing defects across varied hardware, which is essential for real-world deployment in automated inspection pipelines and maintenance planning. The construction of the data set involved adapting the publicly available InsPLAD fault to incorporate segmentation masks for defects in these selected objects. Table 3.1 and Figure 3.1 details the attributes of the InsPLAD-seg data set.

For the annotation process, we utilized the Roboflow annotation tool DWYER; NELSON; HANSEN (2024). Roboflow provides an AI-assisted segmentation feature that allows annotators to select a few seed pixels or regions, from which the tool automatically proposes a segmentation mask based on geometric cues. This proved efficient for well-defined structural components, where boundaries are clear and consistent. However, for defect types such as *missing caps*, the process posed unique challenges, as the defect corresponds to an absent part rather than a visible anomaly. In such cases, annotators labeled the empty region where the component element should have been, effectively marking the gap between existing structures.

For *rust* defects, only visually significant and representative corrosion areas were annotated, excluding minor or superficial discoloration that would not affect component functionality. This distinction is critical, as small dirt marks or weathering patterns can closely resemble rust in appearance. To ensure annotation consistency and domain accuracy, annotators worked in close collaboration with experienced electrical engineers who routinely inspect these components. Their input was essential in determining defect thresholds and in distinguishing between harmless surface artifacts and actual degradations.

Figure 3.2 presents the distribution of pixel-level anomaly density across different component classes in the InsPLAD-seg dataset. It is evident that the *glass-insulator* (GI) class exhibits the lowest anomaly density, with minimal variation across samples. This is consistent with the nature of its primary defect type—*missing caps*—which are spatially localized and typically affect only a small region of the image. In contrast, components such as *vari-grip* (VG) and *polymer-insulator-upper-shackle* (PIUS) display broader distributions and higher

(a) Normal Glass Insulator

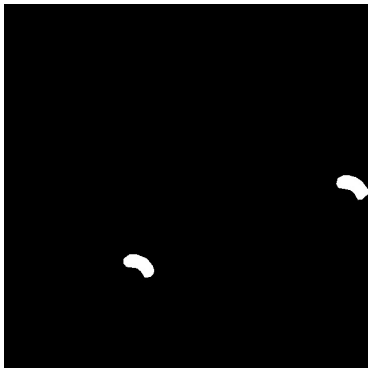
(b) Normal Lightning Rod Suspension

(c) Normal Vari-grip

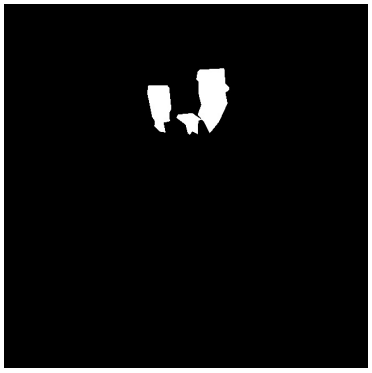
(d) Anomalous Glass Insulator

(e) Anomalous Lightning Rod Suspension

(f) Anomalous Vari-grip



(g) Anomaly Mask: Glass Insulator



(h) Anomaly Mask: Lightning Rod Suspension

(i) Anomaly Mask: Vari-grip

Figure 3.1: Visual Samples from the InsPLAD-seg Dataset. Examples of normal and anomalous components are shown in the first and second rows, respectively. The third row presents the corresponding pixel-level ground truth masks for the anomalies.

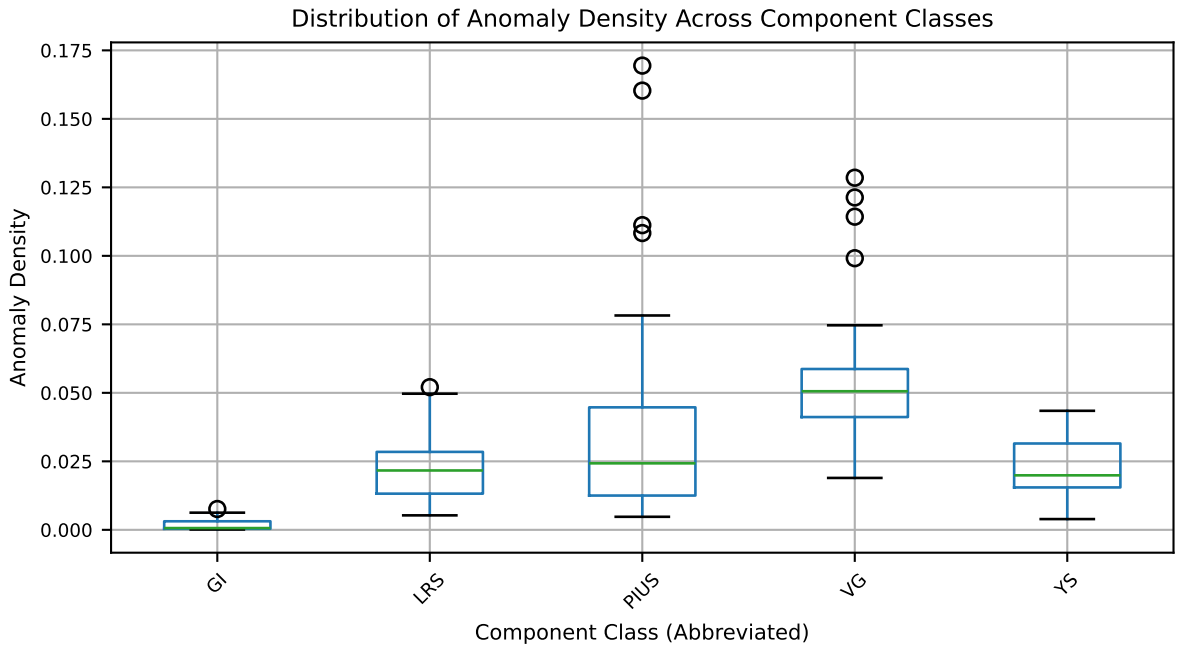


Figure 3.2: Distribution of Anomaly Density Across Classes. Each box represents the variation in pixel-level anomaly coverage within the annotated masks for each class in the InsPLAD-seg dataset. Abbreviations: GI = *glass-insulator*, LRS = *lightning-rod-suspension*, PIUS = *polymer-insulator-upper-shackle*, VG = *vari-grip*, YS = *yoke-suspension*.

average densities, often corresponding to visually extensive defects such as surface corrosion. These differences reinforce the importance of pixel-level analysis, as defect severity and spatial coverage vary substantially not only between classes but also within a single class. The use of domain expertise during annotation was crucial in ensuring accurate and consistent identification of defect boundaries, especially in cases where visual characteristics, such as rust versus dirt, are ambiguous.

The dataset includes several challenges to of real-world conditions accurately:

- Capture Conditions: Images were captured by drones inspecting fully functioning transmission power lines in various scenarios, all in the wild with no controlled background. This introduces variability in the images, such as different angles, occlusions, and background elements.
- Lighting Variations: Pictures were taken at different times of the day, resulting in significant lighting variations, including shadows, glare, and low-light conditions. This variability tests the robustness of detection algorithms under diverse lighting scenarios.
- Corrosion Patterns: The corrosion presented in the anomalous images exhibits different patterns and levels of severity. This diversity in corrosion types helps in training models to recognize a wide range of defects.

Asset category	Anomaly detection		
	Train	Test	
	Flawless	Flawless	Anomalous
Glass Insulator	1104	591	88
Lightning Rod Suspension	462	117	46
Vari-grip	477	114	40
Yoke Suspension	4834	1207	46
Polymer Insulator Upper Shackle	935	235	82
Total Images	**7.812**	**2.264**	**302**

Table 3.1: Description of the InsPLAD-seg anomaly detection and localization dataset. Glass Insulator anomalies are missing caps, while the others are corrosion-related. The number of mask annotations matches the number of anomalous images in each category.

3.2 Attention Mechanisms in Convolutional Neural Networks

Attention mechanisms have become a cornerstone in deep learning, enhancing the ability of Convolutional Neural Networks (CNNs) to emphasize potentially relevant features while downplaying less informative ones. Inspired by human cognitive attention, these mechanisms aim to focus computational resources on aspects of the input that are more likely to contribute to accurate predictions. In CNN architectures, attention modules have been shown to improve feature representation, particularly in tasks such as object detection, image classification, and medical imaging.

In our work, we evaluate three notable attention mechanisms: the **Squeeze-and-Excitation Networks (SENet)** HU; SHEN; SUN (2018), the **Convolutional Block Attention Modules (CBAM)** WOO et al. (2018), and the more recent **Self-Supervised Predictive Convolutional Attentive Block (SSPCAB)** RISTEA et al. (2022b). Each of these modules modulates the flow of information through the network by refining channel-wise dependencies, integrating spatial context, or introducing self-supervised learning to improve generalization. While these modules are designed to guide the network toward more informative regions or features, their effectiveness depends on training dynamics and may vary depending on the dataset and task.

3.2.1 Squeeze-and-Excitation Networks (SENet)

SENet HU; SHEN; SUN (2018) introduces an efficient channel-wise attention mechanism that recalibrates feature maps by learning interdependencies between channels. It consists of three key submodules:

1. **Squeeze Module**: Reduces spatial dimensions while retaining essential feature information.

2. **Excitation Module**: Learns and applies channel-wise attention weights using a Multi-Layer Perceptron (MLP).

3. **Scale Module**: Uses a sigmoid activation function to normalize learned weights and applies them to the original feature maps.

A detailed visualization of the SENet architecture is shown in Figure 3.3.

3.2.1.1 Squeeze Module

The Squeeze Module is responsible for adapting feature maps to optimize channel-wise attention. To achieve this, it employs a **global average pooling (GAP)** operation that computes the average activation for each channel, effectively condensing spatial information into a single descriptor per channel. This descriptor captures the global distribution of activations, ensuring that essential features are retained while reducing computational complexity.

Mathematically, the **squeezed** feature descriptor for a given channel c is computed as:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \tag{3.1}$$

where $x_c(i, j)$ represents the activation at spatial position (i, j) within channel c, and H and W denote the height and width of the feature map, respectively.

3.2.1.2 Excitation Module

Once the squeeze operation generates a compact feature descriptor, the Excitation Module applies a **fully connected (FC) network** that learns the relationships between channels. The excitation function consists of a **two-layer MLP** with a reduction factor r, reducing the channel dimension before expanding it back to its original size:

$$s = \sigma(W_2 \delta(W_1 z)) \tag{3.2}$$

where:

- $W_1 \in \mathbb{R}^{C/r \times C}$ and $W_2 \in \mathbb{R}^{C \times C/r}$ are learnable weight matrices,
- δ is the **ReLU activation function**,
- σ is the **sigmoid function**, ensuring the weights remain in the range $[0, 1]$.

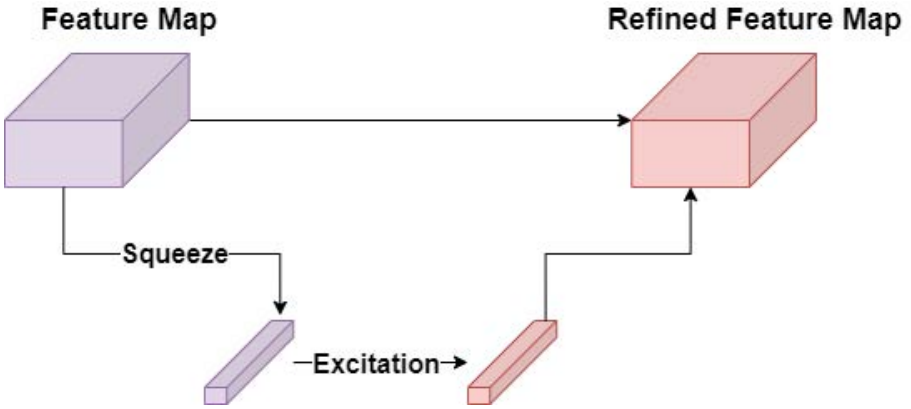


Figure 3.3: SENet module architecture representation. The Feature Map is refined by the Excitation application.

3.2.1.3 Scale Module

Finally, the output of the Excitation Module is used to **scale the original feature maps** via element-wise multiplication:

$$\hat{x}_c = s_c \cdot x_c \tag{3.3}$$

where s_c is the learned channel-wise attention weight, and x_c is the original feature map.

3.2.2 Convolutional Block Attention Modules (CBAM)

CBAM WOO et al. (2018) extends SENet by introducing **channel** and **spatial attention**, sequentially learning **what** and **where** to focus within feature maps (Figure 3.4). It consists of two submodules:

1. **Channel Attention Module** — focuses on inter-channel relationships.
2. **Spatial Attention Module** — focuses on spatially significant regions.

3.2.2.1 Channel Attention Module

Given an input feature map $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$, the channel attention mechanism aggregates spatial information using global average pooling (GAP) and global max pooling (GMP):

$$z_c^{avg} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \tag{3.4}$$

$$z_c^{max} = \max_{i,j} X_c(i,j) \tag{3.5}$$

where z_c^{avg} and z_c^{max} are the average and maximum pooled values for channel c.

These descriptors are forwarded to a shared MLP with a hidden layer, followed by summation and sigmoid activation to compute the attention map:

$$M_c = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot [z^{avg}; z^{max}]\right)\right) \tag{3.6}$$

where:

- $[z^{avg}; z^{max}] \in \mathbb{R}^{2C}$ is the concatenated vector;
- $W_1 \in \mathbb{R}^{\frac{C}{r} \times 2C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are the MLP weights with reduction ratio r;
- $\delta(\cdot)$ is the ReLU activation;
- $\sigma(\cdot)$ is the sigmoid function;
- $M_c \in \mathbb{R}^C$ is the channel attention map.

The output is obtained by multiplying M_c with the input feature map along the channel dimension.

3.2.2.2 Spatial Attention Module

The spatial attention module focuses on the **where** by aggregating information across channels. From the input feature map $\mathbf{X}' \in \mathbb{R}^{C \times H \times W}$ (output of the channel attention), it computes:

$$f^{avg} = \frac{1}{C} \sum_{c=1}^{C} X'_c \tag{3.7}$$

$$f^{max} = \max_c X'_c \tag{3.8}$$

where $f^{avg}, f^{max} \in \mathbb{R}^{H \times W}$ are average and max pooled maps over the channel dimension. These are concatenated and passed through a convolutional layer with a 7×7 kernel:

$$M_s = \sigma\left(f^{7\times7}([f^{avg}; f^{max}])\right) \tag{3.9}$$

where $M_s \in \mathbb{R}^{H \times W}$ is the spatial attention map.

The final output is obtained by applying M_s over $\mathbf{X}'$ via element-wise multiplication along the spatial dimensions.

3.2.3 Self-Supervised Predictive Convolutional Attentive Block (SSPCAB)

The Self-Supervised Predictive Convolutional Attentive Block (SSPCAB) RISTEA et al. (2022b) introduces a **masked prediction mechanism** to enforce global feature learning. By

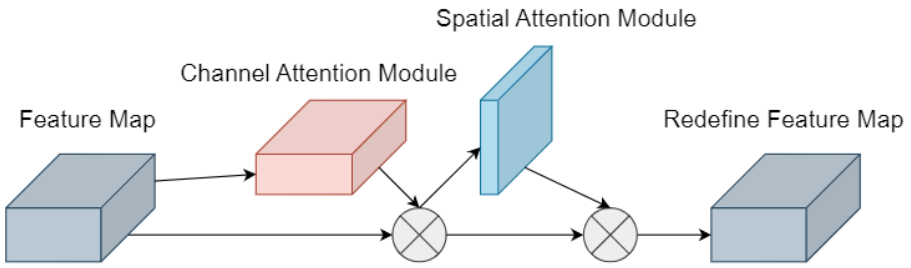


Figure 3.4: CBAM module architecture. The feature map is refined by sequentially applying channel and spatial attention.

intentionally **masking parts of the feature map** and learning to predict them, SSPCAB compels the network to develop a **stronger global contextual understanding**.

SSPCAB integrates:

- A **masked convolutional layer** for self-supervised feature prediction.
- A **Squeeze-and-Excitation (SE) module** for refining learned representations.
- A learning strategy that enhances **robustness** and **feature generalization** in low-data scenarios.

3.3 Anomaly Detection Methods with Attention Modules

This section presents the anomaly detection methods evaluated in this study. The selected models—**DifferNet, SimpleNet, EfficientAD, and Reverse Distillation++ (RD++)**—leverage deep learning techniques to model normality and detect deviations as an indicator of anomalies. In addition, we explore how **attention modules** can enhance feature extraction and improve anomaly localization in these models.

3.3.1 DifferNet

DifferNet RUDOLPH et al. (2021) is a **unsupervised image-based anomaly detection model** that leverages **normalizing flows** to estimate the probability density of image representations. The method extracts features from input images using **AlexNet** as a backbone, mapping them to a latent space via a normalizing flow model. The key principle is that normal data points form a high-likelihood region, while anomalous samples exhibit lower likelihood values.

Given an input image x, its feature representation is obtained through a convolutional encoder $f_\theta : X \to \mathbb{R}^d$, where d is the feature space dimension. These extracted embeddings are transformed by an invertible function g_ϕ, following the normalizing flow transformation:

$$y = g_\phi(f_\theta(x)), \tag{3.10}$$

where y follows a predefined simple distribution, such as a standard Gaussian $\mathcal{N}(0,I)$. The likelihood of a sample is computed using the **change-of-variables formula**:

$$p_X(x) = p_Y(y) \left| \det \frac{\partial g_\phi^{-1}}{\partial y} \right|. \tag{3.11}$$

The model is trained to maximize the likelihood of normal samples, making it possible to classify test images based on their likelihood scores.

To improve DifferNet's effectiveness, we introduce an attention-based variation called **AttentDifferNet**, where **SENet and CBAM attention modules** refine the extracted features. Attention mechanisms are incorporated at multiple depths to amplify relevant spatial and channel-wise information while suppressing irrelevant details. The proposed architecture is illustrated in Figure 3.5.

3.3.2 SimpleNet

SimpleNet LIU et al. (2023) is a **lightweight, efficient anomaly detection framework** designed for industrial inspection. Its training process involves extracting features from normal images and learning a compact representation that can distinguish normal from anomalous data. The process consists of:

1. **Feature Extraction** – A pre-trained convolutional network, such as **ResNet-18**, extracts features from input images.

2. **Feature Adaptation** – The extracted features are transformed to align with the target domain.

3. **Anomalous Feature Synthesis** – Artificial anomalies are generated by applying Gaussian noise to the adapted features.

4. **Discriminator Training** – A discriminator is trained to distinguish between normal and synthetic anomalous features.

During inference, the synthetic anomaly generator is removed, and the anomaly score for an image x is computed as:

$$s(x) = -\log D_\psi(f_\theta(x)), \tag{3.12}$$

where D_ψ represents the trained discriminator. To enhance its ability to detect subtle anomalies, we integrate **attention mechanisms** into the **ResNet-18 backbone**, following the approach of Vieira e Silva et al. SILVA et al. (2024b).

3.3.3 EfficientAD

EfficientAD BATZNER; HECKLER; KÖNIG (2024) is a **computationally efficient anomaly detection model** that combines a **patch description network (PDN)** with an **autoencoder**. The PDN is a shallow convolutional network responsible for extracting localized patch-level features, while the autoencoder reconstructs normal patterns to detect deviations.

The **student-teacher framework** employed by EfficientAD enhances anomaly detection by training a lightweight student network to replicate the knowledge of a deeper teacher model. The anomaly score for an image is a weighted combination of the **PDN score** $s_{PDN}(x)$ and the **autoencoder reconstruction error** $s_{AE}(x)$:

$$s(x) = \alpha s_{PDN}(x) + \beta s_{AE}(x), \tag{3.13}$$

where α, β are weighting parameters.

To further refine feature selection, we incorporate **attention modules within the PDN**, improving its ability to focus on fine-grained anomalies.

3.3.4 Reverse Distillation++ (RD++)

RD++ TIEN et al. (2023) is a **self-supervised anomaly detection model** based on a **wide ResNet-50 decoder** and a novel **feature distillation strategy**. Unlike standard teacher-student models, RD++ introduces **projection layers** after each intermediate teacher block to learn a more compact normal representation.

The model is trained using a multi-component loss function defined as:

$$L = \lambda_{KD} L_{KD} + \lambda_{SSOT} L_{SSOT} + \lambda_{Con} L_{Con} + \lambda_{Recon} L_{Recon}, \tag{3.14}$$

where:

L_{KD} — Knowledge distillation loss, transferring information from teacher to student.

L_{SSOT} — Self-supervised optimal transport loss, compacting normal feature distributions.

L_{Con} — Contrastive loss, encouraging separation between normal and anomalous features.

L_{Recon} — Reconstruction loss, regularizing the feature space around normal samples.

Additionally, RD++ introduces a **pseudo-anomalies mechanism**, in which artificial perturbations are generated during training to improve robustness and help the model detect subtle or underrepresented anomalies.

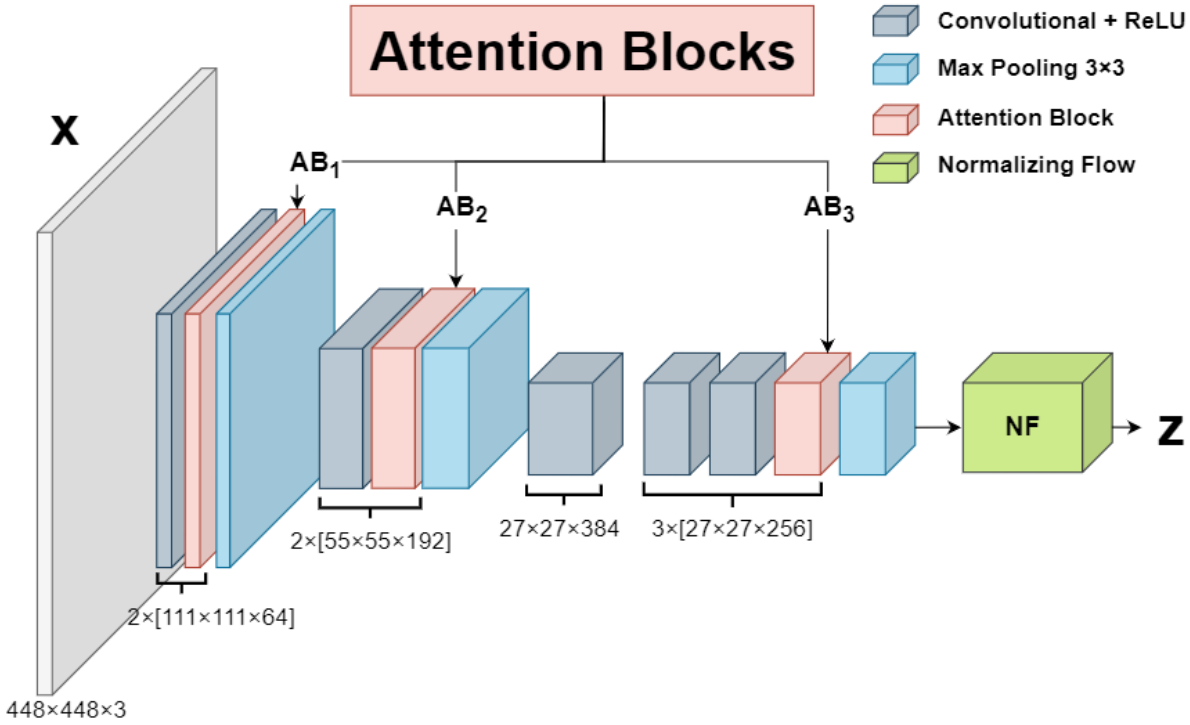


Figure 3.5: Proposed AttentDifferNet architecture.

3.3.5 Integrating Attention Modules in Anomaly Detection

To improve anomaly localization, **attention modules** are integrated into the feature extractors of each model, following the structured placement strategy proposed in **AttentDifferNet** SILVA et al. (2024b). This ensures a consistent enhancement of feature representations across different architectures.

For **ResNet-like architectures**, including SimpleNet and RD++, attention modules are inserted **before the summation with the identity branch** within each residual block. This allows attention-refined features to be properly combined with skip connections, maintaining information flow while improving feature discrimination.

For **Inception-like architectures**, attention modules are positioned **after each Inception module**, ensuring that multi-scale feature maps are effectively modulated.

For models with **shallow CNN-based feature extractors**, such as the **Patch Description Network (PDN) in EfficientAD**, attention mechanisms are integrated at **multiple depths** to refine both low-level and high-level feature representations.

All modifications adhere to the **structured attention integration pattern proposed in AttentDifferNet as in Figure 3.5**, ensuring that attention mechanisms consistently refine feature extraction across different anomaly detection methods.

3.4 Noisy Background Removal with Object Segmentation

To improve the performance of the anomaly detection methods, we propose a straightforward background removal approach to focus solely on the object of interest. The following steps outline our methodology:

1. **Annotation of Training Images for the Segmentation Model:** A subset of 100 images from the training set of the object class is manually annotated. The annotations included only the object of interest, ensuring precise object localization and segmentation.

2. **Training a Segmentation Model:** The annotated dataset was used to train a YOLOv11 segmentation model for 100 epochs. The model was configured to segment and detect the object of interest at a resolution optimized for the dataset. This step generated a segmentation model capable of accurately detecting and isolating the object of interest from images.

3. **Object Extraction from Full Dataset:** The trained YOLOv11 model was applied to the entire dataset (including both training and test sets). For each image, the model extracted only the segmented object of interest, effectively removing the noisy background. The resulting dataset contained only the isolated objects, preserving the class-specific details required for anomaly detection.

4. **Anomaly Detection Training and Evaluation:** Using the new dataset without noisy backgrounds, an anomaly detection model was trained and evaluated. The background removal process ensured that the anomaly detector focused exclusively on the object's intrinsic features, minimizing interference from environmental noise and enhancing detection performance.

This pipeline demonstrates an efficient approach to preparing data for anomaly detection by leveraging segmentation techniques to reduce background noise. Figure 3.6 illustrate the background removal process.

3.5 Evaluation Metrics

This section describes the metrics used to evaluate the anomaly detection methods in our experiments. The primary evaluation metric is the *Area Under the Receiver Operating Characteristic Curve (AUROC)*, which is widely employed in anomaly detection research to measure how well a model distinguishes between normal and anomalous instances. Additionally, we measure *latency* to compare the computational efficiency of different methods.

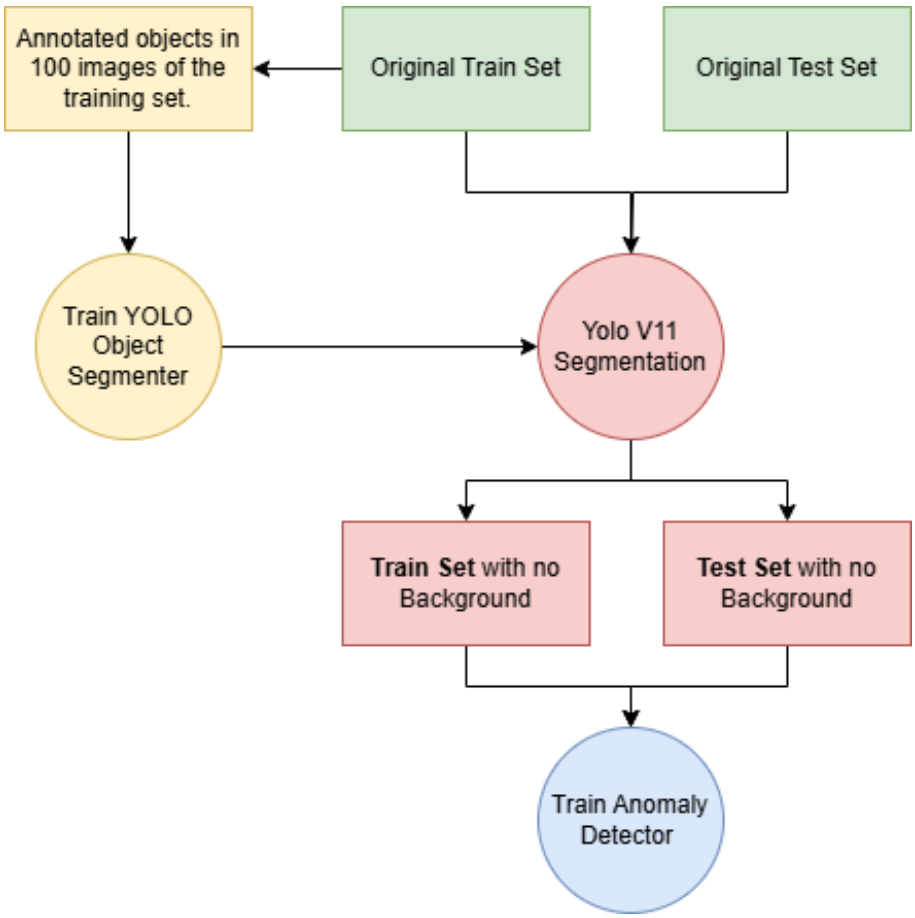


Figure 3.6: Workflow for Preparing Background-Free Datasets Using YOLO Segmentation for Anomaly Detection Training.

3.5.1 AUROC – Area Under the Receiver Operating Characteristic Curve

AUROC quantifies a model's ability to distinguish between normal and anomalous instances by measuring the area under the *Receiver Operating Characteristic (ROC) curve*. The ROC curve plots the *True Positive Rate (TPR)* against the *False Positive Rate (FPR)* as the decision threshold varies. These rates are formally defined as:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.15}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{3.16}$$

where *TP* (true positives) refers to correctly classified anomalous samples, *TN* (true negatives) refers to correctly classified normal samples, *FP* (false positives) are normal samples misclassified as anomalous, and *FN* (false negatives) are anomalous samples misclassified as normal.

The AUROC is computed as the area under the ROC curve, mathematically expressed as:

$$\text{AUROC} = \int_0^1 \text{TPR}(\text{FPR})\, d\text{FPR} \tag{3.17}$$

An AUROC of **1.0** indicates perfect discrimination, while an AUROC of **0.5** corresponds to random guessing.

3.5.1.1 AUROC for Image-Level Detection

At the image level, AUROC evaluates whether an entire image is correctly classified as normal or anomalous. The model assigns a *global anomaly score* to each image, and the ROC curve is constructed by thresholding these scores. The AUROC reflects how well the model separates normal from anomalous images.

3.5.1.2 AUROC for Pixel-Level Detection

At the pixel level, AUROC assesses how accurately the model localizes anomalous regions within an image. The model produces an *anomaly heatmap*, assigning a score to each pixel. The ROC curve is generated by varying a threshold over these pixel-level scores, and computing the corresponding TPR and FPR at each threshold.

3.5.1.3 Justification for AUROC

AUROC is a *threshold-independent* metric, making it particularly suitable for anomaly detection tasks where choosing an optimal threshold is non-trivial. It is widely adopted due to its robustness under class imbalance, which is inherent to most anomaly detection problems. The **MVTec Anomaly Detection Benchmark** BERGMANN et al. (2019b) systematically employs AUROC to evaluate industrial defect detection methods. Similarly, **SimpleNet** ZAVRTANIK; KRISTAN; SKOČAJ (2021) and **RD++** COHEN et al. (2022) report AUROC to benchmark their self-supervised anomaly detection frameworks. Following this standard practice ensures that the results in this study are directly comparable to prior work and reliably reflect detection performance.

3.5.2 Latency Measurement

Latency measurement is one of the main approaches to approximate the computational efficiency of an anomaly detection method. It is defined as the total time required for an input image to be processed and classified. This includes *preprocessing time*, *inference time*, and *postprocessing time*. The total latency for a given input image I is given by:

$$T_{\text{total}} = T_{\text{preprocessing}} + T_{\text{inference}} + T_{\text{postprocessing}} \tag{3.18}$$

where each component is measured in milliseconds (ms). To ensure fair comparisons, latency is measured under identical computational conditions, and results are reported as the *mean, standard deviation, and minimum processing time per image*.

3.5.3 Conclusion

This chapter outlined the methodology for anomaly detection in overhead power line components, leveraging the InsPLAD-seg dataset, attention-enhanced CNN architectures, and state-of-the-art anomaly detection models. We detailed the integration of SENet, CBAM, and SSPCAB to refine feature extraction, as well as the YOLOv11-based segmentation approach for removing background noise, ensuring improved inputs for anomaly detection.

To evaluate model performance, we adopted AUROC as the primary metric for detection accuracy and latency measurements to assess computational efficiency. The next chapter presents experiments and results, analyzing the impact of attention mechanisms, segmentation strategies, and different anomaly detection models on performance and efficiency.

4

Experiments, Results and Analysis

This chapter presents the experimental studies conducted to evaluate the performance and effectiveness of the proposed anomaly detection models. The experiments are designed to systematically analyze the impact of attention mechanisms, the trade-offs between accuracy and computational efficiency, and the influence of object segmentation on anomaly detection.

We begin with an ablation study on attention modules, where their contribution to image- and pixel-level performance is examined. This is followed by an in-depth analysis of anomaly detection models within the Insplad-seg dataset, highlighting quantitative and qualitative evaluations, as well as the effectiveness of attention mechanisms in distinguishing structural from appearance-based anomalies.

Next, we conduct an experimental evaluation of the anomaly detectors, emphasizing the balance between accuracy and efficiency. We also explore the impact of attention modules and examine latency versus accuracy trade-offs. Finally, we investigate the role of object segmentation in enhancing anomaly detection performance through controlled experiments on background removal and segmentation quality.

The results of these experiments provide crucial insight into the behavior of the model in different configurations, providing guidance for future improvements and practical applications in real world scenarios.

To simplify the visualization of the results in the following tables, we use abbreviated names for the models. Table 4.1 provides a reference for these abbreviations.

4.1 Ablation Study on Attention Modules

This section presents an ablation study of attention modules integrated into the RD++ and DifferNet anomaly detectors for fault detection in the wild. The objective of this preliminary experiment was to determine which attention mechanisms (SENet, CBAM, and SSPCAB) are most effective for anomaly detection at both the image and pixel levels. To achieve this, we first tested these three modules in DifferNet and RD++. Based on the results, we decided to discard SSPCAB due to its inferior performance. In the next experiment, we will evaluate SENet and

Table 4.1: Abbreviations of the evaluated methods

Abbreviation	Full Method Name
Diff	DifferNet
Diff-SE	DifferNet with SENet
Diff-CB	DifferNet with CBAM
RD	RD++
RD-SE	RD++ with SENet
RD-CB	RD++ with CBAM
SNet	SimpleNet
SNet-SE	SimpleNet with SENet
SNet-CB	SimpleNet with CBAM
EAD	EfficientAD
EAD-SE	EfficientAD with SENet
EAD-CB	EfficientAD with CBAM

CBAM in EfficientAD and SimpleNet to further investigate their effectiveness.

4.1.1 Experimental Setup

To ensure reproducibility, all experiments were conducted with a fixed random seed, set to `42`. Although alternative seeds were not tested, future work could explore their impact on training stability and performance variance.

Each model configuration—including the base anomaly detection methods and their variants incorporating attention modules—was trained for 100 epochs. This training duration follows the standard protocol used in the original papers of SimpleNet, DifferNet, RD++, and EfficientAD on the MVTec-AD benchmark, ensuring comparability with existing literature.

A learning rate scheduler was not employed in this phase to maintain consistency with the training procedures of the original implementations. Introducing schedulers could be considered in future experiments for further optimization.

Model evaluation was performed every 10 epochs, and the highest AUROC obtained during training was selected as the representative result, as illustrated in Figure 4.1. The train-test split adopted follows the configuration detailed in Table 3.1.

All experiments were conducted on a workstation equipped with an NVIDIA RTX 4060 GPU, ensuring consistent computational resources throughout training and evaluation.

4.1.2 Results and Discussion

This experiment evaluates the effect of attention modules on both image-level and pixel-level anomaly detection performance for DifferNet and RD++. The results are presented in Tables 4.2 and 4.3, with each entry formatted as I-AUROC%/P-AUROC%.

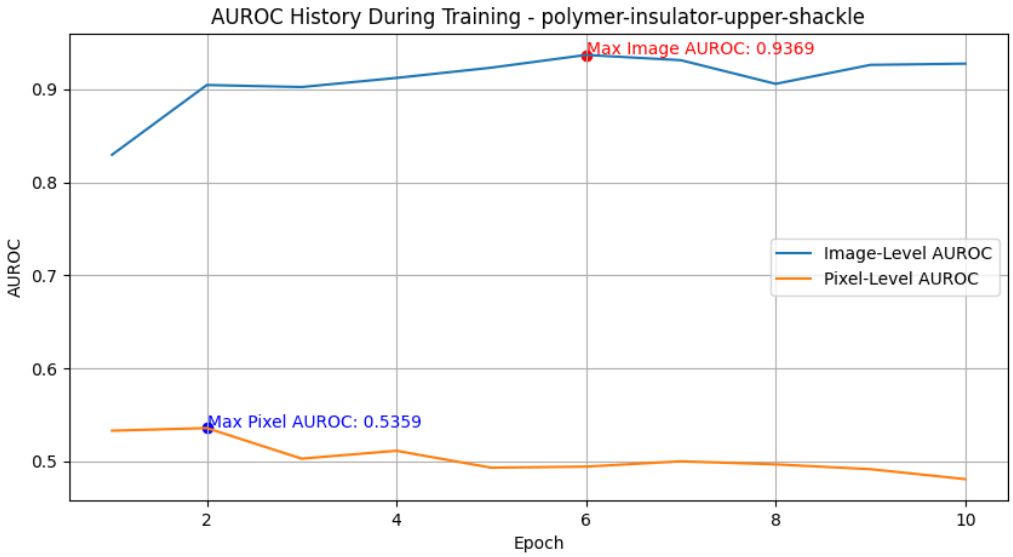


Figure 4.1: Training AUROC Progression on InsPLAD-Seg for the Polymer-Insulator-Upper-Shackle Class. The best Image-Level AUROC is highlighted in red and the best Pixel-LevelAUROC is highlighted in blue.

Table 4.2: Performance of DifferNet variants on the InsPLAD-seg dataset using model abbreviations. Each entry shows Image-Level AUROC / Pixel-Level AUROC (I-AUROC% / P-AUROC%)

Category	Diff	Diff-SE	Diff-CB	DifferNet(SSPCAB)
Glass Insulator	82.81/51.53	**86.57/62.96**	81.03/51.70	64.15/55.98
Lightning Rod Susp.	99.08/**55.76**	**99.62**/53.20	99.33/55.50	85.40/54.95
Vari-Grip	91.20/56.31	**93.52**/76.75	88.99/**77.31**	82.81/60.74
Yoke-Suspension	96.77/**59.89**	**97.38**/53.12	96.86/51.79	80.49/55.80
Polym.-Ins.-Up.-Shack.	92.42/52.96	**94.62/54.47**	92.10/50.38	90.12/52.69
Average	92.46/55.29	**94.34/60.11**	91.66/57.33	80.59/56.034

Table 4.3: Performance of RD++ variants on the InsPLAD-seg dataset using model abbreviations. Each entry shows Image-Level AUROC / Pixel-Level AUROC (I-AUROC% / P-AUROC%)

Category	RD	RD-SE	RD-CB	RD++(SSPCAB)
Glass Insulator	84.09/94.81	**85.13/95.28**	80.26/91.90	80.35/91.13
Lightning Rod Susp.	**94.45**/88.96	91.50/**92.00**	86.60/89.00	66.80/87.40
Vari-Grip	75.70/88.72	**87.21**/89.82	82.68/**92.05**	72.35/89.73
Yoke-Suspension	**92.58**/96.60	89.74/**97.16**	86.05/95.02	80.42/92.87
Polym.-Ins.-Up.-Shack.	**87.68**/89.78	82.74/88.36	76.31/89.25	74.10/**89.61**
Average	**86.90**/91.774	**87.28/92.52**	82.38/91.43	74.80/90.15

DifferNet: The base DifferNet model achieves an average image-level AUROC of **92.46%**, but its pixel-level performance remains lower (55.29%), indicating persistent limitations in capturing fine-grained anomalies. Among the attention modules, SENet leads to the highest

performance at both levels, achieving the best average image-level AUROC (**94.34%**) and the highest pixel-level score (**60.11%**). CBAM, in contrast, does not provide meaningful improvements — it slightly reduces the image-level average to 91.66%, while offering moderate pixel-level gains, particularly in the Vari-Grip class (**77.31%**). The SSPCAB module further reduces image-level performance to 80.59%, but achieves competitive results at the pixel level (56.034%), suggesting that its impact is more beneficial for localization rather than classification.

RD++: RD++ exhibits consistently high performance, especially at the pixel level. The base model already achieves an average of 91.774%, and SENet enhances this further to **92.522%**, along with a slight image-level boost (**87.285%**). While CBAM maintains strong pixel-level results in specific categories, such as Vari-Grip (92.05%), it does not improve overall performance and causes a decline in image-level AUROC (82.388%). SSPCAB again yields the lowest scores, indicating limited compatibility with the RD++ architecture.

Comparison: RD++ clearly outperforms DifferNet at the pixel level, and its integration with SENet provides the most robust performance across both evaluation metrics. DifferNet retains an edge in image-level AUROC but struggles with localization, and the attention modules offer only limited improvements.

The ablation study reveals that SENet is the most effective attention mechanism, particularly when integrated with RD++. It consistently improves or maintains performance in both evaluation levels. CBAM offers some category-specific gains, especially in pixel-level segmentation, but fails to boost the average performance for either model. SSPCAB does not yield benefits in any configuration and is the least effective across the board.

These findings suggest that RD++ is more compatible with architectural enhancements via attention modules, especially SENet. Future work will explore the generalizability of these modules within different detection architectures such as EfficientAD and SimpleNet.

4.2 Anomaly Detection Models in Insplad-seg

This section analyzes the results of integrating attention modules (SENet and CBAM) into different state-of-the-art anomaly detection models at both image and pixel levels.

4.2.1 Quantitative Analysis of Image-Level AUROC Results

The table 4.4 provided shows the image level performance of various anomaly detection techniques in different categories, with and without the application of attention modules (SENet and CBAM). The average AUROC scores offer a comprehensive view of how these attention modules influence the performance of each method.

1. **DifferNet** shows an increase in AUROC from 82.81 to 86.57 when SENet is applied in the Glass Insulator category. However, CBAM leads to a slight decrease to 81.03, suggesting that SENet provides a more robust enhancement for this model. This

pattern holds across most categories, where SENet consistently improves DifferNet's performance, while CBAM does not yield similar gains and, in some cases, even slightly degrades performance.

2. **RD++** demonstrates a similar pattern. For example, in the Vari-Grip category, the AUROC increases from 75.70 without any attention module to 87.21 with SENet and 82.68 with CBAM. However, CBAM tends to have a slightly lower impact compared to SENet in some cases, suggesting that SENet might be more effective in this context.

3. **SimpleNet** also benefits from attention modules, with a notable increase in performance when SENet is applied. For instance, in the Glass Insulator category, the AUROC improves from 78.50 to 81.17 with SENet, although it slightly drops to 78.40 with CBAM.

4. **EfficientAD** shows modest improvements with attention modules, although the gains are less pronounced compared to other techniques. For instance, in the Lightning Rod Suspension category, the AUROC improves from 89.48 with CBAM. Here, CBAM provides a more significant boost than SENet.

Table 4.4: Image-Level AUROC of all evaluated methods on the InsPLAD-seg dataset. Abbreviations defined in Table 4.1.

Category	Diff	Diff-SE	Diff-CB	RD	RD-SE	RD-CB	SNet	SNet-SE	SNet-CB	EAD	EAD-SE	EAD-CB
Glass Insulator	82.81	86.57	81.03	84.09	85.13	80.26	78.50	81.17	78.40	79.13	81.98	79.18
Lightning Rod Susp.	99.08	99.62	99.33	94.45	91.50	86.60	79.70	84.60	82.70	88.33	88.25	89.48
Vari-Grip	91.20	93.52	88.99	75.70	87.21	82.68	79.50	82.40	78.60	85.61	89.65	85.15
Yoke-Suspension	96.77	97.38	96.86	92.58	89.74	86.05	80.40	86.20	83.70	78.63	73.92	79.54
Polym.-Ins.-Up.-Shack.	92.42	94.62	92.10	87.68	82.74	76.31	89.80	84.80	81.14	83.22	77.97	78.53
Average	92.46	**94.34**	91.66	86.90	**87.26**	82.39	81.58	**83.83**	80.91	**82.98**	82.35	82.38

Table 4.5: Pixel-Level AUROC of all evaluated methods on the InsPLAD-seg dataset. Abbreviations defined in Table 4.1.

Category	Diff	Diff-SE	Diff-CB	RD	RD-SE	RD-CB	SNet	SNet-SE	SNet-CB	EAD	EAD-SE	EAD-CB
Glass Insulator	51.53	62.96	51.70	94.81	95.28	91.90	58.20	59.40	53.50	49.48	58.50	50.20
Lightning Rod Susp.	55.76	53.20	55.50	88.96	92.00	89.00	59.40	54.44	70.10	43.74	42.01	44.65
Vari-Grip	56.31	76.75	77.31	88.72	89.82	92.05	62.80	52.20	60.00	49.45	46.45	47.91
Yoke-Suspension	58.21	53.12	51.79	96.60	97.16	95.02	73.60	76.10	77.10	41.74	37.81	50.01
Polym.-Ins.-Up.-Shack.	53.65	54.52	50.38	89.78	88.36	89.25	71.70	73.90	77.20	48.91	30.25	33.48
Average	55.09	**60.11**	57.33	91.77	**92.52**	91.43	65.14	63.21	**67.58**	**46.66**	43.00	45.25

4.2.2 Quantative Analysis of Pixel-Level AUROC Results

The pixel-level results in table 4.5 show varying degrees of improvement when attention modules are applied:

1. **DifferNet** sees a notable improvement with SENet and CBAM in several categories. For example, in the Glass Insulator category, the AUROC increases from 51.53 to 62.96 with SENet and slightly decreases to 51.70 with CBAM. This suggests that SENet may be more effective than CBAM in enhancing DifferNet's performance for certain pixel-level tasks.

2. **RD++** demonstrates significant gains with attention modules across most categories. In the Glass Insulator category, the AUROC jumps from 94.81 without any attention module to 95.28 with SENet, and slightly decreases to 91.90 with CBAM. This trend is consistent in other categories, highlighting SENet's consistent ability to boost performance more effectively than CBAM.

3. **SimpleNet** also benefits from attention modules, particularly SENet, which improves its performance in categories like Yoke-Suspension, where the AUROC increases from 73.60 to 76.10 . Notably, CBAM achieves its highest average pixel-level AUROC in this model configuration.

4. **EfficientAD** shows modest improvements with both SENet and CBAM, although the gains are less substantial compared to other techniques. For instance, in the Lightning Rod Suspension category, the AUROC remains relatively stable, with only slight increases from 43.74 to 44.65 with CBAM.

4.2.3 Qualitative Analysis

To qualitatively assess the effect of attention mechanisms on anomaly localization, we employed Grad-CAM SELVARAJU et al. (2017), an explainable AI (XAI) technique that highlights the most influential regions of an input image based on the gradients of the model's predictions. The visualizations shown in Figure 4.2 illustrate that the integration of SENet in DifferNet substantially improves the model's ability to localize both the object and the associated anomaly, particularly in the Vari-Grip and Glass Insulator categories. In both cases, SENet helps the model concentrate on the relevant object structures by reducing the influence of background noise, as evidenced by more focused and interpretable activation maps.

For the glass-insulator, the SENet-enhanced model demonstrates a more concentrated focus on the structural defect, specifically the missing cap from the insulator chain. This indicates that SENet not only improves object localization but also enhances the model's sensitivity to structural anomalies, leading to more precise anomaly detection.

The figure 4.3 compares anomaly detection techniques applied to glass-insulator and Polymer Insulator Upper Shackle components using regular RD++, RD++ enhanced with SENet, and the ground truth mask. In both rows, RD++ with SENet shows significantly improved performance over regular RD++, with more precise and less noisy anomaly maps that better align with the ground truth. In the first row, SENet reduces false positives and focuses on relevant

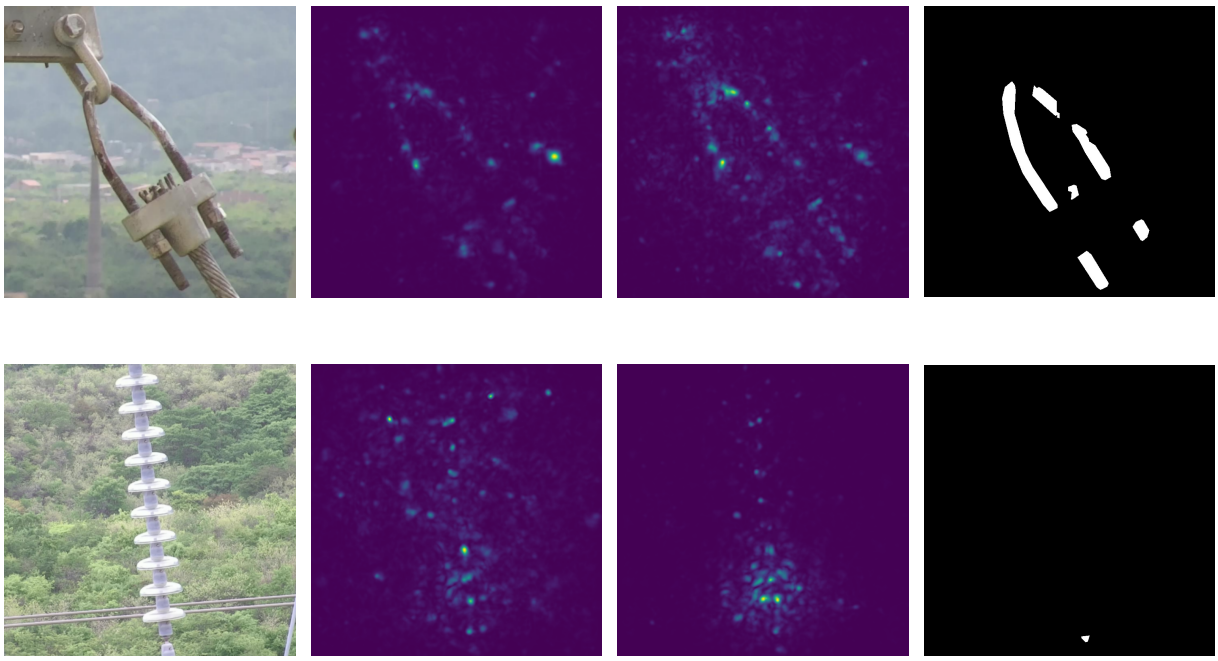


Figure 4.2: Comparison of anomaly detection techniques. First row: Vari-grip component (from left to right: Original image, regular DifferNet, DifferNet with SENet, and ground truth mask). Second row: Glass-insulator component (from left to right: Original image, regular DifferNet, DifferNet with SENet, and ground truth mask).

areas of the insulator, while in the second row, it accurately highlights defect-prone regions that match the ground truth annotations.

4.2.4 Effectiveness of Attention Modules in Detecting Structural vs. Appearance-Based Anomalies

The experimental results suggest that attention mechanisms, such as SENet and CBAM, offer **greater performance gains when detecting structural anomalies** than when addressing appearance-based ones. This difference likely stems from how attention modules enhance discriminative features. In structural anomalies—such as missing components—these changes are often more spatially and semantically distinct, allowing attention layers to focus more effectively on relevant spatial or channel-level signals.

As illustrated in Figure 4.4, the glass insulator missing its cap represents a structural anomaly characterized by the complete absence of a part. In contrast, the corrosion seen on the vari-grip is an appearance-based anomaly, which subtly alters surface characteristics like texture and color without affecting the object's structure. In our experiments, models equipped with attention modules achieved notably higher AUROC scores for structural anomalies (e.g., RD++ with SENet: 85.13 at image level and 95.28 at pixel level), reinforcing the hypothesis that attention helps localize and emphasize spatially distinctive features.

However, the benefit of attention was less evident for appearance-based defects. This

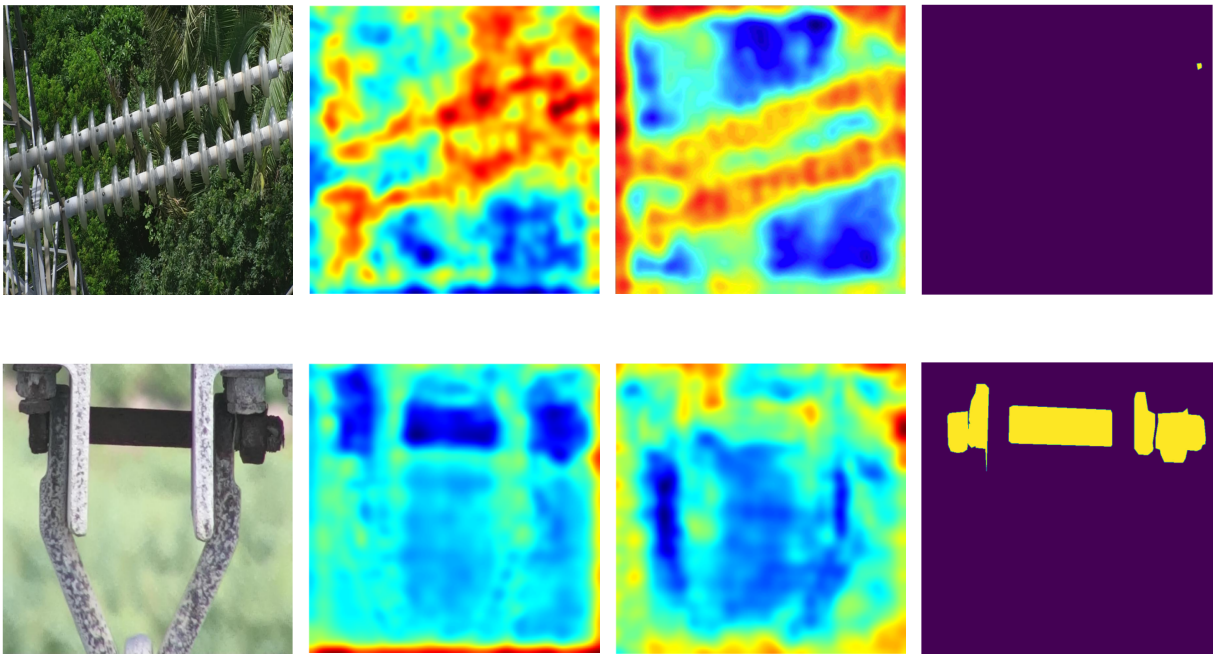


Figure 4.3: Comparison of anomaly detection techniques. First row: Glass-Insulator component (from left to right: Original image, regular RD++, RD++ with SENet, and ground truth mask). Second row: Polymer Insulator Upper Shackle component (from left to right: Original image, regular RD++, RD++ with SENet, and ground truth mask).

may be due to the diffuse and subtle nature of these changes, which do not significantly alter spatial structure and thus pose a challenge for attention modules to isolate. These findings suggest that while attention mechanisms can enhance detection performance, their effectiveness may depend on the anomaly type, and particularly on whether anomalies are spatially structured or visually subtle.

Overall, this distinction underscores the importance of tailoring architectural enhancements—such as attention modules—to the nature of the anomalies expected in a given application context.

4.3 Experimental Evaluation of Anomaly Detectors: Accuracy vs. Efficiency

Anomaly detection models must balance accuracy and efficiency to be practical for real-world applications. While AUROC is commonly used to measure detection performance, inference latency is equally critical in scenarios requiring real-time processing, such as defect inspection LIU (2024); VELESACA; CARRASCO; CARPIO (2024), autonomous driving VINOTH; SASIKUMAR (2025); BASARAN; DRESSLER (2025), medical imaging diagnostics LEE; LEE (2024), cybersecurity intrusion detection BHAROT; BRESLIN (2025); GIRUBAGARI; RAVI (2024), and industrial IoT for predictive maintenance PILLAI (2024); AJAYI

Figure 4.4: Comparison of structural and appearance-based anomalies: the left image shows a glass insulator missing its cap, representing a **structural anomaly**, while the right image depicts a vari-grip affected by corrosion, illustrating an **appearance-based anomaly**.

(2024).

Latency directly impacts the feasibility of deploying anomaly detection models in real-world applications. High-latency models, despite achieving superior accuracy, may not be viable for time-sensitive tasks, such as online quality control in manufacturing or real-time monitoring in medical imaging. Conversely, ultra-fast models may struggle with accuracy, limiting their reliability. Current state-of-the-art methods often prioritize accuracy at the expense of efficiency, making it crucial to evaluate and compare latency alongside AUROC.

This experiment aims to answer the following question: Which anomaly detection method offers the best trade-off between accuracy and inference time? We evaluate several state-of-the-art anomaly detection models on both image- and pixel-level AUROC, incorporating latency as a key performance factor. The results will help determine the most suitable models for real-time deployment.

4.3.1 Experiment Setup

To ensure a fair and reproducible evaluation, all models were tested on the same Insplad-Seg dataset using an NVIDIA RTX 4060 GPU. The evaluated models include:

- DifferNet, RD++, SimpleNet, and EfficientAD, along with their variants incorporating CBAM (Convolutional Block Attention Module) and SENet (Squeeze-and-Excitation Network).

The key metrics analyzed were:

- **AUROC (Image Level)**: Measures the ability to classify entire images as normal or anomalous.
- **AUROC (Pixel Level)**: Evaluates the ability to localize anomalies at the pixel level.
- **Latency**: The mean inference time per image across the entire test set, covering five anomaly classes from the Insplad-Seg dataset.

All models were run using batch inference with batch size 1 and default floating-point precision (FP32) on an NVIDIA RTX 4060 GPU.

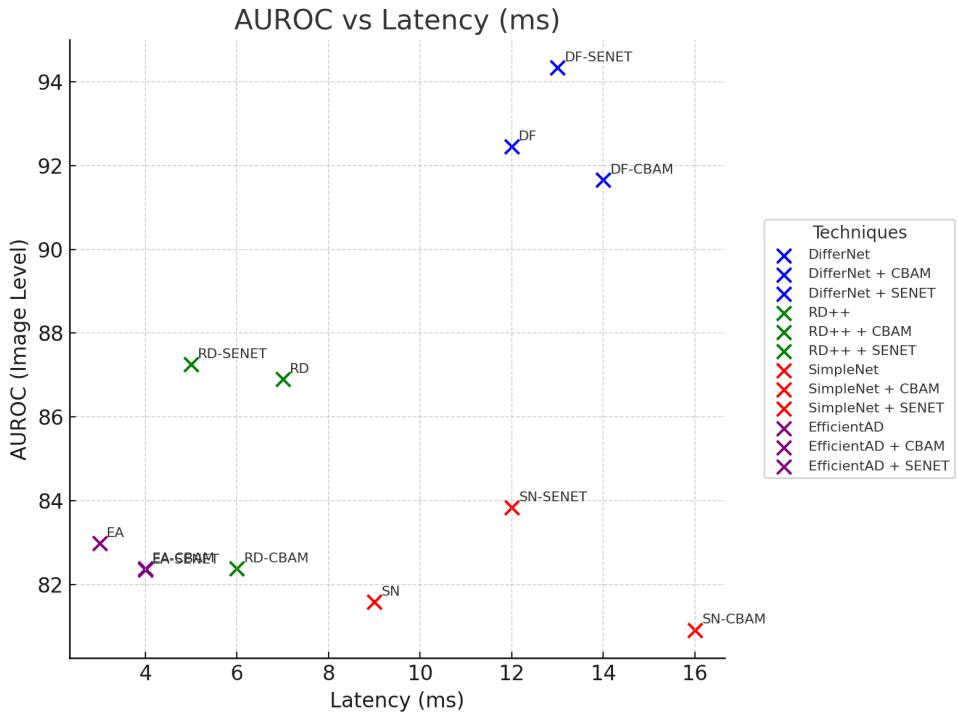


Figure 4.5: AUROC (Image Level) vs. Latency for various anomaly detection methods.

4.3.2 Results

The results of the experiment, as shown in Figures 4.5 and 4.6, reveal significant trade-offs between accuracy and latency among the evaluated models.

4.3.2.1 Image-Level Performance

- DifferNet variants achieved the highest AUROC scores, exceeding 94 with SENET, but had a higher latency (12–16 ms) when compared with the other methods.
- RD++ variants maintained competitive AUROC (86–88) while significantly reducing latency (5–9 ms).

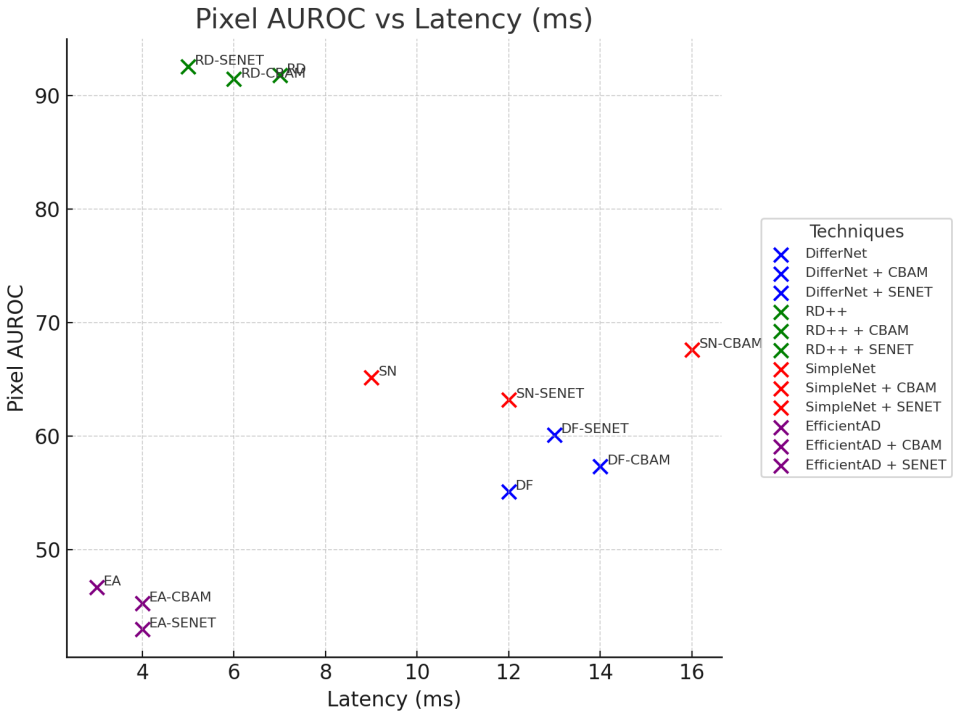


Figure 4.6: AUROC (Pixel Level) vs. Latency for various anomaly detection methods.

- EfficientAD, although it had lower accuracy (AUROC close to 82%), excelled in efficiency ($<$ 6 ms), making it ideal for latency-sensitive applications.

4.3.2.2 Pixel-Level Performance

- RD++ variants outperformed others, achieving AUROC scores of approximately 90 while keeping latency low.
- DifferNet variants had moderate pixel-level accuracy (AUROC 65–75) but suffered from high latency.
- EfficientAD and SimpleNet struggled, with AUROC below 55%, despite their speed.

AUROC was computed from a single test run using fixed seeds for reproducibility. While this eliminates random variations, further studies could explore variance across multiple test runs.

4.3.2.3 Impact of Attention Modules

- CBAM and SENet generally improved AUROC scores but increased latency.
- RD++ benefited the most, with AUROC gains at a latency cost of only 1–2 ms.
- EfficientAD and SimpleNet saw minimal accuracy gains with significant latency overhead, reducing their practicality in real-time applications.

4.3.3 Latency vs. Accuracy Trade-offs

- DifferNet variants prioritize accuracy over latency, making them suitable for batch processing or offline inference.
- EfficientAD variants sacrifice accuracy for fast inference, making them ideal for edge computing applications.
- RD++ finds a balance, delivering high accuracy with low latency, making it ideal for real-time scenarios.

4.3.4 Conclusion

This experiment evaluated anomaly detectors based on accuracy (AUROC) and inference latency, revealing a clear trade-off between the two.

Scenario	Best Model (AUROC)	Best Model (Latency)	Best Trade-off
Image-Level Performance	DifferNet(SE)	EfficientAD	RD++(SE)
Pixel-Level Performance	RD++(SE)	EfficientAD	RD++(SE)

Table 4.6: Comparison of the best-performing models. DifferNet achieved the highest AUROC but incurred the highest latency. EfficientAD had the lowest latency but comparatively lower AUROC, making it suitable for time-critical applications with relaxed accuracy demands. RD++ with SENet provided strong AUROC scores at both image and pixel levels, with only a modest increase in latency, making it the most balanced model for real-time deployment.

Ultimately, RD++ with SENet emerges as the best overall choice, offering a strong balance between high AUROC scores and low inference latency. Its ability to deliver robust image- and pixel-level performance with only a modest increase in computational cost makes it particularly suitable for real-time industrial applications where both accuracy and speed are critical.

4.4 Object Segmentation for Enhanced Anomaly Detection Performance

This section evaluates the impact of applying an object segmentator on the dataset, presenting quantitative and qualitative results that highlight its effectiveness in reducing background interference and improving overall anomaly detection performance.

In anomaly detection tasks, especially in uncontrolled or "wild" environments, background noise can obscure critical features, complicating the detection process. To mitigate this issue, we manually annotated a subset of the data set to create object-specific training data for

segmentation models. Using these annotations, we trained YOLOv11 models, with each class having its dedicated object segmentator tailored to isolate the object of interest.

EfficientAD was selected for this experiment because it exhibited the lowest pixel-level AUROC scores among all evaluated methods (see Table 4.5). This makes it an ideal candidate to test the effectiveness of background removal: models with weaker spatial localization stand to benefit the most from the suppression of irrelevant visual context. In other words, any improvement in anomaly localization after segmentation would be more easily observed in a model where performance is initially limited.

4.4.1 Experiment 1: Evaluating Background Removal on Anomaly Detection Performance

This experiment aimed to evaluate the impact of background removal on anomaly detection performance using the state-of-the-art anomaly detector, EfficientAD. This model was tested across all five classes of the Insplad-Seg dataset in two formats: the original images with their natural backgrounds intact and the pre-processed images where the background was removed using class-specific YOLOv11 object segmentators.

The objective was to determine how background noise influences the effectiveness of EfficientAD in identifying anomalies and whether the removal of background elements improves detection accuracy. The results provide a baseline comparison to assess the benefits of integrating object segmentation into the anomaly detection pipeline.

Table 4.7: AUROC results for EfficientAD with and without background. Abbreviations defined in Table 4.1.

Image-Level AUROC (%)						
Category	**EAD**	**EAD-SE**	**EAD-CB**	**EAD-NB**	**EAD-NB-SE**	**EAD-NB-CB**
Glass Insulator	79.13	81.98	79.18	81.61	81.32	83.59
Lightning Rod Suspension	88.33	88.25	89.48	92.36	93.21	90.46
Vari-Grip	85.61	89.65	85.15	80.24	79.64	83.20
Yoke-Suspension	78.63	73.92	79.54	86.27	83.37	86.17
Polymer Insulator	83.22	77.97	78.53	88.21	86.54	88.39
Average	82.98	82.35	82.38	85.74	84.82	**86.36**
Pixel-Level AUROC (%)						
Category	**EAD**	**EAD-SE**	**EAD-CB**	**EAD-NB**	**EAD-NB-SE**	**EAD-NB-CB**
Glass Insulator	49.48	58.50	50.20	71.68	69.54	74.02
Lightning Rod Suspension	43.74	42.01	44.65	52.40	46.80	63.68
Vari-Grip	49.45	46.45	47.91	77.00	59.65	76.97
Yoke-Suspension	41.74	37.81	50.01	83.21	73.62	83.21
Polymer Insulator	48.91	43.00	45.25	66.41	66.41	61.87
Average	**46.66**	43.00	45.25	71.68	66.41	**72.53**

Table 4.7 presents the AUROC results at the image level and pixel levelpixel level for EfficientAD and its attentionvariants,iants, with and without background removal. Each column represents a specific model configuration, following the abbreviation conventions defined in Table 4.1. Models labeled with the suffix *-NB* were trained on a version of the dataset in which background regions were removed, isolating only the object of interest. The highest and second-highest average AUROC scores within each group are highlighted for ease of comparison.

The results clearly demonstrate that training the anomaly detection models using the background-free dataset significantly outperforms regular training on both evaluation levels. At the image level, the best average AUROC was achieved by EAD-NB-CB (**86.36**%), representing a notable improvement over the regular EfficientAD baseline (EAD, 82.98%). Similarly, in the pixel-level evaluation, EAD-NB-CB again achieves the top score (**72.53**%), compared to only **46.66**% for the original EAD. These findings highlight that background removal effectively suppresses irrelevant features and allows the models to concentrate on the target object, resulting in more accurate anomaly localization and classification.

Interestingly, prior to background removal, the integration of attention modules such as SENet and CBAM yielded minimal improvements to EfficientAD, suggesting limited synergy under noisy input conditions. However, when trained on the background-free dataset, CBAM demonstrates a substantial performance gain, surpassing both the baseline and SENet-augmented variants. These findings highlight that background removal not only enhances anomaly localization but also enables attention mechanisms to operate more effectively, allowing the models to concentrate on the target object by suppressing irrelevant features.

The impact is particularly pronounced in challenging categories such as *Yoke-Suspension* and *Polymer Insulator*, where both image- and pixel-level performance increased substantially after background removal. This suggests that segmentation-based preprocessing is especially beneficial in scenarios where background clutter overlaps with subtle anomaly cues.

4.4.2 Experiment 2: Impact of Segmentation Quality on Anomaly Detection Performance

The second experiment evaluates how the quality of object segmentation impacts anomaly detection performance. While background removal enhances the focus on objects of interest, inaccuracies in segmentation—such as cutting out critical parts of the object or missing regions with anomalies—can negatively affect the overall detection performance. To quantify this, we trained EfficientAD models on the background-free training set segmented using YOLOv11 and tested them on two versions of the same test set: one with perfectly manually annotated objects and the other segmented by the YOLOv11 model. By comparing the AUROC scores between these two scenarios, we aimed to understand the extent to which automatic segmentation might reduce performance. The results reveal a measurable drop in AUROC when using automatically segmented objects compared to manually annotated ones, highlighting that the accuracy of

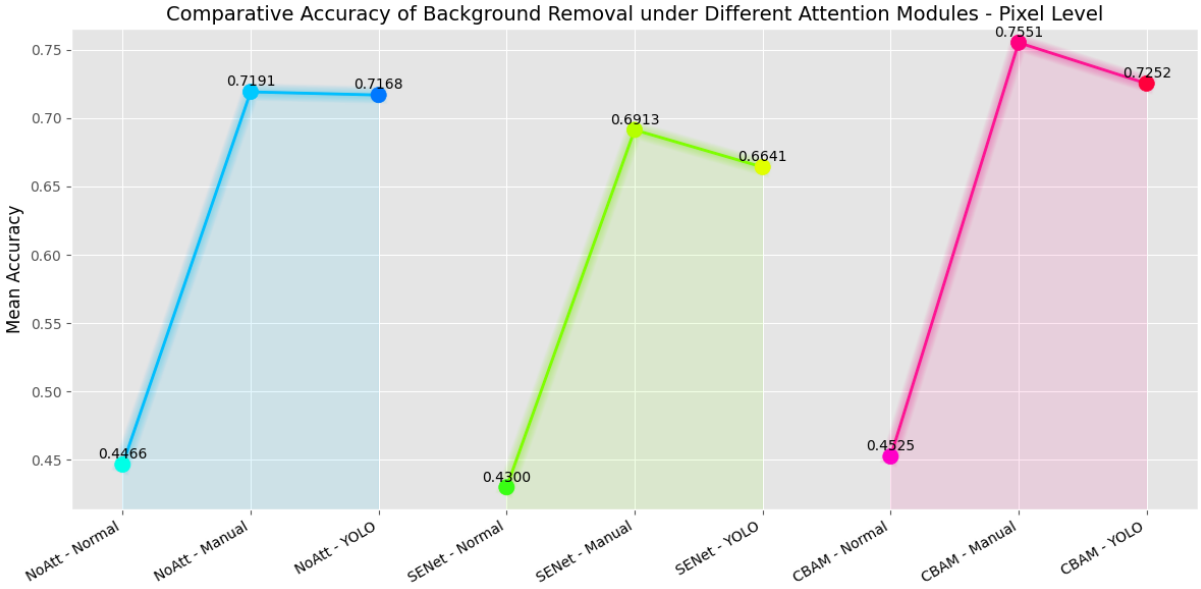


Figure 4.7: Comparative pixel-level accuracy of background removal under different attention mechanisms (No Attention, SENet, and CBAM) combined with three segmentation strategies: basic thresholding ("Normal"), manual segmentation, and YOLO-based segmentation. Results indicate that attention modules generally enhance performance, with CBAM combined with manual segmentation achieving the highest mean accuracy.

segmentation plays a crucial role in anomaly detection. This underscores the importance of improving segmentation quality in automated preprocessing pipelines to minimize potential losses in detection reliability.

Quantitative Analysis

It was anticipated that testing on YOLO-segmented data would result in some performance loss compared to testing on objects with perfectly manual segmentation. This is expected as no segmentation model is flawless and there is always the risk of cutting parts of the object, potentially including regions with anomalies. However, as shown in Figure 4.7, this performance loss is relatively small and represents a worthwhile trade-off for the benefits of automation.

At the **pixel level**, where a finer localization of the anomalies is required, there is a noticeable but modest drop in performance across all attention modules. For instance, the model using CBAM attention shows a decrease in mean accuracy from **75.518%** (manual segmentation) to **72.528%** (YOLO-based segmentation). Similar trends are observed for SENet and the baseline without attention, with accuracy differences consistently below 3%.

These results confirm that while manual segmentation yields the highest accuracy, YOLO-based segmentation introduces only a marginal degradation. The visual trend in Figure 4.7 supports this conclusion, as the YOLO bars remain close to their manually segmented counterparts. The ability to automate object segmentation for background removal simplifies the pipeline, making it feasible for real-world applications without substantially compromising detection

accuracy. This balance of scalability and precision reinforces the viability of YOLO-based segmentation in operational anomaly detection systems.

5

Conclusion

The visual inspection of power line components remains a critical yet challenging task due to the operational scale and variability of real-world environments. This dissertation addressed these challenges by investigating how attention mechanisms and background removal via object segmentation can improve the effectiveness of unsupervised anomaly detection models—aligning with the core research hypothesis and objectives defined at the outset. When looking back at our research objectives we achieved our goals.

To address Objective 1, we reviewed and analyzed several state-of-the-art unsupervised anomaly detection methods, identifying their limitations in uncontrolled environments and motivating architectural enhancements. For Objective 2, we created the *InsPLAD-seg* dataset by annotating pixel-level defect masks, providing a foundation for fine-grained evaluation under realistic conditions.

Fulfilling Objective 3, a benchmarking study was conducted on *InsPLAD-seg*, evaluating four representative anomaly detection methods (DifferNet, RD++, SimpleNet, and EfficientAD) at both image and pixel levels. The results highlighted consistent performance gains when integrating attention modules (Objective 4), particularly for structural anomalies like broken or missing components—validating the hypothesis that attention enhances spatial awareness and focus. While improvements were also observed for appearance-based anomalies, such as corrosion, the gains were less pronounced, suggesting that attention mechanisms are especially effective when anomalies disrupt structural integrity.

Finally, to address Objective 5, we evaluated the effect of background removal using automated object segmentation. The experiments demonstrated that removing wild, cluttered backgrounds improves detection performance, and importantly, that automated masks generated by YOLOv11 yielded results comparable to manual annotations. This supports the feasibility of applying segmentation-based preprocessing in scalable or real-time industrial settings.

Despite these promising findings, some limitations remain. The quality of segmentation directly impacts anomaly detection accuracy, and model generalization across domains or unseen anomaly types requires further investigation.

Future work should explore domain adaptation techniques to generalize this frame-

work across infrastructure types—such as railway or wind turbine inspection—and investigate lightweight, real-time implementations suitable for deployment on UAV platforms or edge devices.

In conclusion, this dissertation achieves all five proposed research objectives by demonstrating that the combination of attention mechanisms and object segmentation can significantly improve unsupervised anomaly detection performance in complex, real-world settings. The resulting approach not only advances the scientific understanding of how these modules interact but also offers practical solutions for industrial inspection at scale.

6

Acknowledgements

This research also received financial and technical support from the ANEEL project from STN - Sistema de Transmissão Nordeste S.A. through the ANEEL R&D Program for the development of the research project under grant code PD-04825-0006/2019.

Finally this work is part of the Universal project "Inspeção visual através de detecção de anomalias em cenários não-controlados", project number: 421727/2023-3, from Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

References

AHMED, M.; MAHMOOD, A. N.; HU, J. A survey of network anomaly detection techniques. **Journal of Network and Computer Applications**, [S.l.], v.60, p.19–31, 2016.

AJAYI, R. Integrating IoT and Cloud Computing for Continuous Process Optimization in Real Time Systems. **International Journal of Research Publication and Reviews**, [S.l.], 2024.

BASARAN, O.; DRESSLER, F. XAInomaly: explainable, interpretable, and trustworthy ai for next generation ultra-reliable low-latency communications (xurllc) in 6g networks. **Technical Report, TU Berlin**, [S.l.], 2025.

BASIT, A.; MANZOOR, H.; AKRAM, M. Machine learning-assisted anomaly detection for power line components: a case study in pakistan. **The Journal of Engineering**, [S.l.], 2024.

BATZNER, K.; HECKLER, L.; KÖNIG, R. EfficientAD: accurate visual anomaly detection at millisecond-level latencies. In: IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV). **Proceedings...** [S.l.: s.n.], 2024. p.128–138.

BERGMANN, P. et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. **IEEE Access**, [S.l.], v.7, p.44264–44276, 2019.

BERGMANN, P. et al. MVTec AD—A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). **Proceedings...** [S.l.: s.n.], 2019. p.9592–9600.

BHAROT, N.; BRESLIN, J. A Transfer Learning-Enhanced TabNet-Based Intrusion Detection System for IoMT Security. **Journal of Cybersecurity**, [S.l.], 2025.

BIAN, K. et al. Detection of transmission towers based on deep learning and aerial images. **IEEE Transactions on Geoscience and Remote Sensing**, [S.l.], v.57, n.12, p.10318–10330, 2019.

BOULIF, A.; ANANOU, B.; OULADSINE, M. A literature review: ecg-based models for arrhythmia diagnosis using artificial intelligence techniques. **Computers in Biology and Medicine**, [S.l.], 2023.

CANO-SOLIS, A.; BALLESTEROS, L.; BRANCH-BEDOYA, J. VEPL: a dataset for vegetation encroachment detection in power line corridors. **arXiv preprint arXiv:2303.10423**, [S.l.], 2023.

CHALAPATHY, R.; CHAWLA, S. Deep Learning for Anomaly Detection: a survey. **arXiv preprint**, [S.l.], 2019.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: a survey. **ACM Computing Surveys (CSUR)**, [S.l.], v.41, n.3, p.1–58, 2009.

CHANG'AN, A.; MA, K.; YANG, F. Multi-classification of arrhythmias using ResNet with CBAM on CWGAN-GP augmented ECG Gramian Angular Summation Field. **Biomedical Signal Processing and Control**, [S.l.], 2022.

CHEN, C. et al. Automatic clearance anomaly detection for transmission line corridors utilizing UAV-Borne LIDAR data. **Remote Sensing**, [S.l.], v.10, n.4, p.613, 2018.

COHEN, N. et al. Remote Supervision for Anomaly Detection: improving unsupervised anomaly detection with weak labels. **arXiv preprint**, [S.l.], v.arXiv:2206.10622, 2022.

DAGA, A. Vibration monitoring: gearbox identification and faults detection. **Core.ac.uk**, [S.l.], 2019.

DEFARD, T. et al. PaDiM: a patch distribution modeling framework for anomaly detection and localization. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR). **Anais...** [S.l.: s.n.], 2021.

DENG, J. et al. ImageNet: a large-scale hierarchical image database. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2009. **Anais...** [S.l.: s.n.], 2009. p.248–255.

DUTTA, T.; SONI, A.; GUPTA, H. Real testbed for autonomous anomaly detection in power grid using low-cost unmanned aerial vehicles and aerial imaging. **IEEE MultiMedia**, [S.l.], v.28, n.3, p.48–57, 2021.

DWYER, B.; NELSON, J.; HANSEN, T. **Roboflow (Version 1.0) [Software]**. 2024.

EVERINGHAM, M. et al. The PASCAL Visual Object Classes (VOC) Challenge. In: 2009 . **Anais...** [S.l.: s.n.], 2010. v.88, p.303–338.

GE, X. et al. Application of unsupervised learning in power delivery system anomaly detection. **Energy Reports**, [S.l.], v.7, p.3482–3489, 2021.

GIRUBAGARI, N.; RAVI, T. Hybrid Intelligent Anomaly Detection System Using Attention-based Deep Learning Approach for Cyber Attacks Prevention. **Indian Journal of Cybersecurity**, [S.l.], 2024.

GOLDSTEIN, M.; UCHIDA, S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. **PLOS ONE**, [S.l.], v.11, n.4, p.e0152173, 2016.

GUAN, H. et al. UAV-lidar aids automatic intelligent powerline inspection. **International Journal of Electrical Power & Energy Systems**, [S.l.], v.129, p.106811, 2021.

GULA, T.; BERTOLDO, J. P. C. Gaussian Image Anomaly Detection with Greedy Eigencomponent Selection. **arXiv preprint arXiv:2308.04944**, [S.l.], 2023.

HAN, C. et al. MADGAN: unsupervised medical anomaly detection gan using multiple adjacent brain mri slice reconstruction. **BMC bioinformatics**, [S.l.], v.22, n.2, p.1–20, 2021.

HE, Y. et al. UAV based sensing and imaging technologies for power system detection, monitoring and inspection: a review. **Journal of Testing and Evaluation**, [S.l.], 2024.

HENDRYCKS, D.; GIMPEL, K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. **International Conference on Learning Representations (ICLR)**, [S.l.], 2018.

HODGE, V. J.; AUSTIN, J. A Survey of Outlier Detection Methodologies. **Artificial Intelligence Review**, [S.l.], v.22, n.2, p.85–126, 2004.

HOEFLER, T. et al. XAI-Guided Insulator Anomaly Detection for Imbalanced Datasets. **arXiv preprint arXiv:2409.16821**, [S.l.], 2024.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-Excitation Networks. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). **Proceedings...** [S.l.: s.n.], 2018. p.7132–7141.

HUANG, Z.; TIAN, Y.; CHEN, L. Deep learning for industrial inspection: a survey. **IEEE Transactions on Industrial Informatics**, [S.l.], v.18, n.5, p.2937–2950, 2022.

JALIL, B. et al. Fault detection in power equipment via an unmanned aerial system using multimodal data. **Sensors**, [S.l.], v.19, n.13, p.3014, 2019.

KIRAN, B. R.; THOMAS, D. M.; PARAKKAL, R. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. **Journal of Imaging**, [S.l.], v.4, n.2, p.36, 2018.

KRIZHEVSKY, A.; HINTON, G. Learning multiple layers of features from tiny images. In: OF THE . **Anais...** [S.l.: s.n.], 2009.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [S.l.], v.86, n.11, p.2278–2324, 1998.

LEE, H. et al. A dataset and evaluation methodology for segmentation of power lines in complex backgrounds. **IEEE Transactions on Power Delivery**, [S.l.], v.32, n.5, p.2290–2299, 2017.

LEE, M.; LEE, H. Magnetically Guided Diagnosis of Current Flow Patterns inside Fault-Simulated Li-Ion Batteries. **Electrochemical Society Meeting Abstracts**, [S.l.], 2024.

LEKIDIS, A.; ANASTASIADIS, A. G.; VOKAS, G. A. Electricity infrastructure inspection using AI and edge platform-based UAVs. **Energy Reports**, [S.l.], v.8, p.5033–5045, 2022.

LI, J. et al. Identification of autism spectrum disorder based on electroencephalography: a systematic review. **Computers in Biology and Medicine**, [S.l.], 2024.

LI, Z. et al. Design and application of a UAV autonomous inspection system for high-voltage power transmission lines. **Remote Sensing**, [S.l.], v.15, n.3, p.865, 2023.

LIU, R. Intelligent Bearing Anomaly Detection for Industrial Internet of Things Based on Auto-Encoder. **IEEE Internet of Things Journal**, [S.l.], 2024.

LIU, Z. et al. SimpleNet: a simple network for image anomaly detection and localization. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). **Proceedings...** [S.l.: s.n.], 2023. p.20402–20411.

MADER, P. et al. MVTec LOCO AD: a benchmark for logical anomaly detection. **arXiv preprint arXiv:2210.09299**, [S.l.], 2022.

MARKOU, M.; SINGH, S. Novelty detection: a review—part 1: statistical approaches. **Signal Processing**, [S.l.], v.83, n.12, p.2481–2497, 2003.

MENDU, B.; MBULI, N. State-of-the-Art Review on the Application of Unmanned Aerial Vehicles (UAVs) in Power Line Inspections: current innovations, trends, and future prospects. **Drones**, [S.l.], v.9, n.4, p.265, 2025.

NEHA SARDANA, H.; KANWADE, R.; TEWARY, S. Arrhythmia detection and classification using ECG and PPG techniques: a review. **Physical and Engineering Sciences in Medicine**, [S.l.], 2021.

PANG, G. et al. Deep Learning for Anomaly Detection: a review. **ACM Computing Surveys**, [S.l.], v.54, n.2, p.1–38, 2021.

PILLAI, A. Using TinyML for the Detection of Irregularities. **ScholarWorks at CalState**, [S.l.], 2024.

PIMENTEL, M. A. F. et al. A Review of Novelty Detection. **Signal Processing**, [S.l.], v.99, p.215–249, 2014.

RISTEA, N.-C. et al. **Self-Supervised Predictive Convolutional Attentive Block for Anomaly Detection**. 2022.

RISTEA, N.-C. et al. Self-Supervised Predictive Convolutional Attentive Block for Anomaly Detection. In: OF THE . **Anais...** [S.l.: s.n.], 2022. p.13566–13576.

RUDOLPH, M. et al. DifferNet: unsupervised anomaly detection with normalizing flows. In: IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV). **Proceedings...** [S.l.: s.n.], 2021. p.1386–1395.

RUDOLPH, M.; WANDT, B.; ROSENHAHN, B. Same Same But DifferNet: semi-supervised defect detection with normalizing flows. **arXiv preprint arXiv:2008.12577**, [S.l.], 2020.

RUDOLPH, M.; WANDT, B.; ROSENHAHN, B. Same same but differnet: semi-supervised defect detection with normalizing flows. **arXiv preprint arXiv:2105.09104**, [S.l.], 2021.

RUFF, L. et al. Deep one-class classification. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 35. **Proceedings...** PMLR, 2018. v.80, p.4393–4402.

SABOOR, A. et al. Latest research trends in gait analysis using wearable sensors and machine learning: a systematic review. **IEEE Transactions on Biomedical Engineering**, [S.l.], 2020.

SALEH, F. et al. Object-centric anomaly detection by attribute-based reasoning. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. **Proceedings...** [S.l.: s.n.], 2013. p.787–794.

SCHLEGL, T. et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. **PLoS one**, [S.l.], v.14, n.3, p.e0212200, 2019.

SELVARAJU, R. R. et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV). **Proceedings...** [S.l.: s.n.], 2017. p.618–626.

SHARMA, N. et al. Recent trends in EEG-based Motor Imagery Signal Analysis and Recognition: a comprehensive review. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, [S.l.], 2023.

SHETE, S. et al. Online-Adaptive Anomaly Detection for Defect Identification in Aircraft Assembly. **arXiv preprint arXiv:2406.12698**, [S.l.], 2024.

SILVA, A. L. B. et al. Attention Modules Improve Modern Image-Level Anomaly Detection: a differnet case study. **arXiv preprint arXiv:2401.08686**, [S.l.], 2023.

SILVA, A. L. B. V. e et al. InsPLAD: a dataset and benchmark for power line asset inspection in uav images. **International Journal of Remote Sensing**, [S.l.], v.44, n.23, p.7294–7320, 2023.

SILVA, A. L. V. e et al. Attention Modules Improve Image-Level Anomaly Detection for Industrial Inspection: a differnet case study. In: IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV). **Proceedings...** [S.l.: s.n.], 2024. p.8246–8255.

SILVA, A. L. V. e et al. Attention Modules Improve Efficient Anomaly Localization for Industrial Inspection in the Wild. , [S.l.], 2024.

SILVA, A. L. V. e et al. Attention Modules Improve Image-Level Anomaly Detection for Industrial Inspection: a differnet case study. In: IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV). **Proceedings...** [S.l.: s.n.], 2024. p.8246–8255.

SINGH, A. et al. A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. **Sensors**, [S.l.], 2021.

TAO, J. et al. Automatic recognition of power line insulators: an image dataset and a benchmark. **IEEE Transactions on Power Delivery**, [S.l.], v.35, n.4, p.1864–1872, 2020.

TIEN, A.; WANG, C.; LEE, S. A novel framework for unsupervised defect detection in industrial environments. **arXiv preprint arXiv:2303.12345**, [S.l.], 2023.

TIEN, T. D. et al. Revisiting Reverse Distillation for Anomaly Detection. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). **Proceedings...** [S.l.: s.n.], 2023. p.24511–24520.

TOMASZEWSKI, D.; RUSZCZAK, B.; MICHALSKI, R. Detection and classification of ceramic power line insulators using convolutional neural networks. **Archives of Electrical Engineering**, [S.l.], v.67, n.3, p.541–554, 2018.

VELESACA, H.; CARRASCO, D.; CARPIO, D. Anomaly Detection in Industrial Production Products Using OPC-UA and Deep Learning. In: RESEARCHGATE CONFERENCE, 2024. **Proceedings...** [S.l.: s.n.], 2024.

VINOTH, K.; SASIKUMAR, P. $\mathrm{VINO}_{EffiFedAV}$: $vinowithefficientfederatedlearningthroughselectiveclientupdatesforreal-timeautonomousvehicleobjectdetection$.**Results in Engineering**, [*S.l.*], 2025.

VITOR, A.; GOEDTEL, A.; CASTOLDI, M. Induction machine fault diagnosis with quadratic time–frequency distributions: state of the art. **IEEE Transactions on Industrial Electronics**, [S.l.], 2023.

WEIMAR, J. et al. MVTec 3D-AD: a 3d anomaly detection dataset and benchmark. **arXiv preprint arXiv:2203.05502**, [S.l.], 2022.

WOO, S. et al. CBAM: convolutional block attention module. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV). **Proceedings...** [S.l.: s.n.], 2018. p.3–19.

YU, J. et al. **FastFlow: unsupervised anomaly detection and localization via 2d normalizing flows**. [S.l.]: arXiv, 2021.

ZAVRTANIK, V.; KRISTAN, M.; SKOčAJ, D. DRAEM—A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection. **Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**, [S.l.], p.8330–8339, 2021.

ZHANG, J. et al. MIAD: a large-scale maintenance inspection anomaly detection dataset. **arXiv preprint arXiv:2301.09647**, [S.l.], 2023.

ZHONG, J. et al. Unmanned aerial vehicle flight data anomaly detection and recovery prediction based on spatio-temporal correlation. **IEEE Transactions on Network and Service Management**, [S.l.], v.18, n.3, p.2983–2994, 2021.

ZHOU, F. et al. A comprehensive survey for deep-learning-based abnormality detection in smart grids with multimodal image data. **Applied Sciences**, [S.l.], v.12, n.11, p.5336, 2022.

ZONG, B. et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. **International Conference on Learning Representations (ICLR)**, [S.l.], 2018.

Appendix