electronics

Article

Towards More Accurate Industrial Anomaly Detection: A Component-Level Feature-Enhancement Approach

Xiaodong Wang [1], Zhiyao Xie [1], Fei Yan [1,*], Jiayu Wang [1], Jiangtao Fan [1], Zhiqiang Zeng [1], Junwen Lu [1], Hangqi Zhang [2] and Nianfeng Zeng [3]

[1] College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China; xdwangjsj@xmut.edu.cn (X.W.); 2222031162@stu.xmut.edu.cn (Z.X.); 2222031150@stu.xmut.edu.cn (J.W.); 2322071010@stu.xmut.edu.cn (J.F.); zqzeng@xmut.edu.cn (Z.Z.); 2010110707@xmut.edu.cn (J.L.)
[2] Xiamen Yaxon Zhilian Technology Co., Ltd., Xiamen 361000, China; zhanghangqi@yaxon.com
[3] E-Success Information Technology Co., Ltd., Xiamen 361024, China; nfzeng18605928867@163.com
* Correspondence: fyan@xmut.edu.cn

Abstract: Industrial visual inspection plays a crucial role in intelligent manufacturing. However, existing anomaly-detection methods based on unsupervised learning paradigms often struggle with issues such as overlooking minor defects and blurring component edges in confidence maps. To address these challenges, this paper proposes an industrial anomaly-detection method based on component-level feature enhancement. This method introduces a component-level feature-enhancement module, which optimizes feature matching by calculating the structural similarity between global coarse-grained confidence features and local fine-grained confidence features, thereby generating enhanced feature maps to improve the model's detection accuracy for minor defects and local anomalies. Additionally, we propose a region-segmentation method based on multi-layer piecewise thresholds, which effectively distinguishes between foreground and background in confidence maps, circumvents background interference and ensures the integrity of structural information of foreground components. Experimental results demonstrate that the proposed method surpasses comparative methods in both logical and structural defect detection tasks, showing significant advantages, especially in fine-grained anomaly detection, with stronger robustness and accuracy.

Keywords: industrial visual inspection; anomaly detection; component-level feature enhancement

Academic Editor: Marcin Witczak

Received: 6 March 2025
Revised: 12 April 2025
Accepted: 15 April 2025
Published: 16 April 2025

Citation: Wang, X.; Xie, Z.; Yan, F.; Wang, J.; Fan, J.; Zeng, Z.; Lu, J.; Zhang, H.; Zeng, N. Towards More Accurate Industrial Anomaly Detection: A Component-Level Feature-Enhancement Approach. *Electronics* **2025**, *14*, 1613. https://doi.org/10.3390/electronics14081613

1. Introduction

Industrial visual inspection, as a crucial component of intelligent manufacturing, is one of the core technologies for achieving product quality control and production automation. The primary goal of industrial visual inspection is to detect product defects during the production process using machine vision technology, thereby improving product quality, reducing production costs, and enhancing production efficiency. Early industrial visual inspection methods mainly relied on rule-based systems using handcrafted features and traditional image processing techniques. Evaluations on the MVTec Logical Constraints Anomaly Detection (MVTec LOCO AD) dataset showed that these methods achieved 65.8% Area Under the Receiver Operating Characteristic curve (AUROC) for logical anomalies and 62.7% AUROC for structural defects. As shown in Figure 1, the baseline anomaly-detection method generates heatmaps where normal components appear as complete blocks (first row), while distinct anomalies (e.g., white paper in the second row) cause missing regions.

However, for visually similar defects (yellow paper in the third row), the method fails to produce clear separations, resulting in scattered low-confidence anomaly scores. These methods exhibit significant limitations when faced with complex and variable industrial scenarios, particularly in handling texture variations and novel defect patterns that require more sophisticated analysis.

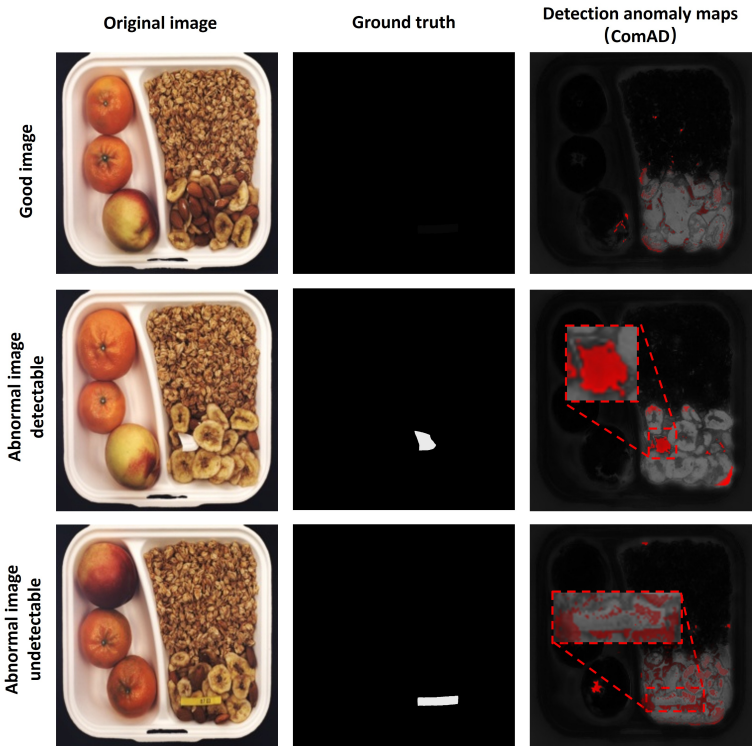

Figure 1. A schematic diagram of local region anomalies, where the first row represents normal samples, and the second and third rows represent abnormal samples. The columns display (from **left** to **right**): the original images, ground truth annotations (white squares mark defective regions) and detection anomaly maps (red lines outline magnified views).

In recent years, with the rapid development of deep learning technologies, significant progress has been made in anomaly-detection methods: Rudolph et al. [1] propose an asymmetric teacher-student framework for industrial defect localization, while Wang et al. [2] address complex industrial inspection challenges through multimodal fusion. For anomaly detection in industrial control systems, Choi et al. [3] introduce an unsupervised learning approach. Yao et al. [4] develop a global-local semantic bottleneck approach for logical anomalies. The latest advancements in graph neural networks have demonstrated advantages in structured industrial data [5]. Notably, Gao et al. [6] further improve cross-domain generalization through feature decoupling techniques, demonstrating robust performance under unseen operating conditions in rotating machinery diagnosis.

Industrial anomaly detection can be categorized into three machine learning paradigms: supervised, semi-supervised, and unsupervised methods. Supervised machine learning algorithms, such as Support Vector Machines (SVM) [7], utilize fully labeled datasets containing both normal and abnormal samples to establish explicit decision boundaries. For instance, Kent et al. [7] demonstrated SVM's effectiveness in detecting sensor anomalies in building automation systems. While achieving high detection accuracy, these methods face practical limitations in industrial settings where acquiring balanced labeled anomaly samples proves challenging. Semi-supervised approaches significantly reduce labeling demands by primarily utilizing normal samples during training AND learning characteristic feature distributions that allow identification of deviations. Unsupervised techniques,

exemplified by Principal Component Analysis (PCA)-based methods like those developed by Mnassri et al. [8], eliminate annotation requirements through intrinsic pattern recognition and density estimation, though often with reduced accuracy for complex anomalies. This classification framework considers not just supervision levels but also the practical constraints of label availability and the fundamental differences in how each paradigm models normal and abnormal features. However, in industrial fields, abnormal samples are often scarce and diverse, making it extremely difficult to obtain comprehensive and balanced training data, which limits the practical application of supervised methods. In contrast, semi-supervised methods primarily rely on normal samples for training, learning the distribution characteristics of normal samples to identify abnormal samples that deviate from the normal distribution. While this approach reduces the reliance on abnormal samples, it still faces certain limitations in detecting unseen complex anomalies.

To address the aforementioned issues, unsupervised anomaly-detection methods have gradually become a research hotspot. These methods do not require the involvement of abnormal samples and can more flexibly adapt to complex scenarios and diverse anomalies by mining latent patterns from normal samples or leveraging self-supervised learning techniques. Unsupervised anomaly-detection methods have been proven particularly suitable for industrial scenarios where data annotation is challenging, and anomaly features are difficult to cover comprehensively. Recent advancements in this field include logical constraint-based methods, as explored in [9], which extend anomaly detection beyond simple structural defects by incorporating semantic consistency checks. Additionally, image resynthesis techniques [10] improve anomaly localization by detecting deviations from expected reconstructions. For efficiency in industrial applications, progressive pruning strategies [11] and noise-guided feature aggregation [12] further enhance unsupervised detection performance in complex environments. They have played a significant role in enhancing industrial visual inspection capabilities and provided new technical means for solving complex anomaly-detection problems in real production environments. Among existing unsupervised anomaly-detection techniques, reconstruction-based methods and feature contrast-based methods are the two most representative approaches.

Reconstruction-based methods rely on Autoencoders (AE) and Variational Autoencoders (VAE) [13], as well as Generative Adversarial Networks (GAN) [14], to perform anomaly detection. Their core idea is to utilize the characteristic that the reconstruction error of normal samples is significantly lower than that of abnormal samples. However, these methods struggle with challenges such as overlooking minor defects and blurring component edges in confidence maps. For example, the Memory-guided Normality for Anomaly Detection (MNAD) [15] proposes a memory-guided anomaly-detection model suitable for local anomaly detection in video sequences, but it inadequately expresses boundary information of anomaly regions during feature extraction. To address the balance between local and global information representation, the Semantic Pyramid Anomaly Detection (SPADE) [16] employs a multi-resolution semantic pyramid to enhance feature representation. This approach not only improves feature expression but also overcomes the limitation of traditional K-Nearest Neighbors (KNN) methods in providing precise anomaly segmentation. Nevertheless, in cases where anomaly regions are morphologically complex and background interference is strong, SPADE struggles to capture minor local features, and the issue of edge blurring remains significant. Facing the challenge of insufficient expression of boundary information in anomaly regions, the Discriminatively Trained Reconstruction Embedding for Anomaly Detection model (DRAEM) [17] achieves precise localization of anomaly regions without complex post-processing by jointly learning the reconstruction representations of abnormal images and normal images. Despite

this, DRAEM's reliance on anomaly simulation training strategies still leaves room for improvement in its applicability to multi-component samples.

To address the limitations of reconstruction-based methods in feature boundary representation and background interference, feature contrast-based methods have further enhanced detection robustness by comparing the feature differences between normal and abnormal samples. The Student-Teacher (S-T) anomaly-detection framework [18] employs a student-teacher architecture, utilizing knowledge distillation to learn the feature distribution deviations of normal samples for anomaly detection. Although this method exhibits strong robustness, its performance degrades in scenarios with limited data or class imbalance between normal and abnormal samples. The Component-aware Anomaly-Detection framework (ComAD) [19], based on contrastive learning and clustering techniques, achieves component-level anomaly detection. However, ComAD demonstrates insufficient adaptability in multi-region anomaly scenarios with complex background interference, and its feature extraction process still requires improvement in separating local anomalies from background information. As shown in Figure 1, the first column displays the original unprocessed input images showing industrial components in their natural state. The anomaly labels (second column) precisely delineate defective regions that deviate from normal patterns, while the third column presents the ComAD-generated anomaly heatmaps with red regions indicating detected anomalies and color intensity representing corresponding anomaly scores. In normal samples (first row), the heatmap of this component appears as a complete block, indicating that the model can correctly identify component features. In the abnormal sample in the second row, a white paper mixed into the nut causes the detection anomaly map to exhibit a significant missing region, thereby correctly identifying it as an abnormal image. In contrast, in the abnormal sample in the third row, a yellow paper mixed into the nut, due to its visual similarity to the nut itself, makes it difficult for the model to distinguish it from normal features. As a result, the detection anomaly map still appears as a complete block, with scattered and dim anomaly distributions, leading ComAD to fail to distinguish it from normal samples and resulting in a detection error.

In summary, the current technological approaches in the field of industrial visual inspection encompass methods based on reconstruction, feature contrast, and knowledge distillation, which have achieved significant progress in anomaly detection and localization. However, numerous challenges remain in practical applications. For instance, anomalies often manifest as minor local defects with blurred boundary information, making precise extraction difficult. Additionally, interference from complex backgrounds or high-resolution images complicates the effective separation of targets from backgrounds, thereby affecting detection accuracy. For example, ComAD is susceptible to edge-blurring effects in multi-region anomaly scenarios, limiting its performance in minor local anomaly detection. These issues are more pronounced in real industrial environments, necessitating further optimization in feature extraction and enhancement strategies to effectively improve the overall robustness and accuracy of detection.

To address the aforementioned challenges, this paper proposes an industrial anomaly-detection method based on component-level feature enhancement. By introducing a component-level feature-enhancement module, the method aims to overcome the limitations of traditional approaches in accurately separating targets from backgrounds in complex scenarios, thereby enhancing the model's capability for fine-grained anomaly detection. Specifically, the method first selects multi-component feature regions and extracts the original features, confidence features, and positional information of the components. The original features of the components encompass visual information such as basic morphology, texture, and edges, which are crucial for identifying anomalies in shape, texture,

and other aspects. The confidence features quantify the importance of feature regions, reflecting the model's attention to specific component areas, while the positional information encodes spatial relationships to ensure consistency across different perspectives and scales. Together, these elements provide critical support for the reliability assessment and spatial localization of components, thereby enhancing the precise capture of target regions and facilitating more accurate detection of fine-grained anomalies.

To further improve feature extraction accuracy, this paper also proposes a region-segmentation method based on multi-layer piecewise thresholds, dividing the confidence map into foreground, background, and transition regions. The transition region serves as a supplement to the foreground, aiding the component-level feature encoder in generating more precise features. This method not only avoids the issue of incomplete foreground structural information but also effectively eliminates background interference and reduces the risk of false detection. Compared to traditional methods (e.g., K-means and Otsu), it achieves higher segmentation accuracy in detecting logical anomalies and structural anomalies. The main contributions of this paper include:

(1) To address the limitations of traditional methods that are prone to edge-blurring effects in multi-region anomaly scenarios, we propose an anomaly-detection method based on component-level feature enhancement. By focusing on potential local anomalies in images and incorporating a structural similarity analysis mechanism, the method enhances the preservation of component details and improves performance in logical anomaly detection. Additionally, through an effective local feature extraction strategy, the method strengthens the model's capability for fine-grained anomaly detection and significantly improves the identification of anomaly patterns.

(2) Addressing the issue of background interference during detection, we propose a region-segmentation method based on multi-layer piecewise thresholds. This method divides the confidence map into foreground, background, and transition regions, not only enhancing the integrity of structural information but also effectively mitigating interference from background regions, thereby further improving detection accuracy.

(3) To address the challenge of optimizing feature matching for enhanced detection of minor defects, we propose a feature-enhancement module based on the Peak Signal-to-Noise Ratio (PSNR). By calculating the structural similarity between original confidence features and deep confidence features in multi-component features, the module optimizes the feature matching and alignment process. Using similarity scores as a basis, it generates enhanced feature maps, effectively improving the model's accuracy and robustness in detecting minor defects and local anomalies.

(4) We validate the superiority of the method on public datasets. Experimental results demonstrate that the method achieves leading performance in both logical defect and structural defect detection.

2. Related Work

This chapter systematically reviews the technological evolution and current research landscape in industrial anomaly detection, establishing a comprehensive academic framework through three interconnected dimensions. Beginning with the fundamental principles of unsupervised detection methods, we analyze their foundational contributions and inherent limitations in detecting both structural and logical anomalies. The discussion then progresses to examine how new paradigms based on large-scale pre-trained models have broken through traditional feature representation bottlenecks. Finally, we discuss how feature enhancement and sample generation techniques improve detection accuracy.

2.1. Unsupervised Anomaly-Detection Methods

Unsupervised anomaly-detection methods have achieved significant progress in the field of industrial visual inspection in recent years, particularly in addressing the more challenging task of logical anomaly-detection. To tackle complex anomaly scenarios in industrial environments, researchers have introduced comprehensive datasets that include both structural anomalies and logical anomalies. Among these, structural anomalies manifest as significant deviations in object appearance, while logical anomalies involve unreasonable arrangements or incorrect combinations of objects. Several methods proposed for this dataset have demonstrated excellent performance in detecting both structural anomalies and logical anomalies.

Logical anomaly-detection methods typically focus on modeling long-range dependencies. The Template-guided Hierarchical Feature Restoration method (THFR) [20] employs deep metric learning to select optimal normal references and performs feature restoration through cross-attention mechanisms, quantifying anomalies via residual analysis. However, although THFR's compensation strategy can transform residual anomaly features into normal features, it may still underperform in detecting subtle edge-blurring anomalies in certain scenarios. Additionally, Zhang et al. [21] proposed a Global Context Compression Block (GCCB) to enhance the global student model's ability to learn long-range dependencies. While this method can roughly outline the shape of anomalies, it still fails to address the issue of edge blurring in small components.

Despite the outstanding performance of the above two methods in logical anomaly detection, they generally overlook anomalies that exhibit edge blurring and small sizes within small components or parts. Such defects are difficult to accurately identify using logical anomaly-detection methods that rely on long-range features. To address this challenge, some studies have attempted to integrate unsupervised detection strategies for both structural anomalies and logical anomalies to improve the detection capability for minor local anomalies. For instance, GCAD [22] detects both logical anomalies and structural anomalies by fusing local and global detection results, providing a comprehensive solution for industrial anomaly detection. EfficientAD [23] inherits the dual-branch idea of GCAD on the MVTec LOCO AD dataset, employing two independent convolutional neural network modules to handle structural anomalies and logical anomalies separately. However, both methods have certain limitations in addressing minor local anomalies. GCAD primarily focuses on overall structural changes in samples containing structural anomalies, with limited capability in identifying minor defects within components. EfficientAD improves upon GCAD in local anomaly detection but emphasizes the logical relationships between anomalous components, failing to fully resolve the issue of minor local anomalies. PUAD [24] combines EfficientAD's approach to handling logical anomalies and addresses some local anomaly issues through reconstruction. However, its method mainly focuses on feature differences in larger regions, making it difficult to precisely locate small and concealed defects.

In summary, while these methods excel in detecting logical anomalies and structural anomalies, they have not sufficiently addressed the detection of extremely subtle minor defects on product surfaces. Minor defects are often mixed within normal components, making them difficult to capture by traditional methods or prone to being misclassified as normal regions, thereby affecting overall detection performance. This limitation highlights the importance of enhancing local detail detection capabilities, which is also one of the core objectives of this study.

2.2. Anomaly-Detection Methods Based on Large Pre-Trained Models

With the increasing complexity and diversity of industrial inspection tasks, as well as the challenges of high annotation costs and scarce anomaly samples, traditional methods relying on task-specific or dataset-specific feature extractors face limitations in generalization capabilities. To address these issues, defect detection methods based on large pre-trained models have demonstrated outstanding performance in various industrial applications, particularly excelling in complex anomaly-detection tasks. For instance, architectures such as ResNet, VGG, and EfficientNet in convolutional neural networks (CNNs) have been widely applied to defect detection tasks. Meanwhile, the DSR method proposed by Zavrtanik et al. [25], based on a dual-decoder structure, utilizes quantized feature space representations and generates anomalies using latent space models pre-trained on ImageNet, achieving superior anomaly detection and localization performance without relying on image-level anomaly synthesis. Additionally, the DTDF [26] method employs pre-trained networks to obtain multi-scale prior embeddings and combines a dual attention mechanism to achieve two-stage reconstruction, effectively enhancing anomaly-detection capabilities.

To further improve the representation of local and structural features in anomaly detection, some methods have introduced self-supervised learning and semantic segmentation techniques into feature modeling. DINO [27], based on self-supervised knowledge distillation, leverages Vision Transformers (ViTs) to extract local and global image features and maintains multi-scale feature consistency through distillation loss, demonstrating excellent performance in enhancing global and local feature representation. ComAD [19] focuses on component-level anomaly detection in industrial visual inspection, dividing images into multiple components using unsupervised semantic segmentation models and capturing logical relationships between components to improve the detection of logical and structural anomalies. Although both DINO and ComAD have achieved success in feature encoding at different levels, the former has limitations in handling edge blurring and minor anomalies, while the latter underperforms in detecting fine-grained local anomalies.

2.3. Anomaly-Detection Methods Based on Feature Enhancement and Sample Generation

In the field of defect detection, the quality of feature representation often directly determines the performance of detection methods. Effective feature enhancement can highlight the saliency of defect regions, thereby improving detection accuracy and robustness. With the advancement of deep learning technologies, enhancement-based methods have been widely applied in complex industrial inspection tasks. For example, Bergmann et al. [18] proposed a student-teacher network framework, where the student network learns to regress the rich feature representations generated by the teacher network and the output differences that the student network fails to generalize effectively are used to localize anomaly regions, thereby achieving feature enhancement and improving anomaly detection and pixel-level segmentation performance. Meanwhile, Jongmin Yu et al. [28] utilized adversarial learning to construct a mapping function from images to the frequency domain, adaptively learning frequency domain features through generative adversarial networks (GANs), effectively enhancing the model's ability to express defect features. However, these methods still face performance challenges in scenarios with scarce defect samples and have limitations in effectively extracting complex features. To address these issues, BLDM [29] employs a hybrid latent diffusion model to generate defect samples in latent space, enhancing defect samples to improve anomaly-detection performance. Nevertheless, these methods underperform in handling long-range dependencies between components. Overall, although existing methods have made significant contributions to feature enhancement and sample generation, more effective solutions are still needed for complex industrial scenarios, especially for fine-grained defects and edge blurring in small components.

3. Methodology

In industrial visual inspection, many anomalies typically appear in local regions of larger normal components. Since these defects are often highly visually similar to the normal component features, detection systems are prone to misclassifying them as normal parts of the components. More complexly, these anomalies may occupy only a small portion of the component, making traditional global feature detection methods ineffective in identifying and capturing these subtle and concealed local anomalies. To address this issue, we designed a multi-component feature region selection method, focusing on anomalies mixed within normal components to enhance the detection capability for local anomaly regions. As shown in Figure 2, the schematic diagram of the proposed method is divided into the training phase (upper part of the figure) and the testing phase (lower part of the figure).

The core innovation of this framework lies in its multi-component feature region selection method, which first divides the image into multiple regions representing the structural features of different components. In industrial visual inspection applications, these regions effectively capture the unique distribution patterns and structural variations of individual components, providing precise and fine-grained input data for subsequent processing stages.

During the training phase, the model first divides the image into multiple local regions and extracts component features for each region. Subsequently, based on the multi-component feature region selection strategy, the feature-enhancement module utilizes a self-attention mechanism to enhance the expression of key information regions in the component features. Specifically, the model first processes the input image I through a pre-trained encoder to generate the feature map F. A greedy sampling algorithm [30] is then employed to extract and stack representative point features from key regions, followed by clustering to produce the multi-kernel feature vector V. Subsequently, the tensor product operation between feature map F and V, combined with Conditional Random Field (CRF) interpolation, yields the component confidence map P_i. Threshold segmentation and contour extraction are applied to obtain the component position encoding R_i, which is then mapped back to the original image to crop the component image I^r and confidence feature P^r. Then, the module combines feature fusion strategies to integrate weighted features from different channels, thereby generating candidate feature images. Finally, by calculating the similarity scores between the confidence images and the candidate feature images and matching the feature map with the highest score, further feature enhancement is achieved.

In the test phase, the test image undergoes multi-component feature region selection and feature enhancement through the feature-enhancement module. Then, an anomaly-detection method based on region, color, and histogram features is used to calculate the similarity between the test image and the normal image and evaluate the anomaly degree by combining the K-Nearest Neighbors (KNN) and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. Through the aforementioned component feature extraction and optimization, the method can more precisely capture the anomalous features of local regions. The following sections provide a detailed introduction to the two aforementioned methods.

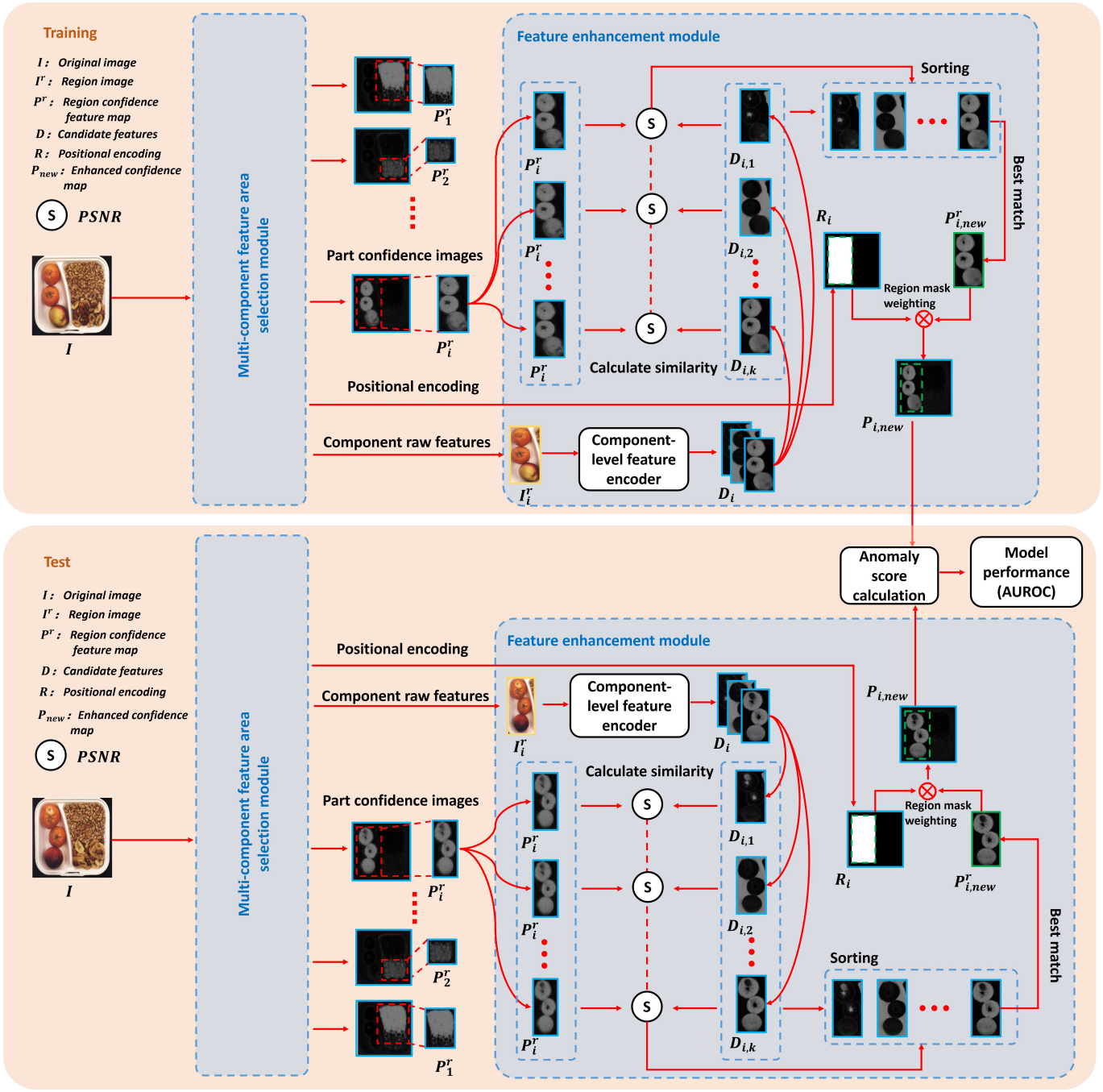

Figure 2. Schematic diagram of component-level feature enhancement.

3.1. Datasets and Comparison Methods

This research employs the MVTec LOCO AD benchmark dataset [9], a publicly available collection for industrial visual anomaly detection. The dataset contains carefully curated normal and abnormal samples specifically designed to assess unsupervised anomaly localization algorithms, featuring 3644 high-resolution images across five representative industrial categories: breakfast box, juice bottle, pushpins, screw bag, and splicing connectors. These categories were selected to mirror actual industrial inspection scenarios. The dataset encompasses two fundamental anomaly types: structural anomalies (including manufacturing defects like scratches, dents, and contamination) and logical anomalies (characterized by violations of functional constraints such as incorrectly positioned or missing components). Each anomaly is meticulously annotated with pixel-precise segmentation masks. The dataset's comprehensive coverage of diverse industrial scenarios

and anomaly types provides a realistic testbed for assessing algorithm performance under challenging conditions typical of industrial applications, particularly for handling complex backgrounds and achieving precise defect localization. The selected product categories present varying levels of complexity in terms of surface properties, object scales, and background environments to thoroughly test algorithm robustness. This dataset serves as an effective benchmark due to its systematic inclusion of both common defect types and challenging inspection scenarios found in real industrial settings. The entire dataset involves 1772 normal images for training and 304 normal images for validation. For the test set, there are a total of 575 normal images, 432 structural anomaly images, and 561 logical anomaly images.

MVTec AD [31] is a dataset for benchmarking anomaly-detection methods with a focus on industrial inspection. It contains over 5000 high-resolution images divided into 15 different object and texture categories. Each category comprises a set of defect-free training images and a test set of images with various kinds of defects, as well as images without defects. For the test set, there are a total of 467 normal images and 1258 abnormal images. The proposed method is solely evaluated on the object categories without considering the homogeneous texture categories.

For comparative evaluation, we compare our method with five established unsupervised anomaly-detection approaches: MNAD [15], SPADE [16], DRAEM [17], S-T [18], and ComAD [19], and MPFnet [32]. These methods represent distinct technical approaches: MPFnet introduces a multi-scale prototype fusion mechanism for enhanced defect localization; ComAD integrates the DINO [27] pre-trained model with KNN clustering for component-based anomaly detection; MNAD implements memory-enhanced normal pattern learning; SPADE employs multi-scale spatial modeling; DRAEM combines denoising and reconstruction strategies; while S-T utilizes teacher-student knowledge distillation. Each method offers unique advantages: MNAD excels in normal pattern memorization, SPADE effectively captures spatial anomalies, DRAEM handles fine-grained defects well, S-T demonstrates robustness with limited data, and ComAD specializes in component-level anomaly identification.

3.2. Multi-Component Feature Region Selection

The core objective of the multi-component feature region selection method is to divide an image into multiple regions, each representing the structural features of different components in the image. In the field of industrial visual inspection, the features of individual components often exhibit unique regional distributions and structural differences. Therefore, this method enables the effective identification of potential anomaly regions, providing more precise and fine-grained input information for subsequent feature-enhancement steps.

The multi-component feature region selection method not only comprehensively considers the overall features of each component in the image but also ensures, through fine-grained partitioning of local regions, that the detection model can focus on subtle changes within components, such as defects, damage, or the intrusion of foreign objects. This approach allows the model to pay greater attention to local regions that are easily overlooked in traditional component feature detection, thereby significantly improving the sensitivity of local anomaly detection. Figure 3 illustrates the detailed workflow of multi-component feature region selection, including key steps such as feature encoding, foreground and background partitioning, and positional information extraction.

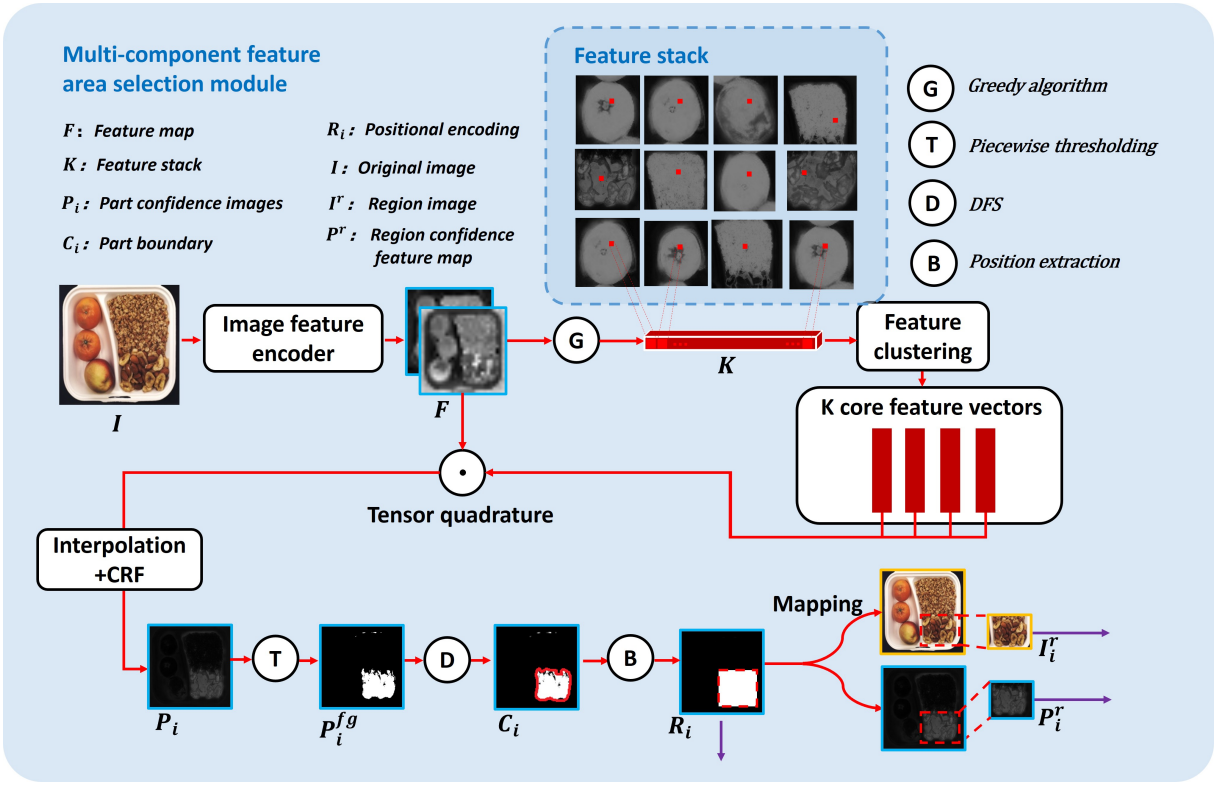

Figure 3. Schematic diagram of multi-component feature region selection.

When initiating multi-component feature region selection, the training image is first input into an image feature encoder to transform it into a feature map. However, a critical step in processing the feature map is selecting representative points. Feature maps typically contain a large amount of information, and directly processing all points not only significantly increases computational load but is also susceptible to interference from redundant information, thereby affecting the model's efficiency and accuracy. Therefore, it is necessary to select the most representative key points to optimize the feature map processing. Specifically, this paper employs a greedy sampling algorithm [30] to select the N most representative key points from the feature map. Compared to random sampling [33], uniform sampling [34], or the K-means clustering algorithm [16], this greedy sampling algorithm offers higher efficiency, particularly when the feature map is large and computational resources are limited, as it maximizes the retention of key information in the feature map with a limited number of representative points.

Specifically, during the training phase, given the original training image I, we generate the corresponding feature map F using a pre-trained image feature encoder, where F is a three-dimensional feature map represented as $F \in \mathbb{R}^{H\times W\times C}$, with H and W being the height and width of the feature map, respectively, and C being the number of channels. The feature vector $f_{i,j} \in \mathbb{R}^C$ at each position (i, j) can be evaluated for its expressive power by calculating its L_2-norm:

$$G(i,j) = \|f_{i,j}\|_2 = \sqrt{\sum_{k=1}^{C} f_{i,j,k}^2}, \tag{1}$$

where $f_{i,j,k}$ represents the value of F at position (i, j) and channel k. The greedy sampling algorithm [30] aims to iteratively select representative points that optimally summarize the key patterns and important regions in the feature map. A set S is defined to store the selected representative points. Initially, S is empty. In each iteration, a point (i, j) is selected

from the remaining candidate points and added to set S, maximizing the expressive power gain of the component representative points. The gain is calculated as:

$$\triangle G(S,(i,j)) = \sum_{(p,q)\in S} \|f_{p,q} - f_{i,j}\|_2 \tag{2}$$

The position (i,j) that maximizes the gain in Equation (2) is selected and added to the set S. This process iterates until the number of points in set S reaches the predetermined N representative points. The final set S contains points that consider both the component's expressive power and focus on local feature variations, thereby effectively capturing the key feature information of the image. Next, the representative point features of all training images are stacked to generate the feature stack matrix K. Each image's representative point set S corresponds to a feature subset of that image, and stacking them yields K, which is the collection of representative point features from all training images. Subsequently, a clustering algorithm (e.g., K-means) is applied to the feature stack K for cluster analysis, dividing these features into multiple categories to obtain the multi-kernel feature vector V, where the k-th component V_k represents the k-th component in the image.

The feature map F is then tensor-multiplied with the multi-kernel feature vector V to generate region response maps related to the feature vector components. These region response maps are interpolated to restore their resolution to the original image size. Additionally, a conditional random field (CRF) is used to further optimize the results, enhancing the boundary accuracy and consistency of local regions. Through these operations, we obtain M component confidence images $P_i (1 \leq i \leq M)$, where i represents the component index and M is the total number of components.

For different components, key local regions are extracted, and corresponding region images I^r and region confidence images P^r are generated, where r represents the region image index. To more finely distinguish different information levels in the image and retain more details, thereby improving the recognition accuracy of anomaly regions and the granularity of subsequent analysis, we employ a piecewise thresholding operation. Unlike single-threshold segmentation methods, piecewise thresholding divides the component confidence image into multiple confidence levels, enabling more precise differentiation between high-confidence, low-confidence, and medium-confidence regions. This approach allows independent processing of regions with different confidence levels, facilitating better identification of potential anomaly regions. The specific steps are as follows:

Operation is performed. Specifically, the confidence value $P(x,y)$ at each pixel position (x,y) in the image is divided into multiple regions according to multiple thresholds T_1, T_2, and T_3, corresponding to different confidence levels. The formula is defined as:

$$P^{fg}(x,y) = \begin{cases} a, & \text{if } P(x,y) \geq T_3 \\ b, & \text{if } T_2 \leq P(x,y) < T_3 \\ c, & \text{if } T_1 \leq P(x,y) < T_2 \\ d, & \text{if } P(x,y) < T_1 \end{cases} \tag{3}$$

In this process, the foreground region $P^{fg}(x,y)$ represents pixel positions with higher confidence levels. After piecewise thresholding, the foreground region is divided into different confidence levels based on the value of $P(x,y)$. Through this piecewise thresholding operation, we can more accurately partition the image into regions of different confidence levels, providing a clear foundation for subsequent anomaly analysis. After completing the partitioning of foreground and background, the next step is to extract the positional encoding R_i of the components, which is used to clarify the relationship between the region

image I^r and its corresponding region confidence feature image P^r. The model extracts the structural information of the foreground region and combines it with a depth-first algorithm to perform a contour search on the foreground $P^{fg}(x, y)$, identifying the parts within the contour as component regions. These component regions are then converted into positional encoding R_i. The specific approach is as follows: For a position (x, y), its four-neighborhood $N(x, y)$ is, If $(x, y) \in P^{fg}$ and has not been visited, a depth-first search is initiated from this point, marking all connected foreground pixels as the same region A_i. For each connected region A_j, its boundary $C_i(A_j)$ is defined as: The region growing algorithm operates as follows: For any given position (x, y), we first define its four-neighborhood $N(x, y)$. When encountering an unvisited foreground pixel $(x, y) \in P^{fg}$, the algorithm initiates a depth-first search to aggregate all connected foreground pixels into a unified region A_i. Subsequently, for each connected region A_j, its boundary $C_i(A_j)$ is determined. The formal mathematical definitions are:

$$N(x, y) = \left\{ (x', y') \mid |x' - x| + |y' - y| = 1 \text{ and } P^{fg}(x', y') = 1 \right\} \tag{4}$$

$$C_i(A_j) = \left\{ (x, y) \in A_j \mid \exists (x', y') \notin A_j, |x' - x| + |y' - y| = 1 \right\} \tag{5}$$

That is, if a position (x, y) belongs to region A_j and at least one of its neighboring positions does not belong to A_j, then this position is considered a boundary point. The boundary set of all regions is $C_i = \{C_i(A_1), C_i(A_2), \ldots, C_i(A_j)\}$. The minimum and maximum values of all coordinates in $C_i(A_j)$ in the horizontal and vertical directions are calculated as:

$$\begin{aligned} x_{\min} &= \min_k x_k, x_{\max} = \max_k x_k, \\ y_{\min} &= \min_k y_k, y_{\max} = \max_k y_k, \end{aligned} \tag{6}$$

where $x_{\min}$, $y_{\min}$, $x_{\max}$, and $y_{\max}$ are the boundary points of the region, representing its range. For each region set, the positional encoding R_i is determined by the four key coordinate points that define the region set:

$$R_i = (x_{\min}, y_{\min}, x_{\max}, y_{\max}) \tag{7}$$

Based on the positional encoding, the position data of the components in the confidence map are extracted. Subsequently, using this positional information as a reference, it is mapped back to the original image and the confidence map. The region corresponding to the foreground region is cropped from the original image as the region image I^r. Simultaneously, the corresponding foreground part is extracted from the confidence map as the region confidence feature image P^r. These two types of features serve as inputs for the region feature enhancement in the next subsection.

3.3. Feature Enhancement

After completing the multi-component feature region selection, we obtain the region images, region confidence features, and corresponding positional information for each component. To enhance the model's ability to focus on component-level features, particularly in detecting minor defects or local anomalies, we need to extract more detailed information from local features. To this end, we introduce a feature-enhancement module, as shown in Figure 4. The goal of the feature-enhancement module is to achieve feature enhancement by comprehensively utilizing self-attention mechanisms, feature fusion strategies, and similarity scores between features based on the output of the multi-component feature region selection module. Specifically, the component images obtained from the multi-component feature region selection module are input into a pre-trained image feature

encoder to generate feature maps. Subsequently, to enable the model to better focus on key local features, the feature-enhancement module introduces a self-attention mechanism, which automatically adjusts the model's attention to local information through feature weighting. This mechanism enhances the model's ability to express key information regions in the image while suppressing irrelevant information in the background. Then, the feature-enhancement module divides the feature maps into multiple groups, with each group serving as a candidate feature image. Finally, by calculating the similarity scores between the region confidence images and the candidate feature images, the best-matching feature map is selected and multiplied with the positional encoding from Section 3.1, thereby enhancing the representation of anomaly regions.

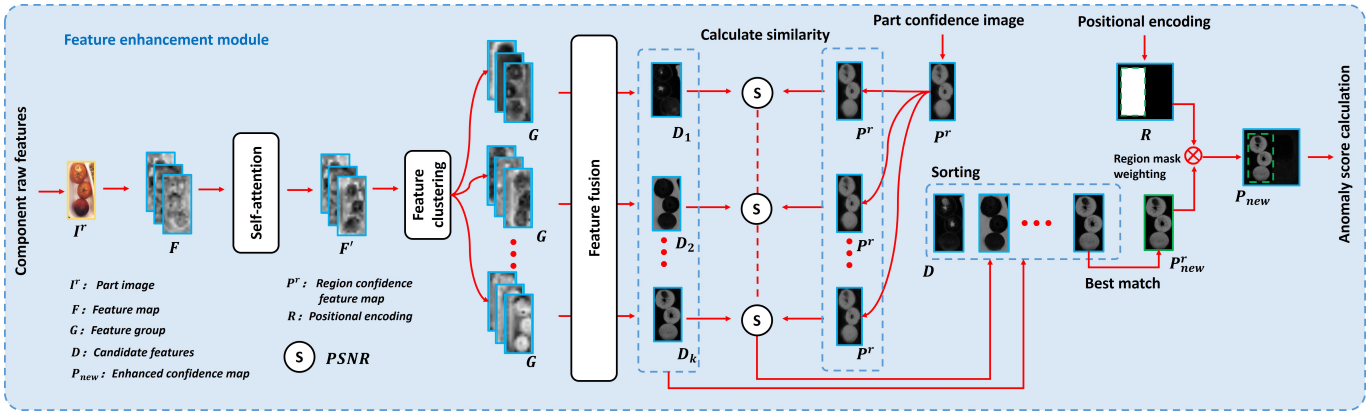

Figure 4. Schematic diagram of feature enhancement, where the region images I^r, region confidence images P^r, and positional encoding R are all derived from the multi-component feature region selection module.

It is important to note that the feature-enhancement module, through refined similarity analysis and efficient region alignment, enables the model to automatically focus on regions exhibiting high consistency in the local feature space, thereby significantly improving sensitivity to minor defects and local anomalies. From a mathematical perspective, the above process can be described as follows: Let the feature-enhancement module be $\mathcal{E}(P^r, I^r, R)$, where P^r is the region confidence image, I^r is the region image, R is the positional encoding, $\Phi(\cdot)$ is the structural similarity score calculation function, and $\delta(I^r)$ is the candidate feature image-generation function. The feature-enhancement process can then be expressed as:

$$\mathcal{E}(P^r, I^r, R) = \arg\max_{\delta(I^r)} \Phi(P^r, \delta(I^r)) \odot R \tag{8}$$

The following is a detailed explanation of the specific workflow of the feature-enhancement process: The region image I^r is input into a pre-trained image feature encoder to generate the feature map $F \in \mathbb{R}^{H\times W\times C}$. Each position in the feature map F contains certain local information. To enhance the model's focus on key local features, inspired by the method of Tongkun Liu et al. [19], this paper employs a self-attention mechanism in the feature encoder. This mechanism calculates dependencies between channels to determine the correlation between each channel and others, adjusting the attention of different channels through weighting operations. Specifically, the input feature map F is first linearly transformed to obtain the query (Q), key (K), and value (V) matrices:

$$Q = W_q F, K = W_k F, V = W_v F \tag{9}$$

The attention weights $A \in \mathbb{R}^{C\times C}$ are computed using trainable projection matrices W_q, W_k, and W_v, where the scaled dot-product operation applies a softmax normaliza-

tion $\sigma(z_i) = e^{z_i}/\sum_{j=1}^{N} e^{z_j}$ that converts raw similarity scores $QK^T/\sqrt{d}$ into probability distributions, yielding the final formulation:

$$A = \sigma\left(\frac{QK^T}{\sqrt{d}}\right) = \frac{\exp\left(QK^T/\sqrt{d}\right)}{\sum_{j=1}^{N}\exp\left(QK_j^T/\sqrt{d}\right)} \tag{10}$$

Next, the feature map $f' = AV$ is weighted according to the attention relationships between channels, capturing and enhancing the similarity and dependencies between channels. To further enhance the model's focus on key local features, clustering is introduced to help the model concentrate on the local features represented by each channel group. Specifically, cosine similarity is used to calculate the correlation between channels in the weighted feature map f'. Let $\widetilde{f'} = flatten(f')$ be the flattened feature of the channels. The similarity matrix $M \in \mathbb{R}^{C\times C}$ between channels in the weighted feature map can be calculated as:

$$M(i,j) = \cos\left(\widetilde{f}_i', \widetilde{f}_j'\right), \tag{11}$$

where $\widetilde{f}_i'$ and $\widetilde{f}_j'$ are the feature vectors of the i-th and j-th channels in $\widetilde{f'}$. Using the calculated similarity matrix M, the channels are clustered. Based on the clustering results, the channels are divided into k groups $\{G_i\}_{i=1}^{k}$. Each group G_i contains N channels and is treated as a feature group. In order to obtain the set of candidate feature images $\{D_i\}_{i=1}^{k}$, each candidate feature image D_i is generated by averaging all channels within its corresponding group G_i:

$$G_i = \{g_1, g_2, \ldots, g_N\}, \quad g_j \in \mathbb{R}^N \tag{12}$$

$$D_i = \delta(I^r) = \frac{1}{N}\sum_{j=1}^{N} g_j \tag{13}$$

In this way, the candidate feature image D_i effectively represents the deep structural information of the current component, providing reliable input for subsequent feature matching, particularly aiding in the detection of minor defects or local anomalies. The region confidence image P^r is mapped to each candidate feature image D_i to ensure alignment in the same space. Based on the mapping results, the structural similarity between the two is calculated. The structural similarity is defined as follows:

For the region confidence image P^r and the candidate feature image D_i, we define a similarity metric function $\Phi(P^r, D_i)$. $\Phi(\cdot)$ can be implemented using candidate metric functions such as MSE, COS, SSIM, and PSNR. In this paper, PSNR is selected as the similarity metric function, specifically defined as:

$$\Phi(P^r, D_i) = PSNR(P^r, D_i) = 10 \times \log_{10}\left(\frac{I_{\max}^2}{MSE(P^r, D_i)}\right) \tag{14}$$

$$MSE(P^r, D_i) = \frac{1}{H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(P^r(h,w) - D_i(h,w)\right)^2 \tag{15}$$

where $I_{max} = 255$ represents the maximum pixel value of the confidence image in our implementation. $H \times W$ is the size of the region confidence image P^r and the candidate feature image D_i. For the k candidate feature images $\{D_i\}_{i=1}^{k}$, we obtain k corresponding structural similarity scores $\{S_i^r\}_{i=1}^{k}$. Specifically, we calculate structural similarity score S_i^r between the region confidence image P^r and the candidate feature image D_i. The structural similarity score S_i^r is calculated as:

$$S_i^r = \Phi(P^r, D_i) \tag{16}$$

All candidate feature images are sorted in descending order based on their similarity scores, where:

$$S_1^r \geq S_2^r \geq \ldots \geq S_k^r \tag{17}$$

The candidate feature image D_{best} with the highest similarity is selected as the best match for the current component, $P_{new}^r = D_{best}$. The best match confidence image P_{new} is then multiplied with the positional encoding R using element-wise multiplication:

$$P_{new} = P_{new}^r \odot R \tag{18}$$

P_{new} is the enhanced confidence map, which retains the information of the best match confidence image while enhancing the representation of anomaly patterns through similarity optimization.

4. Experiments

This section verifies the effectiveness of the proposed component-level feature-enhancement method in industrial visual inspection tasks, evaluating its precision and robustness in fine-grained anomaly detection.

4.1. Results of Anomaly Detection

In this experiment, we explore the performance of the proposed anomaly-detection method, including tasks for logical anomaly and structural anomaly-detection. Additionally, we investigate the impact of the two modules of the proposed method on detection performance, namely the component-level feature enhancement and the piecewise thresholding module. In our implementation, the multi-kernel feature vector V uses $K = 4$ kernels, while the feature groups G are set to $k = 3$ groups.

Table 1 summarizes the experimental results of all methods on the MVTec LOCO AD dataset. From the results in the table, it can be seen that the S-T method significantly outperforms previous methods such as MNAD, SPADE, and DRAEM in terms of the overall score (i.e., the average AUROC values for logical defect and structural defect detection), with a score 3.75% higher than DRAEM. The latest ComAD method further optimizes the S-T method, improving the score by 2.39%. However, our method outperforms all other comparison methods in both logical defect and structural defect detection, achieving an overall score of 81.73%, which is 1.99 percentage points higher than ComAD. In logical defect detection, the ComAD method ranks second with a detection score of 86.38%, demonstrating its excellent detection performance. In contrast, MNAD, SPADE, and DRAEM achieve scores of 60.00%, 70.90%, and 72.80%, respectively, significantly lower than ComAD. The proposed method, by integrating component-level feature enhancement and multi-layer piecewise thresholding, achieves a detection score of 87.92%, which is 1.54 percentage points higher than ComAD, significantly surpassing other comparison methods and showcasing its superior performance in logical defect detection. In structural defect detection tasks, the proposed method demonstrates significant performance advantages over traditional methods such as MNAD and SPADE. Specifically, SPADE achieves a score of 66.80% in structural defect detection, while ComAD leads with a score of 73.10%, an improvement of 6.30 percentage points. In contrast, the proposed method performs even better in structural defect detection, achieving a score of 75.53%, which is 2.43 percentage points higher than ComAD. Considering both logical defect and structural defect detection capabilities, the proposed method achieves the highest overall score of 81.73%, leading the second-best method by nearly 2%.

Table 1. Comparison of results of different anomaly-detection methods on the MVTec LOCO AD dataset.

Method	Logical	Structured	Overall Score
MNAD	60.00	70.20	65.10
SPADE	70.90	66.80	68.85
DRAEM	72.80	74.40	73.60
S-T	66.40	88.30	77.35
ComAD	86.38	73.10	79.74
MPFnet	73.90	84.80	79.40
Ours	87.92	75.53	81.73

Our experimental evaluation covers two distinct industrial inspection benchmarks: MVTec LOCO AD for logical and structural anomalies and MVTec AD for object-level defect detection. The results in Table 2 demonstrate consistent performance improvements across both datasets. On MVTec LOCO AD, our method achieves 87.92% AUROC for logical anomaly detection and 75.53% AUROC for structural defects. For MVTec AD's object categories, we obtain 74.94% AUROC. These results represent an average improvement of 2.07 percentage points over ComAD, confirming the effectiveness of our multi-scale fusion approach across diverse industrial inspection scenarios.

Table 2. Performance comparison on industrial inspection benchmarks (AUROC%).

Datasets	ComAD	Ours
MVTec LOCO AD Logical	86.38	87.92
MVTec LOCO AD Structural	73.10	75.53
MVTec AD Object	72.70	74.94
Average	77.39	79.46

4.2. Validation of the Effectiveness of Component-Level Feature Enhancement

To further validate the effectiveness of the proposed method, this paper conducts ablation experiments to evaluate the performance of the component-level feature-enhancement module in logical defect and structural defect detection tasks. Specifically, the paper compares the effects of including the component-level feature-enhancement module versus removing it, analyzing its impact on overall performance. Table 3 presents a comparative analysis of the experimental results for component-level feature enhancement. Using the ComAD method as a baseline, the performance differences of the proposed module under various test scenarios are analyzed. It can be observed that in logical defect detection, after incorporating the component-level feature-enhancement module, the model can more accurately capture the details of logical defects, especially in samples with complex logical anomalies (e.g., connectors), where the improvement is significant (increasing from the baseline of 84.51% to 92.30%). This demonstrates its advantage in handling complex anomaly patterns.

In structural defect detection, the addition of the component-level feature-enhancement module shows relatively stable advantages across most samples. To more intuitively demonstrate the effectiveness of the component-level feature-enhancement method, we provide visual results of its effects, as shown in Figure 5. The left side displays the original images from the MVTec LOCO AD dataset, the middle shows the component confidence images before enhancement, and the right side shows the component confidence images after enhancement. Taking the pushpin sample in the second row as an example, the confidence map generated by the baseline method suffers from edge blurring, and the grayscale values of non-component regions around the target area are significantly higher, indicating

limitations in component feature representation. In contrast, the proposed component-level feature-enhancement method can more precisely capture component-level structural anomalies, and the generated component confidence images exhibit clearer local feature representation. According to the data in Table 3, the component-level feature-enhancement method achieves a structural defect detection score of 98.05% on the pushpin dataset, a significant improvement of 4.97 percentage points over the baseline method (ComAD) at 93.08%, demonstrating stronger local anomaly-detection capabilities. Although in some samples, such as the Juice Bottle dataset, the proposed method achieves a structural defect detection score of 69.22%, slightly lower than the baseline method (ComAD) at 77.07%, overall, the proposed method shows clear advantages in structural defect detection.

Table 3. Experimental results of component-level feature enhancement (CLE).

Method		Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors	Total
Baseline	Logical	94.53	84.40	88.52	79.97	84.51	86.38
	Structured	69.95	77.07	93.08	61.46	63.92	73.10
CLE	Logical	93.68	82.73	87.18	79.57	92.30	87.09
	Structured	65.82	69.22	98.05	63.25	77.79	74.83

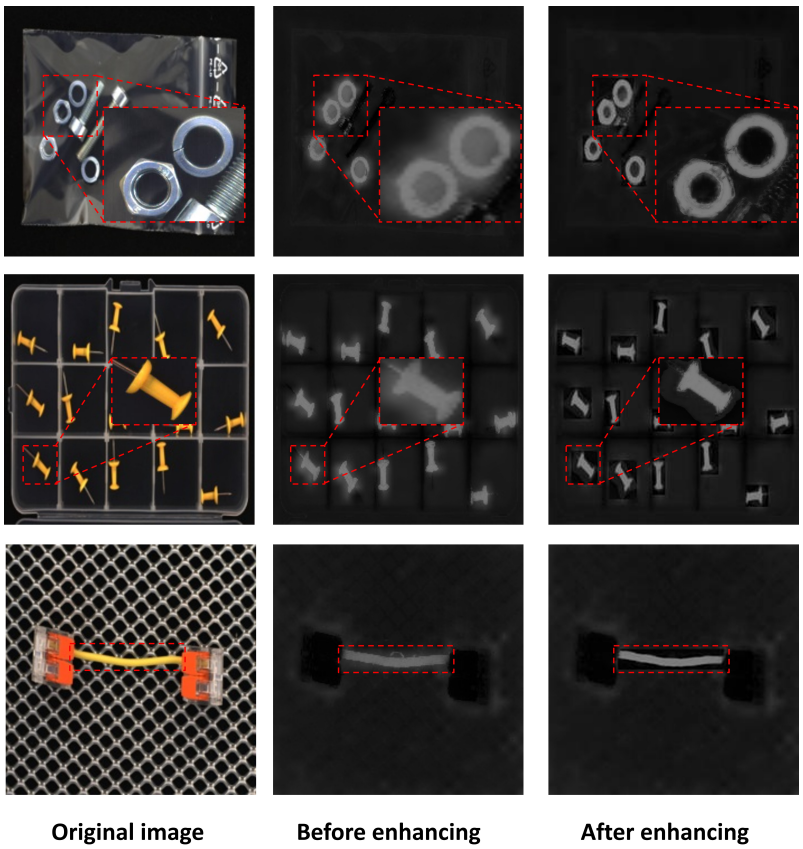

Figure 5. Demonstration of the effectiveness of local feature enhancement. The left side shows the original images from the MVTec LOCO AD dataset, the middle shows the component confidence images before enhancement, and the right side shows the component confidence images after enhancement. The red dashed boxes indicate magnified views of specific components.

Additionally, we conducted a visual analysis of the tendency of unsupervised anomaly-detection tasks to overlook minor defects. As shown in Figure 6, taking the sample in the second row as an example, a pill is mixed into the nuts. After optimizing the feature map using the component-level feature-enhancement method, the experimental results show

significant differences. Before enhancement, the differences between normal components and defects in the image are small, and the feature distinction between the pill defect region and the normal component region is low, causing the detection algorithm to fail to identify the defect location accurately. The defect region is not annotated in the detection results, leading to detection errors. In contrast, in the enhanced feature map, the defect region is clearly delineated, and the pill location appears as a distinct gap in the confidence map. Through comparison with normal images, this gap is accurately identified as an anomaly region, effectively avoiding misjudgment of normal regions and significantly improving detection accuracy and reliability. This indicates that the proposed method can effectively enhance the detection capability for fine-grained defects.

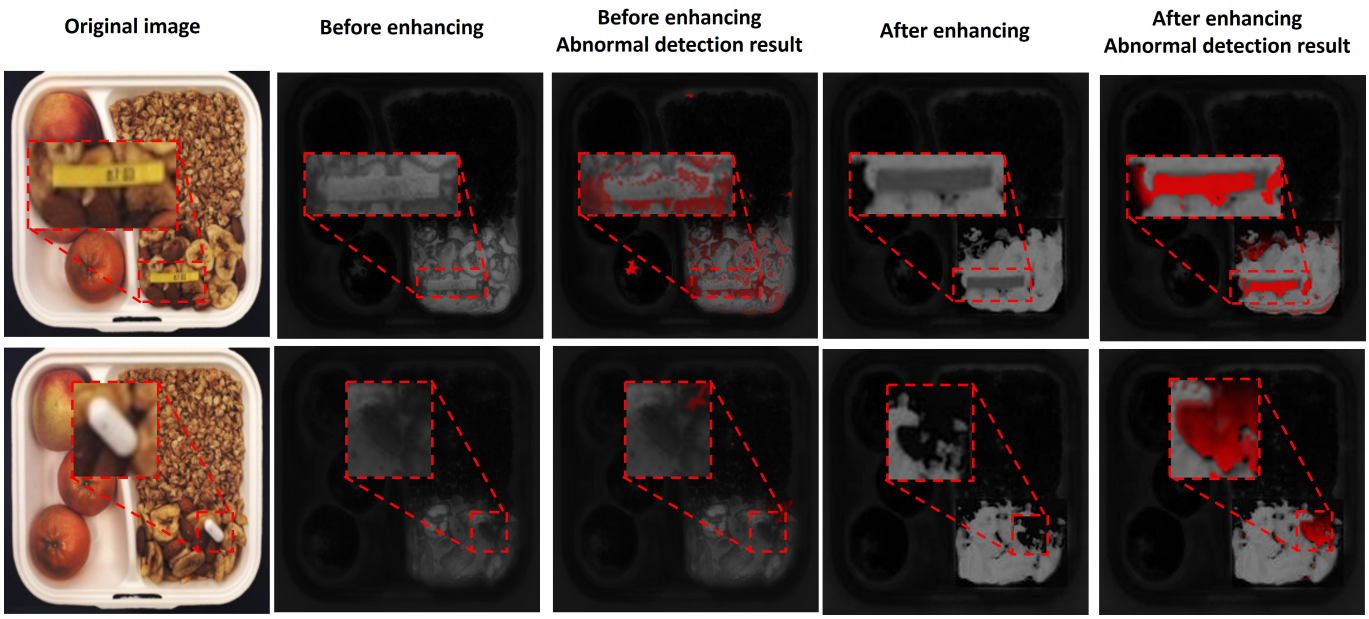

Figure 6. Demonstration of the effectiveness of component-level feature enhancement in detecting minor defects. The left side shows the original image, the middle shows the confidence map and detection results before enhancement, and the right side shows the confidence map and detection results after enhancement. The red dashed boxes indicate magnified views of abnormal part.

4.3. Validation of the Effectiveness of Multi-Layer Piecewise Thresholding

The multi-layer piecewise thresholding strategy proposed in this paper divides the confidence map into three regions: foreground, background, and transition region, and selects the transition region along with the foreground as input to the component-level feature encoder. This dual-region collaborative input approach effectively enhances the accuracy of component-level feature encoding, avoiding the potential issue of structural incompleteness when using only the foreground. Unlike previous methods that solely rely on the foreground region, the proposed method significantly improves region-segmentation accuracy by introducing the transition region. As shown in Figure 7, the left side displays the confidence map of a pushpin sample, while the right side shows an enlarged image of a single pushpin, where different colors represent different region divisions: red for the foreground, blue for the background, and green for the transition region. The first row on the right shows the original image and its corresponding segmentation diagram, while the second row shows the confidence image and its corresponding segmentation diagram. Through this color-coding method, different regions in the image (foreground, background, and transition region) are clearly distinguished, thereby reducing the risk of misjudgment.

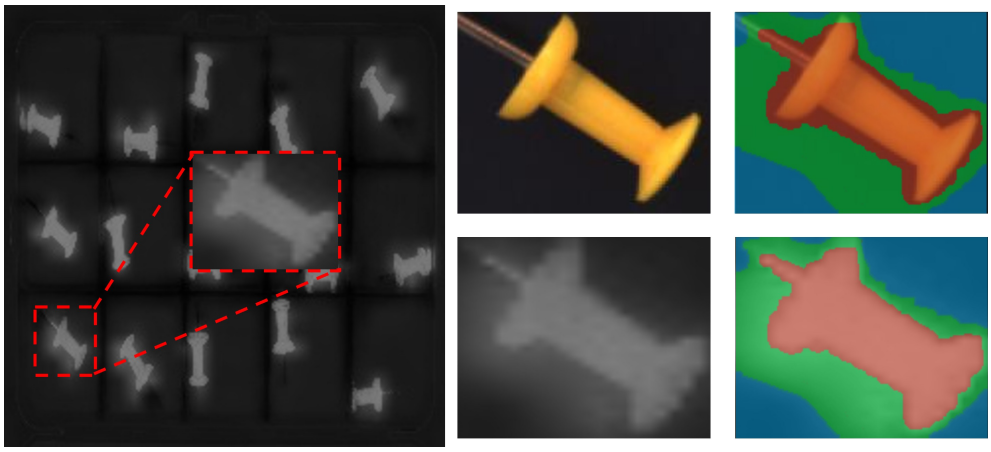

Figure 7. Effect of applying multi-layer piecewise thresholding to a pushpin sample. The red dashed boxes highlight magnified views of individual pushpins.

To further validate the role of the transition region, as shown in Figure 8, this paper uses a breakfast box sample to demonstrate the segmentation results before and after introducing the transition region. Before introducing the transition region, the model mistakenly identified the cereal region in the upper right corner as part of the target component when extracting the "fruit" component. After introducing the transition region, the model further segmented the confidence map, effectively excluding non-target component regions and ensuring the correct division of the target component. This allows for more precise localization and identification of defect regions in subsequent defect detection processes, thereby improving detection accuracy.

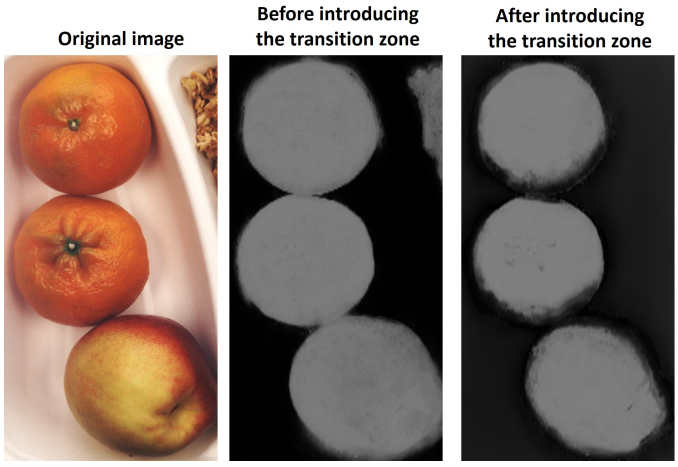

Figure 8. Comparison of component segmentation results before and after introducing the transition region.

4.4. Validation of the Effectiveness of Multi-Layer Piecewise Thresholding

To further validate the effectiveness of the proposed multi-layer piecewise thresholding method, two commonly used region-segmentation methods in the field of image segmentation—K-means and Otsu (maximum inter-class variance method)—are introduced for quantitative comparison. Table 4 summarizes the AUROC (Area Under the Receiver Operating Characteristic Curve) comparison results of these region-segmentation methods in logical anomaly and structural anomaly detection. As shown in Table 4, the multi-layer piecewise thresholding method outperforms both K-means and Otsu in both logical anomaly and structural anomaly detection.

Table 4. Comparison of segmentation accuracy (AUROC) across different methods (All values in %).

Methods	Logical	Structured
K-means	86.47	72.49
Otsu	87.58	72.27
Multi-threshold	87.92	75.53

To further verify the applicability and performance of the method across different defects, Table 5 presents the region-segmentation accuracy evaluated by the AUROC metric between the multi-layer piecewise thresholding method compared to other methods. The data in Table 5 show that the multi-layer piecewise thresholding method significantly improves segmentation accuracy, especially in structural anomaly detection, further demonstrating its robustness and effectiveness in different environments.

Table 5. Comparison of segmentation accuracy (AUROC) across different categories (All values in %).

Methods		Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors	Total
Logical	K-means	92.79	77.84	88.17	86.22	87.32	86.47
	Otsu	91.06	84.94	89.54	86.26	86.08	87.58
	Multi-threshold	92.85	84.48	87.21	86.70	88.34	87.92
Structured	K-means	69.51	62.98	96.28	65.10	68.57	72.49
	Otsu	62.78	73.65	94.28	62.80	67.82	72.27
	Multi-threshold	65.26	78.45	95.74	64.92	73.25	75.53

4.5. Efficiency Analysis of Similarity Calculation Methods

In the feature-enhancement module, the similarity calculation between the confidence image and the candidate feature image is a critical step. To evaluate the efficiency of different similarity measurement methods, we compared and analyzed four commonly used similarity calculation methods: MSE (Mean Squared Error), SSIM (Structural Similarity Index), COS (Cosine Similarity), and PSNR (Peak Signal-to-Noise Ratio). The experimental summary is as follows:

According to the data in Table 6, the AUROC values represent the average performance in logical anomaly and structural anomaly-detection tasks. The AUROC in the table reflects the overall performance of the model under different segmentation methods in logical anomaly and structural anomaly detection. PSNR consistently performs best in logical anomaly detection and achieves near or the highest scores in structural anomaly detection, demonstrating its stable performance.

Table 6. Performance comparison of detection accuracy at logical and structural levels (AUROC %).

Methods		Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors	Total
Logical	MSE	86.03	74.33	72.84	57.89	60.69	72.65
	COS	95.54	81.53	80.88	84.31	69.43	84.66
	SSIM	94.76	81.53	80.58	84.23	72.07	83.94
	PSNR	95.12	81.65	89.60	84.29	69.79	86.54
Structured	MSE	66.97	60.97	61.49	60.69	54.92	61.01
	COS	71.25	67.09	87.31	69.43	67.53	72.52
	SSIM	68.83	67.09	84.27	72.07	66.80	71.81
	PSNR	70.38	67.09	85.89	69.79	68.33	72.30

Across different defects, PSNR exhibits stable performance, indicating its ability to effectively meet the detection requirements of various scenarios. In addition to accuracy, we also compared the computational efficiency of different methods, primarily measured by FPS (Frames Per Second). The FPS data only include the time required for similarity calculation, so the FPS differences between methods are significant, mainly due to the varying complexity of similarity calculations. As summarized earlier, the AUROC in the table reflects the overall performance of the model under different segmentation methods in anomaly detection. Table 7 shows the comparison of efficiency (FPS) and accuracy (AUROC).

Table 7. Comparison of efficiency (FPS) and accuracy (AUROC%).

Criteria	MSE	COS	SSIM	PSNR
FPS	1901.93	57.19	24.81	448.25
AUROC	66.83	78.59	77.88	79.42

Based on the experimental data, PSNR achieves a slight improvement in AUROC (approximately 1.67%) while maintaining a relatively high FPS compared to COS and SSIM methods. PSNR not only provides high accuracy but is also suitable for scenarios requiring both precision and performance, demonstrating balanced overall performance. Therefore, this paper adopts PSNR as the default method for calculating the similarity between the confidence image and the candidate feature image.

4.6. Ablation Experiments

To further evaluate the effectiveness of the method, we conducted an in-depth analysis of the specific contributions of each module. In this ablation experiment, the ComAD method is used as the baseline to evaluate the performance of the component-level feature-enhancement method and its combination with the multi-layer piecewise thresholding method.

According to the experimental results in Table 8, the proposed method achieves a stable overall score of 87.09% in logical defect detection when only the component-level feature-enhancement module is retained. In structural defect detection, the overall score is 74.83%, an improvement of 1.73 percentage points over the baseline method. The component-level feature-enhancement module focuses more on detecting minor defects and local anomalies, demonstrating clear advantages in identifying fine-grained defects, especially on the “Splicing Connectors” and “Pushpins” datasets. Additionally, when the multi-layer piecewise thresholding method is introduced on top of the component-level feature-enhancement module, the logical defect detection score increases to 87.92%, and the structural defect detection score increases to 75.53%, an improvement of 0.7 percentage points compared to using the component-level feature-enhancement method alone. This may be because the transition region introduced by the multi-layer piecewise thresholding method enhances the accuracy of component-level feature encoding, further optimizing region segmentation and anomaly region identification in structural defect detection. Our method achieves strong logical defect detection but shows varied structural detection performance across categories, excelling with Pushpins while under-performing on Breakfast Box and Screw Bag. We acknowledge that while the component-level feature-enhancement method demonstrates strong overall performance, there remains room for improvement in handling highly diverse anomalies. Specifically, the structured anomaly-detection scores of 65.26% AUROC for the Breakfast Box and 64.92% AUROC for the Screw Bag indicate potential limitations when dealing with particularly complex defect patterns.

Table 8. Comparison of anomaly-detection results between component-level feature enhancement (CLE) and multi-layer piecewise thresholding (MLPT) methods and the baseline method (All values in %).

Methods		Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors	Total
ComAD	Logical	94.53	84.40	88.52	79.97	84.51	86.38
	Structured	69.95	77.07	93.08	61.46	63.92	73.10
CLE	Logical	93.68	82.73	87.18	79.57	92.30	87.09
	Structured	65.82	69.22	98.05	63.25	77.79	74.83
CLE + MLPT	Logical	92.85	84.48	87.21	86.70	88.34	87.92
	Structured	65.26	78.45	95.74	64.92	73.25	75.53

5. Conclusions

This paper proposes an industrial anomaly-detection method based on component-level feature enhancement, achieving efficient detection of fine-grained anomalies through the selection and optimization of multi-component feature regions. Experimental results demonstrate that the proposed method outperforms existing mainstream methods in both logical defect and structural defect detection tasks, significantly improving detection accuracy and robustness. Specifically, the method can more precisely capture semantic logical consistency between components and component-level anomaly features, demonstrating its applicability to multi-component samples in complex industrial scenarios and its ability to effectively avoid background interference.

To further enhance the adaptability and robustness of industrial anomaly detection, future research will focus on developing self-adaptive feature evaluation frameworks capable of dynamically adjusting to diverse defect characteristics. This includes: (1) Developing lightweight defect detection architectures through neural network pruning techniques [11], where we will adapt progressive filter pruning methods to preserve critical defect features while improving inference speed; (2) few-shot anomaly-detection protocols to improve recognition of rare defects by learning from limited examples; and (3) context-aware metric optimization to ensure that different defect types are evaluated using the most appropriate criteria. Additionally, we will explore cross-domain generalization techniques to enhance model performance in varying industrial environments, as well as automated threshold adaptation to reduce manual calibration efforts. These advancements aim to bridge the gap between controlled experimental settings and real-world deployment, where defects exhibit high variability in appearance and frequency.

Author Contributions: Conceptualization, F.Y.; methodology, X.W.; validation, Z.Z.; formal analysis, J.W. and J.F.; investigation, N.Z.; resources, J.L.; data curation, H.Z.; writing—original draft preparation, Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by Natural ScienceFoundation of Xiamen (3502Z202473071, 3502Z20227073). National Natural Science Foundation of Fujian Province (Grant Nos. 2023J011428, 2022J011236, 2023J011426, 2022Y0077), Industry-University-Research Collaborative Innovation Project of Fujian Province (Grant No. 2024H6035).

Data Availability Statement: The MVTec AD dataset and the MVTec Logical Constraints Anomaly-Detection dataset (MVTec LOCO AD) can be obtained from https://www.mvtec.com/company/research/datasets/mvtec-ad (accessed on 5 March 2025) and https://www.mvtec.com/company/research/datasets/mvtec-loco (accessed on 5 March 2025).

Conflicts of Interest: The author Hangqi Zhang was employed by the company Xiamen Yaxon Zhilian Technology Co., Ltd. The author Nianfeng Zeng was employed by the company E-Success Information Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Rudolph, M.; Wehrbein, T.; Rosenhahn, B.; Wandt, B. Asymmetric student-teacher networks for industrial anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 2592–2602.
2. Wang, Y.; Peng, J.; Zhang, J.; Yi, R.; Wang, Y.; Wang, C. Multimodal industrial anomaly detection via hybrid fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 8032–8041.
3. Choi, W.H.; Kim, J. Unsupervised learning approach for anomaly detection in industrial control systems. *Appl. Syst. Innov.* **2024**, *7*, 18. [CrossRef]
4. Yao, H.; Yu, W.; Luo, W.; Qiang, Z.; Luo, D.; Zhang, X. Learning global-local correspondence with semantic bottleneck for logical anomaly detection. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 3589–3605. [CrossRef]
5. Lu, J.; Chen, Z.; Deng, X. A graph convolutional neural network model based on fused multi-subgraph as input and fused feature information as output. *Eng. Appl. Artif. Intell.* **2025**, *139*, 109542. [CrossRef]
6. Gao, T.; Yang, J.; Wang, W.; Fan, X. A domain feature decoupling network for rotating machinery fault diagnosis under unseen operating conditions. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110449. [CrossRef]
7. Kent, M.; Huynh, N.K.; Schiavon, S.; Selkowitz, S. Using support vector machine to detect desk illuminance sensor blockage for closed-loop daylight harvesting. *Energy Build.* **2022**, *274*, 112443. [CrossRef]
8. Mnassri, B.; Ananou, B.; Ouladsine, M. Fault detection and diagnosis based on PCA and a new contribution plot. *IFAC Proc. Vol.* **2009**, *42*, 834–839. [CrossRef]
9. Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; Steger, C. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *Int. J. Comput. Vis.* **2022**, *130*, 947–969. [CrossRef]
10. Lis, K.; Nakka, K.; Fua, P.; Salzmann, M. Detecting the unexpected via image resynthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2152–2161.
11. Wang, X.; Zheng, Z.; He, Y.; Yan, F.; Zeng, Z.; Yang, Y. Progressive local filter pruning for image retrieval acceleration. *IEEE Trans. Multimed.* **2023**, *25*, 9597–9607. [CrossRef]
12. Wang, X.; Fan, J.; Yan, F.; Hu, H.; Zeng, Z.; Huang, H. Unsupervised fur anomaly detection with B-spline noise-guided Multi-directional Feature Aggregation. *Vis. Comput.* **2025**, 1–17.
13. Liu, T.; Li, B.; Zhao, Z.; Du, X.; Jiang, B.; Geng, L. Reconstruction from edge image combined with color and gradient difference for industrial surface anomaly detection. *arXiv* **2022**, arXiv:2210.14485.
14. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Langs, G.; Schmidt-Erfurth, U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* **2019**, *54*, 30–44. [CrossRef] [PubMed]
15. Park, H.; Noh, J.; Ham, B. Learning memory-guided normality for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 14372–14381.
16. Cohen, N.; Hoshen, Y. Sub-image anomaly detection with deep pyramid correspondences. *arXiv* **2020**, arXiv:2005.02357.
17. Zavrtanik, V.; Kristan, M.; Skočaj, D. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 8330–8339.
18. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 4183–4192.
19. Liu, T.; Li, B.; Du, X.; Jiang, B.; Jin, X.; Jin, L.; Zhao, Z. Component-aware anomaly detection framework for adjustable and logical industrial visual inspection. *Adv. Eng. Inform.* **2023**, *58*, 102161. [CrossRef]
20. Guo, H.; Ren, L.; Fu, J.; Wang, Y.; Zhang, Z.; Lan, C.; Wang, H.; Hou, X. Template-guided hierarchical feature restoration for anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 6447–6458.
21. Zhang, J.; Suganuma, M.; Okatani, T. Contextual affinity distillation for image anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 1–6 January 2024; pp. 149–158.

22. Zhuang, Z.; Ting, K.M.; Pang, G.; Song, S. Subgraph centralization: A necessary step for graph anomaly detection. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM), SIAM, Saint Paul, MN, USA, 27–29 April 2023; pp. 703–711.
23. Batzner, K.; Heckler, L.; König, R. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 1–6 January 2024; pp. 128–138.
24. Sugawara, S.; Imamura, R. PUAD: Frustratingly simple method for robust anomaly detection. In Proceedings of the 2024 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 27–30 October 2024; pp. 842–848.
25. Zavrtanik, V.; Kristan, M.; Skočaj, D. Dsr—A dual subspace re-projection network for surface anomaly detection. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 539–554.
26. Yao, H.; Luo, W.; Yu, W. Visual anomaly detection via dual-attention transformer and discriminative flow. *arXiv* **2023**, arXiv:2303.17882.
27. Martínez-Ferrer, L.; Jungbluth, A.; Gallego-Mejia, J.A.; Allen, M.; Dorr, F.; Kalaitzis, F.; Ramos-Pollán, R. Exploring Generalisability of Self-Distillation with No Labels for SAR-Based Vegetation Prediction. *arXiv* **2023**, arXiv:2310.02048.
28. Yu, J.; Kim, D.Y.; Lee, Y.; Jeon, M. Unsupervised pixel-level road defect detection via adversarial image-to-frequency transform. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 1708–1713.
29. Li, H.; Zhang, Z.; Chen, H.; Wu, L.; Li, B.; Liu, D.; Wang, M. A novel approach to industrial defect generation through blended latent diffusion model with online adaptation. *arXiv* **2024**, arXiv:2402.19330.
30. Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P. Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14318–14328.
31. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9592–9600.
32. Shao, H.; Peng, J.; Shao, M.; Liu, B. Multiscale Prototype Fusion Network for Industrial Product Surface Anomaly Detection and Localization. *IEEE Sens. J.* **2024**, *24*, 32707–32716. [CrossRef]
33. Ghayab, H.R.A.; Li, Y.; Abdulla, S.; Diykh, M.; Wan, X. Classification of epileptic EEG signals based on simple random sampling and sequential feature selection. *Brain Inform.* **2016**, *3*, 85–91. [CrossRef] [PubMed]
34. Acher, M.; Perrouin, G.; Cordy, M. BURST: Benchmarking uniform random sampling techniques. *Sci. Comput. Program.* **2023**, *226*, 102914. [CrossRef]