

Article

RD-RE: Reverse Distillation with Feature Reconstruction Enhancement for Industrial Anomaly Detection

Youjia Fu * and Antao Lin *

School of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China

* Correspondence: youjia_fu@cqut.edu.cn (Y.F.); linantap@stu.cqut.edu.cn (A.L.)

Abstract

Industrial anomaly detection methods based on reverse distillation (RD) have shown significant potential. However, existing RD approaches struggle to achieve an effective balance between constraining the feature consistency of the teacher–student networks and maintaining differentiated representation capability, which is crucial for precise anomaly detection. To address this challenge, we propose Reverse Distillation with Feature Reconstruction Enhancement (RD-RE) for Industrial Anomaly Detection. Firstly, we design a cross-stage feature fusion student network to integrate spatial detail information from the encoder with rich semantic information from the decoder. Secondly, we introduce a Locally Aware Dynamic Attention (LDA) module to enhance local detail feature response, thereby improving the model’s robustness in capturing anomalous regions. Finally, a Context-Aware Adaptive Multi-Scale Feature Fusion (CFFMS-FF) module is designed to constrain the consistency of local feature reconstruction. Experiments on the MVTec AD benchmark dataset demonstrate the effectiveness of RD-RE, achieving competitive results of 99.0%, 95.8%, 78.3%, and 99.7% on pixel-level AUROC, PRO, and AP and image-level AUROC metrics, and outperforming existing RD-based approaches. These results conclude that the integration of cross-stage fusion and local attention effectively mitigates the representation-consistency trade-off, providing a more robust solution for industrial anomaly localization.

Keywords: anomaly detection; cross-stage feature fusion; locally aware dynamic attention; adaptive multi-scale feature fusion



Academic Editors: Hussain
Mohammed Dipu Kabir and Subrota
Kumar Mondal

Received: 11 December 2025

Revised: 31 December 2025

Accepted: 31 December 2025

Published: 4 January 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Anomaly detection holds significant application value in fields such as intelligent manufacturing and medical image analysis [1,2], with its core tasks encompassing anomaly classification and region segmentation [3]. However, due to the complex and variable shapes of anomalous regions and the myriad types of anomalies, traditional classification and segmentation methods are often inadequate. Consequently, one-class learning methods trained solely on normal samples have garnered increasing attention. Among these, knowledge distillation-based approaches are particularly favored for their superior feature decoupling capabilities and real-time performance [4].

Knowledge distillation functions by constructing teacher–student networks, where the student network mimics the feature representations of the teacher network, utilizing the feature discrepancy between them to localize anomalies. Industrial scenarios demand models that possess high sensitivity to anomalous features while maintaining a lightweight architecture. In early distillation methods [4,5], the teacher and student networks shared

overly similar structures with highly consistent data flows, causing the student network to be prone to “over-generalization” and struggling to effectively capture anomalous features. To address this, Deng and Li proposed the Reverse Distillation (RD) strategy [6], which utilizes a pre-trained encoder as the teacher and a decoder as the student. By employing an asymmetric structure and reverse data flow, RD enhances sensitivity to anomalous features, significantly improving detection accuracy and robustness. Subsequently, Tien et al. [7] proposed RD++, which further bolsters the student network’s reconstruction capability by training a bottleneck layer and introducing synthetic anomalies into the teacher encoder.

Despite the promising performance of RD in anomaly detection tasks, the framework still faces two key challenges: (1) Insufficient filtering capability for fine-grained anomalous features (e.g., micro-cracks, pitting corrosion); and (2) Limited capability in reconstructing normal feature details. When dealing with complex texture details, the student network is prone to misinterpreting normal detailed features as anomalies, leading to a high false-positive rate and compromising overall detection precision.

To address the aforementioned issues, this paper proposes a method named Reverse Distillation with Reconstruction Enhancement (RD-RE), the overall framework of which is illustrated in Figure 1. Adopting the parameter-efficient ResNet18 [8] as the backbone, our method first introduces a cross-stage feature fusion mechanism in the student network to effectively mitigate information loss during feature reconstruction. Second, a Locally Aware Dynamic Attention (LDA) module is designed to enhance the capture of fine-grained features and effectively filter anomalous features. Furthermore, we propose a Context-Aware Adaptive Multi-Scale Feature Fusion (CAAMS-FF) module, which integrates Adaptive Sparse Self-Attention (ASSA) with multi-scale fusion strategies to further refine features and capture long-range dependencies. Finally, the generated multi-scale anomaly map is input into a segmentation sub-network combining Coordinated Attention (CA) [9] and Atrous Spatial Pyramid Pooling (ASPP) [10] to achieve more coherent and accurate anomaly localization. Extensive experiments demonstrate that the RD-RE method surpasses existing RD and mainstream Anomaly Detection (AD) methods on the MVTec AD benchmark dataset, achieving State-of-the-Art (SOTA) performance.

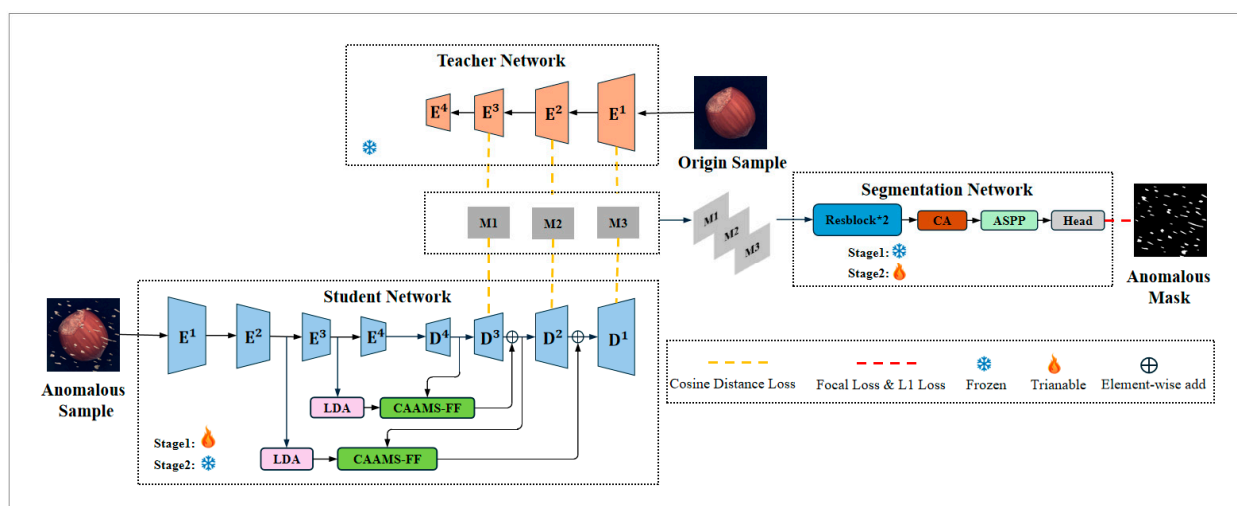


Figure 1. Overall Architecture of the RD-RE Framework.

The main contributions of this paper are summarized as follows:

- (1) A student network architecture with cross-stage feature fusion is proposed. By fusing intermediate encoder features into decoder features, this architecture significantly

improves the model's ability to filter anomalous features and the quality of reconstructing fine-grained normal features.

- (2) Novel LDA and CAAMS-FF modules are proposed. The LDA module utilizes sliding windows and a dynamic channel attention mechanism to refine original features and dynamically generate channel attention weights based on max-pooling and average-pooling references of local features, significantly enhancing sensitivity to micro-defects. The CAAMS-FF module introduces an adaptive sparse self-attention mechanism and multi-scale feature fusion strategies.
- (3) An efficient segmentation sub-network based on CA and ASPP is constructed. This network enhances positional sensitivity, expands the receptive field, and effectively eliminates background interference, making the final anomaly detection results more precise and coherent.

This work provides a new technical avenue for anomaly detection in fields such as industrial quality inspection.

2. Related Work

Industrial anomaly detection has been widely researched and applied in areas such as intelligent manufacturing and quality control in recent years [11,12]. Existing mainstream unsupervised or weakly supervised anomaly detection methods typically do not rely on true anomalous samples. Instead, they are trained solely using normal samples or artificially synthesized pseudo-anomalies, significantly reducing data annotation costs. During the inference phase, these methods localize anomalous regions by comparing the difference between the input image and the image generated by the model. Methods based on Knowledge Distillation (KD) are particularly prominent in industrial scenarios due to their excellent feature decoupling capabilities and high inference efficiency, making them a focal point of this study [13].

2.1. Application of Knowledge Distillation in Anomaly Detection

Knowledge distillation-based anomaly detection methods train a student network to mimic the output of a teacher network, allowing the student network to learn only the feature distribution of normal samples. In the testing phase, anomalies are localized using the discrepancy between the outputs of the two networks. Bergmann et al. [4] utilized a large, pre-trained neural network as the teacher to distill its representation capabilities into a lightweight student network, performing anomaly detection by leveraging the feature differences among multiple randomly initialized student networks. This approach extracts patch features at the pixel level and achieves high-precision detection by comparing the teacher and student outputs. Salehi et al. [5] further proposed a multi-scale knowledge distillation method, employing a large pre-trained model as the teacher to construct a more compact student network. They simultaneously distilled both the magnitude and direction of intermediate layer features, thereby enhancing feature alignment and detection performance. However, traditional forward knowledge distillation methods still exhibit limitations in anomaly suppression and feature consistency, prompting researchers to explore the reverse distillation framework.

2.2. Development of Reverse Distillation in Anomaly Detection

To enhance the model's sensitivity to anomalies and reduce the false-positive rate, Deng et al. [6] proposed a Reverse Distillation (RD) framework based on a teacher encoder and a student decoder. This method uses the teacher network's output directly as the student network's input, enabling the student to learn semantic features more effectively and capture the normal sample distribution information during the reverse reconstruction

of multi-scale features. Tien et al. [7] adopted Simplex noise to construct natural pseudo-anomalies, simulating real anomaly patterns by randomly injecting anomalous regions to improve the model's robustness against various anomaly types. Furthermore, this method added lightweight projection layers to each stage of the teacher network, mapping multi-scale features to a compact representation space. This was designed to inhibit the propagation of anomalous information to the student network, thereby effectively resolving potential issues of feature looseness and insufficient anomaly suppression in reverse distillation.

3. Methodology

To address the issues of fine-grained feature loss and anomaly misclassification prevalent in industrial scenarios, this paper proposes a Reverse Distillation with Reconstruction Enhancement (RD-RE) framework. As illustrated in Figure 1, the framework primarily consists of four core components: a teacher–student network featuring cross-stage feature fusion, a Locally Aware Dynamic Attention (LDA) module, a Context-Aware Adaptive Multi-Scale Feature Fusion (CAAMS-FF) module, and a segmentation sub-network integrating Coordinated Attention (CA) and Atrous Spatial Pyramid Pooling (ASPP). During the training phase, the framework takes an Original Sample and its derived Synthesized Anomalous Sample as inputs to the teacher and student networks, respectively, to learn the reconstruction of normal feature distributions.

Specifically, the input image is first processed by the teacher–student network. The student network utilizes a cross-stage feature fusion strategy to interactively combine fine-grained features from the encoder's lower layers with rich semantic information from the decoder's higher layers, thereby compensating for potential detail loss during the reverse distillation process. Secondly, to suppress the interference of anomalous features and enhance sensitivity to minute defects, we introduce the LDA module to refine local features. Subsequently, the CAAMS-FF module establishes local contextual dependencies, further integrating cross-stage and multi-scale feature representations to ensure the consistency of reconstructed features. Finally, the multi-scale feature discrepancy maps output by the teacher–student network and is fed into the CA + ASPP segmentation sub-network. By enhancing spatial positional awareness and expanding the receptive field, this sub-network ultimately generates precise anomaly classification scores and pixel-level localization results. The detailed design of each component and the model's training and inference procedures are elaborated in the subsequent sections.

3.1. Reverse Knowledge Distillation with Cross-Stage Feature Fusion

In existing reverse knowledge distillation methods, the student network's normal feature reconstruction capability is weak, leading to normal region features being erroneously judged as anomalous region features, which increases the false positive rate of anomaly detection [14,15]. To enhance the student network's normal feature reconstruction capability, the Reverse Knowledge Distillation framework with cross-stage feature interaction that we propose is shown in Figure 1. The student network decoder features are supplemented with detail information by fusing encoder features. First, the intermediate layer features of the student network encoder undergo fine-grained anomaly filtering via the designed LDA module. Then, the filtered feature maps and the intermediate layer features of the decoder are fused through the CAAMS-FF module, enabling the student network to acquire richer spatial information during the distillation process and thereby enhancing the decoder's normal feature reconstruction capability.

To better enhance the robustness of the student network, an anomaly synthesis method is employed to generate an anomalous image I_a for every image I_n from the normal

training set $I \in R^{H \times W \times C}$. Here, we utilize the DRAEM [16] anomaly synthesis method, which generates anomalous images via a Perlin noise generator and the DTD [17] texture anomaly dataset, as shown in Equation (1). The student network S accepts the synthesized anomalous image $I_S = \{I_a\}$. The four-layer feature maps output by the student network encoder are denoted as $F_S^e = \{F_S^{e1}, F_S^{e2}, F_S^{e3}, F_S^{e4}\}$, and the four layers of features output by the student decoder are $F_S^d = \{F_S^{d1}, F_S^{d2}, F_S^{d3}, F_S^{d4}\}$. Let $f(\cdot)$ and $g(\cdot)$ be the mapping functions for the LDA module and the CFFMS-FF module, respectively. Let β be an adjustable factor, selected within the range $[0.15, 1]$, and let A be the anomaly source randomly selected from the DTD anomaly set, and M_p be the binary mask generated by Perlin noise. The calculation formula for the student decoder output feature map can then be expressed as:

$$I_p = \beta(M_p \odot A) + (1 - \beta)(M_p \odot I_n) + (1 - M_p) \odot I_n \quad (1)$$

$$F_S^{d1} = S(I_a) \quad (2)$$

$$F_S^{d2} = S(I_a) \quad (3)$$

$$F_S^{d3} = g(f(F_S^{e2}) + \text{Up}(F_S^{d1})) + F_S^{d2} \quad (4)$$

$$F_S^{d4} = g(f(F_S^{e3}) + \text{Up}(F_S^{d2})) + F_S^{d3} \quad (5)$$

During the training phase, we select the last three layers of the student network decoder and the first three layers of the teacher network encoder for knowledge distillation. The teacher network T accepts a normal image $I_T = I_n$, and its first three layers output feature maps are denoted as $F_T = \{F_T^1, F_T^2, F_T^3\}$.

3.2. Locally Aware Dynamic Attention Module

The feature maps output by the student network encoder may contain potential anomalous features introduced by the input image during the reverse knowledge distillation process. To suppress the contribution of these anomalous channels, traditional channel attention mechanisms typically generate channel weights using Global Average Pooling. However, when anomalous regions are small, scattered, or irregularly shaped, this global smoothing operation easily over-averages the information of local anomalous features, leading to inaccurate channel weight assignment and making it difficult to effectively suppress the propagation of anomalous features. To this end, we propose the Locally Aware Dynamic Attention (LDA) module, as shown in Figure 2. The core idea of LDA is to effectively capture fine-grained features and dynamically suppress anomalous channels through local perception and dynamic weighting.

Specifically, the module employs a sliding window mechanism to subdivide the feature map into multiple local regions. This localized processing avoids the over-dilution of local information caused by global average pooling. Within each local window, the module calculates the difference between Max Pooling and Average Pooling to highlight anomalous features or significant details within that region, thus dynamically generating channel attention weights.

The channel weight generation process can be expressed as:

Window Pooling:

$$F_S^{e'} = \text{Concat}(\text{MaxPool}_{k \times k}(F_S^e)) - \text{Concat}(\text{AvgPool}_{k \times k}(F_S^e)) \quad (6)$$

where k denotes the window size, the feature map is divided into $\frac{H}{k} \times \frac{H}{k}$ non-overlapping windows.

Dynamic Weight Generation:

$$C = \text{Sigmoid}(W_1 \cdot \text{Relu}(W_2 \cdot \text{LayerNorm}(W_3 \cdot F_S^{e'}))) \quad (7)$$

$$F_S^{e''} = F_S^e \otimes C \quad (8)$$

where W_1 , W_2 , and W_3 are all learnable weight parameters, $\otimes$ denotes channel-wise multiplication, which suppresses abnormal channel responses.

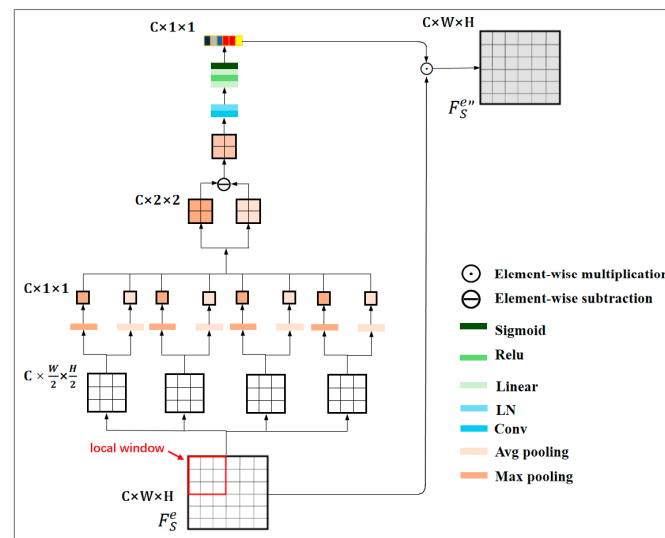


Figure 2. LDA Module Structure Diagram.

3.3. Context-Aware Adaptive Multi-Scale Feature Fusion

The Context-Aware Adaptive Multi-Scale Feature Fusion (CAAMS-FF) module proposed in this paper, with its overall structure shown in Figure 3, and the structure of the ASSA mechanism within it shown in Figure 4, primarily functions to fuse multi-scale feature information, thereby providing more reference information for feature reconstruction. The ASSA mechanism is introduced within the module; this mechanism establishes contextual dependencies, reduces the computational complexity and the influence of redundant features, and finally, a residual block is introduced to achieve a feature reconstruction effect with fine-grained details.

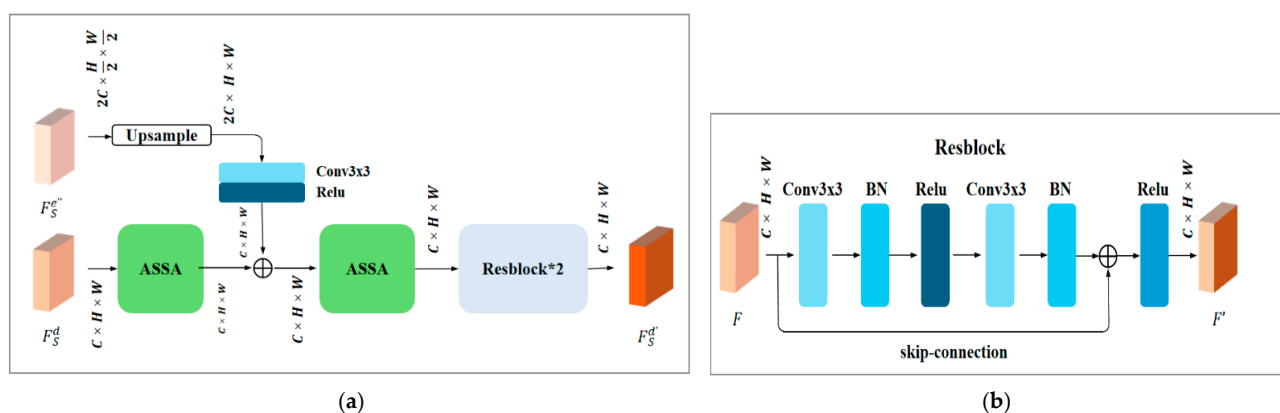


Figure 3. CAAMS-FF Module Structure Diagram. (a) illustrates the overall framework of CAAMS-FF, which includes ASSA, residual blocks, and multi-scale feature fusion. (b) demonstrates the internal structure of the residual block.

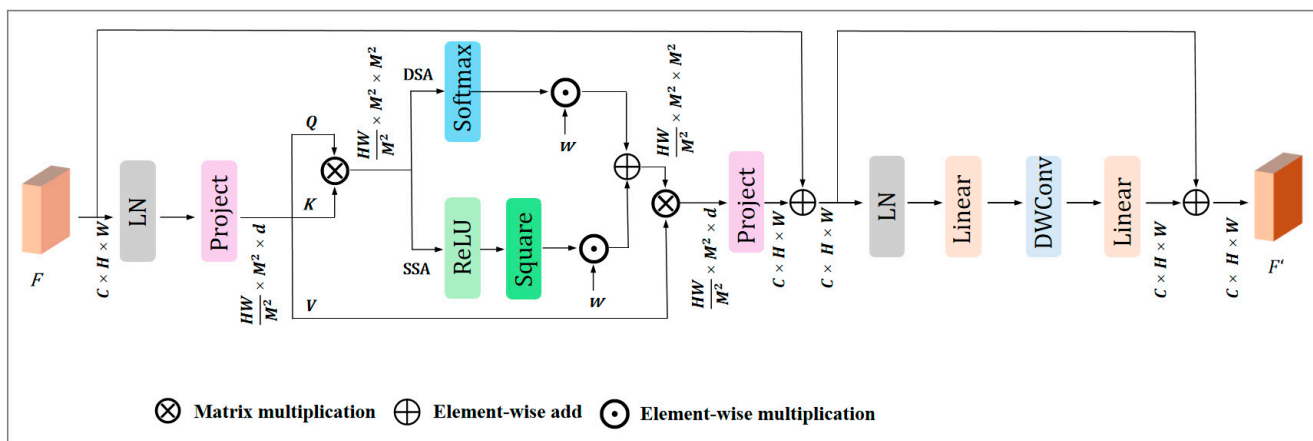


Figure 4. ASSA Module Structure Diagram.

The Adaptive Sparse Self-Attention (ASSA) mechanism consists of two components: Sparse Self-Attention (SSA) and Dense Self-Attention (DSA). SSA is a self-attention mechanism based on the squared ReLU (Rectified Linear Unit), which filters out features negatively influenced by low query-key matching scores. This reduces the impact of irrelevant feature information in the spatial dimension while simultaneously alleviating computational load. DSA, on the other hand, preserves critical information to prevent issues of excessive sparsity. Linear projection and depth-wise separable convolutions are utilized to mitigate the interference of redundant information in the channel dimension, which is crucial for high-quality feature reconstruction.

The computation process of ASSA is as follows:

Linear Projection to Generate Q/K/V:

$$Q_h = F_{win} \cdot W_h^Q \quad (9)$$

$$K_h = F_{win} \cdot W_h^K \quad (10)$$

$$V_h = F_{win} \cdot W_h^V \quad (11)$$

where W_h^Q , W_h^K and W_h^V are the projection matrices for the h-th head, and F_{win} represents the feature map obtained from the non-overlapping windows of size $M \times M$.

Intra-window Attention Weight Calculation:

$$A_h = \text{Softmax}\left(\frac{Q_h K_h^T}{\sqrt{d}} + B\right) \cdot V_h \quad (12)$$

where $d = C/H$ is the dimension per head, and H is the number of heads. To enhance positional awareness, a learnable relative position bias B is introduced into the attention weight calculation.

DSA Computation:

$$DSA = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right) \quad (13)$$

SSA Computation:

$$SSA = \text{Relu}^2\left(\frac{QK^T}{\sqrt{d}} + B\right) \quad (14)$$

Adaptive Self-Attention Computation:

$$A_h = (W_1 \cdot SSA_h + W_2 \cdot DSA_h) \cdot V_h \quad (15)$$

Multi-head Output Concatenation and Fusion:

$$A = \text{Concat}(A_1, A_2, \dots, A_H) \cdot W^O \quad (16)$$

Feature Refinement Computation:

$$F_{out} = \text{LayerNorm}(W_4 \cdot \text{DWConv}(W_3 \cdot F_{in})) \quad (17)$$

where F_{in} represents the input feature, F_{out} represents the output feature, W_3 and W_4 denote learnable parameters, and DWConv stands for depthwise separable convolution.

3.4. Segmentation Sub-Network with CA and ASPP

Traditional knowledge distillation uses the summation of anomaly maps from each layer as the anomaly detection result; however, this method does not yield optimal results. The RD-RE method proposed in this paper feeds the anomaly maps obtained from the teacher–student network into our designed segmentation sub-network to obtain the final anomaly detection and localization results. The segmentation sub-network is composed of a decoder comprising multiple residual networks, and the CA and ASPP modules are introduced before the network head. The CA mechanism explicitly models the horizontal and vertical dependencies of the feature maps through decomposed coordinate encoding, which increases the detection response for edge defects. The ASPP module covers different ranges of receptive fields using convolution kernels with varying dilation rates, effectively capturing contextual information for targets at different scales. The combination of the two forms an optimized path from attention-based selection to multi-scale modeling, significantly improving spatial localization accuracy and multi-scale adaptability while maintaining computational efficiency. Finally, the sub-network outputs the pixel-level anomaly probability mask through a 1×1 convolution and a Sigmoid activation function.

3.5. Model Training and Inference

In the RD-RE framework, the teacher–student network is first trained using cosine similarity distance. Specifically, a synthesized anomalous sample is fed into the student network encoder, while a normal sample is fed into the teacher network. The feature vectors output by the last three layers of the student network decoder are aligned with the feature vectors output by the first three layers of the teacher network. Let $F_T^k \in \mathbb{R}^{C_k \times H_k \times W_k}$ denote the feature map output of the k -th layer of the teacher network, and let $F_S^k \in \mathbb{R}^{C_k \times H_k \times W_k}$ denote the feature map output of the k -th layer of the student network decoder. We define $M_k \in \mathbb{R}^{C_k \times H_k \times W_k}$ as the cosine similarity map of the k -th layer between the teacher and student networks, where $k = 1, 2, 3$.

Then, the total distance loss can be expressed as:

$$M_k(i, j) = \frac{F_T^k(i, j) \odot F_S^k(i, j)}{\|F_T^k(i, j)\|_2 \|F_S^k(i, j)\|_2} \quad (18)$$

$$D_k(i, j) = 1 - \sum_{c=1}^{C_k} M_k(i, j)_c \quad (19)$$

$$L_{\cos} = \sum_{k=1}^3 \frac{1}{H_k W_k} \sum_{i,j=1}^{H_k W_k} D_k(i, j) \quad (20)$$

where i and j denote the spatial coordinates of the feature map, with $i = 1 \dots H_k$ and $j = 1 \dots W_k$.

Subsequently, we freeze the parameters of the teacher–student network and begin training the segmentation sub-network. The multi-scale anomaly maps output by the

teacher–student network serve as the input to the segmentation sub-network, where the binary mask map M of the original image acts as the ground truth. The multi-scale anomaly map is obtained by Equation (19), and each scale’s anomaly map is up-sampled to the same resolution. Concurrently, the binary mask map is down-sampled to M_1 , matching the resolution of the anomaly maps. We employ Focal Loss [18] and L1 Loss to train the segmentation sub-network. Since the anomalous regions often occupy only a small fraction of the image, leading to class imbalance, Focal Loss addresses this by introducing a modulating factor γ . The term $(1 - p(i, j))^\gamma$ directs the model’s focus toward hard-to-classify anomalous regions. L1 Loss computes the absolute difference between the prediction and the ground truth. The joint use of Focal Loss and L1 Loss effectively tackles class imbalance issues and boundary regression accuracy challenges. Let $\hat{Y}$ be the probability map (H_1, W_1) output by the segmentation sub-network.

The total loss calculation can then be expressed as:

$$p(i, j) = M_{ij}\hat{Y}_{ij} + (1 - M_{ij})(1 - \hat{Y}_{ij}) \quad (21)$$

$$L_{focal} = -\frac{1}{H_1W_1} \sum_{i,j=1}^{H_1W_1} (1 - p_{ij})^\gamma \log(p_{ij}) \quad (22)$$

$$L_{l1} = -\frac{1}{H_1W_1} \sum_{i,j=1}^{H_1W_1} |M_{ij} - \hat{Y}_{ij}| \quad (23)$$

$$L_{seg} = L_{focal} + L_{l1} \quad (24)$$

4. Experiments

4.1. Experimental Setup

4.1.1. Datasets

To validate the effectiveness of the RD-RE method, we conducted experiments on two datasets: MVTec AD [19] and VisA [20]. The MVTec dataset is a widely used benchmark dataset in the field of industrial anomaly detection. This dataset contains a total of 5354 color images, encompassing 15 object categories, including 5 texture-type categories and 10 object-type categories. In each category, the training samples consist only of normal samples, while the test samples are composed of both normal and defective samples. The dataset provides pixel-level mask annotations for the defective samples to support the quantitative evaluation of anomaly detection performance. Following the standard MVTec AD protocol, the model for each category is trained exclusively on normal samples (60 to 300 images), and evaluated on an independent test set containing both normal and anomalous samples (40 to 160 images). The dataset statistics are presented in Table A1.

The VisA dataset is a challenging industrial anomaly detection dataset. This dataset comprises 10,821 RGB images, covering 12 object categories, and also provides pixel-level mask annotations for the defective samples.

4.1.2. Implementation Details

For the MVTec AD dataset, the input image resolution is resized to 256×256 . Since the original image resolution of the VisA dataset is higher than that of MVTec AD, and most defects in the anomalous samples are subtle, the input image resolution for VisA is set to 512×512 . The window size k for the LDA module is set to 4 and 2, respectively.

During the training phase, the learning rates for the student network, LDA module, CAAMS-FF module, and segmentation sub-network are set to 0.1, 0.01, 0.01 and 0.1, respectively, with a batch size of 32. All experiments involving the RD-RE network are conducted on a single NVIDIA Tesla V100 32GB GPU.

4.1.3. Evaluation Metrics

For the anomaly classification task at the image level, we adopt the Area Under the Receiver-Operating Characteristic (AUROC) as the evaluation metric. For the anomaly localization task at the pixel level, in addition to using AUROC, we also employ the Average Precision (AP) and Per-Region-Overlap (PRO) evaluation metrics. This is because AUROC may be biased towards cases where the anomalous regions have a large area, whereas PRO and AP are better indicators of the true localization performance level when the anomalous regions are subtle or minute [21]. AUROC is calculated by integrating the area under the curve formed by the True Positive Rate (TPR) and the False Positive Rate (FPR):

$$TPR = \frac{TP}{TP + FN} \quad (25)$$

$$FPR = \frac{FP}{FP + TN} \quad (26)$$

where TP, TN, FP, FN represent true positives, true negatives, false positives, and false negatives, respectively.

AP and PRO provide a more rigorous evaluation of localization performance for subtle defects by focusing on the Precision-Recall trade-off and regional overlap consistency, respectively. Detailed mathematical definitions for AP and PRO can be found in [21].

4.2. Comparison with Current State-of-the-Art Methods

In this section, we compare the RD-RE method with existing State-of-the-Art (SOTA) methods on the MVTec AD and VisA datasets. These SOTA methods include: RD4AD [6], RD++ [7], DeSTSeg [22], CND [23], Patchcore [2], and Diad [24]. Patchcore is a representative method based on feature embedding. This method proposes extracting spatially contextual patch features from images using a pre-trained network, then aggregating intermediate layer features with local neighborhood information. The method uses the maximum distance between a test sample patch and its nearest neighbor in a memory bank as the anomaly score, generating the anomaly map through spatial alignment of the patch features. Diad is a method based on image generation.

This method reconstructs the input image into a normal sample image of itself, referred to as the restored image, via a diffusion model. The final anomaly detection result is obtained by the cosine similarity between the restored image and the input image. RD4AD, RD++, DeSTSeg and CND are all recent methods based on reverse distillation.

4.2.1. MVTec AD Datasets

Table 1 presents the results of anomaly classification on the MVTec AD dataset. Our proposed method achieves State-of-the-Art (SOTA) performance on the image-level AUROC (I-AUROC) metric. Among the texture-type categories, four categories—Carpet, Grid, Leather, and Tile—all achieved 100%, while the Wood category reached 99.7%. For the object-type categories, six categories—Bottle, Cable, Hazelnut, Metal nut, Toothbrush, and Zipper—all scored 100%. The remaining categories, Capsule, Pill, Screw, and Transistor, achieved 99.7%, 97.1%, 99.3%, and 99.5%, respectively. The mean accuracy across all categories reached 99.7%, fully demonstrating the superiority of the RD-RE method's anomaly classification capability.

Table 2 presents the quantitative analysis and comparison results for anomaly localization capability on the MVTec AD dataset. Our method ranks first across the pixel-level AUROC (P-AUROC), AP, and PRO metrics with averages of 99.0%, 78.3%, and 95.8%, respectively, surpassing previous SOTA methods. This fully validates the advanced nature

and effectiveness of the RD-RE method in the anomaly localization task. Figure 5 showcases the qualitative visualization results.

Table 1. IMAGE-LEVEL ANOMALY DETECTION RESULTS (I-AUROC) ON MVTEC AD DATASET.

	Method	PatchCore (CVPR 2022)	RD4AD (CVPR 2022)	RD++ (CVPR 2023)	DeSTSeg (CVPR 2023)	Diad (AAAI 2024)	CND (AAAI 2025)	Ours
Texture	Carpet	98.7	98.5	100.0	100.0	99.4	99.9	100.0
	Grid	98.2	98.0	100.0	100.0	98.5	99.1	100.0
	Leather	100.0	100.0	100.0	100.0	99.8	100.0	100.0
	Tile	98.7	98.3	99.7	97.4	96.8	100.0	100.0
	Wood	99.2	99.2	99.3	97.6	99.7	98.3	99.7
	mean	99.0	98.8	99.8	99.0	98.8	99.5	99.9
Objects	Bottle	100.0	99.6	100.0	100.0	99.7	100.0	100.0
	Cable	99.5	84.1	99.2	97.6	94.8	98.9	100.0
	Capsule	98.1	94.1	99.0	98.7	89.0	98.0	99.7
	Hazelnut	100.0	60.8	100.0	100.0	99.5	100.0	100.0
	Metal nut	100.0	100.0	100.0	100.0	99.1	100.0	100.0
	Pill	96.6	97.5	98.4	96.5	95.7	96.8	97.1
	Screw	98.1	97.7	98.9	93.1	90.7	93.8	99.3
	Toothbrush	100.0	97.2	100.0	100.0	99.7	99.5	100.0
	Transistor	100.0	94.2	98.5	98.7	99.8	96.0	99.5
	Zipper	99.4	99.5	98.6	98.9	99.3	99.1	100.0
	mean	99.2	92.5	99.3	98.4	96.7	98.2	99.5
	MEAN	99.1	94.6	99.4	98.6	97.2	98.6	99.7

The best performing results are highlighted in bold.

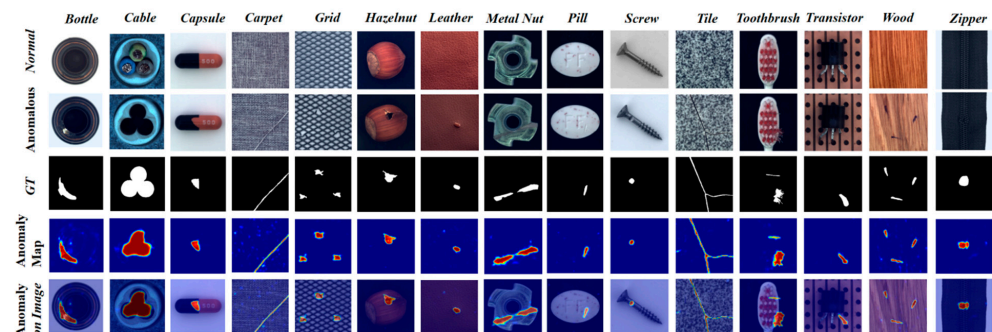


Figure 5. Visualization results of anomaly localization cases using the RD-RE method on the MVTEC AD dataset.

Table 2. QUANTITATIVE RESULTS OF PIXEL-LEVEL ANOMALY LOCALIZATION ON THE MVTEC AD DATASET, EVALUATED USING P-AUROC, AP AND PRO.

	Method	PatchCore (CVPR 2022)	RD4AD (CVPR 2022)	RD++ (CVPR 2023)	DeSTSeg (CVPR 2023)	Diad (AAAI 2024)	CND (AAAI 2025)	Ours
Texture	Carpet	99.1/66.7/96.6	99.0/58.5/95.1	99.2/63.9/97.7	96.1/72.8/93.6	98.6/42.2/90.6	99.3/70.7/97.0	98.6/82.0/97.8
	Grid	98.9/41.0/96.0	96.5/23.0/97.0	99.3/49.5/97.7	99.1/61.5/96.4	96.6/66.0/94.0	98.4/25.8/95.0	99.4/58.6/97.3
	Leather	99.4/51.0/98.9	99.3/38.0/97.4	99.4/51.4/99.2	99.7/75.6/99.0	98.8/56.1/91.3	99.5/50.1/98.7	99.8/77.7/99.3
	Tile	96.6/59.3/87.3	95.3/48.5/85.8	96.4/56.2/92.1	98.0/90.0/95.5	92.4/65.7/90.7	97.7/73.4/94.1	99.2/94.1/97.1
	Wood	95.1/52.3/87.3	95.3/47.8/90.0	95.7/51.8/93.2	97.7/81.9/96.1	93.3/43.3/97.5	96.4/63.3/91.4	98.5/83.5/96.6
	mean	97.8/54.1/93.2	97.1/43.2/93.1	98.1/54.6/96.0	98.1/76.4/96.1	95.9/54.7/92.8	98.3/56.7/95.2	99.1/79.2/97.6

Table 2. Cont.

	Method	PatchCore (CVPR 2022)	RD4AD (CVPR 2022)	RD++ (CVPR 2023)	DeSTSeg (CVPR 2023)	Diad (AAAI 2024)	CND (AAAI 2025)	Ours
Objects	Bottle	98.9/80.1/96.2	96.1/48.6/91.1	98.7/80.0/96.6	99.2/90.3/96.6	98.4/52.2/86.6	99.0/81.8/97.1	99.2/91.2/96.9
	Cable	98.8/70.0/92.5	85.1/26.3/75.1	98.4/63.6/ 93.9	97.3/60.4/86.4	96.8/50.1/80.5	98.2/64.1/92.5	97.0/64.3/88.8
	Capsule	99.1/48.1/95.5	98.8/43.4/94.8	98.9/47.4/96.5	99.1/56.3/94.2	97.1/42.0/87.2	98.2/64.1/92.5	99.3/64.0/97.2
	Hazelnut	99.0/61.5/93.8	97.9/36.2/92.7	99.2/66.5/96.3	99.6/88.4/97.6	98.3/79.2/91.5	98.2/36.9/93.8	99.7/88.9/98.0
	Metal nut	98.8/88.8/91.4	94.8/55.5/91.9	98.0/83.9/93.2	98.6/ 93.5/95.0	97.3/30.0/90.6	98.8/53.3/94.9	99.0/89.9/94.9
	Pill	98.2/78.7/93.2	97.5/80.2/96.1	98.4/79.6/ 97.1	98.7/83.1/95.3	95.7/46.0/89.0	95.5/68.4/89.2	99.3/89.0/97.0
	Screw	99.5/41.4/97.9	99.0/94.7/26.2	99.6/55.5/98.3	98.5/ 58.7/92.5	97.9/60.6/95.0	98.8/53.3/94.9	99.1/53.8/95.9
	Toothbrush	98.9/51.6/91.5	99.0/93.0/49.7	99.1/56.3/94.5	99.3/75.2/94.0	99.0/ 78.7/95.0	99.0/26.2/94.7	99.6/77.4/95.4
	Transistor	96.2/63.2/83.7	94.5/57.3/74.7	94.4/58.3/82.8	89.1/64.8/85.7	95.1/15.6/ 90.0	94.5/57.3/74.1	97.3/75.1/89.2
	Zipper	99.0/64.7/93.3	98.5/53.9/94.1	98.9/60.5/96.4	99.1/85.2/97.4	96.2/60.7/91.6	97.6/56.3/91.9	99.4/85.3/97.7
	mean	98.6/64.7/93.3	96.1/58.9/78.6	98.4/65.2/94.6	97.9/75.6/93.5	97.2/51.5/89.7	97.9/56.3/91.9	98.9/77.9/95.1
	MEAN	98.4/61.2/93.4	96.1/48.6/91.1	98.3/61.6/95.0	97.9/75.8/94.4	96.8/52.6/90.7	98.0/56.4/93.0	99.0/78.3/95.8

The best performing results are highlighted in bold.

4.2.2. VisA Datasets

This study conducted comprehensive experimental validation on the VisA dataset; the quantitative results are shown in Tables 3 and 4. Table 3 presents the anomaly classification results, where our method achieved the highest mean AUROC metric of 98.9%, representing an improvement of 2.3% compared to the current highest method, DeSTSeg, and an improvement of 5.7% compared to the latest method, CND.

Table 3. IMAGE-LEVEL ANOMALY DETECTION RESULTS(I-AUROC) ON VisA DATASET.

Method	PatchCore (CVPR 2022)	RD4AD (CVPR 2022)	RD++ (CVPR 2023)	DeSTSeg (CVPR 2023)	Diad (AAAI 2024)	CND (AAAI 2025)	Ours
Candle	98.7	92.3	96.4	97.1	92.8	93.7	97.9
Capsules	68.8	82.2	92.1	96.6	58.2	83.4	97.2
Cashew	97.7	92.0	97.8	93.7	91.5	94.1	100
Chewinggum	99.1	94.9	96.4	99.6	99.1	98.7	99.4
Fryum	91.6	95.3	95.8	96.7	89.8	96.4	98.2
Macaroni1	90.1	75.9	94.0	96.8	85.7	86.7	99.3
Macaroni2	63.4	88.3	88.0	87.8	62.5	84.4	97.4
Pcb1	96.0	96.2	97.0	96.7	88.1	94.1	100
Pcb2	95.1	97.8	97.2	96.9	91.4	95.9	99.2
Pcb3	93.0	96.4	96.8	98.8	86.2	92.0	99.3
Pcb4	99.5	99.9	99.8	99.5	99.6	99.9	99.7
Pipe fryum	91.6	97.9	99.6	98.9	96.2	98.9	99.0
MEAN	91.0	92.4	95.9	96.6	86.8	93.2	98.9

The best performing results are highlighted in bold.

Furthermore, RD-RE outperformed other methods in 8 categories, achieving 100% in two categories, with a maximum improvement of 9.1% in these categories, establishing it as the most outstanding SOTA method currently available.

Table 4 shows the quantitative analysis results for the anomaly localization task on the VisA dataset. RD-RE also achieved the highest average values across the three metrics: AUROC, AP and PRO. Compared to the latest CND method, the maximum increase was 1.0% for AUROC, 11.8% for AP, and 2.7% for PRO. Our method likewise demonstrates superior

performance under the comprehensive evaluation metrics for anomaly localization. Figure 6 provides the qualitative visualization results of defect segmentation on the VisA dataset.

Table 4. QUANTITATIVE RESULTS OF PIXEL-LEVEL ANOMALY LOCALIZATION ON THE VISA DATASET, EVALUATED USING P-AUROC, AP AND PRO.

Method	PatchCore (CVPR 2022)	RD4AD (CVPR 2022)	RD++ (CVPR 2023)	DeSTSeg (CVPR 2023)	Diad (AAAI 2024)	CND (AAAI 2025)	Ours
Candle	99.2/-/94.0	99.1/25.3/94.9	98.6/18.3/93.8	99.3/50.0/94.3	97.3/12.8/89.4	98.4/16.7/91.9	99.5/53.0/96.4
Capsules	96.5/-/85.5	99.4 /60.4/93.1	99.4 /40.6/95.8	99.4/69.0/97.0	97.3/10.0/77.9	98.4/33.6/88.6	98.7/56.9/96.9
Cashew	99.2 /-/94.5	91.7/44.2/86.2	95.5/31.6/91.2	95.9/53.6/ 97.6	90.9/53.1/61.8	98.1/ 62.9 /87.4	96.5/60.3/92.7
Chewinggum	98.9/-/84.6	98.7/59.9/76.9	98.4/ 63.7 /88.1	99.1/25.3/91.8	94.7/11.9/59.5	99.1/61.3/89.4	99.2 /31.2/ 93.3
Fryum	95.9/-/85.3	97.0/47.6/ 93.4	96.5/22.3/90.0	89.2/46.9/89.4	97.6/58.6 /81.3	97.0/47.3/92.1	93.7/46.4/86.0
Macaroni1	98.5/-/95.4	99.4/2.9/95.3	99.7/16.3/96.9	99.7/38.9/97.9	94.1/10.2/68.5	98.6/7.8/90.5	99.8/40.0/98.7
Macaroni2	93.5/-/94.4	99.7 /13.2/97.4	99.7 /2.7/97.3	99.5/ 28.4 /97.5	93.6/0.9/73.1	98.1/12.7/93.6	99.4/27.6/ 99.1
Pcb1	99.8 /-/94.3	99.4/66.2/ 95.8	99.7/68.5/ 95.8	99.6/72.3/93.4	98.7/49.6/80.2	99.5/70.7/92.6	99.2/ 73.0 /93.2
Pcb2	98.4/-/89.2	98.0/22.3/90.8	98.9 /20.0/90.6	98.2/ 34.0 /87.9	95.2/7.5/67.0	98.4/18.1/88.8	98.7/29.5/ 91.5
Pcb3	98.9/-/90.9	97.9/26.2/ 93.9	99.2/22.4/93.1	98.4/34.2/87.3	96.7/8.0/68.9	98.6/21.7/93.7	99.5/36.8 /92.0
Pcb4	98.3/-/95.7	97.8/31.4/88.7	98.8/31.4/91.9	98.7/58.7/89.1	97.0/17.6/85.0	99.0/40.5/90.5	99.2/60.6/92.3
Pipe fryum	99.3/-/95.7	99.1/56.8/95.4	99.1/37.2/95.6	99.3/76.0/96.3	99.4/72.7/89.9	98.4/61.4/ 97.5	99.6/80.0 /97.2
MEAN	98.0/-/91.2	98.1/38.0/91.8	98.6/34.5/93.3	98.0/48.9/93.3	96.0/26.1/75.2	98.5/37.8/91.4	98.6/49.6/94.1

The best performing results are highlighted in bold.

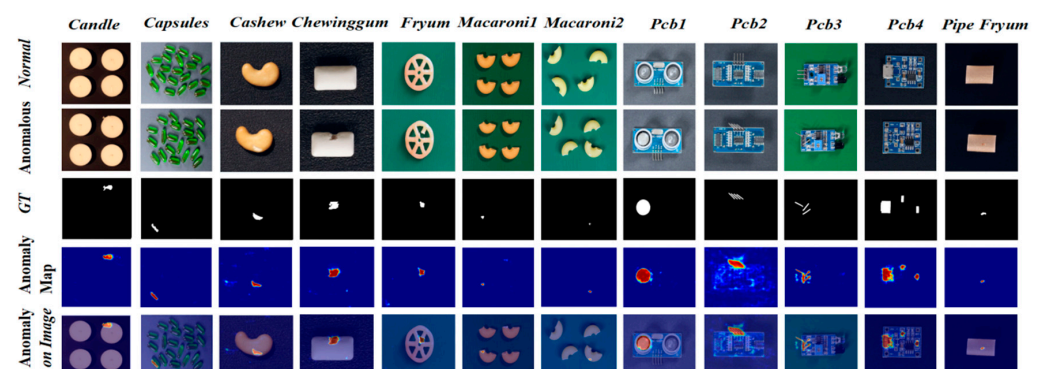


Figure 6. Visualization results of anomaly localization cases using the RD-RE method on the VisA dataset.

4.3. Generalization, Robustness, and Efficiency

4.3.1. Generalization Testing

To evaluate the transferability and practical applicability of RD-RE in real-world industrial scenarios, generalization experiments were conducted on the BTAD (BeanTech Anomaly Detection) dataset. BTAD is a real-world industrial dataset comprising 2830 high-resolution images across three distinct categories of industrial products, encompassing a wide range of complex defects on both product bodies and surfaces. Performance was quantified using AUROC, AP, and PRO metrics. As illustrated in Figure 7, the experimental results demonstrate that RD-RE outperforms competing methods in anomaly localization on the BTAD dataset. These findings validate that the proposed model not only excels in specific tasks but also possesses robust generalization capabilities, effectively addressing diverse challenges in industrial anomaly detection.

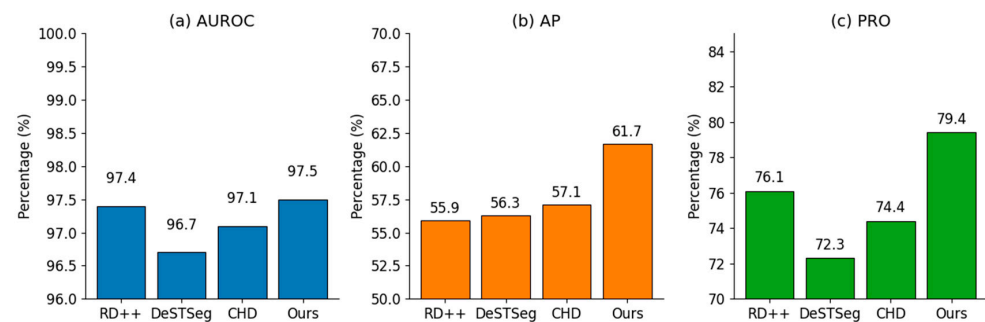


Figure 7. Experimental Results on the BTAD Dataset.

4.3.2. Robustness Testing

To further verify the robustness of RD-RE against industrial noise interference, perturbation experiments were conducted by introducing Additive White Gaussian Noise (AWGN) of varying intensities. Specifically, standard deviations of σ in $\{0.001, 0.005, 0.01, 0.05\}$ were set to simulate different levels of sensor noise or environmental interference encountered in real-world industrial settings. The experimental results are illustrated in Figure 8. The observations indicate that as noise intensity increases, the AUROC metric of RD-RE exhibits exceptional stability, with fluctuations consistently maintained within a 3% margin. These results intuitively validate that the RD-RE method sustains reliable detection performance under complex interference environments, demonstrating a significant advantage in robustness for practical industrial deployment.

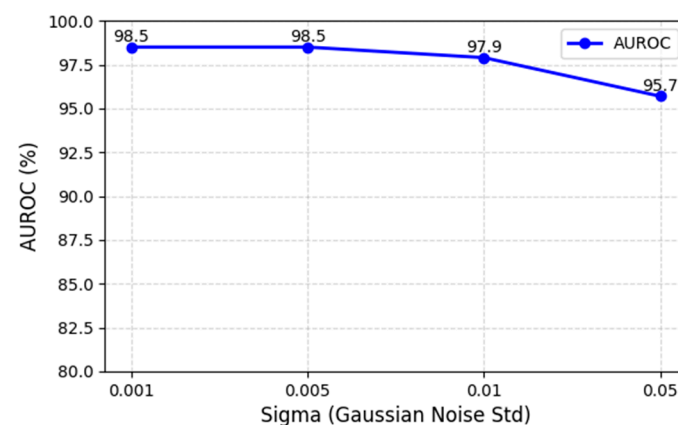


Figure 8. Experimental Results of Robustness Testing.

4.3.3. Model Efficiency

To evaluate the computational efficiency and inference performance of the model, we compared RD-RE with the baseline method DeSTSeg and the state-of-the-art method CHD on a single NVIDIA Tesla V100 32GB GPU. The evaluation focused on key performance indicators, including FLOPs (Floating Point Operations), FPS (Frames Per Second), and GPU memory usage. All experiments were conducted with a batch size of 1 to simulate real-time industrial inference scenarios.

The experimental data, as shown in Figure 9, reveal that RD-RE demonstrates a significant efficiency advantage while maintaining superior detection accuracy. Specifically, it requires only 1107 MB of GPU memory, incurs a computational load as low as 35.77 GFLOPs, and achieves an inference speed of 35 FPS. This performance proves that the proposed method aligns well with the requirements of industrial production lines for high real-time responsiveness and low resource consumption.

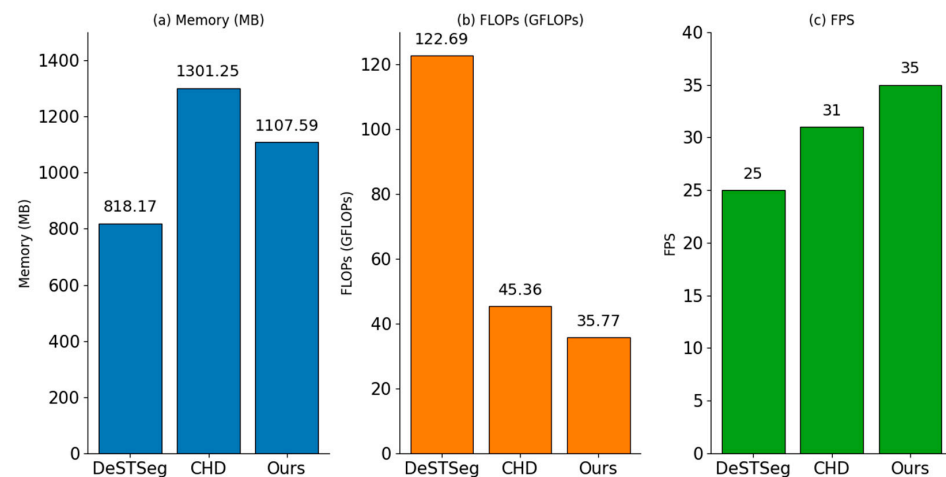


Figure 9. Evaluation of FLOPs, FPS, and GPU memory usage for RD-RE and competing models.

4.4. Ablation Study

4.4.1. The Effectiveness of Each Component in the RD-RE Method

The method proposed in this paper introduces four main innovative designs: 1. Cross-stage Feature Fusion Architecture (CFFA); 2. Locally Aware Dynamic Attention (LDA); 3. Context-Aware Adaptive Multi-Scale Feature Fusion (CAAMS-FF); and 4. Segmentation Sub-network with Coordinated Attention and ASPP (Seg). To validate the contribution of each component within our proposed RD-RE method, we conducted ablation experiments on the MVTec AD dataset; the results are presented in Table 5.

Table 5. ABLATION STUDY ON THE CONTRIBUTIONS OF EACH COMPONENT IN RD-RE ON THE MVTec AD DATASET, EVALUATED USING I-AUROC, P-AUROC, AP AND PRO.

CFFA	LDA	CAAM-FF	Seg	Performance
×	×	×	×	95.6/94.9/57.2/91.2
✓	×	×	×	96.8/96.7/61.9/92.0
×	×	×	✓	98.3/98.1/73.8/93.4
✓	×	×	✓	98.7/98.2/75.2/93.2
✓	✓	×	×	97.8/97.5/65.5/92.3
✓	✓	✓	×	98.4/97.9/72.2/93.1
✓	×	✓	×	98.3/97.6/69.7/92.2
✓	✓	×	✓	98.9/98.4/76.1/94.4
✓	×	✓	✓	99.2/98.6/76.6/94.5
✓	✓	✓	✓	99.7/99.0/78.3/95.8

✓ denotes inclusion; × denotes exclusion; The best performing results are highlighted in bold.

Each component—CFFA, LDA, CAAMS-FF, and Seg—demonstrates a positive effect on enhancing anomaly detection performance. The integrated optimization strategy of CFFA + LDA + CAAMS-FF + Seg resulted in improvements in image-level AUROC from 95.6% to 99.7%, in pixel-level AUROC from 94.9% to 99.0%, in pixel-level AP from 57.2% to 78.3%, and in pixel-level PRO from 91.2% to 95.8%, achieving the highest score across all four evaluation metrics.

4.4.2. Hierarchical Selection for Cross-Stage Fusion

Table 6 presents the quantitative analysis results for different layer fusion selection combinations on the MVTec AD dataset. Considering the four comprehensive evaluation metrics: Image-level AUROC, Pixel-level AUROC, Pixel-level AP, and Pixel-level PRO, it can be concluded that selecting the second and third layers of both the encoder and the decoder achieves the optimal classification and localization performance.

Table 6. ABLATION STUDY ON CROSS-STAGE FUSION LATER SELECTION ON THE MVTEC AD DATASET, EVALUATED USING I-AUROC, P-AUROC, AP ADN PRO.

Layer-1	Layer-2	Layer-3	Performance
×	×	×	98.3/98.1/73.8/93.4
✓	×	×	98.6/98.0/70.7/91.2
×	✓	×	99.1/98.6/72.1/92.6
×	×	✓	99.3/98.2/71.2/91.5
✓	✓	×	98.5/98.1/74.3/93.7
✓	×	✓	98.1/98.9/76.6/95.0
✓	✓	✓	99.7/98.6/76.4/95.2
×	✓	✓	99.7/99.0/78.3/95.8

✓ denotes inclusion; × denotes exclusion; The best performing results are highlighted in bold.

4.4.3. Determining the Window Size in LDA

Figure 10 illustrates the impact of different window size parameters k for the LDA module in layers 2 and 3 on the pixel-level AUROC metric across the MVTec AD dataset. The selected values for k are set as $1/2$, $1/4$, $1/8$, $1/16$ and $1/32$ of the original feature map dimension, where the maximum possible values of k for layers 2 and 3 are 32 and 16, respectively. As shown in Figure 7, for the second layer, the pixel-level AUROC reaches its maximum value when $k = 4$. For the third layer, the maximum value is achieved when $k = 2$. Therefore, the optimal parameters for k in layers 2 and 3 are 4 and 2, respectively.

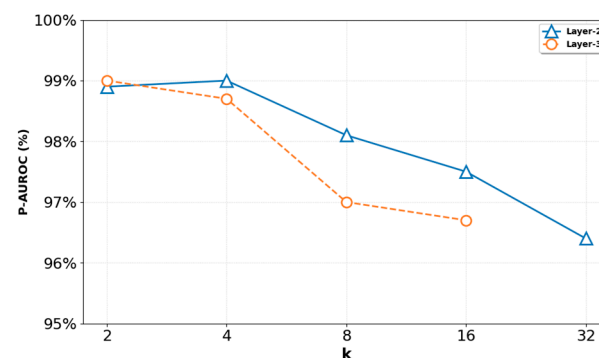


Figure 10. Impact of Parameter k on P-AUROC in LDA Modules of Layer-2 and Layer-3.

4.4.4. Determination of Sparse Window Size in ASSA

Table 7 presents a detailed performance evaluation of the sparse attention mechanism within the ASSA module across various sparse window sizes (Win), focusing on two critical metrics: AUROC and FPS. The experimental results indicate that the window size exerts a substantial influence on both detection accuracy and computational efficiency. A comprehensive comparative analysis reveals that the system achieves an optimal performance trade-off when the window size is set to 8. At this configuration, the AUROC reaches its peak, while the FPS remains sufficiently high to meet the requirements of real-time

detection. These findings validate that a window size of 8 allows ASSA to capture an appropriate sparse receptive field while maintaining rapid inference speeds, making it the optimal parameter choice for practical deployment in this study.

Table 7. PERFORMANCE COMPARISON OF THE ASSA MODULE WITH DIFFERENT SPARSE WINDOW SIZES.

Variety	Sparse			
	Win = 0	Win = 4	Win = 8	Win = 16
AUROC	98.8	96.3	99.0	94.3
FPS	15	28	35	40

The best performing results are highlighted in bold.

4.4.5. Selection of Anomaly Synthesis Strategies

To investigate the impact of different anomaly synthesis strategies on model performance, four representative categories from the MVTec AD dataset were selected: Capsule, Metal Nut, Wood, and Grid. These categories encompass both texture-based and structure-based industrial objects. Three experimental schemes were evaluated: (1) no synthesis strategy; (2) using Simplex noise to construct natural pseudo-anomaly samples; and (3) generating anomaly images by combining Perlin noise with the Describable Textures Dataset (DTD).

Mean AUROC, AP, and PRO were employed as quantitative evaluation metrics, with the results summarized in Table 8. The data demonstrate that the synthesis strategy combining Perlin noise with the DTD collection yields the highest performance across all metrics. This indicates that, compared to simple noise interference, a strategy incorporating DTD texture features generates anomaly samples that more closely align with the distribution of real-world industrial damages. Consequently, this approach significantly enhances the model's ability to discriminate complex defects.

Table 8. PERFORMANCE COMPARISON OF ANOMALY SYNTHESIS STRATEGIES.

Variety	Anomaly Synthesis Strategy		
	None	Simplex Noise	Perlin Noise + DTD
AUROC/AP/PRO	94.3/60.2/89.0	97.2/68.2/92.5	99.1/74.0/96.5

The best performing results are highlighted in bold.

4.4.6. Effectiveness of the L1 Loss Function

Table 9 presents the impact of the L1 loss function on model classification and localization performance during the training of the segmentation sub-network. The results indicate that employing a joint optimization strategy—combining Focal loss and L1 loss—yields significant improvements over using Focal loss alone. Specifically, the image-level AUROC, pixel-level AUROC, pixel-level AP, and pixel-level PRO increased by 0.7%, 0.4%, 3.1%, and 1.7%, respectively. These quantitative gains validate the effectiveness of the L1 + Focal loss joint strategy in enhancing the precision of both anomaly detection and localization.

Table 9. ABLATION STUDY ON THE SEGMENTATION NETWORK L1 LOSS ON THE MVTec AD DATASET, EVALUATED USING I-AUROC, P-AUROC, AP AND PRO.

	I-AUROC	P-AUROC	AP	RPO
ω/o L1 loss	99.0	98.6	75.2	94.1
$\omega/$ L1 loss	99.7	99.0	78.3	95.8

The best performing results are highlighted in bold.

5. Discussion

Despite the competitive performance of RD-RE on the MVTec AD and VisA benchmark, we acknowledge certain limitations inherent to the reconstruction-based paradigm. Specifically, the model may struggle with logical anomalies—such as missing components, misplaced parts, or incorrect assembly orders—where the local texture appears normal but violates high-level semantic constraints. Since our student network is trained to reconstruct normal features, it might inadvertently reconstruct these logical defects if they do not present significant textural deviations, potentially leading to missed detections in complex assembly scenarios.

To address this challenge, future research could explore integrating few-shot learning strategies into the unsupervised framework. In real-world industrial settings, while anomalous samples are rare, obtaining a small number of them (e.g., 1–5 shots) is often feasible. By leveraging these few real anomalous samples, the model could be fine-tuned to explicitly recognize specific logical defect patterns that are difficult to simulate via synthesis, thereby combining the strengths of unsupervised reconstruction and supervised classification.

6. Conclusions

This paper proposes a novel Reverse Distillation method based on Reconstruction Enhancement (RD-RE) to mitigate two core issues in existing reverse knowledge distillation approaches: insufficient filtering of fine-grained anomalous features (such as micro-cracks and point corrosion) and the loss of partial detail features during the feature reconstruction process. The cross-stage feature interaction structure of the student network in the RD-RE method solves the problem of potential detail feature loss during decoder feature reconstruction. The Locally Aware Dynamic Attention (LDA) module and the Context-Aware Adaptive Multi-Scale Feature Fusion (CAAMS-FF) module effectively filter fine-grained anomalous features and establish contextual dependencies. Furthermore, the segmentation sub-network, which integrates Coordinated Attention (CA) and Atrous Spatial Pyramid Pooling (ASPP), generates segmentation results with fine shapes and clear boundaries. Extensive experiments conducted on two benchmark datasets validate the significant advantages of the RD-RE method, providing a new baseline reference for reverse knowledge distillation-based anomaly detection.

Author Contributions: A.L.: Conceptualization; Formal analysis; Methodology; Software; Data curation; Investigation; Validation; Visualization; Writing—original draft. Y.F.; Funding acquisition; Project administration; Resources; Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Chongqing Municipality, grant number No. CSTB2022NSCQ-MSX0786, and the Humanities and Social Sciences Research Project of the Ministry of Education, grant number No. 24YJA870003.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

RD-RE	Reverse Distillation with Feature Reconstruction Enhancement
AD	Anomaly Detection
LDA	Locally Aware Dynamic Attention
CAAMS-FF	Context-Aware Adaptive Multi-Scale Feature Fusion
RD	Reverse Distillation
ASSA	Adaptive Sparse Self-Attention
CA	Coordinated Attention
ASPP	Atrous Spatial Pyramid Pooling
SSA	Sparse Property of the Attention
DSA	Dense Self-Attention
AUROC	area under the receiver-operating characteristic
AP	average precision
PRO	per-region-overlap
CFFA	Cross-stage Feature Fusion Architecture
I-AUROC	Image-area under the receiver-operating characteristic
P-AUROC	pixel-area under the receiver-operating characteristic

Appendix A

Table A1. Summary of the MVTec AD dataset across 15 categories.

Category	Texture									Objects					
	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable	Capsule	Hazelnut	Metal nut	Pill	Screw	Toothbrush	Transistor	Zipper
Train	280	264	245	230	247	99.9	224	219	391	220	267	320	60	213	240
Test	117	78	124	117	79	99.1	150	132	110	115	167	160	42	100	151
Summary	397	342	369	347	326	100	274	351	501	335	434	480	102	313	391

References

1. Liu, J.; Hou, Z.; Li, W.; Tao, R.; Orlando, D.; Li, H. Multipixel anomaly detection with unknown patterns for hyperspectral imagery. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 5557–5567. [[CrossRef](#)] [[PubMed](#)]
2. Roth, K.; Pemula, L.; Zepeda, J.; Sch, B.; Brox, T.; Gehler, P. Towards total recall in industrial anomaly detection. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 14298–14308.
3. Huang, C.; Jiang, A.; Feng, J.; Zhang, Y.; Wang, X.; Wang, Y. Adapting visual-language models for generalizable anomaly detection in medical images. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 11375–11385.
4. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4182–4191.
5. Salehi, M.; Sadjadi, N.; Baselizadeh, S.; Rohban, M.H.; Rabiee, H.R. Multiresolution knowledge distillation for anomaly detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14897–14907.
6. Deng, H.; Li, X. Anomaly detection via reverse distillation from one-class embedding. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 9727–9736.
7. Tien, T.D.; Nguyen, A.T.; Tran, N.H.; Huy, T.D.; Duong, S.T.; Truong, S.Q.H. Revisiting reverse distillation for anomaly detection. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 24511–24520.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717.

10. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
11. Xing, P.; Li, Z. Visual anomaly detection via partition memory bank module and error estimation. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 3596–3607. [[CrossRef](#)]
12. Li, H.; Chen, Z.; Xu, Y.; Hu, J. Hyperbolic anomaly detection. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 17511–17520.
13. Tsai, M.-C.; Wang, S.-D. Self-supervised image anomaly detection and localization with synthetic anomalies. In Proceedings of the 2023 10th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), San Antonio, TX, USA, 23–25 October 2023; Volume 14, pp. 90–95.
14. You, Z.; Cui, L.; Shen, Y.; Yang, K.; Lu, X.; Zheng, Y.; Le, X. A unified model for multi-class anomaly detection. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: San Diego, CA, USA, 2022; Volume 35, pp. 4571–4584.
15. Yao, X.; Li, R.; Zhang, J.; Sun, J.; Zhang, C. Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 24490–24499.
16. Zavrtanik, V.; Kristan, M.; Sko, D. DrÆm—A discriminatively trained reconstruction embedding for surface anomaly detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 8310–8319.
17. Cimpoi, M.; Maji, S.; Kokkinos, I.; Mohamed, S.; Vedaldi, A. Describing textures in the wild. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3606–3613.
18. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Doll, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)] [[PubMed](#)]
19. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Mvtec ad—Comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9584–9592.
20. Zou, Y.; Jeong, J.; Pemula, L.; Zhang, D.; Dabeer, O. Spot-the difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 392–408.
21. Tao, X.; Gong, X.; Zhang, X.; Yan, S.; Adak, C. Deep learning for unsupervised anomaly localization in industrial images: A survey. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–21. [[CrossRef](#)]
22. Zhang, X.; Li, S.; Li, X.; Huang, P.; Shan, J.; Chen, T. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 3914–3923.
23. Wang, X.; Wang, X.; Bai, H.; Lim, E.G.; Xiao, J. Cnc: Cross-modal normality constraint for unsupervised multi-class anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2025; Volume 39, pp. 7943–7951.
24. He, H.; Zhang, J.; Chen, H.; Chen, X.; Li, Z.; Chen, X.; Wang, Y.; Wang, C.; Xie, L. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2024; Volume 38, pp. 8472–8480.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.