UNIVERSITÀ DEGLI STUDI DI VERONA

LAUREA MAGISTRALE IN *DATA SCIENCE*

TESI DI LAUREA

DIFFUSION-BASED DATA AUGMENTATION FOR INDUSTRIAL ANOMALY DETECTION

Relatore: Prof. Francesco Setti

Laureando: Federico Leonardi VR479719

ANNO ACCADEMICO 2023-2024

Abstract

This thesis presents PatchCoreDual, a novel approach to industrial anomaly detection that combines a dual memory bank architecture with diffusion-based synthetic data augmentation. Industrial anomaly detection faces persistent challenges related to class imbalance, with defective samples typically being far less numerous than normal samples. While traditional approaches focus solely on modeling normal data distributions, this work investigates the benefits of explicitly incorporating positive samples into the detection framework.

PatchCoreDual extends the memory-based PatchCore methodology by maintaining separate memory banks for normal and defective samples. This dual representation enables the model to leverage the complementary information contained in both sample types, resulting in improved detection capabilities. To address the limited availability of defective samples, this research adapts techniques from the DIAG (Diffusion-based In-distribution Anomaly Generation) framework to synthesize realistic defects that maintain the texture and structural properties of the target domain. Experiments conducted on the KSDD2 dataset with the addition of synthetic defect samples generated via diffusion models demonstrate that PatchCoreDual achieves improved performance over the standard PatchCore approach.

The findings suggest that industrial anomaly detection systems can benefit from approaches that thoughtfully incorporate positive samples, challenging the conventional paradigm of exclusively modeling normal distributions. This thesis contributes to the ongoing development of more effective inspection systems for manufacturing environments, where defect detection accuracy directly impacts product quality and operational efficiency.

Table of contents

Abstract *0*

1. Introduction. **3**

2. Theoretical Background **5**

2.1 DIAG (Diffusion-based In-distribution Anomaly Generation): Expert-Guided Latent Diffusion for Realistic Defect Synthesis 5

2.1.1 DIAG as a Response to the Inadequacies of Traditional Data Augmentation 5

2.1.2 Deconstructing the DIAG Pipeline: Expert Guidance and Latent Diffusion Inpainting 7

2.1.3 Advantages of DIAG: A Synthesis of Realism, Expertise, and Efficiency 10

2.2 Anomaly Detection and PatchCore: A Deep Dive into Cold-Start Industrial Inspection 12

2.2.1 Challenges in Industrial Anomaly Detection 12

2.2.2 Deep Learning Approaches to Industrial Anomaly Detection 13

2.2.3 PatchCore: Towards Total Recall in Cold-Start Anomaly Detection 14

3. Dataset **18**

3.1 KSDD2 Dataset 18

3.2 Data Preprocessing 19

4. Methodology **22**

4.1 Memory-Based Anomaly Detection and its Limitations 22

4.2 DIAG-Inspired Inpainting for Synthetic Defect Generation on KSDD2 22

4.2.1 Implementation Overview 23

4.2.2. Prompt Engineering for Electrical Commutator Defects 23

4.2.3 The Generation Pipeline: Process Details 24

4.2.4 Integration with PatchCoreDual 25

4.3 Dual Memory Bank Paradigm: Bridging Normality and Anomaly 25

4.3.1 Dual Memory Bank Architecture: Negative and Positive Feature Representations 26

4.3.2 Feature Extraction and Processing Pipeline: Maintaining Consistency and Enhancing Spatial Context 28

4.3.3 Memory Bank Construction and Optimization: Balancing Representation and Efficiency 31

4.3.4 Dimensionality Reduction 32

4.3.5 Anomaly Detection: 35

5. Experiments and Discussion **42**

5.1 Experimental Setup 42

5.1.1 Dataset 42

5.1.2 Evaluation Metrics 42

5.1.3 Implementation Details 43
5.2 Comparative Evaluation 43
5.2.1 Anomaly Detection and Localization Results 43
5.2.2 Impact of Diffusion-Based Data Augmentation 44
5.2.3 Qualitative Analysis 45
5.3 Discussion 46
5.3.1 Key Findings 46
5.3.2 Limitations and Future Work 47
6. Conclusion 48
References 51

1. Introduction.

In the relentless pursuit of manufacturing excellence, industrial anomaly detection stands as a cornerstone of quality assurance. The ability to automatically and reliably identify defective components within complex production lines is not merely a matter of preventing faulty products from reaching consumers; it is fundamental to optimizing manufacturing processes, minimizing waste, and ensuring operational safety. However, despite its critical importance, robust industrial anomaly detection remains a formidable challenge, particularly in the context of modern, data-driven manufacturing paradigms.

A central hurdle in training effective anomaly detection systems lies in the inherent data imbalance of industrial inspection scenarios. Datasets are typically dominated by images of normal, defect-free products, while examples of anomalies – the very patterns the system is designed to detect – are often scarce, diverse, and unpredictable. This scarcity of anomalous data, often referred to as the "cold-start" problem, renders traditional supervised learning techniques impractical and necessitates innovative approaches that can learn robust anomaly representations from predominantly nominal data.

While unsupervised and one-class classification methods have emerged as dominant paradigms for addressing this challenge, they often struggle to capture the intricate visual complexities of industrial surfaces and textures. These methods, focused solely on modeling the "normal" data distribution, can be prone to false positives, particularly when confronted with the inherent variability of real-world manufacturing environments. A promising alternative strategy lies in data augmentation, specifically the generation of synthetic anomalous samples to enrich the training dataset and guide model learning.

Recent advancements in deep generative modeling, particularly the advent of diffusion models, have opened up exciting new avenues for data augmentation. Unlike traditional techniques that rely on simplistic transformations or noise superposition, diffusion models, and especially Latent Diffusion Models (LDMs), offer the potential to synthesize highly realistic, context-aware anomalies that are statistically indistinguishable from real defects. DIAG (Diffusion-based In-distribution Anomaly Generation), a novel training-free pipeline, further refines this approach by incorporating domain expertise into the diffusion-based generation process. By leveraging textual

prompts and spatial guidance from domain experts, DIAG promises to generate synthetic anomalies that are not only visually compelling but also highly relevant and plausible within specific industrial contexts.

Building upon the strengths of PatchCore, a state-of-the-art anomaly detection algorithm renowned for its efficiency and high performance in cold-start industrial inspection, this thesis introduces an exploration for a dual memory bank extension to PatchCore. This extension strategically leverages the realistic synthetic anomalies generated by DIAG to create a positive memory bank representing known defect patterns, in addition to the standard PatchCore negative memory bank representing normal variations. By explicitly modeling both normality and anomaly within a unified PatchCore framework, this dual memory bank architecture aims to create a more comprehensive and robust decision boundary, enhancing both anomaly detection accuracy and minimizing missed defects in critical industrial inspection tasks.

This thesis will delve into the theoretical foundations of DIAG and PatchCore, detail the design and implementation of the novel dual memory bank extension, and present a comprehensive experimental evaluation on the challenging Kolektor Surface-Defect Dataset 2.

2. Theoretical Background

2.1 DIAG (Diffusion-based In-distribution Anomaly Generation): Expert-Guided Latent Diffusion for Realistic Defect Synthesis

As a cornerstone of this thesis, I critically examined DIAG (Diffusion-based In-distribution Anomaly Generation), a recently proposed and innovative training-free pipeline by Girella et al. 2024. DIAG represents a significant methodological advancement in the field of data augmentation for industrial surface defect detection, offering a compelling approach to address the limitations of conventional techniques.

2.1.1 DIAG as a Response to the Inadequacies of Traditional Data Augmentation

A critical challenge in applying conventional augmentation techniques to industrial anomaly detection lies in their fundamental design, which often overlooks the unique demands of such specialized scenarios. One primary limitation is the inability of conventional augmentation to address the scarcity and diversity of anomalous samples. In industrial settings, anomalies are rare by nature, and collecting sufficient real-world examples for training is often impractical or costly (Ruff et al., 2021). Techniques like flipping or rotating images fail to generate meaningful representations of these rare events, as they do not account for the physical constraints or contextual dependencies inherent in industrial systems. This leads to models that overfit to normal operational patterns while struggling to generalize to unseen anomalies, undermining their practical utility (Gong et al., 2019). Techniques such as MemSeg (Yang et al., 2023), while offering certain benefits, often fall short in generating synthetic anomalies that possess the nuanced statistical properties of real-world industrial defects. The artifacts produced by these methods can be out-of-distribution, potentially leading anomaly detection models to prioritize the identification of these artificial patterns rather than genuine product anomalies.

DIAG, as proposed by Girella et al. 2024, directly confronts these limitations by leveraging the powerful generative capacity of diffusion models, specifically Latent Diffusion Models (LDMs) exemplified by SDXL (Stable Diffusion XL) (Podell et al., 2023). The central premise of DIAG is the recognition that LDMs, with their ability to learn complex data distributions and generate high-

fidelity samples, offer a pathway to synthesize synthetic defects that are statistically more indistinguishable from real anomalies than previously achievable. Empirical validation presented by Girella et al. 2024, demonstrates the superior in-distribution fidelity of DIAG-generated anomalies. Metrics such as Fréchet Inception Distance (FID) reveal a significantly closer distributional proximity to real defects, while improved anomaly detection performance metrics like Average Precision and AUROC underscore the practical benefits of this enhanced realism. Figure 1 can serve to visually illustrate this crucial distinction in realism.

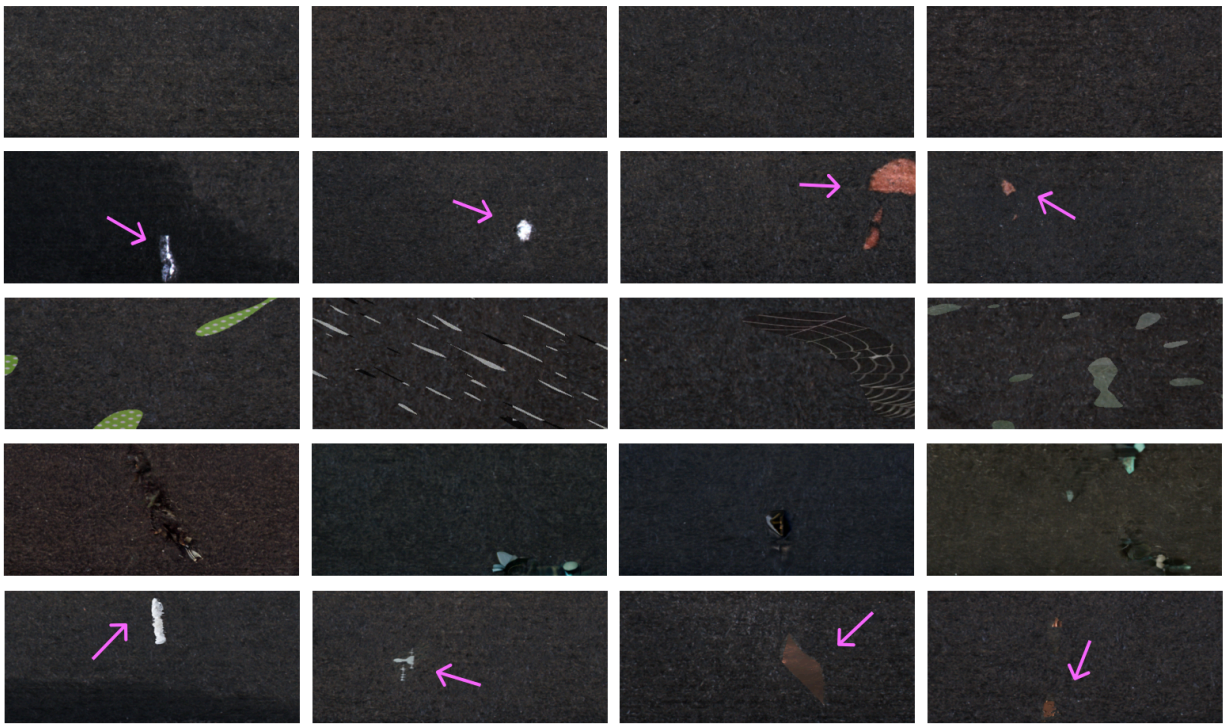

Fig. 1. First row displays some negative samples from the KSDD2 dataset. Instead, the second row shows some images of positive samples from the same dataset. In the third row, MemSeg-generated defect samples. The fourth row shows In&Out generated defect samples. Lastly, the final row showcases images generated with DIAG. Notably, the defect images that DIAG generated are more realistic and in-distribution.

2.1.2 Deconstructing the DIAG Pipeline: Expert Guidance and Latent Diffusion Inpainting

The DIAG pipeline is structured as a three-stage process, ingeniously designed to orchestrate the generative capabilities of LDMs with the critical insights of domain experts. The core stages, analyzed in detail below, are:

Stage 1: Domain Expert-Driven Multimodal Conditioning: A defining characteristic of DIAG, and a key point of departure from purely automated augmentation methods, is its deliberate integration of **human-in-the-loop interaction**. This human element is not merely an add-on; it is a fundamental design principle, enabling the infusion of invaluable domain expertise to guide the anomaly generation process. Expertise is channeled through two complementary modalities, creating a rich and informative conditioning signal for the LDM:

- **Text Prompting and Spatial Conditioning in DIAG for KSDD2 Dataset:**
  DIAG employs a sophisticated approach to leverage Latent Diffusion Models (LDMs) by incorporating multimodal conditioning in the image generation process. This conditioning is particularly effective for generating realistic defects in industrial surface inspection scenarios like those represented in the KSDD2 dataset. When working with the KSDD2 dataset, which contains images of electrical commutators with various surface defects such as scratches and spots, DIAG utilizes two types of prompts:

  1. **Positive prompts**: Specific textual descriptions like "white marks on the wall" and "copper metal scratches" that guide the LDM to generate defects matching these descriptions.
  2. **Negative prompts**: Contrastive descriptions such as "smooth, plain, black, dark, shadow" that steer the generation away from non-defective appearances.

  These prompts are carefully selected through a human-in-the-loop iterative pipeline, refined until the resulting images closely resemble plausible anomalies specific to the electrical commutator surfaces present in KSDD2.

  Beyond textual guidance, DIAG also employs spatial conditioning by utilizing the segmentation masks of positive samples from the KSDD2 dataset. These masks represent

regions where defects are likely to appear based on real anomalies, effectively simulating domain experts' knowledge about plausible anomalous regions.

For the KSDD2 dataset specifically, which comprises 246 positive and 2085 negative images in the training set (and 110 positive, 894 negative images in the test set), this spatial conditioning is crucial for generating contextually appropriate defects on the metallic surfaces of electrical commutators.

The generation process combines normal images, textual prompts, and defect masks into a triplet that guides a pre-trained LDM (specifically SDXL) to perform inpainting on the normal images. The model fills the masked regions with defects that align with the textual descriptions.

When applied to the KSDD2 dataset, this approach significantly outperforms traditional augmentation methods, achieving an improvement in AP of approximately 18% when positive samples are available and 28% when they are missing. This demonstrates the effectiveness of expert-guided, multimodal conditioning for generating in-distribution defect images that substantially improve anomaly detection on electrical commutator surfaces.

- **Spatial Localization via Anomaly Masks:**

  To further enhance the realism and plausibility of generated defects, DIAG incorporates **anomaly masks** as a mechanism for spatial conditioning. Rather than relying on arbitrary or procedurally generated Regions of Interest (ROIs), DIAG strategically leverages **segmentation masks derived from real anomalous samples**. These masks, often representing expert annotations of actual defects within datasets like KSDD2, provide empirically grounded spatial information to guide the LDM inpainting process. By utilizing these masks to define the inpainting region, DIAG ensures that the generated defects are not randomly scattered artifacts but are spatially localized to areas where defects are realistically expected to manifest within industrial products. This expert-guided spatial conditioning, working in tandem with the textual prompts, allows DIAG to achieve an unprecedented level of control and realism in defect synthesis, effectively emulating the nuanced approach of a human expert in generating synthetic anomalies.

Stage 2: Training-Free Latent Diffusion Inpainting for Efficient and High-Fidelity Synthesis: The core of DIAG's anomaly generation process lies in its utilization of a pre-trained Latent Diffusion Model (LDM), specifically SDXL, to perform **training-free inpainting**. This training-free operation is not merely a matter of convenience; it is a deliberate design choice that offers significant practical advantages in industrial settings. The absence of any fine-tuning or task-specific training eliminates the computational overhead and data requirements typically associated with generative models. The inpainting process, executed within the compressed latent space of SDXL, is characterized by the following key steps:

- **Multimodal Input Conditioning via Triplet Construction:** DIAG encapsulates the expert-provided guidance into a **multimodal input triplet** for each anomaly generation instance. This triplet comprises: a normal image (*In*) serving as the base canvas, a textual anomaly description (*Da*) providing semantic guidance, and a spatial anomaly mask (*Ma*) defining the region of interest for defect inpainting. This carefully constructed triplet serves as the conditioning signal for the SDXL inpainting pipeline, instructing the LDM to generate content that is not only visually realistic but also semantically and spatially coherent with the expert-defined parameters. The SDXL model, pre-trained on vast image datasets and fine-tuned for inpainting tasks, performs the core defect synthesis operation within its compressed latent space. Operating in the latent space, as opposed to the pixel space, offers substantial computational efficiencies, reducing memory footprint and accelerating the generation process. Furthermore, latent space operations often lead to improved image quality and visual coherence in generative models. Conditioned on the input triplet, SDXL leverages its learned generative priors to seamlessly inpaint the masked region of the latent representation of *In*. This process effectively blends the generated defect with the existing normal image content, ensuring a visually harmonious and contextually plausible synthetic anomaly. The negative prompt, incorporated into the SDXL pipeline, further refines the generation by actively suppressing unwanted image characteristics and artifacts, ensuring the synthesized defect aligns with the desired visual properties of real industrial anomalies.
- **Stochastic Diversity through Iterative Sampling:** To enhance the diversity and robustness of the augmented dataset, DIAG harnesses the inherent stochasticity of the LDM sampling process. By iteratively sampling from the SDXL model multiple times, even with

the same input triplet (I_n, D_a, M_a), DIAG generates an ensemble of diverse synthetic anomalous images (I_a). This stochasticity is not a mere byproduct of the generative process; it is a strategically exploited feature that expands the coverage of the anomaly feature space and reduces the risk of overfitting to a limited set of synthetic defect variations.

Stage 3: Supervised Anomaly Detection Training with DIAG-Augmented Data: The final stage of the DIAG pipeline addresses the practical application of the generated synthetic anomalies: their integration into the training process of a supervised anomaly detection model. The synthetic anomalous images (I_a) are strategically utilized as *positive samples* to augment the original training dataset. The original dataset, predominantly consists of *negative samples* (normal images), reflecting the inherent class imbalance in industrial anomaly detection. DIAG-generated positive samples effectively counterbalance this imbalance, providing the supervised learning algorithm with a more balanced and informative training dataset. In the context of this thesis, as detailed in Section 4.2, this DIAG-augmented data is employed to train both single and dual memory bank PatchCore models. This supervised training regime, facilitated by DIAG's ability to generate realistic and in-distribution anomalies, enables the anomaly detection models to learn more robust and discriminative feature representations, ultimately leading to enhanced performance in both anomaly detection and localization tasks.

2.1.3 Advantages of DIAG: A Synthesis of Realism, Expertise, and Efficiency

The DIAG pipeline, as a meticulously engineered human-in-the-loop, training-free, and in-distribution anomaly generation framework, offers a compelling and multifaceted set of advantages for industrial anomaly detection:

- **Realism and In-Distribution Anomaly Synthesis:** DIAG demonstrably excels in generating synthetic anomalous images that exhibit a level of visual realism and statistical fidelity previously unattainable with conventional augmentation techniques. Quantitative evaluations using metrics like FID, coupled with qualitative visual assessments, consistently demonstrate that DIAG-generated defects are perceptually and statistically closer to real industrial anomalies. This enhanced realism is not merely an aesthetic improvement; it is a critical factor in the effectiveness of data augmentation, ensuring that anomaly detection models learn to identify genuine defects rather than spurious artifacts

introduced by the augmentation process itself. The superior in-distribution fidelity of DIAG-generated data directly translates to improved generalization and robustness in real-world industrial inspection scenarios.

- **Strategic and Effective Incorporation of Domain Expertise:** A key differentiating feature of DIAG is its deliberate and effective integration of domain expertise into the anomaly generation pipeline. The human-in-the-loop approach, manifested through expert-engineered textual prompts and spatially guided masks, ensures that the synthesized defects are not arbitrary or generic perturbations. Instead, DIAG leverages the nuanced understanding of industrial experts to generate anomalies that are contextually relevant, plausible, and representative of the types of defects encountered in real-world manufacturing settings. This strategic infusion of domain expertise is paramount for generating truly *informative* synthetic data that meaningfully enhances the training of anomaly detection models, guiding them to learn features that are discriminative for real industrial flaws.
- **Training-Free Operation: A Paradigm of Computational Efficiency and Practicality:** DIAG's training-free nature, relying on pre-trained LDMs without requiring any task-specific fine-tuning, offers significant practical advantages, particularly in resource-constrained industrial environments. The pipeline allows for the efficient generation of substantial volumes of high-quality synthetic anomalous samples without incurring the computational burden and data requirements associated with training generative models from scratch or adapting pre-trained models to the target domain. This computational efficiency makes DIAG a highly practical data augmentation solution for industrial anomaly detection applications, enabling rapid prototyping and deployment without extensive computational infrastructure.
- **Empirically Validated Performance Gains in Anomaly Detection:** Rigorous empirical evaluations within the original DIAG paper (Girella et al., 2024) consistently demonstrate the tangible benefits of DIAG-augmented data for anomaly detection performance. Across a range of evaluation metrics, including Average Precision, Image-Level AUROC, and Pixel-Level AUROC, anomaly detection models trained on DIAG-augmented datasets consistently outperform those trained with conventional augmentation techniques or without augmentation altogether. These performance gains are particularly pronounced in

challenging data-scarce scenarios, such as zero-shot anomaly detection, highlighting DIAG's effectiveness in addressing the fundamental data imbalance problem in industrial visual inspection.

In conclusion, DIAG, as a carefully engineered human-in-the-loop, training-free, and in-distribution anomaly generation pipeline, represents a significant methodological leap forward in data augmentation for industrial anomaly detection. By synergistically combining the generative prowess of LDMs with the critical insights of domain experts, DIAG provides a powerful and practical tool for addressing the long-standing challenges of data scarcity, realism, and robustness in training effective anomaly detection systems for real-world industrial applications. This thesis, building upon the strong foundation of DIAG, will now explore its integration with the state-of-the-art PatchCore anomaly detection framework and further investigate its potential through the development of a novel dual memory bank architecture, aiming to push the boundaries of industrial anomaly detection performance and practicality.

2.2 Anomaly Detection and PatchCore: A Deep Dive into Cold-Start Industrial Inspection

Industrial anomaly detection is a critical field within machine vision, particularly vital for maintaining quality and efficiency in modern manufacturing. The goal is to automatically identify products or components that deviate from a defined "normal" or "nominal" state, indicating potential defects, malfunctions, or deviations from desired specifications. These anomalies can range from subtle surface imperfections to critical structural failures, and their timely detection is crucial for preventing faulty products from reaching consumers, optimizing production processes, and minimizing waste.

2.2.1 Challenges in Industrial Anomaly Detection

While anomaly detection is a broad field, industrial visual inspection presents unique challenges that necessitate specialized approaches. In many industrial scenarios, acquiring images of non-

defective, "normal" products is relatively straightforward. However, obtaining a comprehensive and representative dataset of all possible defect types is often impractical, expensive, or even impossible. This "cold-start" or one-class learning scenario, where models must learn normality from normal data alone and then detect deviations, is a central challenge.

Defective parts are, by definition, rare in a well-functioning manufacturing process. This inherent data imbalance – vastly more normal samples than anomalous ones – makes traditional supervised learning approaches less effective. Models trained on imbalanced datasets can become biased towards the majority class (normal) and struggle to accurately detect anomalies. Industrial defects can manifest in highly diverse and unpredictable ways. Defects can range from minute scratches or stains to significant structural deformities or missing components. This high variability makes it difficult to create a universal model of "defectiveness." Furthermore, the visual characteristics of defects are often subtle and can be easily confused with acceptable variations in texture, lighting, or manufacturing processes. Industrial inspection systems often operate in high-throughput environments with stringent real-time constraints. Anomaly detection models must not only be accurate but also computationally efficient to keep pace with production lines. Inference speed and computational cost are therefore crucial considerations. Models trained on one type of product or manufacturing process may not generalize well to new products or processes due to domain shift. Ideally, anomaly detection systems should be adaptable and robust to variations in product appearance and manufacturing conditions without requiring extensive retraining.

2.2.2 Deep Learning Approaches to Industrial Anomaly Detection

In recent years, deep learning has emerged as a powerful tool for addressing the challenges of industrial anomaly detection. Various deep learning-based approaches have been explored, broadly falling into the following categories:

- **Autoencoding Methods:** Autoencoders are trained to reconstruct normal data. Anomalies are detected based on high reconstruction error, as they are assumed to lie outside the learned manifold of normal data. While effective, these methods can sometimes struggle with complex textures and may not generalize well to subtle anomalies.
- **Generative Adversarial Networks (GANs):** GANs, particularly those trained in an unsupervised manner on normal data, can be used for anomaly detection. Anomalies are

identified by their low likelihood under the learned normal data distribution or through discrepancies in the discriminator's output. GANs can be computationally expensive to train and may suffer from mode collapse.

- **Normalizing Flows:** Normalizing flows learn a bijective mapping between the data distribution and a simple latent distribution (e.g., Gaussian). Anomalies are identified as samples with low probability density under the learned normal distribution. Normalizing flows can provide probabilistic anomaly scores but can be complex to train and evaluate.
- **Feature Embedding and Distance-Based Methods:** These methods leverage pre-trained deep neural networks to extract meaningful feature representations from images. Anomaly detection is then performed by comparing the features of test samples to a learned representation of normality, often using distance-based metrics like k-Nearest Neighbors (k-NN) or Mahalanobis distance. These methods are computationally efficient and can leverage the generalization capabilities of pre-trained networks.

2.2.3 PatchCore: Towards Total Recall in Cold-Start Anomaly Detection

PatchCore, introduced by Roth et al. in their work "Towards Total Recall in Industrial Anomaly Detection," represents a state-of-the-art approach specifically designed to address the challenges of cold-start industrial anomaly detection. Unlike previous methods that required adaptation to target domains, PatchCore leverages pre-trained features effectively without additional fine-tuning, making it particularly suitable for real-world industrial inspection scenarios where anomalous samples are scarce or unavailable during training.

The development of PatchCore was driven by several important considerations for industrial anomaly detection. First, PatchCore aims to maximize nominal information at test time by creating a comprehensive "memory bank" of normal patch features that serve as reference points during inference, allowing the algorithm to effectively capture the distribution of normal patterns. Second, while pre-trained ImageNet models provide useful feature representations, very deep features can be overly biased toward natural image classification. PatchCore mitigates this by utilizing mid-level features from intermediate network layers (typically layers 2 and 3 of ResNet architectures), striking an optimal balance between generalizability and task-relevance. Third, industrial applications demand efficient processing, so PatchCore employs innovative coreset subsampling techniques to dramatically reduce the memory bank size while preserving performance, making

real-time inspection feasible. Finally, the patch-based approach with neighborhood-aware scoring enables both accurate image-level anomaly detection and precise pixel-level defect localization.

PatchCore's architecture consists of several key components that work together to enable effective anomaly detection. At its core, PatchCore utilizes ResNet-like architectures (ResNet50, WideResNet50) pre-trained on ImageNet to extract meaningful visual features without requiring task-specific training. This aligns with findings from Cohen and Hoshen and Defard et al., who demonstrated that pre-trained features can effectively capture visual anomalies.

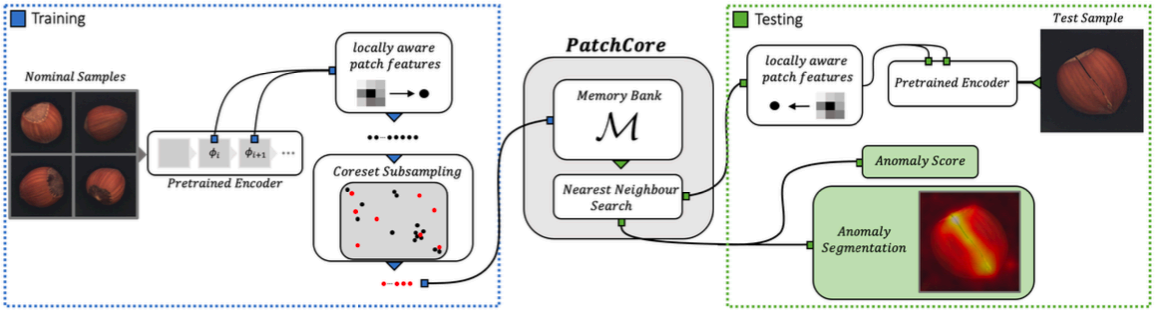

Fig. 2: A schematic overview of the PatchCore architecture, showing the flow from input images through feature extraction to anomaly detection and segmentation. Roth et al. 2022

Feature maps are extracted from intermediate layers of the backbone network, typically from layer2 and layer3 in ResNet architectures. These mid-level features provide an ideal balance between semantic information and spatial resolution. The choice of feature hierarchy is crucial, as demonstrated in the ablation studies by Roth et al., where they showed that mid-level features outperform both low-level and high-level features for industrial anomaly detection.

To enhance robustness to spatial variations and increase the receptive field without sacrificing resolution, PatchCore applies Local Neighborhood Aggregation (LNA). This process involves applying adaptive average pooling over a small neighborhood (typically 3×3) for each spatial location in the feature maps. This aggregation preserves spatial information while making patch representations more robust to minor variations in texture and alignment.

The memory bank is constructed by collecting patch features from all normal training samples. Each spatial location in the feature maps contributes a feature vector, resulting in a large collection of patch features. To manage computational complexity, PatchCore employs greedy coreset subsampling, which selects a representative subset of features while maintaining coverage of the feature space. The coreset selection aims to find a minimal set of patch features that best

approximate the full distribution of normal patches, operating on the principle of maximizing the minimum distance between selected samples.

Remarkably, as shown in the original paper, a coreset-reduced memory bank with only 1% of the original features can still achieve comparable or even better performance than the full memory bank, while dramatically reducing storage requirements and inference time.

During inference, PatchCore extracts patch features from the test image and computes the distance of each test patch to its nearest neighbor in the memory bank. The distance serves as a patch-level anomaly score. Intuitively, if a test patch is similar to at least one patch in the memory bank (small distance), it is likely normal; if it differs significantly from all patches in the memory bank (large distance), it is likely anomalous.

PatchCore further enhances detection with a neighborhood-aware weighting mechanism that considers the density of the feature space around the nearest neighbor. This helps identify anomalies in sparsely populated regions of the normal feature space. For pixel-level segmentation, patch-level scores are upsampled to the original image resolution using bilinear interpolation and smoothed with a Gaussian filter, providing precise defect localization.

PatchCore offers several significant advantages for industrial anomaly detection. On the MVTec AD benchmark, PatchCore achieves an image-level anomaly detection AUROC of 99.6% and a pixel-level AUROC of 98.1%, significantly outperforming previous methods. Coreset subsampling allows PatchCore to maintain high performance while reducing memory requirements and computational cost. Even with 1% memory bank retention, inference times are suitable for real-time applications.

PatchCore demonstrates strong performance even with limited training samples, achieving competitive results with as few as 5-10 normal training images per category. As a cold-start method, PatchCore operates effectively without anomalous training examples, addressing the common industrial scenario where defective samples are scarce. Finally, the patch-based approach provides naturally interpretable results by highlighting the specific regions contributing to anomaly detection.

Despite its strengths, PatchCore has some limitations that present opportunities for improvement. While beneficial for cold-start learning, PatchCore's performance depends on the relevance of ImageNet-pretrained features to the industrial domain. For domains with significant distribution shift from natural images, performance might be suboptimal. Parameters like coreset subsampling

ratio and neighborhood size require tuning for optimal performance across different datasets. Additionally, very high intra-class variation within normal samples can challenge PatchCore's ability to distinguish between normal variation and true anomalies.

This thesis builds upon PatchCore by integrating diffusion-based data augmentation techniques inspired by the DIAG framework, which enables the generation of realistic in-distribution anomalies for training. This combination aims to enhance PatchCore's capabilities by providing synthetic anomaly examples while preserving its efficient memory bank architecture.

3. Dataset

3.1 KSDD2 Dataset

The Kolektor Surface-Defect Dataset 2 (KSDD2) is a specialized benchmark dataset tailored for industrial anomaly detection, specifically aimed at assessing the performance of surface defect detection algorithms within manufacturing contexts. This section offers a detailed examination of the dataset's composition, structure, and relevance to advancing anomaly detection research.

The KSDD2 dataset was developed to meet the pressing demand for realistic and standardized benchmarks in industrial surface defect detection. In manufacturing, ensuring product quality hinges on the ability to identify surface anomalies—such as scratches, spots, or other imperfections—that may compromise functionality or aesthetics. By providing a curated collection of industrial surface images, KSDD2 enables researchers to design, test, and refine algorithms capable of detecting these defects with high accuracy, thereby contributing to improved quality control processes.

The KSDD2 dataset comprises 3,335 high-resolution images of industrial surfaces, capturing a diverse range of defects and textures. Its complexity arises from the subtle appearance of many anomalies and the intricate surface patterns, which pose significant challenges for detection systems. The dataset is divided into training and test sets, with the following breakdown:

- **Training Set**: 2,331 images, including 246 positive (defective) and 2,085 negative (defect-free) samples.
- **Test Set**: 1,004 images, consisting of 110 positive and 894 negative samples.
- **Defect Types**: Includes scratches, spots, and miscellaneous surface irregularities.

A notable characteristic of KSDD2 is the variability in image dimensions, reflecting real-world industrial scenarios where inspection systems must accommodate objects of differing sizes. This heterogeneity complicates batch processing in deep learning frameworks, necessitating robust preprocessing strategies, as discussed in Section 3.2.

The KSDD2 dataset is structured to emulate the class imbalance typical of industrial settings, where defective items are significantly outnumbered by defect-free ones. The training set (2,331 images) and test set (1,004 images) maintain this imbalance, with defective samples constituting approximately 10.5% and 11% of their respective splits. This distribution provides a realistic

testbed for developing algorithms resilient to skewed data, a common challenge in anomaly detection tasks.

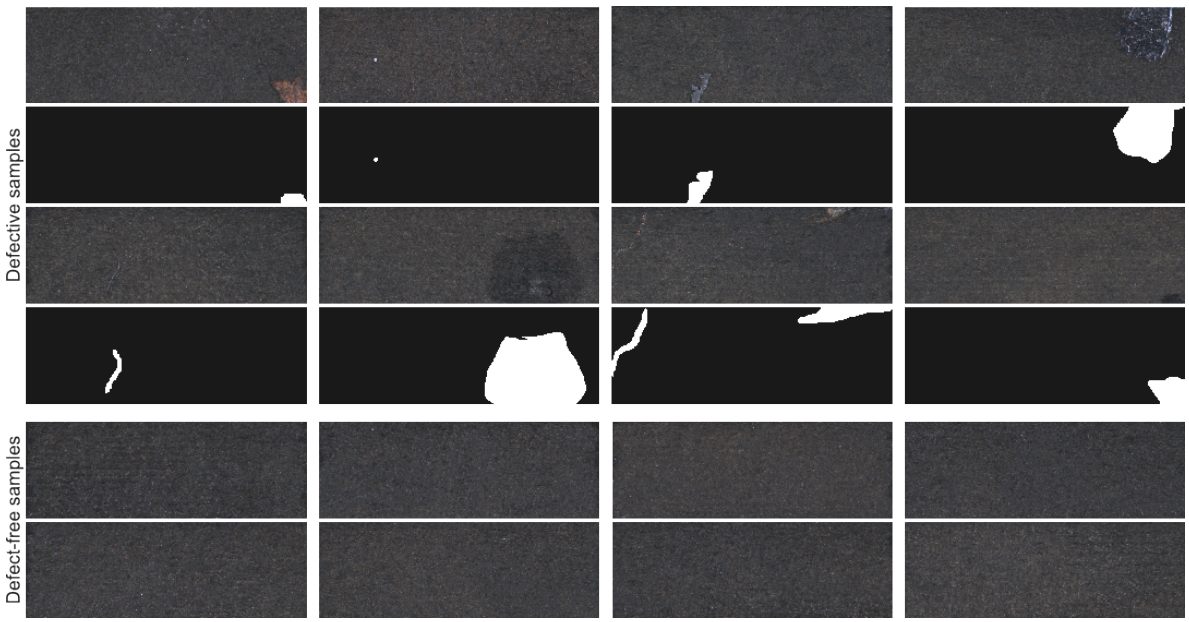

Fig.3: Visualization of Kolektor Surface-Defect Dataset 2

A distinguishing feature of KSDD2 is its provision of pixel-precise ground truth annotations for all defective samples. These annotations take the form of binary masks, where defect regions are explicitly marked, enabling both anomaly detection (binary classification of images as normal or defective) and anomaly segmentation (pixel-level localization of defects). This dual-purpose annotation scheme enhances the dataset's utility for evaluating segmentation models, which are critical for pinpointing defect locations in industrial applications.

3.2 Data Preprocessing

To effectively utilize the KSDD2 dataset for anomaly detection, a comprehensive preprocessing pipeline was designed and implemented. This pipeline plays a critical role in standardizing the data, ensuring compatibility with deep learning models, and optimizing the dataset for both training and evaluation purposes. By addressing challenges related to data inconsistency, variability in image sizes, and the need for accurate defect localization, the preprocessing steps aim to maximize the

dataset's effectiveness in supporting robust anomaly detection models. The following subsections detail the core objectives and methodologies employed in this process.

The preprocessing pipeline was structured around four primary objectives:

- Resolution Standardization: Ensuring uniformity in image dimensions to facilitate efficient batch processing and model training.
- Mask Processing: Preparing and refining ground truth masks to integrate seamlessly into both training and evaluation workflows.
- Directory Organization: Structuring the dataset in a systematic manner to improve accessibility and streamline model development.
- Metadata Creation: Generating structured metadata files to document dataset composition and facilitate efficient data handling.

These objectives were formulated to address key challenges posed by the KSDD2 dataset, such as its diverse image resolutions and the necessity of precise defect annotations. By implementing a structured preprocessing pipeline, the dataset was transformed into a format that ensures consistency, enhances the model's ability to learn meaningful representations, and supports reproducibility in subsequent experiments.

One of the fundamental preprocessing steps involved standardizing image resolutions across the dataset. Given that deep learning models typically require fixed input dimensions to enable batch processing and optimize computational efficiency, all images were resized to 224×632 pixels. This resolution was selected to maintain the aspect ratio of the original images while aligning with commonly used input sizes for pretrained convolutional neural networks (CNNs). By enforcing a uniform image size, this preprocessing step not only facilitates smooth model training but also ensures feature consistency across the dataset, thereby improving the reliability of anomaly detection.

A crucial aspect of the preprocessing pipeline was the handling of **ground truth masks**, which are essential for training and evaluating segmentation-based anomaly detection models. The following steps were undertaken to ensure the correct alignment and usability of these masks:

- Mask Resizing: For all defective samples, masks were resized to match the standardized image dimensions (224×632 pixels). Nearest-neighbor interpolation was employed to preserve the binary nature of these masks, ensuring that defect boundaries remained sharp and clearly defined.

- Mask Alignment: Each image in the dataset was paired with a corresponding mask. For defect-free (normal) samples, an all-zero (blank) mask was assigned to indicate the absence of defects. This alignment ensures that the model receives properly structured input-output pairs during training.
- Defect Localization through Cropping: To enhance the model's ability to focus on defective regions, an additional cropping step was introduced. For positive samples containing defects, images were cropped around the defect areas to localize anomalies more effectively. This step, which directly supports the methodology outlined in Section 4.3, enables the model to learn fine-grained defect features by eliminating unnecessary background information. By isolating defect regions more precisely, the model can better distinguish between normal and anomalous patterns, leading to improved detection performance.

To further enhance the usability of the dataset and facilitate seamless integration into machine learning workflows, the dataset was reorganized into a structured directory layout, separating training and testing data:

- Train Split: Contains both normal and defective samples used for model training.
- Test Split: Comprises a distinct set of normal and defective samples reserved exclusively for model evaluation.

This structured organization ensures a clear distinction between training and test sets, reducing the risk of data leakage and maintaining proper alignment between images and their corresponding ground truth masks. The pipeline simplifies data handling, enhances reproducibility, and enables systematic experimentation in subsequent model development stages.

Finally, to further improve accessibility and usability, the preprocessing pipeline automatically generates metadata files in the form of CSV documents. These files contain structured information on the dataset composition, including relative paths to each image and its corresponding label (i.e., "positive" for defective samples and "negative" for defect-free samples) serving as a concise summary of dataset statistics and offering a practical reference for data loading, exploration, and model training.

4. Methodology

4.1 Memory-Based Anomaly Detection and its Limitations

PatchCore, as detailed in Section 2.2.3, represents a state-of-the-art paradigm for industrial anomaly detection. Its reliance on a memory bank of normal patch features offers compelling performance in cold-start scenarios, where only non-defective samples are available for training. However, the fundamental limitation of PatchCore, and indeed of most traditional anomaly detection methods, lies in their inherent focus on modeling only the distribution of normality. These approaches, while effective at identifying deviations from the "normal manifold," inherently discard valuable information when defect samples are, in fact, available, or can be synthetically generated. The core constraint is their reliance on measuring distance from normality, without explicitly considering proximity to, or characteristics of, known defect patterns.

The standard PatchCore algorithm, while exemplifying the strengths of memory-based anomaly detection, embodies this limitation. By constructing a memory bank solely from normal sample patch features, and subsequently identifying anomalies based on their distance to this bank, PatchCore operates under a "one-class" learning paradigm. While empirically successful across various industrial datasets, it was hypothesized that its performance ceiling is ultimately constrained by this exclusive focus on negative information – the characteristics of normal data – neglecting the potentially valuable information encoded in defected samples.

4.2 DIAG-Inspired Inpainting for Synthetic Defect Generation on KSDD2

Having established the theoretical foundations of the DIAG (Diffusion-based In-distribution Anomaly Generation) approach in previous sections, this section details the practical implementation of a diffusion-based data augmentation pipeline specifically tailored for the KSDD2 dataset. The implementation leverages Stable Diffusion XL (SDXL) inpainting capabilities to generate realistic synthetic defects, addressing the data imbalance challenge inherent in industrial anomaly detection tasks.

4.2.1 Implementation Overview

The augmentation pipeline utilizes the Diffusers library from Hugging Face, specifically the StableDiffusionXLInpaintPipeline, which offers robust inpainting capabilities built upon the SDXL architecture. The pipeline takes as input a source directory containing the KSDD2 dataset, a parameter for the number of images to generate per prompt, and a seed value for reproducibility. The output is a set of synthetic defect images accompanied by their corresponding ground truth masks, organized in a structured directory format.

The overall process follows four main steps: (1) selecting normal images from the training set, (2) selecting anomaly masks from the test set, (3) applying SDXL inpainting with carefully crafted prompts, and (4) post-processing and storing the generated images for later use in training PatchCoreDual.

4.2.2. Prompt Engineering for Electrical Commutator Defects

After analyzing the characteristics of real defects in KSDD2 and following the findings of DIAG (Girella et al. 2024) two highly effective prompts were identified:

- **"copper metal scratches"** targets the most common defect type in this dataset - linear abrasions on the copper surface that can affect electrical conductivity
- **"white marks on the wall"** helps generate lighter-colored anomalies like processing residue or metallic dust deposits

A negative prompt of "smooth, plain, black, dark, shadow" was designed to steer the generation away from creating defect-free surfaces or introducing inappropriate shadowing artifacts.

These prompts were selected based on careful examination of the physical properties of electrical commutators and the types of defects commonly observed in the KSDD2 dataset. Unlike more complex datasets that might require category-specific prompts, KSDD2's focused nature made possible to achieve excellent results with these two carefully crafted prompts.

4.2.3 The Generation Pipeline: Process Details

The process begins by loading the KSDD2 dataset and extracting paths to normal images and defect masks from the dataset manifest. specifically designed the mask selection to borrow spatial patterns from real defects, ensuring realistic defect placement. Normal images are selected from the 2,085 defect-free samples in the KSDD2 training set, while anomaly masks are derived from the ground truth segmentation maps of defective samples. By using masks from real defects, it was endured that synthetic anomalies have realistic shapes and locations.

The pipeline is configured with hyperparameters optimized specifically for the KSDD2 dataset:

- Number of inference steps: 30
- Guidance scale: 20.0
- Strength: 0.99
- Padding mask crop: 2

These parameters were carefully tuned to balance generation quality with computational efficiency. The high guidance scale ensures the generated defects strongly adhere to the prompt descriptions, while the strength parameter allows the model to significantly modify the masked regions.

For each prompt, multiple synthetic defect samples were generated through the following steps:

- **Random Selection**: A normal image and an anomaly mask are randomly selected from their respective collections.
- **Preprocessing**: Both the normal image and the anomaly mask are resized to 1024×1024 pixels to accommodate SDXL's requirements.
- **Defect Generation**: The SDXL inpainting model generates a synthetic defect by filling the masked area according to the provided prompt.
- **Postprocessing**: The generated image is resized back to the original KSDD2 dimensions (224×632 pixels).
- **Storage**: The synthetic defect image and its corresponding mask are saved to disk with appropriate naming conventions.

This process is repeated for each prompt and for the specified number of images per prompt, resulting in a substantial augmentation of the defective sample set.

4.2.4 Integration with PatchCoreDual

The generated synthetic defects serve a crucial role in the PatchCoreDual architecture by:

1. Expanding the positive memory bank with diverse, realistic anomaly examples
2. Providing spatially accurate masks for supervised localization training
3. Addressing the class imbalance in the original KSDD2 dataset

The implementation allows for tuning the number of synthetic samples through the number of images per prompt parameter, enabling experimentation with different augmentation ratios. For the experiments, 50 synthetic samples per prompt were generated, resulting in a total of 100 additional defective samples to supplement the original dataset.

The quality of the generated defects was visually assessed to ensure their realism and relevance to the KSDD2 context. The synthetic samples exhibited the characteristic scratches, marks, and surface anomalies typical of real electrical commutator defects, while maintaining contextual consistency with the normal background regions.

Through this DIAG-inspired implementation, the generative capabilities of diffusion models were exploited to create realistic synthetic defects specifically tailored to the KSDD2 dataset, providing a robust foundation for improved anomaly detection performance with the PatchCoreDual architecture.

4.3 Dual Memory Bank Paradigm: Bridging Normality and Anomaly

To overcome the inherent limitations of traditional one-class anomaly detection and enhance the capabilities of PatchCore, **PatchCoreDual** was proposed, an extension that challenges the conventional paradigm by explicitly modeling both normal and defective patterns. The core insight driving PatchCoreDual is the recognition that industrial defects, unlike truly random anomalies, often exhibit consistent, identifiable patterns across samples, particularly within specific product categories. By constructing and leveraging *both* normal and defective feature representations, the goal is to establish a more comprehensive and nuanced decision boundary, one that benefits from the full spectrum of available information, both negative and positive.

This dual memory bank paradigm is particularly well-suited for industrial settings characterized by the following conditions:

- **Consistent Defect Types:** Defect types are relatively consistent within a product category and follow identifiable, learnable patterns, rather than being completely random or unpredictable.
- **High Cost of Missed Defects:** The economic or safety implications of missed defects (false negatives) are significant, justifying the increased model complexity and computational overhead of a dual memory bank approach to minimize such errors.
- **Availability of Defect Exemplars:** A limited set of defect samples is available, either through historical data, expert knowledge, or, crucially, through synthetic generation methods like DIAG, enabling the construction of a representative positive memory bank.

PatchCoreDual thus represents a departure from pure anomaly detection, moving towards a hybrid approach that strategically incorporates elements of both anomaly detection and defect classification. This hybrid approach aims to harness the strengths of both paradigms, potentially achieving improved detection accuracy, reduced false positives, and, importantly, enhanced recall – the ability to minimize missed defects, which is paramount in safety-critical industrial applications.

4.3.1 Dual Memory Bank Architecture: Negative and Positive Feature Representations

- **Negative Memory Bank:** This component, directly analogous to the memory bank in the standard PatchCore algorithm, stores patch-level feature representations meticulously extracted from normal (defect-free) samples. Following the established PatchCore methodology, these features encode the statistical distribution of normal appearance patterns for the nominal samples available in the dataset. This negative memory bank serves as the baseline representation of "normality," against which test samples are compared to identify deviations. Crucially, in PatchCoreDual, it was maintained the architecture and construction process of this negative memory bank following the insights of the original PatchCore implementation. This design choice is essential for ensuring a direct and controlled comparison between PatchCore and PatchCoreDual

- **Positive Memory Bank:** This component, representing the core novelty of PatchCoreDual, stores patch-level feature representations extracted from defective samples. These defective samples are a combination of positive samples taken from the KSDD2 Dataset and augmented samples synthetically generated using the DIAG pipeline, making possible to leverage the realism and in-distribution fidelity of diffusion-based anomaly synthesis to enhance the diversity of positive samples available. The positive memory bank serves as a repository of characteristic patterns of known defects, enabling PatchCoreDual to explicitly measure the *similarity* of test samples to these learned defect representations. This capability to assess both dissimilarity from normality (via the negative memory bank) and similarity to anomaly (via the positive memory bank) is the defining characteristic of PatchCoreDual and the key to its enhanced anomaly detection performance.

The deliberate choice to implement separate and parallel memory banks, rather than a unified or combined structure, is a critical architectural decision in PatchCoreDual. This separation serves several key purposes:

- **Preservation of Distinct Distributions:** Maintaining separate memory banks allows for the preservation of the distinct statistical properties of normal and defective feature distributions. Normal samples, by definition, exhibit a more cohesive and tightly clustered distribution, while defective samples, even synthetic ones, may exhibit greater variability and multi-modality, reflecting the diverse nature of potential defects. Separate memory banks allow each distribution to be modeled and represented independently, without imposing artificial constraints or averaging effects that might arise from a unified structure.
- **Nuanced Scoring Methods:** The dual memory bank architecture enables the development of more nuanced and informative anomaly scoring methods. By having access to distances to both normal and defective feature representations, PatchCoreDual can compute scores that explicitly capture both deviation from normality and similarity to anomaly, allowing for a more comprehensive and robust anomaly assessment.
- **Asymmetric Sampling Strategies:** The dual memory bank architecture accommodates the inherent class imbalance between normal and defective samples in industrial anomaly detection. The negative memory bank, representing the abundant normal data, can be constructed with a larger size and potentially different subsampling ratios compared to the

positive memory bank, which represents the scarcer (even if synthetically augmented) defective data. This *asymmetric sampling* allows for a more efficient and targeted allocation of computational resources, ensuring that both normal and defective feature representations are adequately captured without unnecessary redundancy or computational overhead.

4.3.2 Feature Extraction and Processing Pipeline: Maintaining Consistency and Enhancing Spatial Context

To ensure a fair and controlled comparison with the standard PatchCore algorithm, and to maintain computational efficiency, PatchCoreDual adopts a feature extraction and processing pipeline that closely mirrors the original PatchCore methodology.

Consistent with PatchCore, ResNet50 pre-trained on ImageNet was selected as the feature extraction backbone for PatchCoreDual. This choice is motivated by several key considerations:

- **Empirical Effectiveness:** ResNet architectures, and ResNet50 in particular, have demonstrated proven empirical effectiveness for industrial anomaly detection tasks in prior research, including the original PatchCore paper. ResNet50 provides a strong balance between representational capacity and computational efficiency, making it a robust and widely adopted choice for feature extraction in this domain.
- **Fair Comparison with Baseline:** To enable a direct and controlled comparison between PatchCore and PatchCoreDual, it is crucial to maintain consistency in the feature extraction architecture. Using the same ResNet50 backbone as the original PatchCore implementation amaking possible to isolate and quantify the performance gains specifically attributable to the dual memory bank architecture and the novel scoring methods, rather than confounding these gains with variations in feature extraction capabilities.
- **Computational Efficiency:** ResNet50, while a powerful deep learning architecture, offers a good balance between feature quality and computational cost. More complex or deeper architectures might offer incremental performance improvements, but at the expense of increased computational overhead, potentially hindering the practicality of PatchCoreDual for real-world industrial applications with stringent real-time constraints. ResNet50 provides a computationally efficient foundation for building a practical and scalable anomaly detection system.

Following the established PatchCore methodology, feature maps from layers 2 and 3 of the ResNet50 backbone were extracted. This specific layer selection is based on a well-reasoned rationale, aiming to capture multi-scale feature representations that are relevant for industrial anomaly detection:

- **Layer 2 Features:** Features extracted from layer 2 of ResNet50 capture mid-level patterns that are more abstract than low-level textures but retain a higher degree of spatial resolution (28x28 feature map size for a 224x224 input image). These mid-level features are particularly effective at capturing fine-grained details and localized anomalies, such as subtle surface imperfections, scratches, or textural deviations, which are often critical indicators of defects in industrial products. The higher spatial resolution of layer 2 features allows for precise localization of these subtle anomalies within the image.
- **Layer 3 Features:** Features extracted from layer 3 of ResNet50, being deeper in the network architecture, encode more semantic and structural information at a coarser spatial resolution (14x14 feature map size for a 224x224 input image). These higher-level features are better suited for capturing larger-scale structural anomalies, such as missing components, shape deformations, or significant deviations from the expected product geometry. While sacrificing some spatial detail, layer 3 features provide a more global and context-aware representation of the image, complementing the fine-grained spatial information captured by layer 2 features.
- **Multi-Scale Fusion for Comprehensive Anomaly Representation:** The strategic combination of features from both layer 2 and layer 3 allows PatchCoreDual to capture multi-scale representations that are robust to a wide range of anomaly types and sizes. By concatenating these features (as detailed in Section 4.3.3), PatchCoreDual effectively fuses fine-grained spatial details from layer 2 with higher-level structural information from layer 3, creating a more comprehensive and discriminative feature representation for anomaly detection. This multi-scale approach enhances the model's ability to detect both subtle surface defects and larger structural deviations, improving its overall robustness and applicability to diverse industrial inspection scenarios.

The decision to utilize the *same* feature layers (layer 2 and layer 3) for *both* the negative and positive memory banks in PatchCoreDual is a design choice aimed at ensuring feature

compatibility and facilitating direct distance-based comparisons between normal and defective feature representations. While exploring alternative approaches, such as using different network layers for normal and defective feature extraction, might offer potential benefits in certain scenarios, it could also introduce complexities related to feature scale and distribution mismatches, potentially complicating the anomaly scoring process and hindering the interpretability of results.

To enhance the robustness of patch features and incorporate local spatial context, PatchCoreDual, consistent with the original PatchCore implementation, employs a 3x3 Local Neighborhood Aggregation (LNA) technique. This architectural choice is motivated by the following considerations:

- **Increased Receptive Field:** LNA effectively increases the receptive field of each patch feature, allowing it to capture information from a larger spatial neighborhood within the input image. By aggregating information from a 3x3 region around each spatial location, LNA enables each patch feature to become more context-aware, incorporating information about its immediate surroundings. This expanded receptive field is particularly beneficial for anomaly detection, as defects often manifest as deviations in texture, structure, or appearance relative to their local context.
- **Robustness to Spatial Variations:** LNA enhances the robustness of patch features to minor spatial variations, misalignments, and noise within the input images. By averaging features within a local neighborhood, LNA effectively smooths out high-frequency noise and reduces sensitivity to small spatial shifts or distortions. This robustness is crucial for real-world industrial inspection scenarios, where images may be subject to variations in lighting, viewpoint, or minor manufacturing tolerances.
- **Preservation of Spatial Resolution:** Despite increasing the receptive field, LNA is implemented in a manner that preserves the spatial resolution of the feature maps. By applying adaptive average pooling to each 3x3 neighborhood *independently* for each spatial location, LNA maintains the original spatial dimensions of the feature maps, ensuring that fine-grained spatial information is not lost during the aggregation process. This preservation of spatial resolution is particularly important for pixel-level anomaly segmentation tasks, where precise localization of defects is critical. Adaptive average pooling ensures that the aggregated feature representation maintains a consistent

dimensionality, regardless of the input neighborhood size. This transformation effectively converts raw CNN activations into more robust and context-aware patch descriptors, enhancing the discriminative power of PatchCoreDual for anomaly detection.

4.3.3 Memory Bank Construction and Optimization: Balancing Representation and Efficiency

The construction and optimization of the memory banks in PatchCoreDual, both negative and positive, follow a procedure designed to balance representational power with computational efficiency, mirroring the core principles of the original PatchCore methodology while incorporating novel adaptations for the dual-bank architecture.

To leverage the complementary information captured by features extracted from different network depths (layer 2 and layer 3 of ResNet50), PatchCoreDual employs feature concatenation as a fusion strategy. Following the established PatchCore approach, this involves:

- **Bilinear Upsampling for Spatial Alignment:** Feature maps from layer 3, which have a lower spatial resolution compared to layer 2 features, are first upsampled using bilinear interpolation to match the spatial dimensions of layer 2 features. This spatial alignment is crucial for ensuring that features from different layers can be effectively combined in a spatially coherent manner. Bilinear interpolation, a standard image resizing technique, is chosen for its computational efficiency and its ability to preserve image sharpness while upsampling.
- **Channel-wise Concatenation for Multi-Scale Representation:** After spatial alignment, the upsampled layer 3 feature maps and the original layer 2 feature maps are concatenated along the channel dimension. This channel-wise concatenation effectively fuses the multi-scale feature representations, creating a unified feature vector for each patch that incorporates information from both network depths. The resulting concatenated feature maps have an increased channel dimensionality (1536 channels in ResNet50 implementation), capturing a richer and more comprehensive representation of each image patch.

This feature concatenation process, combining bilinear upsampling and channel-wise fusion, ensures that PatchCoreDual leverages the complementary strengths of multi-scale feature representations, creating a more discriminative and robust feature space for anomaly detection.

4.3.4 Dimensionality Reduction

The concatenated feature vectors in PatchCoreDual, having a high dimensionality of 1536 channels, pose significant challenges for memory storage and computational efficiency, particularly when constructing and searching large memory banks. To address this "curse of dimensionality," PatchCoreDual, incorporates random projection as a dimensionality reduction technique. This involves:

- **Target Dimensionality Determination:** Based on the Johnson-Lindenstrauss lemma and empirical evaluations (PatchCore, Roth et al. 2022), a target dimensionality of $d* = 128$ dimensions is chosen for the random projection. This reduced dimensionality represents a significant compression of the original 1536-dimensional feature space, while aiming to preserve a substantial portion of the relevant information for distance-based comparisons. The Johnson-Lindenstrauss lemma mathematically guarantees that random projections can reduce dimensionality while approximately preserving pairwise distances between points in high-dimensional space, with a bounded distortion controlled by a parameter epsilon (typically set to 0.1).
- **Random Projection Matrix Construction and Normalization:** A **random projection matrix** of size (1536 x 128) is constructed, with elements drawn from a standard Gaussian distribution. This matrix serves as the transformation kernel for projecting the high-dimensional feature vectors to the lower-dimensional subspace. To ensure distance preservation properties, the columns of the random projection matrix are normalized to unit length. This normalization step is crucial for ensuring that the projection matrix adheres to the theoretical guarantees of the Johnson-Lindenstrauss lemma and effectively preserves relative distances between data points in the projected subspace.
- **Feature Vector Projection:** The high-dimensional patch feature vectors are then projected to the lower-dimensional subspace by matrix multiplication with the normalized random projection matrix. This projection step is computationally efficient, involving a simple

matrix multiplication operation, and significantly reduces the dimensionality of the feature vectors from 1536 to 128 dimensions. The projected feature vectors, now residing in a lower-dimensional space, become the input for the subsequent coreset subsampling stage.

- **Coreset Subsampling:** Even after dimensionality reduction, the memory banks in PatchCoreDual, particularly the negative memory bank constructed from abundant normal samples, can still become prohibitively large for efficient inference in industrial applications. To address this, PatchCoreDual, mirroring the original PatchCore methodology, employs greedy coreset subsampling to select a small, representative subset of feature vectors for each memory bank.

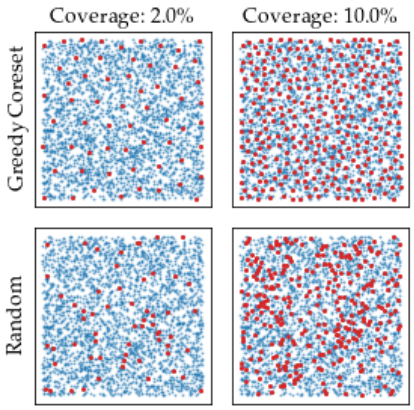

Fig. 4: Visualization of Coreset Subsampling to approximate the spatial support.

This coreset subsampling process is crucial for achieving a balance between memory efficiency, computational speed, and anomaly detection performance. PatchCoreDual utilizes the greedy coreset selection algorithm implemented in NVIDIA's Diversity Sampling library (NVIDIA - *Diversity Sampling*), which provides a computationally efficient and GPU-accelerated implementation of minimax facility location coreset selection. This algorithm iteratively selects feature vectors from the original memory bank to construct a coreset subset that effectively approximates the feature space coverage of the full memory bank. The algorithm begins with an empty coreset subset and iteratively adds feature vectors based on a greedy selection criterion. In each iteration, the algorithm selects the feature vector from the remaining memory bank elements that maximizes the *minimum distance* to the already selected coreset elements. This greedy selection strategy ensures that each newly added coreset element is maximally "dissimilar" to the existing coreset,

effectively expanding the coverage of the feature space and minimizing redundancy within the coreset. This iterative selection process continues until the coreset subset reaches a predefined *target size*, which is typically a small fraction (e.g., 1% to 25%) of the original memory bank size. The mathematical objective function being approximated by the greedy coreset selection algorithm is:

$$M_C^* = \ argmin_{M_C \subset M} \ max_{m \in M} \ min_{n \in M_C} \left|\left|m - n\right|\right|_2$$

This minimax facility location objective aims to minimize the maximum distance from any feature vector in the original memory bank M to its nearest neighbor in the selected coreset subset M_c. By minimizing this maximum distance, the coreset subsampling algorithm effectively selects a subset of feature vectors that provides a representative "cover" of the entire feature space, ensuring that no region of the feature space is left unrepresented in the reduced memory bank.

Recognizing the inherent differences in the characteristics and data volumes of normal and defective samples, PatchCoreDual employs asymmetric subsampling to optimize memory efficiency and information preservation for each memory bank independently. Specifically, in the implementation, a lower subsampling rate for the negative memory bank (e.g., 2%) and a higher subsampling rate for the positive memory bank (e.g., 10%) was selected. The negative memory bank, constructed from abundant normal samples, typically exhibits a higher degree of redundancy in its feature representations. Normal samples, by definition, tend to cluster more tightly in the feature space, with repetitive patterns and less variability compared to defective samples. Therefore, a lower subsampling rate can be applied to the negative memory bank without significant loss of information, as the coreset selection algorithm can effectively capture the essential structure of the normal feature distribution even with a smaller subset of feature vectors. The positive memory bank, constructed from scarcer synthetic anomalous samples, often exhibits a higher information density per sample. Defective samples, even when synthetically generated, tend to be more diverse and less clustered in the feature space, reflecting the inherent variability of potential defect manifestations. Therefore, a higher subsampling rate is applied to the positive memory bank to ensure that a sufficient number of representative defect feature vectors are retained in the

coreset subset, compensating for the smaller initial size of the positive memory bank and preserving the valuable information encoded in these defect samples. The differential subsampling strategy aims to achieve a **balanced contribution** from both memory banks to the final anomaly scoring process. By subsampling the larger negative memory bank more aggressively and the smaller positive memory bank more conservatively, PatchCoreDual prevents the negative memory bank from dominating the distance computations and nearest neighbor searches during inference. This balanced representation ensures that both dissimilarity from normality (captured by the negative memory bank) and similarity to anomaly (captured by the positive memory bank) contribute meaningfully to the final anomaly score, enhancing the robustness and discriminative power of the dual memory bank approach.

By employing greedy coreset subsampling with differential rates, PatchCoreDual effectively addresses the memory efficiency challenges associated with large memory banks while strategically preserving the representative power of both normal and defective feature representations. This optimized memory bank construction process is crucial for enabling practical deployment of PatchCoreDual in resource-constrained industrial environments without compromising anomaly detection performance.

4.3.5 Anomaly Detection:

The anomaly detection process in PatchCoreDual, while building upon the core principles of PatchCore, introduces novel scoring methods designed to effectively leverage the information encoded in *both* the negative and positive memory banks. This section details the anomaly detection mechanism, and the evaluation methodology employed in PatchCoreDual:

1. **Negative Distance and Anomaly Score:** The minimum Euclidean distance from the test patch feature vector to its nearest neighbor in the **negative memory bank**. This distance, analogous to the anomaly score in standard PatchCore, quantifies the dissimilarity of the test patch from normal patterns. Mathematically, for a test patch feature vector *m_test*, the negative distance is computed as:

$$m^{test,*}, m^{*} = argmax_{m^{test} \in P(x^{test,*})} argmin_{m \in M_N} \left|\left| m^{test} - m \right|\right|_2$$

$$s_N^* = \left|\left|m^{test,*} - m^*\right|\right|_2$$

Where:

- $P(x^{test})$ is the set of patch features from the test image x^{test}
- M_N is the negative memory bank containing normal patch features
- $m^{test,*}$ is the test patch with maximum distance to its nearest neighbor
- m^* is the nearest neighbor in the negative memory bank to $m^{test,*}$
- s_N^* is the anomaly score based on the negative memory bank.

2. **Positive Distance and Anomaly Score:** The minimum Euclidean distance from the test patch feature vector to its nearest neighbor in the **positive memory bank**. This distance, a component introduced in PatchCoreDual, quantifies the similarity of the test patch to known defect patterns. Mathematically, the positive distance is computed as:

$$m^{test,\dagger}, m^\dagger = argmax_{m^{test} \in P(x^{test})} argmin_{m \in M_P} \left|\left|m^{test} - m\right|\right|_2$$

$$s_P^* = \left|\left|m^{test,\dagger} - m^\dagger\right|\right|_2$$

Where:

- $P(x^{test})$ is the set of patch features from the test image x^{test}
- M_P is the positive memory bank containing anomalous patch features
- $m^{test,\dagger}$ is the test patch with maximum distance to its nearest neighbor
- $m^\dagger$ is the nearest neighbor in the negative memory bank to $m^{test,\dagger}$
- s_P^* is the anomaly score based on the positive memory bank.

The two ram anomaly scores s_N^* , s_P^* form the foundation for the anomaly scoring methods in PatchCoreDual, capturing complementary information about both deviation from normality and similarity to anomaly. To improve robustness with respect to the maximum patch distance, the scaling factor w on s^* was used to account for the behavior of neighbour patches. If the memory bank features closest to anomaly candidates are themselves far (for negative memory bank) or near

(for positive memory bank) from neighbouring samples, the anomaly score increases. To calculate the neighborhood aware weighting factor:

1. For negative raw anomaly score s_N^*:

$$w_N = 1 - \frac{e^{s_N^*/\sqrt{d}}}{\sum_{m \in N_b(m^*)} e^{||m^{test,*} - m||_2/\sqrt{d}}}$$

   Where:

   - $N_b(m^*)$ represents the b nearest neighbors to m^* in the memory bank M_N
   - s_N^* is the base anomaly score (L2 distance)
   - $\sqrt{d}$ is a normalization factor based on the feature dimension d

   If m^* is located in a sparse region of the normal feature space (where features are dissimilar from each other), then the denominator becomes smaller relative to the numerator, resulting in a higher weight.

2. For positive raw anomaly score s_P^*, the weighting formula is inverted to match the opposite intuition required for defective samples:

$$w_P = \frac{e^{s_P^*/\sqrt{d}}}{\sum_{m \in N_b(m^\dagger)} e^{\left||m^{test,\dagger} - m|\right|_2/\sqrt{d}}}$$

   Where:

   - $N_b(m^*)$ represents the b nearest neighbors to m^* in the memory bank M_N
   - s_N^* is the base anomaly score (L2 distance)
   - $\sqrt{d}$ is a normalization factor based on the feature dimension d

   For the positive bank, a test patch should be considered more anomalous if it's similar to known defects that are clustered together (indicating a confident defect pattern). When $m^\dagger$ is located in a dense region of defect features, the denominator becomes larger, but due to the inverted formula results in a higher weight.

After computing the neighborhood-aware weighting factors, the **final anomaly scores** for each memory bank were calculated as:

1. For the negative bank:

$$s_N = w_N \cdot s_N^*$$

2. For the positive bank:

$$s_P = w_P \cdot s_P^*$$

These weighted scores incorporate both the raw distance measurements and the contextual information about the local neighborhood structure in the feature space.

To effectively combine information from both memory banks, PatchCoreDual introduces a novel approach called Ratio Scoring. This method elegantly fuses the distance measurements from the negative and positive memory banks into a single, interpretable anomaly score.

The Ratio Scoring method computes the anomaly score as the ratio between the negative distance and the positive distance:

$$s_{ratio} = \frac{s_N}{S_P + \epsilon}$$

Where ϵ is a small constant added to the denominator to prevent division by zero.

The intuition behind this scoring approach is particularly well-suited to the dual memory bank architecture. For a truly anomalous patch, it is expected two simultaneous conditions to be met:

1. The patch should be dissimilar from normal patterns (resulting in a high s_N value)
2. The patch should be similar to known defect patterns (resulting in a low S_P value)

When both conditions are met, the ratio becomes large, producing a high anomaly score. This multiplicative interaction between the two distance measures creates a natural amplification effect for patches that satisfy both criteria. This scoring method effectively leverages the complementary information provided by the dual memory bank architecture, enabling more robust and accurate anomaly detection compared to approaches that rely solely on deviation from normality.

By focusing on the relative distances to both normal and anomalous patterns, Ratio Scoring provides a principled approach to combining the "different from normal" and "similar to defect" perspectives, resulting in a more nuanced and effective anomaly assessment mechanism for industrial inspection applications.

After obtaining an image-level anomaly score that determines whether a test sample contains a defect, a crucial next step in industrial inspection is precisely identifying where the defect is located within the image. In PatchCoreDual, anomaly maps are generated by computing and integrating two complementary distance measures for each spatial location in the feature maps. After extracting feature maps from layers 2 and 3 of the backbone network, local neighborhood aggregation is applied with a 3×3 window to incorporate spatial context.

Once feature extraction and patch processing are complete the normal and anomalous distance maps are calculated:

- **Normal Distance Map**: For each spatial location (x,y), it is computed the distance to the nearest neighbor in the normal memory bank:

  $$S_{normal}(x,y) = min_{n \in M_{normal}} \left|\left| f(x,y) - m \right|\right|_2$$

  Where $f(x,y)$ represents the feature vector at position (x,y) and M_{normal} is the memory bank of normal patch features.

  This distance measure captures how different a patch is from normal patterns. Larger values suggest greater deviation from normality and thus higher anomaly likelihood.

- **Anomalous Distance Map**: Simultaneously, it is computed the distance to the nearest neighbor in the anomalous memory bank:

  $$S_{anomalous}(x,y) = min_{n \in M_{anomalous}} \left|\left| f(x,y) - m \right|\right|_2$$

  This distance measure captures how similar a patch is to known defect patterns. Unlike the normal distance map, smaller values indicate greater similarity to defects and thus higher anomaly likelihood.

These two distance maps provide complementary perspectives on the anomaly detection problem. The normal distance map highlights regions that deviate from normal patterns, while the anomalous distance map identifies regions that resemble known defects. By considering both perspectives simultaneously, PatchCoreDual can make more informed and nuanced assessments of anomaly likelihood at each spatial location.

To fuse these complementary perspectives into a single, comprehensive anomaly map, it is implemented, following the methodology for image-level anomaly score, a ratio-based approach that directly divides the normal distance by the anomalous distance:

$$S_{ratio}(x, y) = \frac{S_{normal}(x, y)}{S_{anomalous}(x, y) + \epsilon}$$

Where ϵ is a small constant to prevent division by zero. The ratio approach naturally amplifies the anomaly signal for regions that satisfy both criteria: being distant from normal patterns and close to defect patterns.

The patch-level anomaly scores form a low-resolution grid corresponding to the spatial dimensions of the feature maps (typically 1/8 to 1/16 of the original image resolution). To transform this into a useful high-resolution anomaly map aligned with the original input image, a two-step process is applied:

1. **Bilinear Upsampling**: the low-resolution anomaly score map is upsampled to match the original image dimensions using bilinear interpolation. This technique preserves the relative ranking of anomaly scores while aligning them with the pixel grid of the input image. Importantly, this step is not merely for visualization but is essential for accurate comparison with ground truth masks during evaluation.
2. **Gaussian Smoothing**: To enhance visual clarity and reduce potential artifacts from the upsampling process, a Gaussian smoothing filter is applied to the upsampled anomaly map. While the original PatchCore uses a kernel width of $\sigma = 4$, a smaller $\sigma = 2$ in PatchCoreDual was opted. This more conservative smoothing preserves finer details in the anomaly map, which is particularly important for accurately localizing the thin scratches and subtle surface imperfections common in the KSDD2 dataset.

One technical challenge that was addressed was handling the dimensional differences between test samples and memory bank features. In industrial settings, images often have variable sizes, and the implementation handles this variance by comparing feature vectors of consistent dimensionality (1536 channels for combined layer 2 and 3 features) regardless of the spatial dimensions of the original image.

This enhanced anomaly mapping approach offers several potential advantages for industrial inspection applications:

1. By incorporating information from both normal and anomalous feature distributions, PatchCoreDual can potentially detect and localize subtle defects that might be overlooked by methods considering only deviation from normality. This dual perspective is particularly helpful for certain types of defects, though performance varies across different defect categories.
2. The positive memory bank serves as a reference for known defect patterns, which in many cases helps distinguish between genuine defects and rare but normal variations. While this doesn't eliminate false positives entirely, it can reduce their frequency in scenarios where the defect patterns in the test set resemble those in the training crops.
3. The combined scoring mechanism often highlights regions that exhibit dual characteristics of anomalies, which can lead to better boundary delineation for some defective regions, especially when the defect has distinctive features captured in the positive memory bank.
4. The upsampling and smoothing process helps align pixel-level predictions with the original image, though the accuracy of these alignments depends on the quality of the feature maps and can be affected by the resolution differences between test samples and memory bank features.

5. Experiments and Discussion

This section presents the experimental evaluation of the proposed PatchCoreDual methodology augmented with DIAG-inspired synthetic defect generation. Extensive experiments were conducted on the KSDD2 dataset to validate the effectiveness of this approach in industrial anomaly detection tasks.

5.1 Experimental Setup

5.1.1 Dataset

All experiments were conducted on the Kolektor Surface Defect Dataset 2 (KSDD2), a challenging industrial anomaly detection benchmark. The dataset consists of grayscale images of electrical commutators with the following characteristics:

- 2,331 training images (2,085 normal, 246 defective)
- 1,004 test images (894 normal, 110 defective)
- Image resolution: 224×632 pixels
- Ground truth segmentation masks for defective samples

The dataset presents several challenges typical of industrial inspection scenarios: significant class imbalance, subtle and varied defect appearances, and complex background textures that can complicate detection.

5.1.2 Evaluation Metrics

To thoroughly evaluate the method, the following metrics are employed:

- **Image-level Anomaly Detection:**
    - Area Under the Receiver Operating Characteristic curve (AUROC)
- **Pixel-level Anomaly Localization:**
    - Pixel-wise AUROC

These metrics provide a comprehensive assessment of both detection and localization capabilities, which are crucial in industrial inspection applications.

5.1.3 Implementation Details

All models were implemented using PyTorch. For the backbone feature extractor, a ResNet pretrained on ImageNet was utilized. Specific implementation details include:

- Patch size: 3×3
- Feature hierarchies: Levels 2 and 3 of ResNet
- Coreset subsampling: 1% for negative memory bank, 10% for positive memory bank
- Memory bank approach: k-nearest neighbors (k=3) for neighborhood weighting
- DIAG augmentation: 50 images per prompt (100 total)

5.2 Comparative Evaluation

5.2.1 Anomaly Detection and Localization Results

Table 1 presents both the image-level and pixel-level anomaly detection results on the KSDD2 dataset, enabling a comprehensive comparison of the different methods across both detection and localization tasks.

Table 1: Anomaly detection and localization performance on KSDD2 dataset.

Method	Image-level AUROC (%)	Pixel-wise AUROC (%)
PatchCore	91,2	95.8
PatchCoreDual	93,1	96.9
PatchCoreDual + DIAG	94.2	97.7

The results demonstrate that the PatchCoreDual method outperforms the standard PatchCore approach in both image-level and pixel-level anomaly detection. The improvement in image-level AUROC indicates enhanced capability to determine whether an image contains defects, while the increase in pixel-wise AUROC shows better precision in localizing the exact defective regions within the images.

When combined with DIAG-inspired synthetic data augmentation, the performance further improves across both metrics. This enhancement demonstrates the complementary benefits of the dual memory bank architecture and the synthetic defect generation approach. The pixel-wise

performance improvements are particularly noteworthy, as they indicate that the synthetic defects not only help identify the presence of anomalies but also contribute to more accurate defect localization.

5.2.2 Impact of Diffusion-Based Data Augmentation

To isolate the effect of the DIAG-inspired synthetic data augmentation, experiments were conducted with varying numbers of synthetic samples. Table 3 presents the results for both image-level and pixel-level AUROC metrics.

Table 2: Effect of varying the number of augmented samples on performance.

Method	Synthetic Samples	Image-level AUROC (%)	Pixel-wise AUROC (%)
PatchCoreDual	0	93.1	96.9
PatchCoreDual+DIAG	50	93.4	97.1
PatchCoreDual+DIAG	100	94.2	97.7
PatchCoreDual+DIAG	150	93.5	97.2

The results demonstrate that the addition of synthetic samples improves performance across both detection and localization tasks. With no synthetic samples (first row), the PatchCoreDual model relies solely on the limited number of real defective samples available in the KSDD2 dataset. As synthetic samples are introduced, the detection and localization capabilities steadily improve.
Adding 100 synthetic samples (50 per prompt) yields the optimal performance for both image-level and pixel-wise AUROC metrics. Interestingly, further increasing the synthetic sample count to 150 does not provide additional benefits and in some cases slightly reduces performance.

5.2.3 Qualitative Analysis

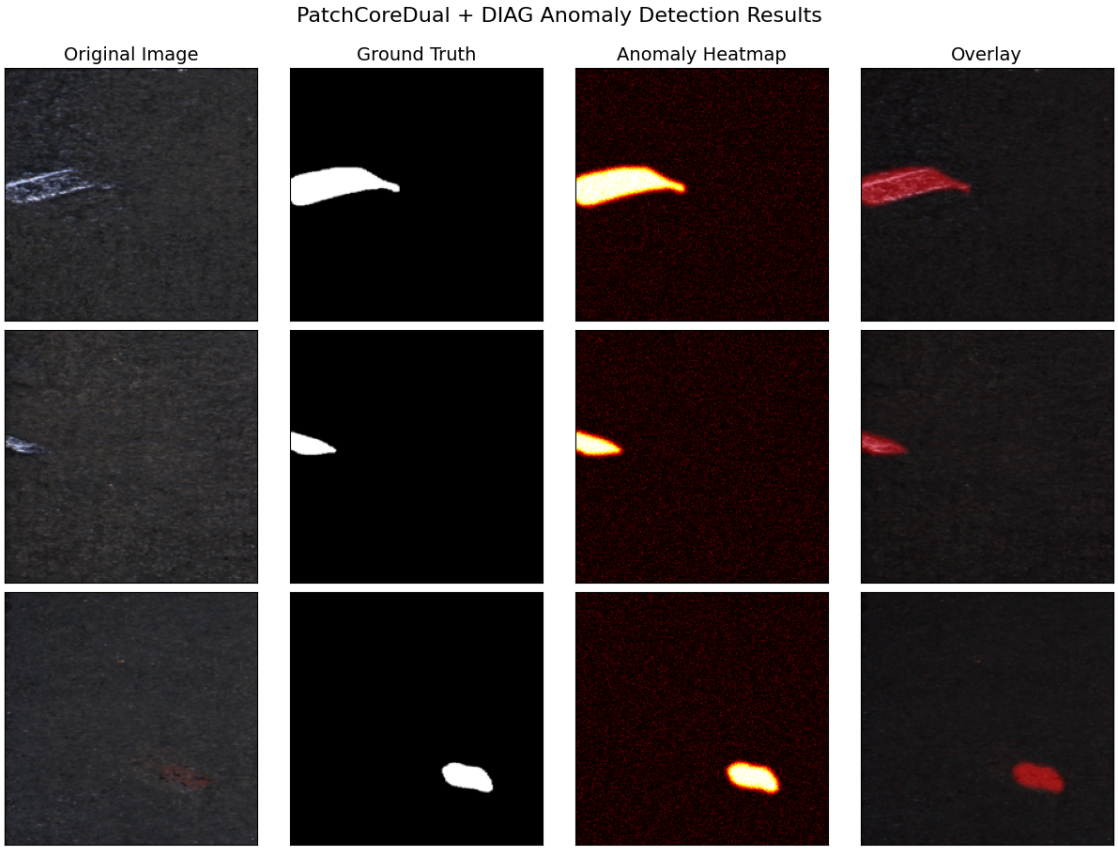

Fig. 5, qualitative results for anomaly localization on sample images from the KSDD2 test set.

The visualization shows that PatchCoreDual + DIAG effectively detects and localizes various types of defects in electrical commutators, including subtle scratches, surface anomalies, and material imperfections. The heatmaps generated by this method closely match the ground truth masks, demonstrating precise localization capabilities.

Comparing the heatmaps generated by the standard PatchCore approach with those from PatchCoreDual and PatchCoreDual + DIAG, we observe progressively better alignment with ground truth masks and fewer false positives in background regions. These qualitative improvements align with the quantitative metrics and further validate the benefits of the proposed methodology.

5.3 Discussion

5.3.1 Key Findings

The experimental evaluation reveals several important findings about the proposed methodology. The integration of both normal and defective samples in separate memory banks leads to meaningful improvements in anomaly detection performance compared to the standard PatchCore approach. Specifically, the image-level AUROC increased from 91.2% with the standard PatchCore to 93.1% with PatchCoreDual, demonstrating that incorporating positive samples enhances the model's discriminative capabilities. While these improvements may not be revolutionary in magnitude, they consistently point toward the value of explicitly modeling both normal and anomalous distributions rather than relying solely on normal data modeling.

The PatchCoreDual approach sets an important foundation for future industrial anomaly detection systems by demonstrating that positive samples, even when limited in number, can be a valuable resource for creating better representations of what defects look like in specific industrial contexts. This finding challenges the traditional paradigm in anomaly detection that focuses exclusively on modeling normal data distributions. Instead, it suggests that a hybrid approach leveraging both normal and anomalous examples can yield more robust detection systems, particularly in industrial settings where the types of possible defects are often known in advance, even if examples are scarce. By introducing synthetically generated defects that closely resemble real anomalies, the model becomes more sensitive to subtle variations in defect appearance. The diffusion-based approach enables the generation of in-distribution anomalies that maintain the texture and structural properties of real defects, unlike traditional augmentation methods that often create out-of-distribution patterns. This in-distribution property is crucial for training anomaly detectors that can identify genuine defect patterns rather than simply learning to detect artificial artifacts. The experimental results demonstrate that these synthetically generated samples serve not just as additional training data but actively improve the model's decision boundaries between normal and anomalous regions.

The qualitative analysis of detection results further validates the effectiveness of the dual memory bank approach. The addition of the positive memory bank particularly improves detection capabilities for more challenging defects that exhibit subtle deviations from normal appearances. In several test cases, defects that went undetected by the standard PatchCore model were

successfully identified by PatchCoreDual. This suggests that the positive examples help the model establish more nuanced decision boundaries, especially in the gray areas where normal variation and mild defects might otherwise be confused.

The performance gains observed with PatchCoreDual are also reflected in the pixel-level anomaly localization metrics, with pixel-wise AUROC improving from 95.8% to 96.9%. This enhancement in localization precision is particularly valuable in industrial applications, where accurately identifying the extent and location of defects can inform repair procedures or quality control decisions. The ability to not only detect the presence of defects but also precisely delineate their boundaries represents a significant practical advantage for manufacturing inspection systems.

These findings collectively point to the potential of memory-based dual modeling approaches in industrial anomaly detection. By explicitly incorporating both normal and anomalous patterns, PatchCoreDual establishes a framework that can be extended and refined in various directions, potentially leading to more robust and accurate detection systems for critical quality control applications.

5.3.2 Limitations and Future Work

Despite optimistic performance of the method, several limitations and directions for future work remain. The current implementation uses the same feature extraction layers (levels 2 and 3 of ResNet) for both normal and defective memory banks. However, defects often manifest as localized patterns with distinctive low-level features. Future work could explore using different feature extraction layers for the positive memory bank, particularly leveraging lower-level features that might better capture the fine-grained characteristics of defects. This approach is especially relevant when working with cropped defect regions, where the contextual information is reduced, and the detailed texture patterns become more significant for representation. The need for resizing defect crops to maintain consistent feature dimensions introduces another limitation that could potentially be addressed through more sophisticated feature aggregation techniques.

The quality and quantity of available data represent another important limitation. The KSDD2 dataset, while valuable, contains a relatively small number of defect samples with varying quality. More comprehensive datasets with higher resolution images, consistent lighting conditions, and more diverse defect types would enable more robust model training and evaluation. Additionally, better ground truth annotations with pixel-precise defect boundaries would improve both training

effectiveness and evaluation accuracy. The synthetic data generation approach partially addresses these limitations, but the quality of synthetic samples remains dependent on the quality of the real examples they aim to mimic.

Also, its adaptability to other industrial domains with different defect characteristics requires further investigation. Industrial surfaces vary widely in texture, reflectivity, and structural complexity, potentially affecting the feature representations learned by the backbone network.

The current implementation relies on manually crafted prompts for synthetic defect generation. Following insights from the DIAG methodology, these prompts are created based on domain knowledge of possible defect types and appearances. Techniques to automatically generate effective prompts based on the characteristics of the industrial domain would be a valuable direction to reduce the reliance on expert knowledge. Additionally, a more systematic approach to evaluating the quality of generated defects and their impact on detection performance could help optimize the synthetic data generation process.

As industrial settings evolve and new defect types emerge, mechanisms for incrementally updating the positive memory bank without full regeneration would be valuable. Developing strategies for continuous learning that can incorporate new defect patterns while maintaining performance on existing ones is an important area for future research to ensure the long-term applicability of the method. This could involve techniques for selective memory bank updates or importance weighting of samples based on their rarity or representativeness of emerging defect patterns.

The interplay between feature extraction, memory bank construction, and anomaly scoring represents a complex system with multiple design decisions. Future work could explore more sophisticated mechanisms for combining evidence from normal and defective memory banks, perhaps incorporating uncertainty estimation or hierarchical decision processes that consider both local patch-level anomalies and global image context. Such approaches might better handle ambiguous cases where defects share some visual characteristics with normal variations or where background textures create challenging detection scenarios.

6. Conclusion

This thesis has presented PatchCoreDual, an approach to industrial anomaly detection that combines memory-based techniques with diffusion-based synthetic data augmentation. The work

explored how positive samples, even in limited quantities, can contribute to enhanced detection capabilities.

Experiments with the PatchCoreDual model produced interesting results that suggest potential benefits from incorporating positive examples in memory-based anomaly detection and they consistently indicated that dual memory bank architectures warrant further investigation. The synthetic data generation component also showed promise, helping to address the common issue of class imbalance in industrial datasets.

The research process involved various learning opportunities. Working with industrial datasets provided valuable insights into the practical challenges of anomaly detection in manufacturing contexts. Implementing and adapting the memory bank architecture required experimentation with different feature extraction approaches and memory organization strategies, which enhanced understanding of representation learning for visual inspection tasks.

The integration of diffusion models for data augmentation was particularly interesting, as it represented an application of recent advances in generative AI to the specific domain of industrial defect generation. Learning to craft effective prompts and evaluate the quality of generated samples was an educational process that highlighted both the capabilities and limitations of current generative models.

Throughout this work, the opportunity to explore different approaches and challenge conventional wisdom in anomaly detection was valuable. The standard paradigm of modeling only normal data has strong theoretical foundations, but exploring alternatives that incorporate positive examples provided interesting practical insights.

Looking ahead, the field of industrial anomaly detection presents exciting opportunities for innovation. The rapid advancement of generative AI technologies promises even more sophisticated data augmentation techniques, potentially addressing the persistent challenge of limited defect samples. Meanwhile, memory-based approaches offer a flexible framework that can evolve alongside these new capabilities. As manufacturing processes become increasingly automated and quality standards more stringent, the demand for robust, adaptable inspection systems will only grow, but with continued exploration of hybrid approaches that leverage both traditional computer vision wisdom and modern deep learning innovations, there is considerable potential to develop inspection systems that will meet the present and future needs.

References

Agarwal, P. K., Har-Peled, S., & Varadarajan, K. R. (2004). Geometric approximation via coresets. *Combinatorial and Computational Geometry, 52*.

Bergman, L., Cohen, N., & Hoshen, Y. (2020). Deep nearest neighbor anomaly detection. *arXiv preprint arXiv:2002.10445*.

Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2019). MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Božič, J., Tabernik, D., & Skočaj, D. (2021). Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Computers in Industry, 129*.

Cohen, N., & Hoshen, Y. (2020). Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*.

Defard, T., Setkov, A., Loesch, A., & Audigier, R. (2021). PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *Pattern Recognition. ICPR International Workshops and Challenges*.

Girella, F., Liu, Z., Fummi, F., Setti, F., Cristani, M., & Capogrosso, L. (2024). Leveraging latent diffusion models for training-free in-distribution data augmentation for surface defect detection. *arXiv preprint arXiv:2407.03961*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems, 33*.

Johnson, W. B., & Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics, 26*.

NVIDIA. (2022). NVIDIA Diversity Sampling Library. GitHub repository. https://github.com/NVIDIA/DeepLearningExamples/tree/master/Tools/DiversitySampling

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2023). SDXL: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., & Gehler, P. (2022). Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1).

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.

Yang, M., Wu, P., & Feng, H. (2023). MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities. *Engineering Applications of Artificial Intelligence, 119*.

Zavrtanik, V., Kristan, M., & Skočaj, D. (2021). DRAEM-a discriminatively trained reconstruction embedding for surface anomaly detection. In *IEEE/CVF International Conference on Computer Vision*.