2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

A Coarse-Precise Refinement Learning-Based Knowledge Distillation Network for Anomaly Detection

Chaoyang Li[1,2[0000-0002-7435-7267]], Haozheng Zhang[1,2[0009-0004-3009-5570]], Fei Wang[1,2[0009-0006-4030-8255]], Chengkun Li[1,2[0009-0003-2400-7637]] and Yanhong Yang[1,2,3(✉) [0000-0003-4547-4659]]

[1] The Engineering Research Center of Learning-Based Intelligent System (Ministry of Education), Tianjin University of Technology, Tianjin 300384, China;
[2] Key Laboratory of Computer Vision and Systems (Ministry of Education), Tianjin University of Technology, Tianjin 300384, China;
[3] Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China;
yyh@email.tjut.edu.cn

Abstract. Anomaly detection serves a crucial role in large-scale industrial manufacturing. Knowledge distillation (KD)-based approaches have demonstrated excellent performance, yet their efficacy is constrained by the identical symmetric structures. In this study, we propose an enhanced KD-based architecture with a dual-learning mechanism, called DLKD, to precisely characterize normal samples and improve detection performance. Specifically, we first introduce a coarse decoder into the student network to preliminarily reconstruct the teacher features, in which the SSM-based global feature reconstruction block (GFRB) and CNN-based local feature reconstruction block (LFRB) effectively model global and local information. A precise refinement learner is subsequently provided to finely tune the coarse reconstructed features. Extensive experiments on two publicly available anomaly detection datasets demonstrate the effectiveness and potential of the proposed DLKD. This work further explores KD-based methods for anomaly detection and provides a unique yet robust baseline for the community.

Keywords: Anomaly detection, Knowledge distillation, Dual-learning, Feature reconstruction.

1 Introduction

Visual anomaly detection (AD) is a challenging problem, as fully specifying expected defect variations is both costly and complicated. In most real-world scenarios, images with anomalies are rare, and formerly unseen defect types may appear at unpredictable times. So, constructing the manifold of normal data for out-of-distribution detection has become the mainstream approach [11]. Unsupervised learning paradigm, denoted as one-class classification, which uses only normal data, achieves promising results on this challenging problem but come with a latency trade-off.

Knowledge distillation (KD)-based frameworks have been extensively explored since they enable knowledge to be transferred within a teacher-student (T-S) pair. Since the student network has never encountered anomaly samples, it is expected that the feature representations it generates for anomalous samples will differ from those produced by the teacher network. Consequently, the anomalies can be detected in the inference phase. To increase the discriminating capability, US [2] uses an ensemble learning method to guide training on normal data. STFPM [13] applies KD across various levels of the feature pyramid, effectively aggregating differences from multiple levels and demonstrating strong performance. In RD [3], a reverse distillation paradigm is realized to address the non-distinguishing filter problem in conventional anomaly detection models. The drawback of KD-based models is that they are unable to capture long-range dependencies, and as a result, they may miss anomalies in complex patterns [6]. Furthermore, since there are no constraints on abnormal samples during the training process, the model generalizes well for anomaly reconstruction, resulting in suboptimal performance.

To enhance discrepant representations, pseudo-anomalies [14,17] are incorporated into the unsupervised AD prototype. The unsupervised task is converted to a supervised learning task. DRAEM [16] superimposes unrelated distributions of texture images onto normal images to simulate anomalous images. Subsequently, MemSeg [15] introduces a foreground simulation strategy aimed at avoiding anomalies in the background of images. CutPaste [7] presents a method that randomly selects a small rectangular region from the original image and pastes this content to different locations on normal images to generate synthetic anomalies. The pseudo-anomaly mechanism in [7] does not need additional auxiliary datasets, significantly reducing reliance on external datasets and enhancing the flexibility of anomaly simulation. Despite the inherent randomness in both pseudo and real anomalies, the significant differences between them make it challenging to effectively generalize the task of locating pseudo anomalies to real scenarios.

In this context, we propose a novel knowledge distillation paradigm for anomaly detection, named DLKD, to address the aforementioned problems. DLKD focuses on two primary tasks: accurately characterizing normal samples and filtering out anomalous signals. Specifically, leveraging the pseudo-anomaly mechanism, we introduce Gaussian noise into the shallow feature space of normal images and employ these as pseudo-anomalous data to train the student network. This enables the student network to amplify the differences in the T-S network's representations within anomalous regions when making inference. However, the teacher features reconstructed by the deconvolutional layer fail to accurately represent the true teacher features, leading to the significant degradation of detection performance. Therefore, we introduce the coarse-precise refinement learning mechanism into the student network to enable two-stage fine-grained reconstruction of the teacher features. In the coarse decoding stage, we combine the global modeling capabilities of the state space model (SSM) [5] with the local correlation strengths of CNNs to preliminarily reconstruct the teacher features. Subsequently, to enhance the robustness of knowledge transfer, the anomaly maps generated by the coarse defect-removal version guide the subsequent precise refinement

learning for final reconstruction. Compared to the existing KD-based approaches, the key contributions of DLKD are as follows:

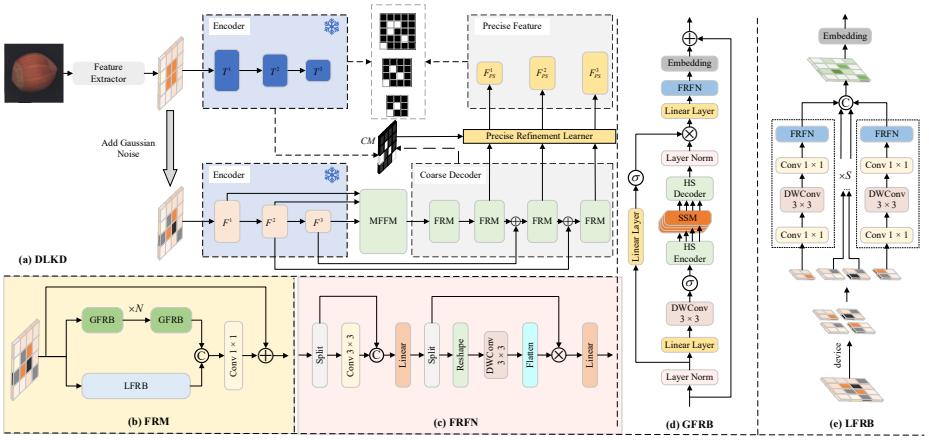

Fig. 1. (a) is the overall architecture of the dual learning knowledge distillation network (DLKD) for industrial defect anomaly detection; (b) is the architecture of the feature reconstruction module (FRM); (c) is the architecture of the feature refinement feed-forward network (FRFN); (d) is the pipeline of the global feature reconstruction block (GFRB); (e) is the details of the local feature reconstruction block (LFRB).

1. We introduce a novel knowledge distillation network with pseudo-anomaly mechanism for anomaly detection, effectively enhancing the differentiation of anomalous regions when making inference on anomalous samples.
2. A coarse-precise refinement learning mechanism is incorporated into the student network to accurately characterize normal samples. First, the coarse decoder is introduced to restore the defect-free feature both globally and locally, while the precise refinement learner (PRL) is designed to enable fine-grained feature reconstruction.
3. We also integrate the SSM-based global feature reconstruction block (GFRB) with the CNN-based local feature reconstruction block (LFRB) to facilitate the modeling and interaction between global and local information. Extensive experiments on the two publicly available datasets demonstrate the competitive performance of the proposed DLKD, both quantitatively and qualitatively.

2 PROPOSED METHOD

2.1 Overview

In this section, we provide a detailed introduction to the proposed DLKD framework for anomaly detection. **Fig. 1** (a) illustrates the overall architecture of the DLKD, which consists of two main components: a pre-trained ResNet-based teacher encoder for extracting shallow and pseudo-anomalous features at different scales, and a student

network for integrating multi-scale features and performing coarse-precise feature refinement to learn the teacher features.

Our method is based on the assumption that the decoder will ignore the abnormal information when performing inference on anomalous samples, provided it utilizes the pseudo-anomalous features as input for feature reconstruction. This enables the T-S network to maximize the differential representation of anomalous regions during inference. Unlike image-level abnormal simulation, we add Gaussian noise to the shallow features of input images, to generate pseudo-anomalous features. This pseudo-anomalies mechanism effectively minimizes visual interference and better simulates potential anomalies, further enhancing the model's ability to recognize anomalous information.

2.2 Coarse-Precise Refinement Learning

Coarse Learning. The large perceptual field of the model focuses more attention on the normal information while suppressing the representation of abnormal information [3]. To enhance our method's ability to capture normal information, we introduce the MFFM to integrate pseudo-abnormal features as input for coarse decoding learning. Moreover, we introduce the feature reconstruction module (FRM) in the decoder, consisting of a global feature reconstruction block (GFRB) based on SSM and a local feature reconstruction block (LFRB) based on CNNs. Through integrating both general and local features, the model's sensitivity to anomalous patterns is further improved.

Motivated by [10], we design the GFRB to model features globally using a hybrid scanning mechanism. Additionally, to enhance anomaly filtering, we introduce the feature refinement feed-forward network (FRFN) [19] and a low-dimensional embedding block to suppress anomalous information.

The low-dimensional embedding block consists of several convolutional layers. The computational process of the GFRB can be expressed as follows:

$$X_{i+1} = \mathrm{Embed}\left(\mathrm{FRFN}\left(\mathrm{GFM}(X_i)\right)\right) + X_i \tag{1}$$

where X_i represents the input feature, and GFM denotes the global feature modeling based on SSM.

To filter out anomalies and reconstruct a multi-scale normal representation during knowledge transfer of the decoder, we introduce the LFRB to capture detailed and subtle anomalies. Therefore, we first divide the input feature into several patches, and then separately transfer them into the CNNs with the FRFN, to focus more on minor anomalies. The overall process of LFRB can be formulated as follows:

$$Y_i^{n\prime} = \mathrm{FRFN}\left(\mathrm{Conv}\left(\mathrm{DWConv}_{3\times3}\left(\mathrm{Conv}(Y_i^n)\right)\right)\right) \tag{2}$$

$$F_L = \mathrm{Embed}\left(\mathrm{Concat}\left(Y_i^{1\prime}, Y_i^{2\prime}, \ldots\, Y_i^{s\prime}\right)\right) \tag{3}$$

where $Y_i^{n\prime}$ represents the output of the n^{th} feature map Y_i^n. F_L denotes the output of the LFRB, and Conv is a 1×1 convolutional layer.

Subsequently, we concatenate the local and global features along the channel dimension for feature aggregation. At end, a channel-wise convolutional layer with a skip connection is employed to restore the channel number, matching the input dimension.

$$F_{out} = \mathrm{Conv}\big(\mathrm{Concat}(F_G, F_L)\big) + F_{in} \tag{4}$$

where F_G represents the global feature, F_{in} denotes the input of FRM. A skip connection is also introduced to prevent the loss of multi-scale information during coarse decoder learning [20].

Precise Refinement Learning. In order to further refine the coarse multi-scale features, we introduce the precise refinement learner (PRL) to enhance feature representation. The multi-scale anomaly map M_i is obtained through calculating the cosine similarity between the multi-scale features of the coarse decoder and the teacher network. The PRL selectively refines the anomalous regions of the anomaly map to obtain higher-fidelity features. Since PRL focuses on local features regions for detailed learning, we maintain design simplicity and effectiveness by making it identical to the LFRB. The process of the PRL is expressed as follows:

$$F_{PS}^{i} = Precise\left(F_{CS}^{i} \cdot (1 - M_i)\right) \cdot M_i + F_{CS}^{i} \cdot (1 - M_i) \tag{5}$$

where F_{CS}^{i} represents the output of the coarse decoder at the i^{th} scale, and F_{PS}^{i} denotes the final output processed by the PRL.

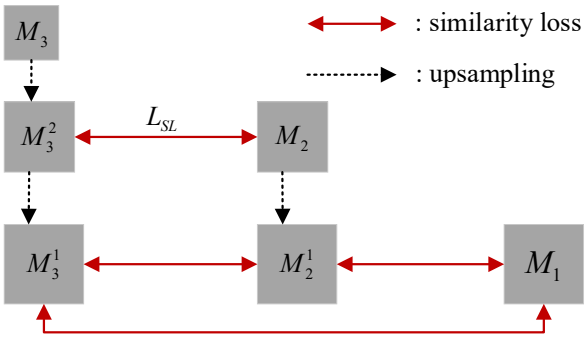

Fig. 2. The computational details of the anomaly similarity loss function.

2.3 Training objectives

Anomaly detection training is conducted under the following three loss functions: (i) the knowledge distillation-based loss $\mathcal{L}_{CKD}$ and $\mathcal{L}_{PKD}$ for filtering preliminary anomalies and restoring multi-scale features in the coarse decoder and the precise refinement learner; (ii) the contrast loss $\mathcal{L}_{Con}$ for further learning to remove anomalies; (iii) the anomaly similarity-based loss $\mathcal{L}_{AS}$ to constrain the output of the student network to be

more consistent with the teacher network's output. In summary, the total loss function can be formulated as follows:

$$\mathcal{L} = \mathcal{L}_{CKD} + \mathcal{L}_{PKD} + \alpha\mathcal{L}_{Con} + \beta\mathcal{L}_{AS} \tag{6}$$

where α and β are the weighting parameters.

Table 1. Anomaly detection results in terms of AUROC at image-level on the MVTec dataset.

Class	STFPM	DRAEM	RD	MB-PFM	Destseg	Noco	Ours
Carpet	95.59	95.39	98.80	**99.92**	99.16	99.36	99.88
Grid	99.58	**100**	**100**	98.33	**100**	98.58	**100**
Leather	94.02	**100**	**100**	**100**	**100**	**100**	**100**
Tile	98.63	**100**	99.20	99.57	**99.96**	99.46	99.24
Wood	99.39	**99.74**	99.30	99.30	97.63	99.65	99.12
Bottle	**100**	97.78	**100**	**100**	**100**	**100**	**100**
Cable	93.59	92.45	96.60	99.06	97.58	91.66	**99.68**
Capsule	87.16	96.33	98.40	93.66	97.69	93.78	**98.88**
Hazelnut	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Metal nut	99.90	99.80	**100**	**100**	99.56	99.71	99.95
Pill	95.72	**98.17**	96.30	96.94	91.16	96.48	97.95
Screw	92.15	97.91	**98.30**	91.76	90.67	87.70	98.20
Toothbrush	85.56	99.44	99.70	88.06	99.44	**100**	99.44
Transistor	94.42	87.75	96.00	97.46	99.17	95.25	**100**
Zipper	95.19	99.89	98.60	96.93	**100**	98.50	99.82
Average	95.39	97.63	98.72	97.40	98.13	97.34	**99.48**

Distillation loss: Firstly, we obtain the coarse 2D anomaly map M_C^K following by the below formula:

$$M_C^K(h,w) = 1 - \frac{\left(T_E^k(h,w)\right)^T \cdot F_{CS}^k(h,w)}{\| T_E^k(h,w) \| \| F_{CS}^k(h,w) \|} \tag{7}$$

where T_E^k and F_{CS}^k represent the outputs of the teacher network and coarse decoder, respectively. (h,w) is the location of the k^{th} feature map.

Through combining it with multi-scale knowledge distillation, taking the coarse decoder as an example, the loss function can be expressed as follows:

$$\mathcal{L}_{CKD} = \sum_{k=1}^{K} \left\{ \frac{1}{H_k W_k} \sum_{h=1}^{H_k} \sum_{w=1}^{W_k} M_C^K(h,w) \right\} \tag{8}$$

where K is the number of layers. The$\mathcal{L}_{PKD}$ is consistent with $\mathcal{L}_{CKD}$.

Contrast loss: We employ the cosine embedding-based loss with margin f to serve as the contrast loss for exploring deeper representation of normal features. It can be expressed as follows:

$$\mathcal{L}_{Con} = \frac{1}{K}\sum_{k=1}^{K} max(0, \cos(F_E^k, F_{PS}^k) - f) \tag{9}$$

where F_E^k represents the multi-scale output extracted from the encoding of the pseudo-anomaly features. F_{PS}^k denotes the output of the precise learner.

Anomaly similarity loss: As shown in the **Fig. 2**, we utilize bilinear interpolation to up-sample the small-scale anomaly map to match the size of the large-scale anomaly map. We compute the similarity loss $\mathcal{L}_{SL}$ through calculating the differences between maps for multi-scale feature refinement, which can be expressed as follows:

$$\mathcal{L}_{SL} = \frac{1}{HW}\sum_{w=1}^{W}\sum_{h=1}^{H}\left(M_i^j - M_j\right)^2 \tag{10}$$

where M_i^j represents the anomaly map M_i up-sampled to the same size as M_j. The final $\mathcal{L}_{AS}$ is the sum of all the similarity loss $\mathcal{L}_{SL}$.

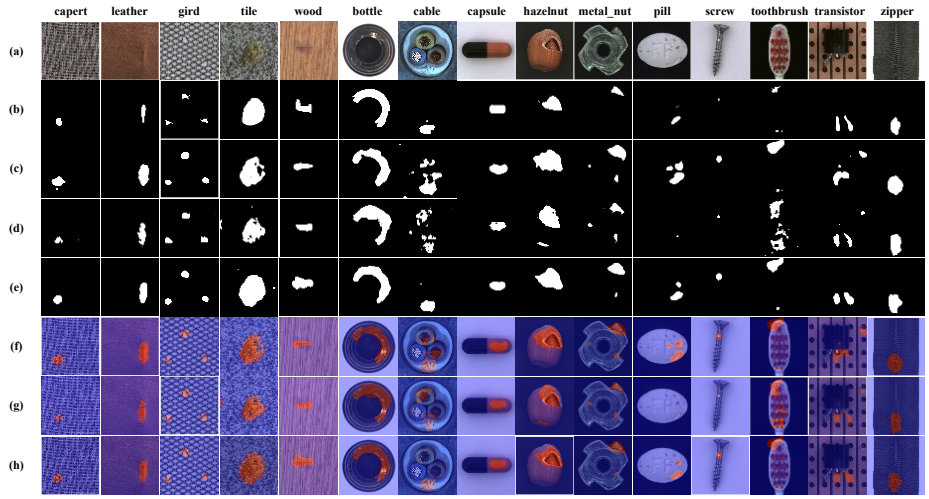

Fig. 3. Qualitative results of different methods for pixel-level anomaly segmentation on the MVTec AD dataset. (a) Anomaly image. (b) Ground truth. (c) Segmentation results of RD. (d) Segmentation results of Destseg. (e) Segmentation results of our proposed method. (f) Heat map of RD. (g) Heat map of Destseg. (h) Heat map of our proposed method.

Table 2. Quantitative results of various pixel-level anomaly detection methods on the MVTec AD dataset.

	Pixel-AUROC			PRO		
Method	Textures Avg.	Objects Avg.	Total Avg.	Textures Avg.	Objects Avg.	Total Avg.
STFPM	98.04	96.56	97.05	94.12	92.88	93.30
DRAEM	98.30	95.11	96.18	94.89	89.21	91.10

RD	97.74	97.88	97.83	95.06	93.50	94.02
MB-PFM	97.83	96.99	97.27	94.57	92.23	93.01
Destseg	**98.37**	97.69	97.92	**96.37**	92.04	93.48
Noco	97.66	98.01	97.89	92.66	91.80	92.08
ours	97.61	**98.29**	**98.06**	95.14	**94.00**	**94.38**

3 Experiments

3.1 Dataset and Metrics

Dataset: To evaluate the performance of our method, we conduct comparative experiments on two publicly available industrial datasets: MVTec AD [1] and BTAD [8]. The MVTec AD dataset consists of 15 categories, including five texture classes and ten object classes. Each category contains approximately 60 to 400 normal samples for training, along with a mix of normal and anomalous samples for testing. The anomalous images in the test set are annotated with binary labels for pixel-level visual evaluation. The BTAD dataset contains three industrial products, with a total of 2,540 samples. The training set includes only normal instances, while the test set is a mix of normal and anomalous samples. Additionally, the test set also has pixel-level annotated images for visual evaluation.

Metrics: The area under the receiver operator curve (AUROC) [1] is utilized to quantitatively evaluate the performance of our proposed DLKD and other state-of-the-art approaches. Specifically, the image-level AUROC is employed to evaluate the general anomaly detection performance, while both the pixel-level AUROC and Per-Region-Overlap (PRO) [2] curves are utilized to evaluate anomaly localization performance.

3.2 Implementation Details

In this study, all methods are implemented on a Linux operating system with an Nvidia RTX 3090Ti GPU, utilizing Python 3.9 and PyTorch 2.1.1. The Adam optimizer is utilized for parameter optimization with a weight decay of 0.0001, and the batch size is set to 8. The learning rate is set to 0.005 for the MFFM and coarse decoder, and 0.002 for the precise refinement learner. All input images are resized to 256×256 without additional augmentation to maintain consistency. The pre-trained ResNet18 is employed to serve as the teacher encoder for feature extraction. The precise refinement learner consists of three parallel branches, each built by stacking two LFRBs with skip connections, to refine the outputs from the coarse decoder. The Gaussian noise intensity is 0.2. The weight parameters α and β are empirically set to 0.02 and 0.05, respectively.

Table 3. Quantitative results of different methods for image-level and pixel-level anomaly detection on the BTAD dataset.

Method	01	02	03	Avg

VT-DAL	95.82/84.26/43.53	70.83/86.31/22.93	66.49/84.14/27.18	77.71/84.90/31.21
DRAEM	95.14/83.01/46.90	76.72/78.65/45.56	99.62/95.20/44.90	90.49/85.62/45.79
PatchCore	**97.90**/**97.00**/**73.50**	82.10/95.90/44.10	**99.90**/99.20/72.60	93.30/**97.37**/63.40
DLKD	94.56/95.77/68.39	**86.11**/**96.41**/**66.23**	99.59/**99.64**/**86.33**	**93.42**/97.27/**73.65**

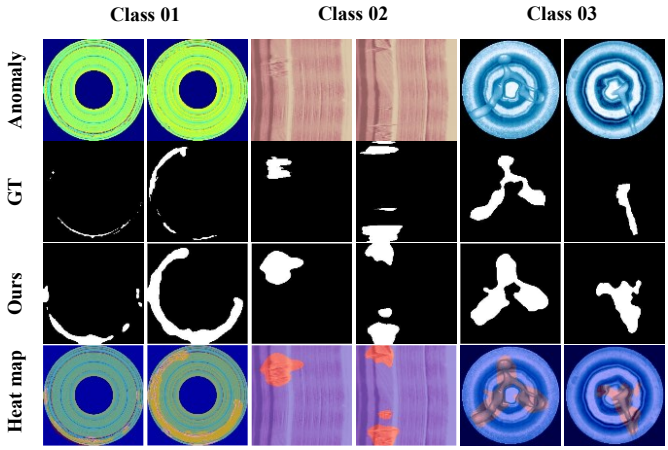

Fig. 4. Qualitative results of our method for pixel-level anomaly segmentation on the BTAD dataset.

3.3 Comparisons with State-of-the-Art Methods

To demonstrate the superior performance of our DLKD, we conduct comparative experiments with state-of-the-art methods, including STFPM [13], RD [3], DRAEM [16], MB-PFM [12], Destseg [18], and NOCO [4].

Results on the MVTec AD dateset. As shown in **Table 1**, the proposed DLKD outperforms other state-of-the-art methods in most categories, particularly in grid, leather, bottle, hazelnut, and transistor, where DLKD achieves 100% AUROC in image-level anomaly detection. It is noteworthy that our approach achieves best average AUROC in both texture and object class anomaly detection, highlighting the competitive capability of DLKD in capturing anomalous details for detection. Compared to other knowledge distillation-based methods, such as STPFM, RD, and Dsetseg, our DLKD outperforms them with average AUROCs of 4.09%, 0.76%, and 1.35% respectively. This further demonstrates the promising generalization of our method in detecting different anomaly samples.

Table 2 presents the comparison results between the proposed DLKD and other methods in anomaly localization. Our method achieves superior average scores across all categories, surpassing the well-performing RD by 0.23% and 0.36%, respectively. This demonstrates that our DLKD can effectively adapt to diverse anomaly localization

tasks through the coarse-precise refinement learning. Destseg excels at detecting texture anomalies but performs poorly on object anomalies. Furthermore, we provide several visualization examples of our method and other methods (RD and Destseg) on the MVTec AD dateset for a detailed comparison. As shown in **Fig. 3**, compared to other methods, our approach achieves more accurate anomaly localization and segmentation, demonstrating its superiority in perceiving anomaly details.

Results on the BTAD dataset. Table 3 presents the quantitative results of different anomaly detection methods on the BTAD dataset, including VT-ADL [8], DRAEM [16], and PatchCore [9]. Compared to other methods, DLKD achieves competitive performance across most metrics, demonstrating its effectiveness in local and global spatial modeling and enhancing the anomaly detection performance. **Fig. 4** presents qualitative results of the proposed method on the BTAD dataset. It is obvious that our approach can accurately recognize and locate the anomalous regions, extensively showcasing its robustness and applicability to various types of anomalies.

3.4 Ablation Study

In this section, we conduct a series of ablation experiments on the MVTec AD dataset to validate the effectiveness of the functional components and loss functions in our method.

Table 4. Ablation study of module components on the MVTec AD dataset.

Exp.	LFRB	GFRB	PRL	Image-AUROC	Pixel-AUROC	PRO
1	√			98.68	97.80	93.65
2	√	√		98.91	98.00	94.15
3	√	√	√	**99.48**	**98.06**	**94.38**

Ablation Studies on the Module Components. We first utilize only the LFRB-based coarse decoder to refine the reconstructed feature, removing the precise refinement learner to validate the effectiveness of the LFRB. Results in **Table 4** indicate the CNN-based LFRB can provide effective local feature learning. Building on the aforementioned architecture, we add the SSM-based GFRB to the coarse decoder to verify the necessity of the GFRB in modeling global dependencies. As indicated in **Table 4**, the inclusion of GFRB leads to a notable improvement across all metrics, significantly enhancing the capability of our methods to capture long-range spatial features and enhance detection performance. When the precise refinement learner is incorporated into the coarse decoder, significant improvements are observed across all metrics. It can be seen that the injection of the precise refinement learner enables the T-S model to achieve superior performance in anomalies detection and localization.

Table 5. Ablation study of loss functions on the MVTec AD dataset.

$\mathcal{L}_{PKD}$	$\mathcal{L}_{CKD}$	$\mathcal{L}_{Con}$	$\mathcal{L}_{AS}$	Image-AUROC	Pixel-AUROC	PRO

√	√			99.42	98.03	94.17
√	√	√		99.45	**98.06**	94.24
√	√	√	√	**99.48**	**98.06**	**94.38**

Ablation Study on Loss Functions. To validate the effectiveness of the loss function in improving both detection and localization performance, we remove the $\mathcal{L}_{Con}$ and $\mathcal{L}_{AS}$ respectively. As reported in **Table 5**, the inclusion of $\mathcal{L}_{Con}$ enhances the model's ability to filter abnormal features while focusing more on precise feature refinement. With the constraint of $\mathcal{L}_{AS}$, DLKD efficiently reconstructs more consistent multi-scale features, enabling precise localization of the anomalous regions.

4 Conclusion

In this paper, we propose a novel knowledge distillation framework for anomaly detection, referred to dual-learning knowledge distillation (DLKD), to effectively filter out anomalous signals and precisely characterizes normal samples. We introduce the coarse-precise refinement learner into the student network to perform a two-stage coarse-to-fine generation of the teacher features, providing a compact representation for the student network. The coarse decoder integrates global modeling prowess of SMMs with CNN-based local feature correlations, enhancing the student network's ability to reconstruct normal features. Our method demonstrates outstanding performance in both anomaly detection and real-time inference. We hope that our approach will stimulate further advancements in this field.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 62406224, in part by the Natural Science Foundation of Tianjin under Grant 23JCQNJC00820, and Grant 23JCYBJC00360.

References

1. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: MVTec AD—A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9592–9600 (2019).
2. 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, pp. 4183–4192 (2020).
3. 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, pp. 9737–9746 (2022).
4. Gao, H., Luo, H., Shen, F., Zhang, Z.: A Norm-Constrained Method for Semantic and Structured Anomaly Detection Tasks. IEEE Trans. Instrum. Meas. (2024).
5. Gu, A., Dao, T.: Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv preprint arXiv:2312.00752 (2023).

6. He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., Xie, L.: MambaAD: Exploring State Space Models for Multi-Class Unsupervised Anomaly Detection. arXiv preprint arXiv:2404.06564 (2024).
7. 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, pp. 9664-9674 (2021).
8. Mishra, P., Verk, R., Fornasier, D., Piciarelli, C., Foresti, G.L.: VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization. In: 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), pp. 1-6. IEEE (2021).
9. 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, pp. 14318-14328 (2022).
10. Ruan, J., Xiang, S.: VM-UNet: Vision Mamba UNet for Medical Image Segmentation. arXiv preprint arXiv:2402.02491 (2024).
11. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep One-Class Classification. In: International Conference on Machine Learning, pp. 4393-4402. PMLR (2018).
12. Wan, Q., Gao, L., Li, X., Wen, L.: Unsupervised Image Anomaly Detection and Segmentation Based on Pretrained Feature Mapping. IEEE Transactions on Industrial Informatics 19(3), 2330-2339 (2023). https://doi.org/10.1109/TII.2022.3182385
13. Wang, G., Han, S., Ding, E., Huang, D.: Student-Teacher Feature Pyramid Matching for Anomaly Detection. arXiv preprint arXiv:2103.04257 (2021).
14. Yan, X., Zhang, H., Xu, X., Hu, X., Heng, P.A.: Learning Semantic Context from Normal Samples for Unsupervised Anomaly Detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 3110-3118 (2021).
15. Yang, M., Wu, P., Feng, H.: MemSeg: A Semi-Supervised Method for Image Surface Defect Detection Using Differences and Commonalities. Engineering Applications of Artificial Intelligence 119, 105835 (2023).
16. 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, pp. 8330-8339 (2021).
17. Zavrtanik, V., Kristan, M., Skočaj, D.: Reconstruction by Inpainting for Visual Anomaly Detection. Pattern Recognition 112, 107706 (2021).
18. 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, pp. 3914-3923 (2023).
19. Zhou, S., Chen, D., Pan, J., Shi, J., Yang, J.: Adapt or Perish: Adaptive Sparse Transformer with Attentive Feature Refinement for Image Restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2952-2963 (2024).
20. Zhu, J., Yan, P., Jiang, J., Cui, Y., Xu, X.: Asymmetric Teacher-Student Feature Pyramid Matching for Industrial Anomaly Detection. IEEE Transactions on Instrumentation and Measurement (2023).