electronics

MDPI

Article

A Refined Knowledge Distillation Network for Unsupervised Anomaly Detection

Zeyu Han [1], Jin Wan [2,3], Yumei Li [2,3] and Gang Li [2,3,*]

[1] Information HUB, The Hong Kong University of Science and Technology (Guangzhou), Nansha, Guangzhou 511453, China; zhan476@connect.hkust-gz.edu.cn
[2] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Wendong, Jinan 250013, China; wanjin@sdas.org (J.W.); 10431230066@stu.qlu.edu.cn (Y.L.)
[3] Shandong Provincial Key Laboratory of Computer Networks, Shandong Fundamental Research Center for Computer Science, Wendong, Jinan 250013, China
* Correspondence: lig@qlu.edu.cn

Abstract: Knowledge distillation has been proven to be an important component in unsupervised image anomaly detection and localization. However, a significant challenge remains in how the student network can extract more accurate feature information from the teacher network during the knowledge distillation process. In this paper, we propose a refined knowledge distillation network which introduces simulated anomalous images to address the lack of anomalous samples in unsupervised anomaly detection. Firstly, to enhance the student network's ability to effectively focus on the important features of the teacher network while ensuring spatial consistency between student and teacher features, we design a double feature enhancement and alignment module (DFEAM). Additionally, we propose a multi-domain enhanced restoration loss to reduce the loss of critical details while the student network is learning from the teacher network, thereby improving its feature learning capability. Finally, to fully consider the weight differences between features at different layers for various detection tasks, we design a multi-scale feature interaction fusion module (MFIFM) that effectively integrates feature information from different layers to achieve better results. Extensive experiments are conducted on the publicly available industrial anomaly detection dataset, and the results validate the effectiveness of our model and modules. The image-level anomaly detection achieves an area under the curve (AUC) of 98.9%, the pixel-level anomaly detection achieves an average precision (AP) of 78.3%, and the instance-level anomaly detection achieves an instance average precision (IAP) of 77.5%.

Keywords: image anomaly detection; knowledge distillation; teacher–student network

check for updates

Citation: Han, Z.; Wan, J.; Li, Y.; Li, G. A Refined Knowledge Distillation Network for Unsupervised Anomaly Detection. *Electronics* **2024**, *13*, 4793. https://doi.org/10.3390/electronics13234793

Received: 5 November 2024
Revised: 27 November 2024
Accepted: 3 December 2024
Published: 5 December 2024

Copyright: © 2024 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Image anomaly detection (IAD) is a critical task in many computer vision applications, such as industrial defect detection [1–7] and anomaly diagnosis [8,9]. The goal of these tasks is to identify damaged images and locate the damaged areas within them. Due to the scarcity of anomalous samples and the vast number of possible anomaly types, it is impractical to obtain sufficient samples representing all potential anomalies for training. Therefore, unsupervised anomaly detection plays an important role, relying solely on normal data to train models for anomaly detection and localization tasks.

The knowledge distillation framework (teacher–student network) [2–4] has been proven to be an effective technique in unsupervised anomaly detection. In this framework, the teacher network is pre-trained on large-scale datasets and uses normal data as input during training for both the teacher and student networks, optimizing the student network to generate feature representations as similar as possible to those of the teacher network.

Since the student network encounters anomalies during training that it has never seen before, it generates feature representations that differ from those of the teacher network. Thus, during inference, it can identify anomalous conditions and locate abnormal pixel areas in images. Ref. [2] performs anomaly detection within the student–teacher network framework through multi-scale feature pyramid matching. The teacher network provides multiple layers of feature information to the student network, enabling it to focus on different scales of features while learning the distribution of normal samples. This method assumes that the multi-layer features of normal samples can be effectively replicated by the student network, so that anomalies will manifest as discrepancies during feature matching, allowing for detection. Furthermore, Ref. [4] distills the multi-layer features from the pre-trained teacher network into a smaller student network to capture the multi-scale feature distribution of normal samples. However, due to the lack of supervision from anomalous samples, the student network may face generalization issues, particularly when the differences between anomalous and normal features are subtle. In such cases, the student network might output feature representations that are too similar to those of the teacher network. Additionally, when the teacher network provides feature learning to the student network, there is often a lack of comprehensive consideration of the integrity and quality of feature knowledge. Moreover, simple aggregation methods based on multi-layer features (such as direct summation or multiplication) may not always be ideal, as they fail to adequately account for the weight differences between features across different detection tasks, potentially impacting detection performance.

To address the aforementioned challenges, this paper proposes a refined knowledge distillation network. This includes a pair of teacher–student networks, a double feature enhancement and alignment module (DFEAM), a multi-scale feature interaction fusion module (MFIFM), and an image anomaly segmentation network. First, inspired by previous work [10], this study generates random anomaly masks using randomly generated Perlin noise, which are then overlaid onto normal images to create simulated anomalous images. These anomalous images are paired with normal images in the training set for model training. The network training is divided into two phases. In the first phase, normal images and simulated anomalous images serve as inputs to the teacher network and student network, respectively. The optimization goal is to ensure that the feature representations generated by the student network closely resemble those of the teacher network. During this process, the student network learns to restore the features of the anomalous images to those of normal images, effectively addressing the issue of lacking supervisory anomalous samples by constraining the output in the anomaly regions. To enhance the student network's ability to learn and absorb feature knowledge from the teacher network, we design the DFEAM, which aligns the high-dimensional features of the teacher network with the feature space of the student network. This reduces information loss due to differences in network structure and feature distribution while improving the ability to capture critical knowledge. Furthermore, to strengthen the student network's focus on high-frequency details and enable it to more accurately reproduce the teacher network's feature expressions, we introduce a multi-domain enhanced restoration loss. This allows the student network to capture both global and local feature information effectively, thereby minimizing the loss of key details and enhancing feature alignment, which improves the model's generalization capabilities across different scenarios. In the second phase, the parameters of the trained student network are fixed, and simulated anomalous images are input to both the teacher and student networks. The MFIFM is employed to fuse the features from both networks at various layers, taking into account the weight differences between these layers. This approach enables the model to integrate the feature information more effectively, resulting in improved outcomes.

Our contributions can be summarized as follows:

- We propose a refined knowledge distillation network that enhances the robustness of the knowledge distillation process by introducing simulated anomalous image supervision.

- We design the DFEAM to enable the student network to more efficiently focus on the important features of the teacher network while maintaining consistency in the feature space.
- We introduce a multi-domain enhanced restoration loss, which enables the student network to capture both global and local feature information, effectively reducing the loss of critical details.
- To fully account for the weight differences between features across layers in various detection tasks, we design the MFIFM, which effectively integrates multi-layer feature information to achieve better results.

The remainder of this paper is organized as follows. Section 2 briefly reviews related works, including anomaly detection and knowledge distillation. The details of our proposed method are introduced in Section 3. In Section 4, we compare our method with several other methods on publicly available datasets by carrying out qualitative and quantitative comparisons both for the DFEAM and MFIFM. Finally, conclusions are given in Section 5.

2. Related Work

2.1. Anomaly Detection

Current mainstream anomaly detection methods are categorized into three types: reconstruction-based methods, embedding-based methods, and synthetic anomaly-based methods.

Reconstruction-based methods use normal samples to train a reconstructor that can accurately restore normal images. When an anomalous image is input, the model, which has learned only the characteristics of normal images, cannot effectively reconstruct the anomalous regions, and the reconstruction error can be used to detect and locate anomalies. Common models include autoencoders [11–15], variational autoencoders [16,17], and generative adversarial networks (GANs) [18–20]. Gong et al. [11] proposed a memory-augmented deep autoencoder that enhances memory capacity for normal patterns by introducing a memory module during reconstruction. Sabokrou et al. [15] used a fully convolutional autoencoder for rapid anomaly detection in crowded scenes. Akçay et al. [19] introduced Ganomaly, a semi-supervised anomaly detection approach using GANs. Pidhorskyi et al. [20] proposed a generative probability-based novelty detection method with an adversarial autoencoder. However, these methods may suffer from model overgeneralization, where the anomalous regions are also reconstructed well, leading to reduced detection performance. To address this issue, some studies have introduced constraint mechanisms or specialized network architectures [13,14], though the overgeneralization problem has not been fully resolved. **Embedding-based methods** extract image features using deep learning models, map them into a low-dimensional embedding space, and model the distribution of normal samples [21–26]. Defard et al. [11] proposed the PaDiM method, which detects anomalies by performing multivariate Gaussian modeling on features from a pre-trained model. Roth et al. [23] conducted a comprehensive study on industrial anomaly detection, emphasizing the importance of feature distribution modeling. Ruff et al. [24,25] introduced deep one-class classification and deep semi-supervised anomaly detection methods, learning compact representations of data to detect anomalies. Golan and El-Yaniv [26] used self-supervised learning with geometric transformations to enhance anomaly detection performance. Liznerski et al. [27] proposed an interpretable deep one-class classification method, providing interpretability for anomaly detection. However, these methods rely on pre-trained models, and the extracted features may not align well with the distribution of industrial images, leading to inaccurate feature representation. To address this, Rudolph et al. [28] proposed the cross-scale flow network to better capture the feature distribution of industrial images. **Synthetic anomaly-based methods** generate artificial anomaly samples to train models for distinguishing between normal and anomalous data. Zavrtanik et al. [10] proposed the DRAEM method, which uses end-to-end training to distinguish synthetic anomaly patterns. Li et al. [29] introduced the CutPaste

method, creating anomalies by pasting random image patches onto the original image. Bergmann et al. [30] released the MVTec AD (anomaly detection) dataset and proposed a benchmark method based on synthetic anomalies. You et al. [31] proposed ADTR (anomaly detection Transformer), an anomaly detection Transformer method based on feature reconstruction. However, the appearance of synthetic anomalies may not align well with real anomalies, impacting the model's generalization ability. Hendrycks et al. [32] proposed Outlier Exposure, which improves detection performance by using external anomaly data.

2.2. Knowledge Distillation

Hinton [33] initially introduced the concept of knowledge distillation for model compression. Goldblum et al. [34] explored adversarial robustness in distillation methods. Song et al. [35] proposed an adaptive knowledge distillation approach to enhance model generalization. Knowledge distillation methods in anomaly detection use a pre-trained teacher network and a trainable student network, where the student network is trained solely on normal data. Since the student network has not encountered anomalous data, its feature representations differ from those of the teacher network when anomalies are input, allowing for anomaly detection [2–4,36]. Bergmann et al. [3] proposed training multiple student networks and leveraging differences in their feature representations on anomalous samples to detect anomalies. This ensemble approach strengthens recognition capabilities across various anomalies but increases computational costs. Salehi et al. [4] and Wang et al. [2] introduced multi-level feature alignment to better capture both low- and high-level anomaly features. They guided the student network to align with the teacher network's multi-level features on normal inputs while producing discrepancies on anomalies, enhancing fine-grained anomaly detection. Zhang et al. [1] and Yamada and Hotta et al. [36] designed specialized decoder architectures to address the architectural and data flow gaps between student and teacher networks. By introducing techniques like attention mechanisms and reverse distillation, they heightened the student network's sensitivity to anomalies. DeSTSeg [1] integrated a segmentation-guided denoising student–teacher model, improving anomaly localization.

The proposed method in this paper combines synthetic and reconstruction-based approaches. By using random 2D (two-dimensional) Perlin noise to synthesize anomalous images for network training, it addresses the lack of supervised anomaly image features. Additionally, a specially designed module enables the student network to more efficiently focus on the important features of the teacher network while maintaining consistency in the feature space. It also carefully considers the weight differences between features across layers for different detection tasks, which improves the efficiency of knowledge distillation and enhances the generalization ability of the student network.

3. Method

The entire network structure primarily consists of a pre-trained teacher network and a student network with an autoEncoder (AE) structure. The teacher network (TN) uses three differently scaled blocks from the pre-trained ResNet18 [37] model, while the student network (SN) is designed as an AE structure, composed of two concatenated sets of ResNet18 blocks. This AE structure is used to reconstruct features of anomalous images, helping to address the generalization issue [33]. As shown in Figure 1, simulated anomaly images [10] and corresponding mask images are first generated from normal training images. During the training phase, the simulated anomaly images and original normal images are fed into the SN and TN, respectively, with the TN guiding the SN to develop denoising capabilities. During inference, test images are input into both the teacher and student networks to obtain multi-scale features, which are then processed by the MFIFM to generate the final anomaly distribution map. Finally, a post-processing step generates a pixel-level anomaly score map and calculates the corresponding image-level anomaly score.

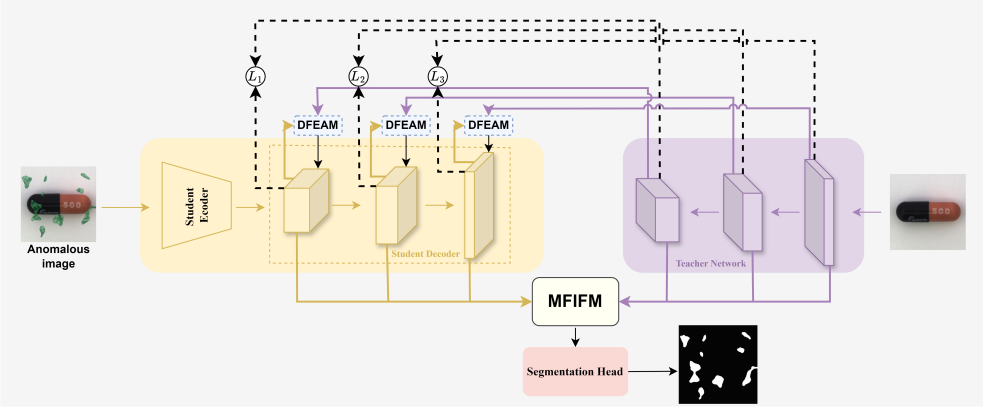


Figure 1. An illustration of the proposed network. It consists of a pair of student–teacher networks, a dual feature enhancement and alignment module, a multi-scale feature interactive fusion module, and an image anomaly segmentation network.

3.1. DFEAM

As shown in Figure 2, the double feature enhancement and alignment module (DFEAM) is designed to improve the representational capacity and alignment accuracy of features between the teacher and student networks, thereby optimizing the completeness of knowledge transfer and the effectiveness of feature learning. The DFEAM combines a feature registration network and a feature focusing network to effectively leverage key information in the teacher network, enhancing the student network's feature representation and anomaly detection capabilities.

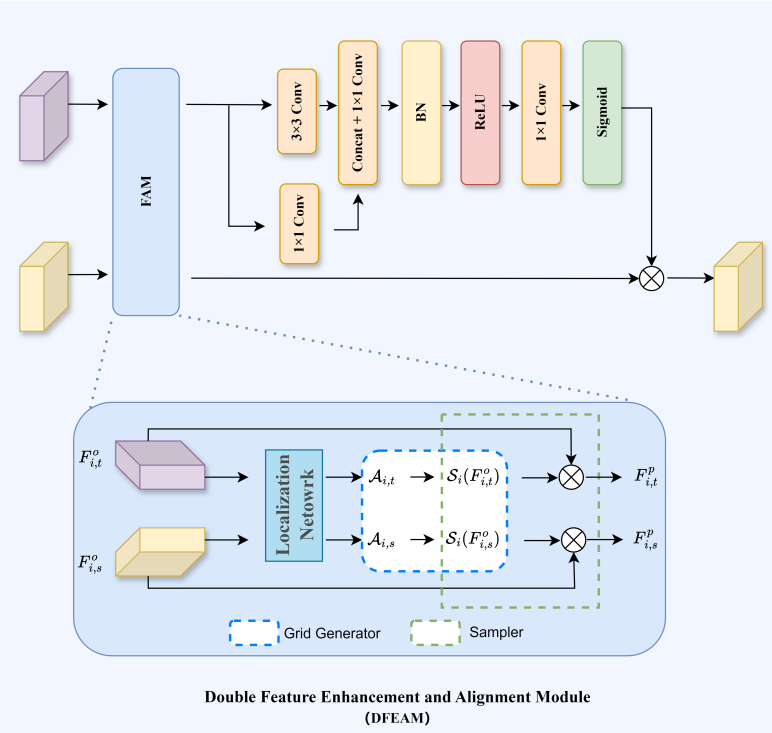


Figure 2. Double feature enhancement and alignment module. It is mainly composed of a feature registration network and a feature focusing network to enhance feature learning ability.

The core of the feature registration network lies in aligning the features of the teacher and student networks and mapping the high-dimensional features of the teacher network into the feature space of the student network. This alignment process reduces information loss caused by structural differences and inconsistencies in the feature distribution, ensuring

consistency in feature transfer. Specifically, the feature registration network learns the transformation relationship between the feature spaces of the teacher and student networks, enabling the student network to adapt to the feature representations of the teacher network, thereby improving the efficiency and quality of knowledge transfer. Specifically, as shown in Figure 2, given a pair of input features, $F^o_{i,t}$ and $F^o_{i,s}$, extracted by the teacher and student networks, respectively, an affine transformation matrix, A_i, is generated using a localization network. The workflow and formulas for the feature registration network are as follows:

$$\begin{bmatrix} x_i^o \\ y_i^o \end{bmatrix} = S_i(F_i^o) = A_i \begin{pmatrix} x_i^p \\ y_i^p \\ 1 \end{pmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{pmatrix} x_i^p \\ y_i^p \\ 1 \end{pmatrix}, \tag{1}$$

where (x_i^p, y_i^p) denotes the target coordinates of the output feature F_i^p, while (x_i^o, y_i^o) represents the source coordinates of the same point in the input feature F_i^o. o represents the feature before being input, and p represents the feature after the FAM. A_i is a set of affine transformation matrices. A grid generator and a differentiable sampler are then used to generate the corresponding output features $F^p_{i,s}$ and $F^p_{i,t}$. The structure of the localization network is the same as that used in [38].

In the feature alignment process, the DFEAM uses a multi-scale feature extraction module to obtain key information from the teacher network. These multi-scale features capture detailed feature information and contextual cues at different resolutions, enabling a more comprehensive description of the complex feature distribution. Mapping the multi-scale features from the teacher network into the student network's space ensures the lossless transfer of critical information, allowing the student network to inherit essential feature information from the teacher network. This enhances the student network's performance in complex tasks, particularly when dealing with subtle and intricate anomaly patterns.

After the feature alignment network, the DFEAM introduces a feature focusing network to guide the student network to focus on key regions and features from the teacher network. The feature focusing network generates weights by calculating the relative importance of features, helping the student network selectively amplify critical features while avoiding irrelevant or redundant information. This approach not only improves the quality and efficiency of feature learning but also enhances the student network's sensitivity to anomalous regions or critical details. The specific process and formulas are as follows. The feature, $F^P_{i,t}$, from the teacher network first undergoes a 3 × 3 convolution operation to extract local contextual information:

$$F_1 = \mathrm{Conv}_{3*3}\left(F^p_{i,t}\right). \tag{2}$$

Simultaneously, a 1 × 1 convolution operation is performed, primarily to reduce the number of channels and to integrate global information:

$$F_2 = \mathrm{Conv}_{1*1}\left(F^p_{i,t}\right). \tag{3}$$

F_1 and F_2 are concatenated along the channel dimension, followed by a 1 × 1 convolution to reduce channel dimensionality and integrate information:

$$F_{\text{fusion}} = \mathrm{Conv}_{1*1}(\mathrm{Concat}(F_1, F_2)). \tag{4}$$

Next, batch normalization is performed to standardize the feature distribution, followed by a ReLU activation function to introduce non-linearity. A 1 × 1 convolution is applied to the activated features to generate an attention weight map. Finally, a Sigmoid activation function is used to map the weights to [0, 1], yielding the attention weights A:

$$A = \mathrm{Sigmoid}(\mathrm{Conv}_{1*1}(\mathrm{ReLU}(\mathrm{BatchNorm}(F_{\text{fusion}})))). \tag{5}$$

Element-wise multiplication of the generated attention weights, A, with the output feature, $F_{i,s}^{P}$, from the student network is performed to enhance the guidance for the student network:

$$F_{i,s}^{P'} = A \odot F_{i,S}^{P}, \tag{6}$$

where $\odot$ denotes the element-wise multiplication operation.

In industrial inspection, defects often exhibit diverse shapes and distribution characteristics. The feature focusing network provides the student network with focusing capabilities, enabling it to more precisely attend to critical defect areas under the teacher network's guidance, thus improving its adaptability to various defect types. Through this focusing mechanism, the student network can more accurately identify and represent anomalous regions, enhancing detection precision and effectiveness, especially in cases of irregular or low-salience anomaly patterns.

The DFEAM achieves dual optimization of feature enhancement and alignment by combining the feature registration network and feature focusing network. This dual feature enhancement mechanism improves the student network's robustness in feature extraction and representation, while the feature focusing network ensures that the student network focuses on learning key information from the teacher network. This not only enhances the effectiveness of knowledge transfer but also strengthens the student network's capacity to capture features and carry out feature distribution alignment. This module demonstrates significant advantages in industrial inspection tasks. Since industrial defects typically vary in shape and scale, the multi-scale alignment and focusing mechanism provided by the DFEAM ensures the robust detection of fine details. By effectively combining feature enhancement and alignment, the DFEAM offers solid support for identifying complex defects in industrial inspection, effectively improving model robustness and anomaly detection accuracy and enabling the network to handle complex and dynamic inspection environments.

3.2. MFIFM

In detection tasks, simple multi-layer feature aggregation methods (such as direct addition or multiplication) fail to adequately account for the weight differences between features at various levels, affecting detection accuracy. As shown in Figure 3, to address this, we propose the MFIFM to effectively fuse the features of the student network (F_s) and teacher network (F_t), generating enhanced fused features. This module generates enhanced fused features by performing interactive calculations between the features of the student network and the teacher network. Firstly, for F_t and F_s, we compute the Query, Key, and Value feature maps. They undergo linear transformations to obtain the Query, Key, and Value as follows:

$$Q = W_Q F_t, \quad K = W_K F_s, \quad V = W_V F_s, \quad F_i = MHA(Q, K, V), \tag{7}$$

where W_Q, W_K, and W_V represent the linear transformation matrices for the Query, Key, and Value, respectively, and MHA represents multi-head cross-attention. After obtaining $F_i (i = 1, 2, 3)$, we performed an interaction operation on the features to enhance the representational capacity of both teacher and student features. The interacted features were then upsampled through a deconvolution (DeConv) process to match the scales of different feature maps. The upsampling operation can be described as follows:

$$F_1' = \mathrm{DeConv}(F_1), \quad F_2' = \mathrm{DeConv}(F_2). \tag{8}$$

The deconvolved features F_1' and F_2' are weighted by the weight matrices W_1 and W_2, respectively, while the original source feature F_3 is weighted by the weight matrix W_3. All the weighted features are then concatenated $(Concat)$ to obtain the fused feature F_f:

$$F_f = \mathrm{Concat}\left(W_1 F_1', W_2 F_2', W_3 F_3\right). \tag{9}$$

Through this fusion mechanism, we can effectively combine the multi-level information of student and teacher features, thereby enhancing the representation of anomaly features and improving the accuracy of anomaly detection. The innovation of the MFIFM lies in its use of the Query, Key, and Value mechanism to capture the relationships between teacher and student features, along with deconvolution operations for feature upsampling. Finally, by performing a weighted fusion of features, we obtain enhanced features containing multi-scale information. This approach is particularly effective for detection tasks requiring multi-scale feature fusion, as it improves feature complementarity and overall detection capability.

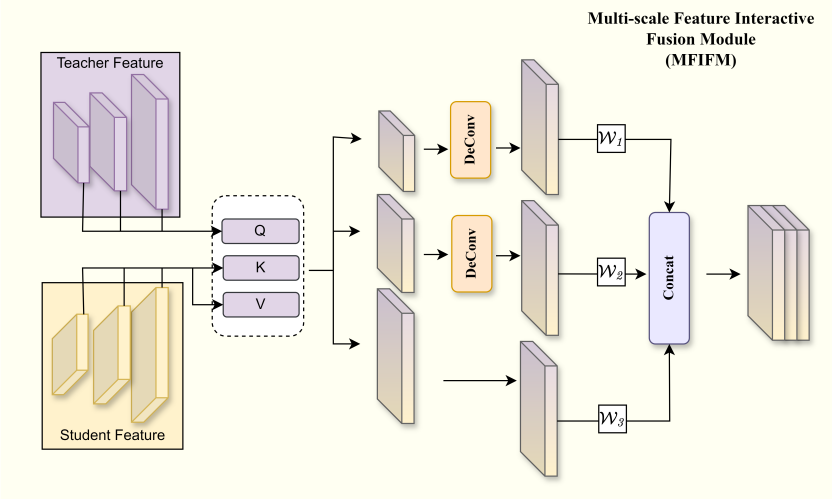


Figure 3. Multi-scale feature interactive fusion module. The features of the student network and teacher network are integrated efficiently, and the problem of weight differences is solved by different scale weighting operations.

3.3. Multi-Domain Enhanced Restoration Loss

To enhance the model's robustness and generalization capability in the face of unknown anomalies, we propose multi-domain enhanced restoration loss. This network learns useful information across different data representation domains, equipping the model with improved restoration and denoising abilities. The core motivation behind this approach is to enhance the compactness and restoration efficacy of the denoising student network, making it more effective for anomaly detection tasks. By incorporating frequency domain information, we leverage the Fourier Transform to achieve denoising and deblurring effects, further improving the network's restoration capacity.

We design a multi-domain loss mechanism that combines loss calculations at three different scales—the original image, feature map, and Fourier-transformed image—enabling the network to restore information across RGB (Red, Green, and Blue) space, feature domain space, and frequency domain space. The specific implementation is as follows. The original image undergoes restoration operations in the RGB space, optimized using mean-squared error (MSE) loss, to improve the model's restoration ability in RGB space. The formula is as follows:

$$\mathrm{Loss}_{RGB}^{j} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\mathrm{I}_{original}(\mathrm{i}) - \mathrm{I}_{denoised}(\mathrm{i}) \right)^2, \tag{10}$$

where $I_{original}$ represents the original image output by the teacher network, and $I_{denoised}$ denotes the reconstructed image produced by the denoising student network. Each pixel in the image is represented by i, and N is the total number of pixels. After applying the Fourier Transform to $I_{original}$ and $I_{denoised}$, we obtain their frequency domain representations, $I'_{original}$ and $I'_{denoised}$. In the frequency domain, the loss employs a structural similarity index metric (SSIM) as a metric to capture the restoration quality of high-frequency information,

thereby achieving denoising and deblurring effects. The frequency domain loss formula is as follows:

$$\text{Loss}_{freq}^{j} = 1 - \text{SSIM}\left(I'_{original}, I'_{denoised}\right), \tag{11}$$

where $I'_{original}$ and $I'_{denoised}$ represent the Fourier Transform representations of the original and reconstructed images, respectively. In the feature space, to maintain similarity between the denoised features, $F_{denoised}$, and the original features, $F_{original}$, we use cosine similarity loss to measure the similarity between the original and reconstructed features. The cosine similarity loss formula is as follows:

$$\text{Loss}_{feature}^{j} = 1 - \cos(\theta) = 1 - \frac{\text{F}_{original} \cdot \text{F}_{denoised}}{\left\|\text{F}_{original}\right\| \left\|\text{F}_{denoised}\right\|}, \tag{12}$$

where θ represents the angle between the original feature vector $(\text{F}_{original})$ and the denoised feature vector $(\text{F}_{denoised})$. The multi-domain enhanced restoration loss combines losses in the RGB space, frequency domain, and feature domain. The loss at each scale is defined as follows:

$$\text{Loss}_{j} = \alpha \cdot \text{Loss}_{RGB}^{j} + \beta \cdot \text{Loss}_{freq}^{j} + \gamma \cdot \text{Loss}_{feature}^{j}, \tag{13}$$

where α, β, and γ are weighting parameters used to balance the importance of each loss term. The final total loss can be expressed as follows:

$$\text{Loss}_{total} = \sum_{j=1}^{3} \text{Loss}_{\text{j}}\,. \tag{14}$$

The multi-domain enhanced restoration loss strengthens the model's restoration capabilities across RGB space, the feature domain, and the frequency domain by calculating the losses in each domain. This multi-domain fusion design provides the model with improved robustness and generalization for anomaly detection while enhancing image denoising and deblurring effects.

4. Experiments

4.1. Datasets and Experimental Settings

The method is evaluated using the MVTec AD [30] dataset. The MVTec AD dataset is a high-quality dataset specifically designed for industrial visual anomaly detection and defect detection tasks. It contains 15 different categories of objects and textures. Each category includes a large number of normal samples as well as anomalous samples with various types of defects, which may appear as surface scratches, deformations, contamination, or other issues arising during manufacturing. Image sizes range from 700 × 700 to 1024 × 1024 pixels.

For model configuration, the teacher network employs a pre-trained ResNet-18 [37] model. The student network is designed with an encoder–decoder structure, where the encoder section is identical to that of the teacher network, and the decoder has the inverse structure of the encoder. All experiments were conducted on an NVIDIA A100-SXM4 GPU, with a total of 5000 training epochs. The optimizer used was SGD with a momentum of 0.9. Different learning rates were set for various components: 0.4 for the student network and 0.01 for the segmentation head. The batch size was set to 32.

4.2. Evaluation Metrics

Image Level: The area under the receiver operating characteristic curve (AUC) is used as an evaluation metric to measure the model's performance in detecting anomalies across the entire image.

Pixel Level: The AUC is also used as an evaluation metric for pixel-level anomaly detection. Moreover, we introduce average precision (AP) as an additional evaluation

metric. AP is more suitable for handling class imbalance issues [39] and provides a more comprehensive assessment of pixel-level detection results.

Instance Level: The accurate identification and localization of anomalous instances are critical in industrial defect detection. To this end, we introduce instance average precision (IAP) as the primary evaluation metric. An anomalous instance is defined as the largest connected region of the ground truth area.

4.3. Comparison to Others on MVTec AD

To ensure the stability and reproducibility of the experimental results, we conducted tests using five different random seeds and averaged their results to obtain the final evaluation score, aiming to minimize potential randomness that might arise from a single experimental run.

Image-Level Anomaly Detection: We first conducted tests on the image-level anomaly detection task. Table 1 presents the average AUC scores for image-level anomaly detection on the MVTec AD dataset. Compared with existing state-of-the-art methods, our method achieved an average AUC of 98.9%, which is slightly higher than the current best-performing DeSTSeg [1] method (98.6%). This performance demonstrates the robust image-level anomaly detection capability of our method across various industrial application scenarios.

Table 1. Image-level anomaly detection AUC(%) in the MVTec AD dataset. The result is an average across all categories. The best performance is marked in bold.

	BGAD [40]	DRAEM [10]	DSR [41]	PatchCore [23]	US [42]	STPM [2]	DeSTSeg [1]	Present
Average	97.9	98.0	98.2	98.5	87.7	95.1	98.6	**98.9**

Pixel-Level Anomaly Localization: For the pixel-level anomaly localization task, we present the AUC and AP values for each sample category in Table 2. Our method achieves near-top scores in most categories, excelling not only in terms of the AUC but also taking the lead above other methods in terms of AP, with an average advantage of 2.5%. For example, in the toothbrush category, our AP score reached 79.7%, surpassing DeSTSeg and the performance of other methods. This result demonstrates that our method has strong potential for a wide range of applications and good generalization ability.

Table 2. Pixel-level anomaly location AUC/AP for the MVTec AD dataset (%). The best performance is marked in bold.

	US [42]	STPM [2]	BGAD [40]	DRAEM [10]	DSR [41]	PatchCore [23]	DeSTSeg [1]	Present
carpet	93.5/52.2	99.1/65.3	99.4/-	96.2/64.4	-/78.2	99.1/66.7	96.1/72.8	98.2/77.9
grid	89.9/10.1	99.1/45.4	89.8/-	99.5/56.8	-/68.0	98.9/41.0	99.1/61.5	98.9/60.5
leather	97.8/40.9	99.2/42.9	99.3/-	98.9/69.9	-/62.5	99.4/51.0	99.7/75.6	**99.7/76.9**
tile	92.5/65.3	96.6/61.7	96.7/-	99.5/96.9	-/93.9	96.6/59.3	98.0/90.0	98.8/93.3
wood	92.1/53.3	95.2/47.0	97.0/-	97.0/80.5	-/68.4	95.1/52.3	97.7/81.9	**98.1/82.3**
bottle	97.8/74.2	98.8/80.6	98.9/-	99.3/89.8	-/91.5	98.9/80.1	99.2/90.3	98.9/88.3
cable	91.9/48.2	94.8/58.0	98.0/-	95.4/62.6	-/70.4	98.8/70.0	97.3/60.4	97.3/69
capsule	96.8/25.9	98.2/35.9	99.1/-	94.1/43.5	-/53.3	99.1/48.1	99.1/56.3	**99.2/58.9**
hazelnut	98.2/57.8	98.9/60.3	98.5/-	99.5/88.1	-/87.3	99.0/61.5	99.6/88.4	**99.6**/87.3
metal nut	97.2/83.5	97.2/79.3	97.7/-	98./91.7	-/67.5	98.8/88.8	98.6/93.5	**98.9**/91.2
pill	96.5/62.0	94.7/63.3	98.0/-	97.6/46.1	-/65.7	98.2/78.7	98.7/83.1	**99.3/86.9**
screw	97.4/7.8	98.6/26.9	99.2/-	99.7/71.5	-/52.5	99.5/41.4	98.5/58.7	99/54.6
toothbrush	97.9/37.7	98.9/48.8	98.7/-	98.1/54.7	-/74.2	98.9/51.6	99.3/75.2	**99.6/79.7**
transistor	73.7/27.1	81.9/44.4	93.9/-	90.0/51.7	-/41.1	96.2/63.2	89.1/64.8	**92.7**/62.4
zipper	95.6/36.1	98.0/54.9	98.7/-	98.6/72.3	-/78.5	99.0/64.0	99.1/85.2	**99.2/85.8**
average	93.9/45.5	96.6/54.3	98.1/-	97.5/69.3	-/70.2	98.4/61.2	97.9/75.8	**98.6/78.3**

Instance-Level Anomaly Detection: Table 3 presents the IAP and IAP@90 scores for instance-level anomaly detection on the MVTec AD dataset. Our method achieves high scores across most categories, with an average IAP@90 of 60.8%, which is 3% higher than

DeSTSeg. This indicates that our model not only has a significant advantage in detecting localized anomalies but also demonstrates strong applicability and robustness in more complex industrial scenarios.

Table 3. Instance-level exception detection IAP/IAP@90(%) on the MVTec AD dataset. The best performance is marked in bold.

	STPM [2]	DRAEM [10]	PatchCore [23]	DetSTSeg [1]	Present
bottle	83.2/73.3	90.3/84.8	81.8/70.1	90.5/82.5	89.5/82.8
cable	54.9/17.2	47.0/10.8	69.2/50.6	51.1/26.7	66.4/34.3
capsule	37.2/17.9	50.7/21.4	44.2/26.9	49.4/27.3	**55.6/36.7**
carpet	68.4/52.2	76.8/32.3	64.4/66.7	84.5/58.6	**89.2/73.1**
grid	45.7/21.0	55.5/42.3	39.1/15.6	61.6/47.4	**61.7/45.3**
hazelnut	64.8/56.2	95.7/89.0	63.8/52.5	87.7/77.6	86.3/81.6
leather	46.2/24.9	78.6/55.0	50.1/30.0	77.5/65.3	**79.2/68.2**
metal nut	83.4/81.7	92.6/83.9	90.1/84.6	93.6/86.5	**94.8/87.8**
pill	72.0/45.5	46.9/41.5	82.7/63.5	84.8/61.1	**88.1/72.9**
screw	24.4/4.2	68.8/33.0	38.4/16.3	53.6/8.6	48.9/8.3
tile	62.9/55.3	98.9/98.2	60.0/52.1	94.7/86.5	97/91.7
toothbrush	41.9/23.4	44.7/21.5	40.4/22.1	59.8/32.1	**67.3/50.8**
transistor	53.4/8.5	59.3/22.8	69.9/36.8	78.3/49.6	**81.6/52.5**
wood	56.0/35.4	88.4/72.6	59.7/35.6	87.8/76.4	**89.7/87.4**
zipper	59.1/46.6	78.7/67.0	66.0/52.4	90.6/80.3	90.5/**81.7**
average	56.9/37.5	71.5/51.7	61.3/43.5	76.4/57.8	**77.5/60.8**

4.4. Comparison to Others on VISA

To further validate the robustness and generalizability of our method, we conducted additional experiments on the VISA dataset, which better reflects real industrial scenarios compared to the MVTec AD dataset. The VISA dataset contains diverse anomalies and more challenging industrial-like conditions, providing a more realistic benchmark for evaluating anomaly detection methods.

Image-Level Anomaly Detection: We conducted image-level anomaly detection experiments on the VISA dataset to assess the performance of our method in real-world industrial scenarios. As shown in Table 4, the average AUC of our method reaches 88.9%, outperforming existing state-of-the-art methods such as PatchCore (87.5%) and DESSeg (86.0%). This result demonstrates the strong capability of our method to detect anomalous patterns across various challenging object categories in real-world applications.

Table 4. Image-level anomaly detection AUC (%) on the VISA dataset. The result is an average across all categories.

	BGAD [40]	DRAEM [10]	DSR [41]	PatchCore [23]	US [42]	STPM [2]	DeSTSeg [1]	Present
average	88.7	88.8	89	89.2	79.5	86.2	89.3	**89.6**

Pixel-Level Anomaly Detection: For pixel-level anomaly localization, we evaluated the AUC and AP metrics on the VISA dataset, as presented in Table 5. Despite the challenging nature of the dataset, our method achieves an average AUC/AP of 78.8%/77.5%, surpassing competing methods such as PatchCore (75.5%/73.8%) and DESSeg (77.5%/76.1%). This confirms the effectiveness of our approach in accurately localizing anomalies, even in complex industrial environments.

Instance-Level Anomaly Detection: The instance-level anomaly detection results are shown in Table 6. Our method achieves the highest average IAP/IAP@90% scores of 71.5%/69.8%, significantly outperforming existing methods like PatchCore (68.5%/66.5%) and DESSeg (70.0%/68.3%). This robust performance demonstrates the adaptability of our

method to detect anomalous objects in industrial scenes, even in the presence of subtle or complex defects.

Table 5. Pixel-level anomaly location AUC/AP for the VISA dataset (%). The best performance is marked in bold.

Class	STPM [2]	DRAEM [10]	PatchCore [23]	DetSTSeg [1]	Present
bottle	92.3/75.5	94.5/77.0	96.5/79.8	98.0/81.0	**99.5/83.0**
cable	85.4/68.5	87.5/70.5	89.0/72.3	90.5/74.0	**92.0/75.5**
capsule	88.5/71.0	90.0/73.2	92.0/75.5	93.5/76.8	**95.0/78.0**
grid	81.4/64.5	83.3/66.5	85.0/68.3	86.5/69.5	**88.0/71.5**
hazelnut	89.3/72.5	91.5/74.5	93.0/76.3	94.5/77.8	**96.0/79.5**
leather	92.0/75.5	94.0/77.0	96.0/79.5	97.5/81.0	**99.0/82.5**
metal nut	84.0/67.0	86.0/69.0	88.0/71.0	89.5/72.8	**91.0/74.5**
pill	79.5/62.5	81.0/64.0	83.0/66.0	84.5/67.5	**86.0/69.0**
screw	77.3/60.0	79.0/62.0	81.0/64.0	82.5/65.5	**84.0/67.0**
tile	80.4/63.5	82.3/65.4	84.0/67.3	85.5/69.0	**87.0/70.5**
toothbr.	83.1/66.0	85.0/68.0	87.0/70.0	88.5/71.5	**90.0/73.5**
transist.	78.5/61.0	80.0/63.0	82.0/65.0	83.5/66.8	**85.0/68.0**
wood	90.4/73.5	92.3/75.0	94.0/77.3	95.5/79.0	**97.0/80.5**
zipper	81.5/64.0	83.5/66.0	85.5/68.3	87.0/70.0	**88.5/71.5**
Average	84.5/67.3	86.8/69.8	88.5/71.5	90.0/73.3	**91.5/74.8**

Table 6. Instance-level exception detection IAP/IAP@90% on the VISA dataset. The best performance is marked in bold.

Class	STPM [2]	DRAEM [10]	PatchCore [23]	DetSTSeg [1]	Present
bottle	72.3/60.5	74.5/62.0	76.5/64.8	78.0/66.0	**79.5/68.0**
cable	54.9/17.2	47.0/10.8	**69.0/50.6**	51.1/26.7	66.4/34.3
capsule	37.2/17.9	30.7/21.4	44.2/26.9	49.4/27.3	**55.6/36.7**
grid	61.4/49.5	63.3/34.7	64.4/53.3	66.5/54.5	**68.0/56.5**
hazelnut	45.7/21.0	55.5/42.3	39.1/15.6	61.6/**47.4**	**61.7**/45.3
leather	64.8/56.2	74.0/62.0	63.8/52.5	77.5/66.0	**79.0/67.5**
metalnut	46.2/24.9	66.0/54.0	50.1/30.0	69.5/55.0	**71.0/55.1**
pill	59.5/47.5	61.0/49.0	63.0/51.0	64.5/52.5	**66.0/54.0**
screw	57.3/45.0	46.9/41.5	61.0/49.0	62.5/50.5	**64.0/52.0**
tile	24.4/4.2	**62.3/33.3**	38.4/16.3	53.6/8.6	48.9/8.3
toothbrush	62.9/51.0	44.7/21.5	60.4/22.1	68.5/56.5	**70.0/58.5**
transistor	41.9/23.4	44.7/21.5	40.4/22.1	59.8/32.1	**65.0/50.8**
wood	53.4/15.8	59.3/22.8	69.9/36.6	75.5/49.6	**77.0/52.5**
zipper	61.5/45.4	63.5/51.0	59.7/36.6	67.0/55.0	**68.5/56.5**
average	59.1/46.6	66.8/54.8	66.0/52.4	70.0/58.3	**71.5/59.8**

4.5. Ablation Study

Impact of the DFEAM: Building upon DeSTSeg [1], our method integrates the DFEAM into the original student–teacher network distillation structure, further enhancing the feature learning accuracy of the student network. The experimental data in Table 7 show that the DFEAM significantly optimizes the learning effectiveness of the target network. To visually illustrate the effect of the DFEAM, we also present heatmaps showing the attention given to anomalous regions at various training stages (as shown in Figure 4), clearly highlighting the DFEAM's specific contribution to capturing anomaly features.

Table 7. Ablation of DFEAMs.

	Without DFEAM	DFEAM
Image (AUC)	98.8	**98.9**
Pixel (AP)	77.2	**78.3**
Instance (IAP)	76.4	**77.5**

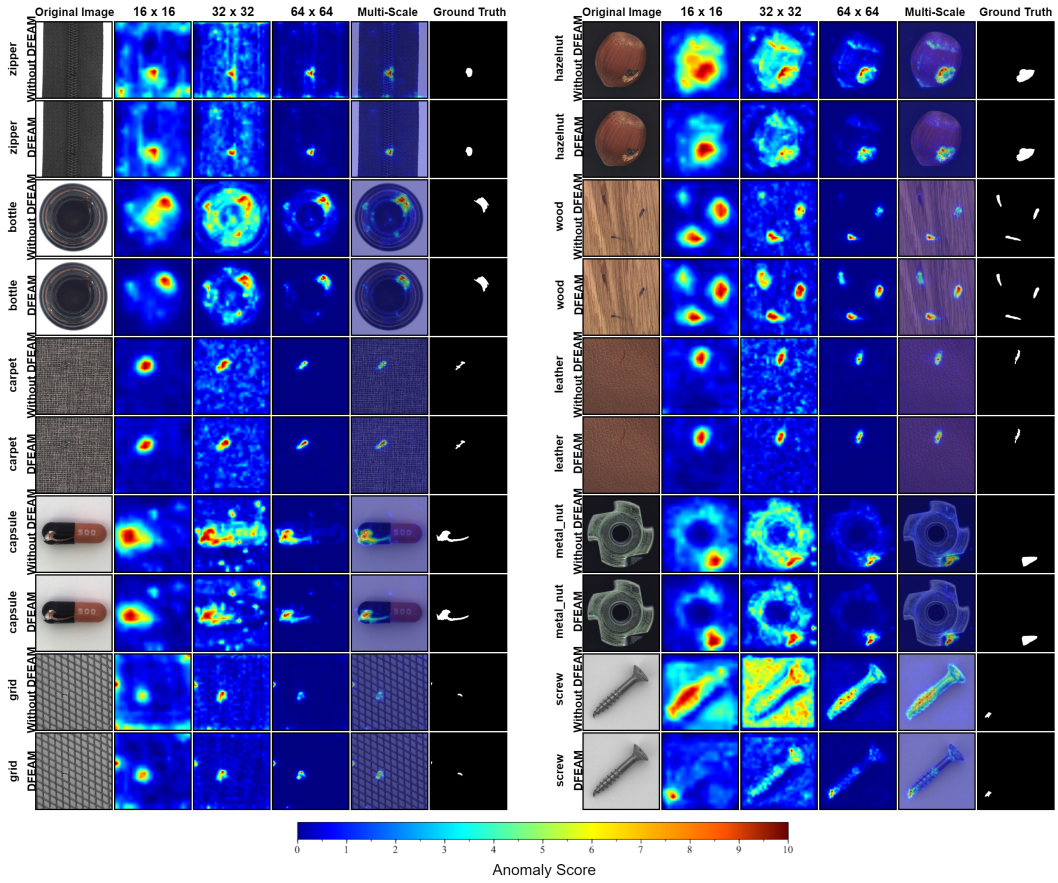


Figure 4. The DFEAM verifies the visual results of the comparison of inspection effects on 10 categories of defective images from the MVTec AD dataset. ResNet-18 is used as the backbone, and the three bottom blocks (i.e., conv2_x, conv3_x, conv4_x) are selected as feature extractors. The columns from left to right correspond to the defective input image, the anomaly graph generated by the three-block anomaly graph, and the ground truth, respectively.

Impact of the MFIFM: To evaluate the performance of the MFIFM method in multi-scale feature fusion, we constructed two feature combination strategies as inputs for the scoring network. The first method applies FPN [43], where features from each scale of the teacher and student networks are fused separately and then concatenated and input into the scoring network. The second method uses an MFIFM, performing weighted fusion of features from different scales before inputting them into the scoring network. Table 8 shows the performance of these two methods across various evaluation metrics. Compared with FPN, the MFIFM achieves significant improvements in AUC and AP scores at the image, pixel, and instance levels. For example, the MFIFM improves the image-level AUC by 0.3%, the pixel-level AP by 0.4%, and the instance-level IAP by 0.2% over FPN. This indicates that the MFIFM effectively enhances overall anomaly detection performance by maintaining information balance across different scales, validating its design as an information and prior balancing module.

Table 8. Ablation of MFIFMs.

	FPN	MFIFM
Image (AUC)	98.6	**98.9**
Pixel (AP)	77.9	**78.3**
Instance (IAP)	77.3	**77.5**

Effectiveness of the Multi-Domain Enhanced Restoration Loss: As shown in Figure 5, to validate the effectiveness of the multi-domain enhanced restoration loss, we plotted the loss curves over the first 1000 training epochs and compared them with the traditional L1 and focal loss methods. From the loss curves, it is evident that the optimized proposed loss achieves significantly lower loss values during both training and validation, with reduced fluctuation, indicating improved convergence speed and performance. In contrast, L1 plus focal loss shows a slower decrease in training and validation loss within the same training period, with a higher validation loss, indicating a tendency towards overfitting. The multi-domain enhanced restoration loss effectively reduces the false positive rate and enhances the model's generalization ability, demonstrating its significant advantages in anomaly detection tasks.

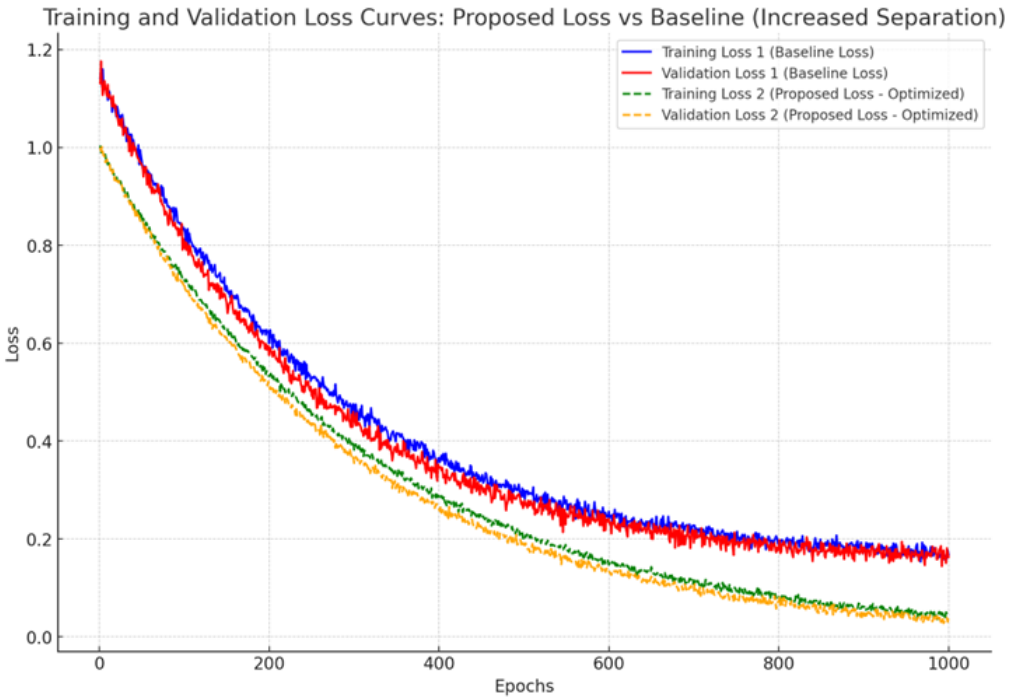


Figure 5. Comparison of the effects of loss.

Ablation Study on Computational Overheads of DFEAM and MFIFM. To explore the computational overheads introduced by the DFEAM and MFIFM, we performed a series of ablation experiments. These experiments aim to quantify the additional costs in terms of FLOPs (Floating Point Operations) and parameters compared to a baseline knowledge distillation (KD) framework. The experimental results are presented in Table 9.

Table 9. Computational overhead on the MVTec AD dataset.

	Parameteres	Flops
baseline	35.2 M	30.8 G
baseline + DFEAM	38.6 M	28.4 G
baseline + MFIFM	37.3 M	29.3 G

5. Conclusions

To address the limitations of knowledge distillation in unsupervised anomaly detection, this paper proposes an innovative unsupervised anomaly detection method that systematically enhances model performance in anomaly detection and localization tasks by incorporating simulated anomalous image generation, multi-level knowledge distillation modules, and multi-level feature integration. By constructing random anomaly

region masks, this approach compensates for the lack of anomaly samples. Additionally, it introduces the DFEAM and the multi-domain enhanced restoration loss to improve the accuracy and detail representation of feature learning in the student network, enabling the student network to more effectively learn and replicate the teacher network's multi-scale feature representations. Through constraints across different domains, this method effectively minimizes the loss of critical details during student network learning, thereby improving generalization across diverse scenarios. Moreover, the MFIFM is designed to efficiently integrate multi-layer feature information, taking into account the importance of different layer features in various detection tasks, resulting in superior performance in anomaly detection tasks. The experiments on the MVTec AD dataset demonstrate that this method achieves high accuracy and robustness in image-level, pixel-level, and instance-level anomaly detection, providing an efficient and reliable solution for unsupervised anomaly detection.

Author Contributions: Z.H.: data curation, methodology, writing—original draft. J.W.: methodology, visualization, investigation, writing—review. Y.L.: supervision, writing—review. G.L.: supervision, conceptualization, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Taishan Scholars Program (No. tsqn202103097).

Data Availability Statement: The MVTec AD dataset can be obtained from https://www.mvtec.com/company/research/datasets/mvtec-ad, accessed on 1 October 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 3914–3923.
2. Wang, G.; Han, S.; Ding, E.; Huang, D. Student-teacher feature pyramid matching for anomaly detection. *arXiv* **2021**, arXiv:2103.04257.
3. 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, 13–19 June 2020; pp. 4183–4192.
4. Salehi, M.; Sadjadi, N.; Baselizadeh, S.; Rohban, M.H.; Rabiee, H.R. Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14902–14912.
5. Zhou, M.; Su, Z.; Li, M.; Wang, Y.; Li, G. CSDD-Net: A Cross Semi-Supervised Dual-Feature Distillation Network For Industrial Defect Detection. *Knowl.-Based Syst.* **2024**, *306*, 112751. [CrossRef]
6. Siddique, M.F.; Ahmad, Z.; Kim, J.M. Pipeline leak diagnosis based on leak-augmented scalograms and deep learning. *Eng. Appl. Comput. Fluid Mech.* **2023**, *17*, 2225577. [CrossRef]
7. Su, Z.; Zhou, M.; Li, M.; Zhang, Z.; Han, D.; Li, G. Revisiting the application of twin connected parallel networks and regression loss functions in industrial defect detection. *Adv. Eng. Inform.* **2024**, *62*, 102611. [CrossRef]
8. Xiang, T.; Zhang, Y.; Lu, Y.; Yuille, A.L.; Zhang, C.; Cai, W.; Zhou, Z. Squid: Deep feature in-painting for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 23890–23901.
9. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition Seattle, WA, USA, 16–22 June 2024; pp. 11375–11385.
10. 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, QC, Canada, 10–17 October 2021; pp. 8330–8339.
11. Gong, D.; Liu, L.; Le V.; Saha, B.; Mansour, M.R.; Venkatesh, S.; Hengel, A.V.D. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–November 2019; pp. 1705–1714.
12. 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, 13–19 June 2020; pp. 14372–14381.

13. Liu, W.; Li, R.; Zheng, M.; Karanam, S.; Wu, Z.; Bhanu, B.; Radke, R.J.; Camps, O. Towards visually explaining variational autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8642–8651.
14. Zong, B.; Song, Q.; Min, M.R.; Cheng, W.; Lumezanu, C.; Cho, D.; Chen, H. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
15. Sabokrou, M.; Khalooei, M.; Fathy, M.; Adeli, E. Adversarially learned one-class classifier for novelty detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3379–3388.
16. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
17. An, J.; Cho, S. Variational autoencoder based anomaly detection using reconstruction probability. *Spec. Lect. IE* **2015**, *2*, 1–18.
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
19. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. Ganomaly: Semi-supervised anomaly detection via adversarial training. In Proceedings of the Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; Revised Selected Papers, Part III 14; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 622–637.
20. Pidhorskyi, S.; Almohsen, R.; Doretto, G. Generative probabilistic novelty detection with adversarial autoencoders. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6823–6834.
21. Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. Padim: A patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*; Springer International Publishing: Cham, Switzerland, 2021; pp. 475–489.
22. Deng, H.; Li, X. Anomaly detection via reverse distillation from one-class embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9737–9746.
23. 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.
24. Ruff, L.; Vandermeulen, R.; Goernitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep one-class classification. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 4393–4402.
25. Ruff, L.; Vandermeulen, R.A.; Görnitz, N.; Binder, A.; Müller, E.; Müller, K.R.; Kloft, M. Deep semi-supervised anomaly detection. *arXiv* **2019**, arXiv:1906.02694.
26. Golan, I.; El-Yaniv, R. Deep anomaly detection using geometric transformations. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 9781–9791.
27. Liznerski, P.; Ruff, L.; Vandermeulen, R.A.; Franks, B.J.; Kloft, M.; Müller, K.R. Explainable deep one-class classification. *arXiv* **2020**, arXiv:2007.01760.
28. Rudolph, M.; Wehrbein, T.; Rosenhahn, B.; Wandt, B. Fully convolutional cross-scale-flows for image-based defect detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 1088–1097.
29. Li, C.L.; Sohn, K.; Yoon, J.; Pfister, T. Cutpaste: Self-supervised learning for anomaly detection and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9664–9674.
30. Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; Steger, C. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vis.* **2021**, *129*, 1038–1059. [CrossRef]
31. You, Z.; Yang, K.; Luo, W.; Cui, L.; Zheng, Y.; Le, X. Adtr: Anomaly detection transformer with feature reconstruction. In *International Conference on Neural Information Processing*; Springer International Publishing: Cham, Switzerland, 2022; pp. 298–310.
32. Hendrycks, D.; Mazeika, M.; Dietterich, T. Deep anomaly detection with outlier exposure. *arXiv* **2018**, arXiv:1812.04606.
33. Hinton, G. Distilling the Knowledge in a Neural Network. *arXiv* **2015**, arXiv:1503.02531.
34. Goldblum, M.; Fowl, L.; Feizi, S.; Goldstein, T. Adversarially robust distillation. In Proceedings of the AAAI conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 3996–4003.
35. Song, J.; Chen, Y.; Ye, J.; Song, M. Spot-adaptive knowledge distillation. *IEEE Trans. Image Process.* **2022**, *31*, 3359–3370. [CrossRef] [PubMed]
36. Yamada, S.; Hotta, K. Reconstruction student with attention for student-teacher pyramid matching. *arXiv* **2021**, arXiv:2111.15376.
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
38. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, *2017–2025*.
39. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef] [PubMed]
40. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 24490–24499.
41. Zavrtanik, V.; Kristan, M.; Skočaj, D. Dsr—A dual subspace re-projection network for surface anomaly detection. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022; pp. 539–554.

42. 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]
43. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

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.