DTSNet: A Denoising Teacher-Student Network with Reverse Distillation for Anomaly Detection

Taixiang Lin[1], Shuyuan Lin[1*], Yanjie Liang[2], Rong Chen[3], Yang Lu[4]
[1] College of Cyber Security, Jinan University, Guangzhou, China
[2] Peng Cheng Laboratory, Shenzhen, China
[3] School of Information Engineering, Xizang Minzu University, Xianyang, China
[4] Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Informatics, Xiamen University, Xiamen, China

***Abstract*—Knowledge distillation has emerged as a promising method for unsupervised anomaly detection. However, the overgeneralization of the student network often reduces the distinction between teacher and student representations for anomalous samples, leading to detection failures. To address this problem, we propose a Denoising Teacher-Student Network (DTSNet), which integrates two teacher networks: a normal teacher and an anomalous teacher. These networks guide the student network in feature-space denoising, enabling it to restore anomalous features and amplify the representational disparity for anomalies. Furthermore, we propose an Attention-guided Perturbation Reconstruction (APR) module, which facilitates the student network to focus on critical pixel regions, enhancing its feature representation capability. Experimental results demonstrate that the proposed DTSNet outperforms several state-of-the-art methods on the MVTec AD, VisA, and BTAD datasets. Source code is available at http://www.linshuyuan.com.**

***Index Terms*—Anomaly detection, knowledge distillation, self-supervised**

I. Introduction

Image anomaly detection (AD) refers to the task of identifying and localizing regions within an image that deviate from normal patterns. It has broad applications in fields such as industrial quality inspection [1], medical screening [2] and video surveillance [3]. Nevertheless, in practical scenarios, anomalous samples are often scarce and challenging to collect. Consequently, AD is redominantly addressed in an unsupervised manner, relying solely on normal samples for training.

Recent studies [4]–[6] have shown that pretrained convolutional neural networks, as versatile visual feature extractors, achieve state-of-the-art performance in AD tasks. Among the various techniques, knowledge distillation based on teacher-student (TS) networks has emerged as an effective framework [7], [8]. In this framework, the student network learns to replicate the feature representations of normal images from the teacher network, which is typically pretrained on large-scale datasets such as ImageNet [9]. During inference, anomaly detection and localization rely on analyzing the feature differences between student and teacher networks. Regions with minor differences are classified as normal, whereas regions with substantial differences are identified as anomalies. However, a critical limitation of this approach is the overgeneralization of the student network, causing it to extract features from anomalous samples that closely resemble those of the teacher network, thereby compromising detection performance.

To address this issue, the reverse distillation (RD) [10] introduces a reverse distillation paradigm, in which the encoder acts as the teacher and the decoder serves as the student. This design partially mitigates the limitations of insufficient output discrepancies arising from identical data flows. However, due to the absence of explicit constraints on anomalous samples, it fails to guarantee consistent features differences between the TS networks for such samples. Existing methods [11], [12] attempt to overcome this limitation by employing memory modules to enhance the student network's retention of normal data, thereby reducing the likelihood of generating representations for anomalous samples. However, the inclusion of memory modules not only increases computational overhead but also reduces the model's inference speed, limiting their practicality in real-world applications.

To amplify the representation differences of anomalous features between the teacher and student networks, we propose to explicitly denoise and refine the student's feature representations via an anomaly synthesis paradigm. To achieve this, we propose a dual-teacher framework consisting of a normal teacher network and an anomalous teacher network. These teacher networks collaboratively guide the student network to restore anomalous feature, thereby establishing a more discriminative decision boundary. Additionally, to enhance student's feature representation capacity, we propose an attention-guided perturbation reconstruction module. This module encourages the student network to focus on the relationships between noise-perturbed key regions and their surrounding contexts, rather than simply mimicking the teacher's output features in the visible regions. As a result, the student network can generate robust and comprehensive feature representations, improving its ability to differentiate anomalies.

Our contributions are summarized as follows:

- We propose a denoising teacher-student network (i.e., DTSNet) based on reverse knowledge distillation, de-

*Corresponding author (swin.shuyuan.lin@gmail.com). This work was supported in part by the National Natural Science Foundation of China (Grant Nos. U22A2095, 62476112, 62202249, 62431004, 62376233); in part by the Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2024A1515011740, 2025A1515010181); in part by the Fundamental Research Funds for the Central Universities (Nos. 21624404, 23JNSYS01); in part by the General Program Foundation of Xizang Minzu University (Grant No. 24MDY04); in part by the Natural Science Foundation of Fujian Province (Grant No. 2024J09001).

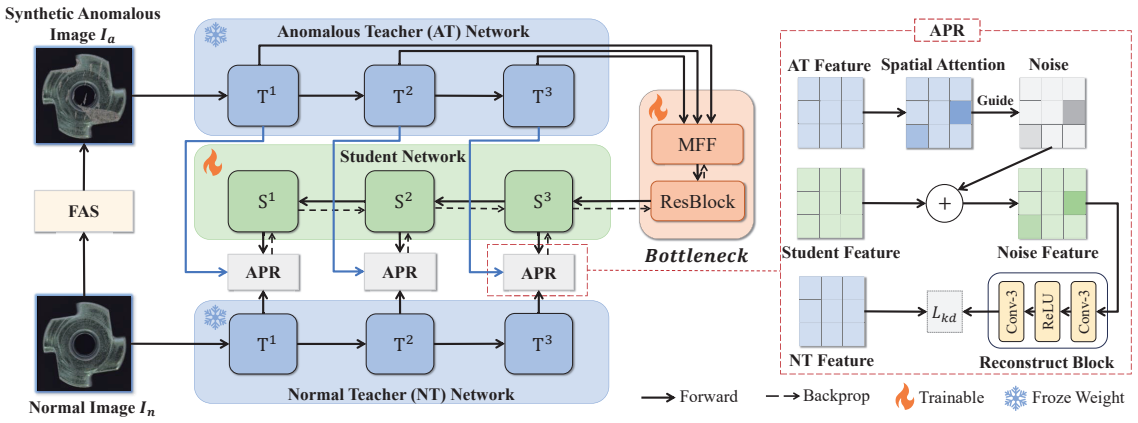

Fig. 1. Overview of DTSNet during training. The student network receives features extracted by the anomalous teacher and learns to restore anomalous features under the guidance of the normal teacher. The knowledge distillation process is guided by the APR module, which incorporates noise into the student features. The noise distribution is controlled by the spatial attention of the teacher features. These noisy features are then reconstructed using teacher-guided convolutional blocks, enabling the student to extract critical knowledge from the noise.

signed to address the overgeneralization issue commonly observed in traditional knowledge distillation models.

- We propose an Attention-Guided Perturbation Reconstruction (i.e., APR) module, which enables the student network to focus on critical pixel regions in noisy environments, enhancing its feature representation capacity.
- We conduct extensive experiments on three AD benchmarks, demonstrating the superiority and effectiveness of the proposed DTSNet.

II. Proposed Method

A. Overall Framework

The overall framework of the proposed DTSNet is illustrated in Fig. 1. Starting with the original TS structure of reverse distillation, we design a denoising TS network. Specifically, during the training process, the student network receives features extracted by the anomalous teacher network and learns to restore the anomalous features under the guidance of the normal teacher network. Subsequently, the APR module perturbs the important pixels of the student network to encourage it to generate feature representations that are closer to those of the normal teacher network. During the inference phase, the teacher network captures anomalous feature, while the student network transforms them into normal features. The discrepancy between these features is utilized to evaluate pixel-level anomaly scores.

B. Foreground-Aware Anomaly Synthesis

Anomaly synthesis strategies have been widely applied in AD. A common strategy involves combining textures from the Describable Textures Dataset (DTD) [13] with Perlin noise [14] masks to generate local texture anomalies [15]. However, in industrial scenarios, anomalies often occur only in the foreground where objects are located. Synthesizing anomalies across the entire image is likely to result in significant discrepancies between synthetic anomalies and real anomalies. To address this limitation, we propose a foreground-aware synthesis (FAS) strategy to enhance anomaly diversity and better suit industrial scenarios. As shown in Fig. 2, we first use

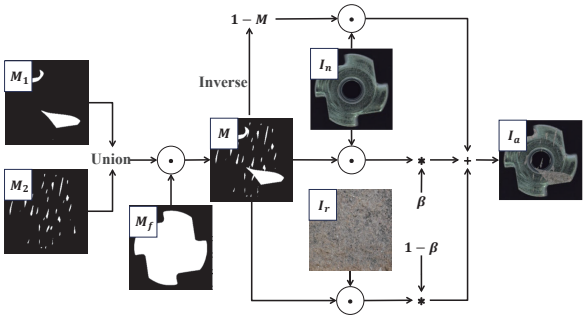

Fig. 2. Illustration of the proposed FAS, where the union operation is employed to generate the final anomaly mask. Visually inconsistent patterns are integrated into normal samples to create synthetic anomalies.

IS-Net [16] to extract the target foreground region M_f, and then generate two binary masks, M_1 and M_2, using Perlin noise. To increase diversity, we construct the final anomaly mask M by combining the intersection or union of M_1 and M_2 [17]:

$$M = \begin{cases} M_1 \cap M_2 \odot M_f, & \text{if } 0 \le p \le \alpha \\ M_1 \cup M_2 \odot M_f, & \text{if } \alpha < p \le 1 \end{cases}, \tag{1}$$

where $\odot$ denotes the element-wise multiplication; α is set to 0.5, and $p \sim U(0,1)$. Finally, we take a linear combination of the normal image I_n and a random image I_r from the DTD, replacing the mask regions to generate the final anomalous image. The synthetic anomaly image I_a can be obtained by the proposed FAS as follows:

$$I_a = \beta(M \odot I_n) + (1-\beta)(M \odot I_r) + \overline{M} \odot I_n, \tag{2}$$

where $\overline{M}$ is the inverse of M, and β is the opacity parameter, randomly selected within the range $[0.15, 1]$, to enhance the fusion of normal and anomalous regions.

C. Reverse Distillation

Following previous work [10], [18], we use the first three blocks of a WideResNet50 pretrained on ImageNet as the

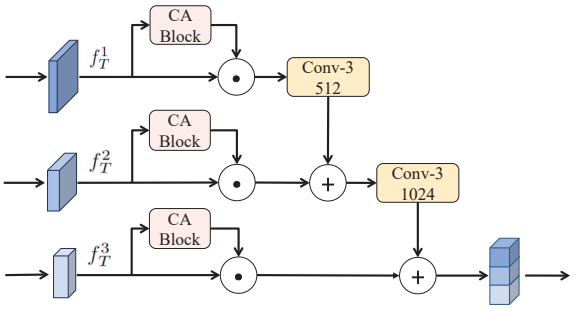

Fig. 3. Illustration of the proposed MFF.

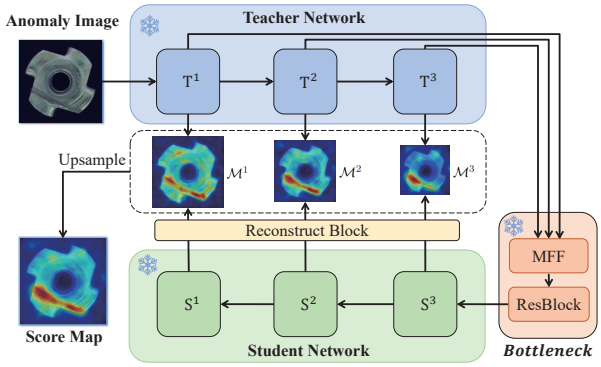

Fig. 4. Online inference of the proposed DTSNet combines the teacher network, the student network, and the bottleneck layer frozen.

teacher network to extract comprehensive representations from both normal and anomalous images. The output feature maps are denoted as $f_{T,n}^k$, $f_{T,a}^k \in \mathbb{R}^{C_k \times H_k \times W_k}$, where H_k, W_k, and C_k are the height, width, and number of channels of the output feature of block k. The student network is a reversed symmetrical counterpart of the teacher network and is randomly initialized, aiming to reconstruct the outputs of the teacher network. The output feature maps of the student network are represented as $f_S^k \in \mathbb{R}^{C_k \times H_k \times W_k}$.

Since the activations of the last layer of the teacher network contain high-level semantic information, directly transferring them to the student network may impede the reconstruction of low-level features. Given the outstanding performance of multiscale feature fusion in target detection [19], an intuitive approach is to employ a channel attention mechanism combined with a multiscale fusion strategy to efficiently integrate the visual and semantic information from each layer of the teacher network. To this end, we propose a multiscale feature fusion (MFF) module as illustrated in Fig. 3. Unlike RD, this module applies coordinate attention (CA) [20] to weight features from different hierarchical levels, then aligns the feature maps using a 3×3 convolution, and finally performs multiscale feature fusion via element-wise addition. Subsequently, the fused features are compressed into a more compact feature space using the fourth residual block of the ResNet, assisting the student network in better reconstructing low-level features.

D. Attention-Guided Perturbation Reconstruction

Reconstructing key features in noisy environments is challenging for the student network, especially due to its capacity limitations compared to the teacher network. To address this issue, we propose the APR module, which synthesizes noise following the principles of noisy feature distillation [21] and leverages spatial attention to guide the student network's focus toward critical pixel regions. Specifically, we introduce unbiased Gaussian noise $\epsilon_k \sim \mathcal{N}(0, \sigma^2)$ to perturb the output feature f_S^k of the student network. Due to the typically significant capacity gap between the teacher and student networks, the latter often struggles to identify key information when processing a large quantity of similar information. Therefore, we filter the output from the teacher network through a spatial attention mask [22] to adjust the noise distribution, helping the student network to focus more on important pixels. In particular, we first calculate the absolute mean value of each pixel across the channel dimension in the teacher network's output feature $f_{T,a}^k$:

$$G_T^k = \frac{1}{C_k} \sum_{c=1}^{C_k} |f_{T,a,c}^k|, \tag{3}$$

where $G_T^k \in \mathbb{R}^{H_k \times W_k}$ represents the spatial attention map. Subsequently, the attention mask is defined as follows:

$$A_T^k = H_k \cdot W_k \cdot \text{softmax}\left(\frac{G_T^k}{\tau}\right), \tag{4}$$

where τ is a temperature hyperparameter that adjusts the distribution [23]. Next, the attention mask is used to adjust the noise standard deviation on a pixel-wise basis. To align the noise distribution more closely with the student features, we compute the standard deviation $\vartheta(f_S^k)$ of the student features. The adjusted noise is then added to the student network's output features as follows:

$$\sigma' = \sigma \cdot \vartheta(f_S^k) \cdot A_T^k, \tag{5}$$

$$\tilde{f}_S^k = f_S^k + \epsilon_k', \tag{6}$$

where $\epsilon_k' \sim \mathcal{N}\left(0, (\sigma')^2\right)$ and σ is the hyperparameter for adjusting the standard deviation of the noise. Finally, the perturbed features are restored by a reconstruction block guided by the teacher network as follows:

$$f_R^k = \mathcal{R}(\tilde{f}_S^k), \tag{7}$$

where $\mathcal{R}$ represents the reconstruction block composed of two 3×3 convolutional layers.

In our TS knowledge transfer model, the student network's decoding process can be regarded as an anomaly feature restoration process. To measure the consistency between the feature representations of the teacher and student network, we compute their cosine similarity as follows:

$$L_{\text{kd}} = \sum_{k=1}^{3} (1 - \frac{\mathcal{F}(f_{T,n}^k)^\top \cdot \mathcal{F}(f_R^k)}{\|\mathcal{F}(f_{T,n}^k)\| \cdot \|\mathcal{F}(f_R^k)\|}), \tag{8}$$

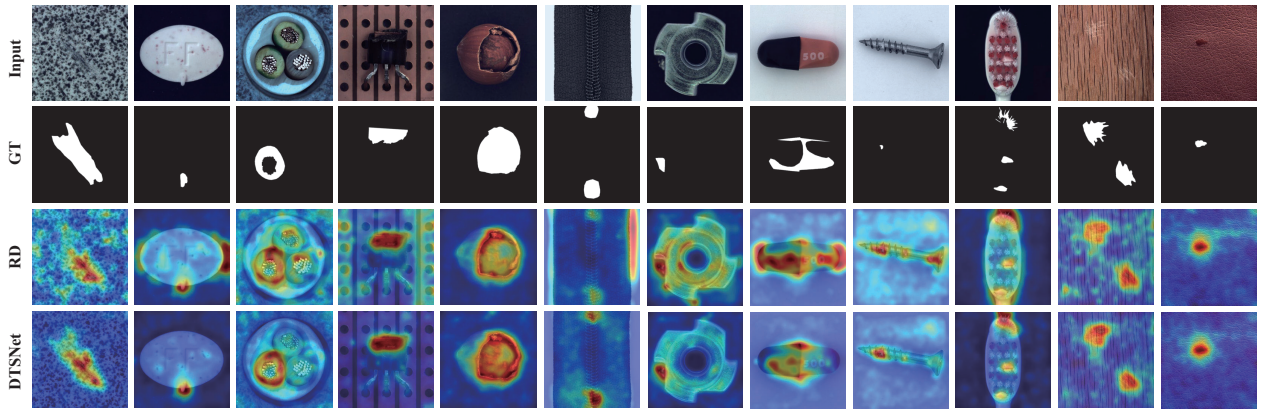

Fig. 5. Qualitative results on MVTec AD [1] dataset. Compared to RD [10], the proposed DTSNet demonstrates superior localization accuracy across various types of anomalies.

where $\mathcal{F}(\cdot)$ denotes the operation of flattening the features into a one-dimensional vector.

E. Inference

The inference process of the proposed DTSNet is shown in Fig. 4. Specifically, when the query image exhibits anomalies, the teacher network is capable of capturing these anomalous features, while the student network restores them in the feature space. Consequently, the teacher and student networks generate inconsistent feature representations. For anomaly localization, we compute the cosine similarity along the channel axis between f_T^k and f_R^k to generate the 2D anomaly map $\mathcal{M}^k$ as follows:

$$\mathcal{M}^k(h,w) = 1 - \frac{(f_T^k(h,w))^\top \cdot f_R^k(h,w)}{\|f_T^k(h,w)\| \cdot \|f_R^k(h,w)\|}, \tag{9}$$

where the values in $\mathcal{M}^k$ indicate the degree of anomaly at each point of the k-th feature map. Next, we upsample $\mathcal{M}^k$ to match the size of the original image and generate the final score map S_{AL} by combining the anomaly maps from the three stages as follows:

$$S_{\text{AL}}(h,w) = \sum_{k=1}^{3} \Psi(\mathcal{M}^k), \tag{10}$$

where Ψ denotes the bilinear upsampling operation. For anomaly detection, the image-level anomaly score is represented by the maximum value in the score map.

III. Experiments

A. Datasets

We conduct experiments on three commonly used unsupervised AD benchmarks. **MVTec AD** [1] is the most widely used industrial AD dataset, consisting of 10 object categories and 5 texture categories, with a total of 4,096 normal and 1,258 anomalous images. **VisA** [26] is a challenging industrial AD dataset that covers 12 object categories, with 9,621 normal and 1,200 anomalous images. **BTAD** [27] contains 2,540 images of real-world products with three different types of defects.

B. Experimental Settings

Following the previous work [10], all images are resized to 256 pixels. No additional image augmentation techniques are applied, except for anomaly synthesis. We train for 10–400 epochs using the Adam optimizer, with a learning rate of 0.005 for the student network and bottleneck module and, 0.001 for the reconstruction module. The noise adjustment hyper-parameter σ is set to 0.1. The experiments are implemented in PyTorch and conducted on a single Nvidia RTX 3090 GPU.

C. Evaluation Metrics

We employ the area under the receiver operating characteristic curve at the image level (I-AUROC) and at the pixel level (P-AUROC) as evaluation metrics for anomaly detection and localization. In addition, for the pixel-level anomaly localization, we also use the per-region-overlap (PRO) [7] score to provide a more comprehensive assessment of localization performance.

D. Anomaly Detection and Localization

The anomaly detection and localization results on the MVTec AD dataset are reported in Table I. The proposed DTSNet achieves competitive performance across all categories, reaching the highest I-AUROC in 6 out of 15 categories. The average I-AUROC is 99.3%, which is 0.8% higher than that obtained by RD and only 0.2% lower than that of SimpleNet. For the pixel-level anomaly localization, DTSNet outperforms all the competing methods in both the P-AUROC and the more robust PRO score. Notably, the PRO score is 1.2% higher than the second-best method (i.e. RD), highlighting its superior anomaly localization capabilities. Some representative samples of anomaly localization are visualized in Fig. 5.

Table II presents the results of the proposed DTSNet on the VisA and BTAD datasets. For the VisA dataset, the proposed DTSNet achieves a state-of-the-art performance, with I-AUROC, P-AUROC, and PRO of 97.1%, 99.0%, and 95.5%, respectively. Similarly, for the BTAD dataset, the proposed DTSNet outperforms existing methods(i.e. DRAEM, PatchCore, SimpleNet, AST and RD), particularly in the PRO metric, showing an improvement of 3.5% over the previous

TABLE I
ANOMALY DETECTION AND LOCALIZATION ON MVTEC AD. RESULTS ARE SHOWN FOR I-AUROC/P-AUROC/PRO METRICS DEFINED IN SECTION III-C (IN %), WITH THE HIGHLIGHTED IN BOLD.

Category	DRAEM [15]	PatchCore [4]	SimpleNet [6]	AST [24]	RD [10]	DTSNet
Carpet	93.3/92.2/92.9	98.7/99.0/96.6	99.7/97.7/88.4	97.3/97.0/89.4	98.8/99.0/97.1	**100/99.3/97.9**
Grid	**100/99.7/98.3**	98.2/98.7/95.9	99.2/94.8/86.9	98.8/96.4/85.1	**100**/99.2/97.3	**100**/99.3/97.7
Leather	**100**/98.8/97.4	**100**/99.3/98.9	**100**/99.2/96.7	**100**/97.5/94.7	**100**/99.4/**99.1**	**100/99.5/99.2**
Tile	**100/99.6/98.2**	98.7/95.6/87.4	99.9/93.8/88.0	99.9/92.7/83.5	99.3/95.6/90.6	98.6/96.4/92.2
Wood	99.6/95.8/90.3	99.2/95.0/89.6	**100**/93.7/83.1	99.9/87.0/76.9	99.1/95.3/90.8	99.7/**96.2/93.7**
Bottle	96.6/**99.3**/96.8	**100**/98.6/96.1	**100**/98.0/91.1	**100**/91.6/83.6	**100**/98.7/96.7	**100**/98.8/**97.1**
Cable	94.4/96.2/81.0	99.5/98.4/92.6	**100**/97.4/90.9	97.3/94.2/83.0	95.9/97.1/90.5	99.2/**98.6/94.6**
Capsule	96.3/93.0/82.7	98.1/98.8/95.5	97.6/**98.9**/92.5	**98.6**/98.0/92.1	97.6/98.6/95.8	97.3/98.7/**96.0**
Hazelnut	**100/99.6/98.5**	**100**/98.7/93.9	99.7/97.3/81.3	99.9/97.2/86.7	**100**/98.9/95.4	**100**/99.0/96.2
Metal_nut	98.8/**99.0/97.0**	**100**/98.4/91.3	**100**/98.7/88.4	98.4/91.9/75.9	**100**/97.3/92.4	**100**/98.3/93.0
Pill	98.0/97.9/88.4	96.6/97.4/94.1	98.6/98.4/93.4	**98.9**/96.3/84.6	95.9/98.2/96.3	98.7/**98.5/97.2**
Screw	**99.6/99.7**/95.0	98.1/99.4/97.9	98.4/99.3/96.7	**99.6**/98.2/94.0	97.7/99.6/98.0	99.0/99.6/**98.5**
Toothbrush	99.7/98.3/85.6	**100**/98.7/91.4	**100**/98.5/92.6	95.8/98.2/86.4	99.2/**99.1/94.5**	99.5/**99.1**/94.3
Transistor	94.0/85.5/70.4	**100**/96.3/83.5	**100/96.8/93.7**	98.1/94.7/77.5	96.0/92.9/78.6	98.2/94.1/83.0
Zipper	**100**/98.3/96.8	99.4/**98.8/97.1**	99.9/**98.8**/95.6	98.8/96.0/88.7	98.2/98.3/95.5	99.0/98.4/95.8
Average	98.0/96.9/91.3	99.1/98.1/93.5	**99.5**/97.4/90.6	98.8/95.1/85.5	98.5/97.8/93.9	99.3/**98.3/95.1**

TABLE II
PERFORMANCE COMPARISON ON THE VISA AND BTAD DATASETS. "I", "P" AND "O" REFER TO THE THREE METRICS OF I-AUROC, P-AUROC, PRO, RESPECTIVELY. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

	VisA			BTAD		
	I	P	O	I	P	O
DRAEM [15]	88.7	94.6	73.1	89.0	87.1	61.6
PatchCore [4]	94.8	98.5	89.8	92.7	97.4	64.0
SimpleNet [6]	96.5	97.8	90.9	95.0	96.8	74.6
RealNet [25]	96.3	98.4	93.5	95.7	**98.0**	74.2
AST [24]	92.4	92.9	71.5	95.2	96.5	66.2
RD [10]	95.7	98.5	93.2	95.2	97.1	72.3
DTSNet	**97.1**	**99.0**	**95.5**	**96.1**	97.6	**78.1**

TABLE III
ABLATION STUDY ON DIFFERENT MODULES IN DTSNET.

Module			Performance		
Baseline	Bottleneck	APR	I	P	O
✓			98.1	97.3	93.6
✓	✓		98.8	97.7	94.7
✓	✓	✓	**99.3**	**98.3**	**95.1**

TABLE IV
ABLATION STUDY ON THE COMPONENTS OF THE FAS STRATEGY.

FAS	Performance		
	I	P	O
w/o Diversity	**99.3**	98.2	95.0
w/o Foreground	99.1	98.0	94.9
All	**99.3**	**98.3**	**95.1**

TABLE V
ABLATION STUDY ON DIFFERENT BACKBONES.

Backbone	Performance		
	I	P	O
ResNet18	98.5	97.4	93.7
ResNet34	98.9	97.6	93.7
ResNet50	99.1	98.1	94.8
WideResNet50	**99.3**	**98.3**	**95.1**

TABLE VI
ABLATION STUDY ON FUSION OF SCORE MAPS AT DIFFERENT SCALES.

Score Map			Performance		
$\mathcal{M}^1$	$\mathcal{M}^2$	$\mathcal{M}^3$	I	P	O
✓			92.8	94.6	89.8
	✓		98.5	97.1	93.6
		✓	97.7	97.4	91.0
	✓	✓	98.7	98.1	94.2
✓	✓	✓	**99.3**	**98.3**	**95.1**

best method (i.e. SimpleNet). These results demonstrate the effectiveness and generalization capability of the proposed DTSNet across different datasets.

E. Ablation Studies

We conducted experiments to evaluate the impact of the Bottleneck and the APR module on the model's AD performance, with the results presented in Table III. As shown in Table III, the reverse TS network with the FAS strategy was used as the baseline model. The Bottleneck module improves the student network's capability to learn both low-level and high-level features by integrating multiscale features into the compact feature embeddings. Meanwhile, the APR module further improves the network's AD performance by effectively transferring knowledge about important pixel regions in noisy environments, thereby increasing the robustness of the student's feature representations.

We also examined the effectiveness of the FAS strategy, as shown in Table IV. During training, we individually removed the diversity mask generation and foreground extraction components and compared these configurations with the complete strategy. The results indicate that removing either component led to a slight performance decline, highlighting the importance of both components in the overall strategy.

We further provided a qualitative comparison of various backbone networks used us teacher models in Table V. The

results indicate that as the depth and width of the network increase, the model's ability to recognize anomalies gradually improves, owing to the stronger representation capabilities of deeper and wider networks. Notably, even with smaller networks such as ResNet18, our method is still able to maintain competitive performance.

We also investigated the effectiveness of different network layers in AD, as shown in Table VI. Among single-layer features, $\mathcal{M}^2$ achieves the best performance by balancing local texture and global structural information. However, due to the randomness of anomaly locations, the single-layer features alone are insufficient to effectively capture all anomalies. In contrast, multiscale fusion proves beneficial by enabling the identification of a wider range of anomalies.

IV. Conclusion

In this paper, we propose a denoising teacher student network based on reverse distillation framework for anomaly detection. Building on the reverse distillation paradigm, we introduce an anomalous teacher network to guide the student network in learning feature associated with anomalies, enhancing the representational disparity for anomalies and improving detection performance. To enhance the representational capacity of the student network, we develop an attention-guided perturbation reconstruction module which directs the student to focus on critical pixel regions. Extensive experiments demonstrate that the proposed DTSNet achieves state-of-the-art performance on the MVTec AD, VisA, and BTAD datasets, surpassing existing methods in anomaly detection.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger, "Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9592–9600.

[2] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth, "f-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *Medical Image Analysis*, vol. 54, pp. 30–44, 2019.

[3] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao, "Future frame prediction for anomaly detection–a new baseline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6536–6545.

[4] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler, "Towards total recall in industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14318–14328.

[5] Tal Reiss, Niv Cohen, Liron Bergman, and Yedid Hoshen, "Panda: Adapting pretrained features for anomaly detection and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2806–2814.

[6] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang, "Simplenet: A simple network for image anomaly detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20402–20411.

[7] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4183–4192.

[8] Guodong Wang, Shumin Han, Errui Ding, and Di Huang, "Student-teacher feature pyramid matching for anomaly detection," *arXiv preprint arXiv:2103.04257*, 2021.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[10] Hanqiu Deng and Xingyu Li, "Anomaly detection via reverse distillation from one-class embedding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9737–9746.

[11] Zhihao Gu, Liang Liu, Xu Chen, Ran Yi, Jiangning Zhang, Yabiao Wang, Chengjie Wang, Annan Shu, Guannan Jiang, and Lizhuang Ma, "Remembering normality: Memory-guided knowledge distillation for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 16401–16409.

[12] Hewei Guo, Liping Ren, Jingjing Fu, Yuwang Wang, Zhizheng Zhang, Cuiling Lan, Haoqian Wang, and Xinwen Hou, "Template-guided hierarchical feature restoration for anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 6447–6458.

[13] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3606–3613.

[14] Ken Perlin, "An image synthesizer," *ACM Siggraph Computer Graphics*, vol. 19, no. 3, pp. 287–296, 1985.

[15] V. Zavrtanik, M. Kristan, and D. Skočaj, "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8330–8339.

[16] Xuebin Qin, Hang Dai, Xiaobin Hu, Deng-Ping Fan, Ling Shao, and Luc Van Gool, "Highly accurate dichotomous image segmentation," in *Proceedings of the European Conference on Computer Vision*, 2022, pp. 38–56.

[17] Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang, "A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization," *arXiv preprint arXiv:2407.09359*, 2024.

[18] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong, "Revisiting reverse distillation for anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 24511–24520.

[19] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

[20] Qibin Hou, Daquan Zhou, and Jiashi Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13713–13722.

[21] Chaokun Shi, Yuexing Hao, Gongyan Li, and Shaoyun Xu, "Knowledge distillation via noisy feature reconstruction," *Expert Systems with Applications*, vol. 257, pp. 124837, 2024.

[22] Zhendong Yang, Zhe Li, Xiaohu Jiang, Yuan Gong, Zehuan Yuan, Danpei Zhao, and Chun Yuan, "Focal and global knowledge distillation for detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4643–4652.

[23] Geoffrey Hinton, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[24] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt, "Asymmetric student-teacher networks for industrial anomaly detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 2592–2602.

[25] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou, "Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16699–16708.

[26] Y. Zou, J. Jeong, L. Pemula, D. Zhang, and O. Dabeer, "Spot-the-difference self-supervised pre-training for anomaly detection and segmentation," in *Proceedings of the European Conference on Computer Vision*, 2022, pp. 392–408.

[27] P. Mishra, R. Verk, D. Fornasier, C.Piciarelli, and G. Foresti, "Vt-adl: A vision transformer network for image anomaly detection and localization," in *Proceedings of International Symposium on Industrial Electronics*, 2021, pp. 01–06.