ADPS: Asymmetric Distillation Post-Segmentation Method for Image Anomaly Detection

Peng Xing, Zechao Li

Abstract

Knowledge distillation-based anomaly detection methods generate same outputs for unknown classes due to the symmetric form of the input and ignore the powerful semantic information of the output of the teacher network since it is only used as a “reference standard”. Towards this end, this work proposes a novel Asymmetric Distillation Post-Segmentation (ADPS) method to effectively explore the asymmetric structure of the input and the discriminative features of the teacher network. Specifically, a simple yet effective asymmetric input approach is proposed to make different data flows through the teacher and student networks. The student network enables to have different inductive and expressive abilities, which can generate different outputs in anomalous regions. Besides, to further explore the semantic information of the teacher network and obtain effective discriminative boundaries, the Weight Mask Block (WMB) and the post-segmentation module are proposed. WMB leverages a weighted strategy by exploring teacher-student feature maps to highlight anomalous features. The post-segmentation module further learns the anomalous features and obtains valid discriminative boundaries. Experimental results on three benchmark datasets demonstrate that the proposed ADPS achieves state-of-the-art anomaly segmentation results.

1. Introduction

Anomaly detection (AD) has a very wide range of applications, such as industrial defect detection [3, 8, 12, 13, 15, 30, 34], medical image diagnosis [2, 25, 35, 40], and the emerging field of automatic driving [6, 41]. However, anomaly detection is different from object detection because the anomaly is difficult to define and it is impossible to obtain all abnormal classes. Anything that differs from normal is considered to be anomalous. When new and unpredictable anomaly classes appear, traditional object detection methods fail. Therefore, unsupervised anomaly detection (UAD) [3,26,33] is proposed to learn only from normal samples. The challenges of UAD are how to fully explore the normal patterns and how to make the anomalous patterns “highlight” in the recognition phase.

Recent studies have shown that knowledge distillation-based anomaly detection methods are very potential [3, 4, 34]. These methods constrain the feature maps of the student network to align the teacher network. When features cannot be aligned, the corresponding regions are detected as the abnormal ones, as shown in Figure 1 (a). However, it does not always work in practice [13]. Firstly, the general teacher-student model uses the same network as the backbone, while their inputs are the same. Therefore, the student network tends to simulate the parameters of the teacher network, which leads to the representation ability of unknown categories by the teacher network being transferred to the student network. When anomalous samples arrive, the difference between the features extracted by the teacher and student networks will disappear. Although smaller student networks (MKDAD) [34] and Reverse Distillation (RD) [13] are proposed to alleviate this problem, they still perform poorly due to MKDAD weakens expressiveness and RD relies on the alignment differences of low-resolution feature maps to obtain decision boundaries. In addition, these methods require to select different thresholds for different datasets and subtle threshold fluctuations produce significant negative effects. The high sensitivity to thresholds indicates that these methods cannot obtain effective feature discriminant boundaries. It can be found from the heatmaps detection in Figure 2, where there are discrimination ambiguities and many misjudged anomalous regions...
for these approaches. Finally, these methods do not fully exploit the teacher network, taking it only as a reference standard while ignoring its powerful knowledge.

Towards this end, a novel unsupervised anomaly detection method, Asymmetric Distillation Post-Segmentation (ADPS) is proposed. As shown in Figure 1 (b), it explores the knowledge distillation with asymmetric input and further exploits the semantic information of the teacher network to obtain discriminative boundaries. Specifically, the proposed asymmetric input method generates semantically identical but differently shaped data streams from the original input data as the new input data. ADPS utilizes it in the different stage layers of the student network. It weakens the student network to learn the inductive and expressive abilities of the teacher network for unlearned categories (abnormal categories). Therefore, the teacher-student networks of ADPS generate different outputs for abnormal samples. Meanwhile, it preserves the deeper structure of the student network, which guarantees its powerful expressiveness. Moreover, the Weighted Mask Block (WMB) is proposed to highlight anomalies explicitly. It utilizes the feature maps to generate the weight mask and leverages it to highlight the abnormal features in the original feature maps of the teacher network. Finally, a post-segmentation module is developed to learn the feature maps after weighting. It is expected to further exploit the powerful knowledge in the teacher network to extract accurate decision boundaries with the help of WMB.

Extensive experiments are conducted on three public benchmark datasets. The experimental results demonstrate that the proposed ADPS method achieves state-of-the-art anomaly segmentation performance. It outperforms the recent state-of-the-art knowledge distillation-based method RD [13] by 17% for anomaly segmentation on the challenging MVTec AD dataset [3]. The main contributions are summarized as follows:

1. This work proposes a novel Asymmetric Distillation Post-Segmentation (ADPS) method for unsupervised anomaly detection. It develops asymmetric input distillation structures and further explores the knowledge of the teacher network, rather than just using it as a “reference standard”.

2. A simple yet effective asymmetric input approach is proposed to enable the student network to ensure powerful expressive abilities. Meanwhile, the student network learns from different data flows to obtain different expressive abilities, thus generating different responses to anomalies.

3. A Weight Mask Block (WMB) is proposed to highlight explicitly the abnormal features in teacher feature maps and a post-segmentation module is developed to fully explore the weighted feature maps to obtain accurate segmentation results.

2. Related Work

Many anomaly detection methods have been proposed, including AutoEncoder (AE) [21]-based approaches, generative model-based approaches, deep feature embedding-based approaches, and knowledge distillation-based approaches.

The AE-based anomaly detection model depends on that the train AE to reconstruct only normal images [5,15,16,23,27,32]. However, AE has a powerful pixel-level reconstruction ability, resulting in anomalous regions that can be well reconstructed even if they are not learned. Therefore, some researchers introduced more complex self-supervised tasks enabling the model to learn semantic reconstruction [15,32] or memory modules storing normal sample features while only reading normal features [16,19,27,43]. However, these methods do not work well in practice for anomaly segmentation due to the large error in the reconstructed normal regions.

To overcome the drawbacks of AE-based methods, some researchers used the generative model to explicitly or implicitly model the feature distribution of normal data. If the generated sample differs from the input sample, the input is detected as an abnormal sample. Schlegl et al. [36] first applied GAN [17] to localize anomalies. GANomaly [1] and F-AnoGAN [35] improve the generator to impose constraints on reconstructing potential spatial features. DefGAN [51] improves the discriminator by employing multiple discriminators. Unfortunately, these methods do not perform well for anomalous samples with tiny anomaly regions.

Deep feature modeling-based methods first construct a feature space for the input normal image and then decide
Figure 3. Overview of the proposed Asymmetric Distillation Post-Segmentation (ADPS) method. The input of the teacher network $T$ is $I$. Meanwhile, the asymmetric input approach generates semantically identical but differently shaped $I_S$ as the input to the student network $S$. Then, $T$ and $S$ extract multi-scale feature maps $T_i$ and $S_i$, respectively. WMB utilizes $T_i$ and $S_i$ to calculate the weight mask $W_i$ as well as weights $T_i$ to obtain $C_i$. $C_i$ is input to the post-segmentation module to get the final result $M$. From $M$, the anomaly score and localization image of $I$ can be calculated.

Whether the features are abnormal or not by spatial boundaries [9–12, 31, 47]. SPADE [11] uses a pre-trained Wide-ResNet50 [48] to extract patch features and match whether the test image patch has $k$ adjacent normal patches. A patch distribution modeling (PaDiM) framework [12] was proposed by first using a pre-trained model to extract features, and then modeling normality with a multivariate gaussian distribution [14, 18] for each location. These methods fail to achieve good anomaly segmentation results.

For the knowledge distillation-based methods, U-Std [4] was first proposed to use a knowledge distillation model for UAD. Further, a multi-resolution knowledge distillation method (MKDAD) [34] was proposed to overcome the drawback that U-Std only uses the output of the last layer. Wang et al. proposed the multi-scale feature map approach by introducing the Student-Teacher Feature Pyramid Matching (STPM) model [42]. Different from MKDAD which uses the gradient of the loss function, STPM uses the difference between the multi-feature maps as localization maps. However, models with symmetric or similar structures are unable to extract differentiated features. RD [13] employs a pre-trained model as an encoder and allows the student decoder to align the encoder output. However, the difference maps from low resolution lacked detailed information, and none of them explore the teacher network effectively. Towards this end, ADPS develops an asymmetric input approach that allows different data flows to pass through the teacher-student network, which enables the student network to learn different inductive abilities than the teacher network. Meanwhile, WMB and post-segmentation module are developed to explore the advanced semantic information of the teacher network to obtain accurate decision boundaries.

3. The Proposed Approach

3.1. Overview

As shown in Figure 3, the proposed Asymmetric Distillation Post-Segmentation (ADPS) method includes a knowledge distillation model, a Weight Mask Block (WMB), and a post-segmentation module. APDS develops an asymmetric input approach in the knowledge distillation model which allows the inputs of the teacher network and the student network to be the same semantics, but with different shapes. The student network needs to learn the feature alignment of different shapes of data flows, which will further enhance the difficulty of student network learning. The feature maps of teacher-student networks are extracted and generated as a weight mask by WMB to weight the feature maps of the teacher network in an explicit way. The post-segmentation module takes the weighted feature maps as input to further explore their potential semantic information segmenting the detailed anomalous regions.

3.2. Asymmetric Input Approach

To expand the difference between the student network $S$ and the teacher network $T$ for anomalous features, asymmetric input method is proposed. It consists of two parts: the asymmetry of the input layer and the asymmetric input
of the stage layers. In the input layer, each input sample \( I \) is divided equally into non-overlapping \( k^2 \) small parts to construct \( I_S = \{ I_1, I_2, ..., I_{k^2} \} \). In this way, \( I \) and \( I_S \) guarantee the same semantics but different shapes. \( I \) is used as the input to \( T \) and \( I_S \) is used as a set of inputs to \( S \). To further make the asymmetric inputs of different stage layers, the output \( S_i \) of the different stages are similarly divided into \( k^2 \) overlapping parts. Therefore, the input of the \( i + 1 \) layer in the student network is \( S_i' = \{ S_i^1, S_i^2, ..., S_i^{k^2} \} \). Importantly, \( S_i' \) is reshaped to the size of \( S_i \) at the feature alignment to fully learn the representation of normal features in \( T_i \).

The inputs of \( T \) and \( S \) are the same, which results in \( S \) tending to simulate the parameters of \( T \) to align the extracted features while being insensitive to the input data. \( T \) transfers strong representational ability to \( S \), especially the inductive ability regarding unlearned classes. The asymmetric input approach makes \( T \) and \( S \) have different input data flows, which are semantically identical and differ only in shape. The same semantics allows the trained models to align normal samples even if shapes are different. Since \( S \) and \( T \) are learned from different data flows, \( S \) learns different expressive capabilities from \( T \). When unknown anomalous samples arrive, \( T \) and \( S \) will produce different responses. Meanwhile, it avoids the use of smaller student networks to ensure that the student network has powerful representation abilities.

### 3.3. Weight Mask Block

Recent knowledge distillation methods are highly sensitive to segmentation thresholds. The main reasons are that (1) the inappropriate distillation approach prevents the model from producing a differentiated representation (mentioned in Section 3.2) and (2) the difference map cannot further expand the differentiated representation of normal and abnormal features to obtain accurate decision boundaries. Towards this end, a Weight Mask Block (WMB) is proposed to explicitly guide the anomalous features of the feature map. Instead of directly using the difference map after thresholding as the final result, WMB is used to initially estimate the anomalous regions of the feature map and to assign anomaly weights.

As shown in Figure 3, the inputs of WMB are \( T_i \) and \( S_i \) and its outputs are \( C_i \), where \( T_i, S_i \) and \( C_i \in \mathbb{R}^{h \times w \times c} \). First, the Cosine similarity function is used to measure whether the feature maps are aligned or not. Specifically, the similarity of the \( x, y \) positions of the feature maps is

\[
W_i^{x,y} = \frac{T_i^{x,y} \cdot S_i^{x,y}}{\|T_i^{x,y}\| \times \|S_i^{x,y}\|},
\]

where \( T_i^{x,y} \in \mathbb{R}^{1 \times 1 \times c} \), \( S_i^{x,y} \in \mathbb{R}^{1 \times 1 \times c} \). Then, the similarity of each position is calculated to constitute the matrix \( W_i = \{ W_i^{x,y} | x \in [1, h], y \in [1, w] \} \) of \( T_i \) and \( S_i \). Larger values of \( W_i \) indicate more consistent features extracted by the teacher-student network. Since the anomalous features are not aligned, \( 1 - W_i \) represents the rough localization of the anomalous features and serves as the weight mask. Finally, WMB reweights the original feature map \( T_i \) utilizing the weight mask \( 1 - W_i \) to highlight the anomalous features in \( T_i \) to provide explicit guidance information for subsequent exploration. The output result is the reweighted feature \( C_i \),

\[
C_i = (1 - W_i) \cdot T_i.
\]

Therefore, \( C_i \) contains comprehensive feature information and the anomalous features are highlighted by WMB, which facilitates the exploration of anomalous features by the post-segmentation module (mentioned in Section 3.4).

### 3.4. Post-Segmentation Module

To fully explore the pre-trained knowledge in the teacher network, rather than just taking it as a “reference standard”, while solving the threshold sensitivity problem, the post-segmentation module is proposed. It learns normal and abnormal features directly, while it will enlarge the difference between normal and abnormal features under the segmentation loss constraint to learn more accurate decision boundaries.
Table 2. The anomaly segmentation results in terms of $\text{AUROC}_\text{seg}$ and $\text{AP}_\text{seg}$ (%) on the MVtec AD dataset. The proposed ADPS method outperforms recent methods by nearly 10% in terms of $\text{AP}_\text{seg}$ and achieves new SOTA. The best results are marked in bold.

| Metric | Methods   | Textures | Objects |
|--------|------------|----------|---------|
|        |            | Carpet  | Grid    | Leather | Wood    | Tile    | mean | Bottle | Cable | Capsule | Haze... |
| $\text{AUROC}_\text{seg}$ | U-Std [4]  | 93.5    | 89.9    | 97.8    | 92.1    | 92.5    | 93.2 | 97.8   | 91.9   | 96.8    | 98.2    | 92.7   | 96.5   | 97.4    | 97.9    | 73.7   | 95.6   | 94.3    | 93.9    |
|        | RIAD [50]  | 96.3    | 98.8    | 99.4    | 85.8    | 89.1    | 93.9 | 98.4   | 84.2   | 92.8    | 96.1    | 92.5   | 95.7   | 98.8    | 98.9    | 87.7   | 97.8   | 94.3    | 94.2    |
|        | Cutpaste [22] | 98.3    | 97.5    | 99.5    | 95.5    | 90.5    | 96.3 | 97.6   | 90.2   | 97.4    | 97.3    | 93.1   | 95.7   | 96.7    | 98.1    | 93    | 99.3   | 95.8    | 96.0    |
|        | SPADE [11] | 97.5    | 93.7    | 97.6    | 88.5    | 87.4    | 92.9 | 98.4   | 97.2   | 99.0    | 99.1    | 98.1   | 96.5   | 98.9    | 97.9    | 94.1   | 96.5   | 97.6    | 96.5    |
|        | PaDim [12] | 99.0    | 98.6    | 99     | 94.1    | 94.1    | 97.0 | 98.2   | 96.7   | 98.6    | 98.1    | 97.3   | 95.7   | 94.4    | 98.8    | 97.6   | 98.4   | 97.4    | 97.4    |
|        | DRAEM [49] | 95.5    | 99.7    | 98.6    | 96.4    | 99.2    | 97.9 | 99.1   | 94.7   | 94.3    | 99.7    | 95.5   | 97.6   | 97.6    | 98.1    | 90.9   | 98.8   | 97.0    | 97.3    |
|        | MKDAD [34] | 95.6    | 91.8    | 98.1    | 82.8    | 84.8    | 90.6 | 96.3   | 82.4   | 95.9    | 94.6    | 86.4   | 89.6   | 96.0    | 96.1    | 76.5   | 93.9   | 90.8    | 90.7    |
|        | RD [13]    | 98.9    | 99.3    | 99.4    | 95.6    | 95.3    | 97.7 | 98.7   | 97.4   | 98.7    | 98.9    | 97.3   | 98.2   | 99.3    | 99.1    | 92.5   | 98.2   | 97.8    | 97.8    |
| $\text{AP}_\text{seg}$ | ADPS       | 99.5    | 99.2    | 99.9    | 99.3    | 99.6    | 99.5 | 99.5   | 94.6   | 94.1    | 99.6    | 97.5   | 93.8   | 98.7    | 99.1    | 92.2   | 99.6   | 97.4    | 98.1    |

Its input is the weighted multiscale feature map $C = \{ C_1, C_2, ..., C_n \}$, where $n$ represents the number of multiscale layers. Therefore, the post-segmentation module fully exploits the semantic information of the extracted features from the teacher network and enhances the differentiated representation of normal and abnormal features. Meanwhile, WMB provides explicit guidance information, which helps the post-segmentation module to obtain accurate discriminative feature boundaries. The output $M$ is a mask with the size of $h \times w$. The value domain ranges from $[0, 1]$. Values closer to 1 indicate anomalous pixels (ADPS utilizes decision threshold of 0.5 for all datasets).

3.5. Loss Function

To achieve anomaly segmentation as well as to balance the positive and negative samples of segmentation loss, the problem is formulated as

$$ L_d = (1 - Y) \cdot (1 - \frac{T_i \cdot S_i}{||T_i|| \times ||S_i||}) + Y \cdot \frac{T_i \cdot S_i}{||T_i|| \times ||S_i||}, \quad (3) $$

where $Y$ represents the GroundTruth (GT) label of training samples, $Y = \{ y_{i,j} \in \{0, 1\} | i \in [1, h], j \in [1, w] \}$. The post-segmentation module introduces the Focal loss $L_s$ [24].

$$ L_s = \begin{cases} 
-\log(p_{t,i,j}) & y_{i,j} = 1, \\
-p_{t,i,j}^{\gamma} \log(1-p_{t,i,j}) & y_{i,j} = 0, 
\end{cases} \quad (4) $$

where $p_{t,i,j}$ represents the prediction value. To incorporate the distill loss $L_d$ and segmentation loss $L_s$, the overall loss $L$ is defined as follows:

$$ L = L_d + \lambda L_s, \quad (5) $$

where $\lambda$ is set to measure the importance of the two losses.

4. Experiments

4.1. Experimental settings

Datasets. (1) MVtec AD [3] contains 15 categories of high-resolution images from real industrial scenes, which includes five categories of texture images and 10 categories of object images. The training dataset contains only normal images and the testing dataset has multiple anomalous patterns. It is very challenging due to its tiny anomalies and high resolution. (2) KolektorSDD [38] were captured in a controlled industrial environment in a real-world case. The dataset consists of 399 images with the size of $500 \times 1250$. (3) KolektorSDD2 [7] is a surface defect detection dataset containing more than 3000 images with a wide range of real industrial defect types. It has different types of defects including scratches, dots, surface defects, etc. The image size is approximately $230 \times 630$.

Implementation Details. The ADPS method introduces WideResNet50 [48] as the backbone for teacher-student networks. The feature maps of stage 1 to stage 3 are extracted. The number $k$ of parts is set to 8. The post-segmentation module utilizes the decoder of UNet [29] as backbone. In the training phase, the learning rates of the distillation part and the segmentation module are set to 0.0001 and the batch size is set to 32. The Adam optimizer is utilized. A total of 300 training epochs are performed. The learning rates are decayed at $[240, 270]$ epochs, and the decay rate is set to 0.2. The size of the input image is set to a specific resolution $256 \times 256$. The compared methods contain knowledge distillation-based methods including U-Std [4], MKDAD [34], and RD [13], self-supervised learning-based methods including DAAD+ [19], RIAD [50], Cutpaste [22], DRAEM [49], AnoSeg [37] and SGSF [45], as well as deep feature embedding-based methods including SPADE [11], PaDim...
The anomaly segmentation task usually consists of two subtasks: anomaly classification and anomaly segmentation. For anomaly classification, the maximum value of the segmentation map \( M \) after the average-pooling layer is served as the anomaly score of the sample. Following previous studies [12,13], the Area Under the Receiver Operating Characteristic (AUROC) is adopted as the evaluation metric, which is denoted as \( \text{AUROC}_{\text{cla}} \). For anomaly segmentation, this work not only report the AUROC result but also introduce pixel-level Average Precision (AP) results, denoted as \( \text{AUROC}_{\text{seg}} \) and \( \text{AP}_{\text{seg}} \), respectively. The reason is that the false positive rate (FPR) in \( \text{AUROC}_{\text{seg}} \) is mainly determined by a very large number of normal pixels [39]. Even if the \( \text{AUROC}_{\text{seg}} \) is very high, the method could not achieve fine-grained localization results. The advantage of \( \text{AP}_{\text{seg}} \) is that it is more suitable for highly unbalanced categories, especially in the field of anomaly detection.

### 4.2. Results

Anomaly classification results on the MVTec AD dataset are shown in Table 1. ADPS achieves the state-of-the-art performance for anomaly classification on 9 out of 15 categories and 100% on 4 categories. Meanwhile, the mean results of ADPS are state-of-the-art for both “Texture” and “Object” images. In addition, ADPS shows the significant performance improvement in several categories compared to the methods using WideResNet50, with about 3-10% improvement of anomaly classification ability, which verifies that the WMB and post-segmentation modules of ADPS explore semantic information more efficiently and improve anomaly detection.

The anomaly segmentation results on the MVTec AD dataset are shown in Table 2. ADPS achieves the new state-of-the-art performance.

### Table 3. Anomaly detection results for the KolektorSDD and KolektorSDD2 datasets. The proposed method achieves the best performance.

| Method       | KolektorSDD | KolektorSDD2 |
|--------------|-------------|--------------|
|              | \( \text{AUROC}_{\text{cla}} \) | \( \text{AUROC}_{\text{seg}} \) | \( \text{AP}_{\text{seg}} \) | \( \text{AUROC}_{\text{cla}} \) | \( \text{AUROC}_{\text{seg}} \) | \( \text{AP}_{\text{seg}} \) |
| Semi-orthogonal [20] | 96.0 | - | 98.1 | - | 96.0 | - | 98.1 | - |
| PaDim [12]  | 94.5 | - | 95.6 | - | 94.5 | - | 95.6 | - |
| U-Std [4]   | 89.6 | - | 95.0 | - | 89.6 | - | 95.0 | - |
| SGSF [45]  | - | 93.5 | 91.5 | 51.6 | - | 93.5 | 91.5 | 51.6 |
| RD [13]    | - | 94.8 | 98.2 | 47.7 | - | 94.8 | 98.2 | 47.7 |
| ADPS       | 96.3 | 95.6 | 99.2 | 72.5 | - | 95.6 | 99.2 | 72.5 |

### Table 4. The anomaly detection results of the recent anomaly detection methods with teacher-student networks in terms of \( \text{AUROC}_{\text{cla}} \), \( \text{AUROC}_{\text{seg}} \) and \( \text{AP}_{\text{seg}} \) (%) on the MVTec AD dataset. The proposed APDS achieves the best performance.

| Metrics | U-Std [4] | MKDAD [34] | RSTPM [46] | STPM [42] | ADPS |
|---------|-----------|------------|------------|-----------|------|
| \( \text{AUROC}_{\text{cla}} \) | 87.7 | 87.7 | 96.9 | 95.5 | 97.4 |
| \( \text{AUROC}_{\text{seg}} \) | 90.0 | 91.4 | 97.7 | 97.0 | 98.1 |
| \( \text{AP}_{\text{seg}} \) | 45.5 | - | - | 77.4 | - |
of-the-art performance in terms of $\text{AUROC}_{\text{seg}}$ and $\text{AP}_{\text{seg}}$. It outperforms the knowledge distillation-based U-Std and MKDAD by 5-10% on $\text{AUROC}_{\text{seg}}$, and outperforms U-std by 32% on $\text{AP}_{\text{seg}}$, which indicate ADPS has excellent anomaly segmentation capability. ADPS even far outperforms the remaining anomaly detection methods by 10%-30% on $\text{AP}_{\text{seg}}$, which shows that the proposed method is better able to explore anomalous semantic information and obtain precise decision boundaries. As shown in Figure 4, ADPS achieves accurate segmentation not only for large anomalous regions but also for subtle anomalous regions. Especially for the anomaly boundaries, the misjudged segmentation regions of ADPS is low compared with that of RD. Benefiting from WMB and the post-segmentation module to further explore the semantic information, more reasonable decision boundaries are obtained. In summary, the results validate that ADPS can effectively solve the anomaly localization problem.

The results of anomaly detection on the KolektorSDD and KolektorSDD2 datasets are shown in Table 3. ADPS achieves state-of-the-art performance on both datasets. Figure 5 shows the results of anomaly segmentation on the KolektorSDD2 dataset. Anomaly patterns with different defects can be detected with a low pixel-level mislocalization rate.

In addition, we compare ADPS and anomaly detection methods using the teacher-student network. The performance on the MVTec AD dataset is shown in Table 4. ADPS has an obvious advantage in both anomaly classification and anomaly segmentation tasks, which demonstrates the effectiveness of the proposed teacher-student network with asymmetric input approach. In addition, ADPS further explores the semantic information of the features extracted by the teacher network, which is crucial for discriminating anomalies.

### 4.3. Ablation analysis

**Validity of knowledge distillation and post-segmentation module:** To validate the effectiveness of each component, three baselines: ADPS, W/O. PS (ADPS without the post-segmentation module) and W/O. S (ADPS without the student network) are compared.

The results on the MVTec AD dataset with three baselines are shown in Figure 6. W/O. PS method lacking the post-segmentation module decreases by 7% on $\text{AUROC}_{\text{cla}}$ and 20% on $\text{AP}_{\text{seg}}$, which shows that the post-segmentation module is able to further explore the pre-training knowledge and obtain effective decision boundaries. W/O. S lacks the weight mask generated by the student network through the WMB. Therefore, it lacks the abil-
Figure 9. Illustration of the different T-S features fusion methods for \( C_i \).

Table 6. Anomaly detection results of different T-S feature fusion methods on the MVTec AD dataset. WMB achieves the best performance.

| Method(1) | \( \text{AUROC}_{\text{cls}} \) | \( \text{AUROC}_{\text{seg}} \) | \( \text{AP}_{\text{seg}} \) |
|----------|-----------------|-----------------|-----------------|
| ADPS     | 98.2            | 99.5            | 83.1            |
| Method(2)| 99.6            | 99.0            | 81.7            |
|          | 94.6            | 59.7            | 69.1            |
| Method(3)| 96.4            | 96.9            | 73.5            |
|          | 94.5            | 92.8            | 66.8            |
|          | 96.1            | 94.9            | 71.8            |

Asymmetric Input Approach: The ablation related to the asymmetric input approach of ADPS is provided on the MVTec AD dataset. In the input layer, different values of \( k \) (1, 2, 4, 8, 16) are explored. The anomaly detection results are reported in Figure 8. First, the performance without the asymmetric input approach \( (k = 1) \) is lower than that with the asymmetric input approach \( (k = 4 \text{ and } k = 8, \text{ etc.}) \). The reason is that the asymmetric input approach makes \( T \) and \( S \) learn different expressive capabilities and WMB provides more effective guidance information. Second, too large \( k \) has a negative impact. The reason may be that too many chunks lose the structural information of the image.

Besides, ADPS introduces the asymmetric input approach into the feature level of different stages, the anomaly detection results are shown in Table 5. It can be observed that the introduction of the asymmetric approach at the feature level has a beneficial effect for the anomaly classification, especially for “Texture” image. For the anomaly segmentation task, it can be observed that the introduction of asymmetric input methods in both the input layer and the shallow layer is effective.

Weight Mask Block: To validate the effect of the WMB module by explicit guidance, as shown in Figure 9, different fusion methods for \( C_i \) are proposed and compared. The method (1) directly uses the feature difference map as input to the post-segmentation module, while method (2) utilizes convolution block to first downscale and then concatenate the feature maps, while WMB to generate weight mask to highlight explicitly anomalous features of \( T_c \). The performance of anomaly detection on the MVTec AD dataset is shown in Table 6. The results show that the proposed WMB is the best. The reason is that the generated mask has powerful guidance information, which assists the post-segmentation module to locate the anomalies more effectively. The rest of the methods are implicitly fused features and do not provide effective guidance.

In addition, the effect of different weight mask generation methods in WMB is explored. Besides the Cosine similarity function, we introduce the mean squared error (MSE) to generate the weight masks. To ensure consistency, the distillation network is constrained by the corresponding loss function. The results for different groups are shown in Table 7. It can be observed that MSE is more effective for anomaly detection on “Texture” images, while having poor results on “Object” images. The combination of cosine similarity loss and cosine similarity generating weight mask is used for all images with the competitive performance.

5. Conclusion

In this paper, the ADPS approach is proposed to address the UAD task, which explores the distillation pattern of asymmetric inputs and fully exploits the semantic feature information of the teacher network rather than just taking the teacher network as a “reference standard”. Guided by the explicit anomaly mask provided by WMB, the post-segmentation module further expands the difference between normal and anomalous features as well as obtains precise decision boundaries. Extensive experiments are conducted on three benchmark datasets to validate the UAD capability of ADPS. In the future, we will develop better explicit guidance methods to further enhance the anomaly detection capability.
[33] Mohammadreza Salehi, Hossein Mirzaei, Dan Hendrycks, Yixuan Li, Mohammad Hossein Rohban, and Mohammad Sabokrou. A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. *arXiv preprint arXiv:2110.14051*, 2021. 1

[34] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *CVPR*, pages 14902–14912, 2021. 1, 2, 3, 4, 5, 6

[35] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *MIA*, 54:30–44, 2019. 1, 2

[36] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *ICIP in medical imaging*, pages 146–157. Springer, 2017. 2

[37] Jouwon Song, Kyeongbo Kong, Ye-In Park, Seong-Gyun Kim, and Suk-Ju Kang. Anoseg: Anomaly segmentation network using self-supervised learning. *arXiv preprint arXiv:2110.03396*, 2021. 1, 2, 3, 4, 5

[38] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *JIM*, 31(3):759–776, 2020. 5

[39] Xian Tao, Xinyi Gong, Xin Zhang, Shaohua Yan, and Chandranath Adak. Deep learning for unsupervised anomaly localization in industrial images: A survey. *TIM*, 2022. 4

[40] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In *ECCV*, pages 485–503, 2020. 1

[41] Tomas Vojír, Tomáš Šipka, Rahaf Aljundi, Nikolay Chumerin, Daniel Olmeda Reino, and Jiří Matas. Road anomaly detection by partial image reconstruction with segmentation coupling. In *ICCV*, pages 15651–15660, 2021. 1

[42] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for unsupervised anomaly detection. *arXiv preprint arXiv:2103.04257*, 2021. 3, 6

[43] Tian Wang, Xing Xu, Fumin Shen, and Yang Yang. A cognitive memory-augmented network for visual anomaly detection. *IEEE JAS*, 8(7):1296–1307, 2021. 2

[44] Jhih-Ciang Wu, Ding-Jie Chen, Chiou-Shann Fuh, and Tyng-Luh Liu. Learning unsupervised metaformer for anomaly detection. In *ICCV*, pages 4369–4378, 2021. 4

[45] Peng Xing, Yanpeng Sun, and Zechao Li. Self-supervised guided segmentation framework for unsupervised anomaly detection. *arXiv preprint arXiv:2209.12440*, 2022. 4, 5, 6

[46] Shinji Yamada and Kazuhiro Hotta. Reconstruction student with attention for student-teacher pyramid matching. *arXiv preprint arXiv:2111.15376*, 2021. 6

[47] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *ACCV*, 2020. 3

[48] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *BMVC*, 2016. 3, 5, 6

[49] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8330–8339, 2021. 5

[50] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *PR*, 112:107706, 2021. 4, 5

[51] Dongkai Zhang, Shihin Gao, Long Yu, Gaoqiang Kang, Xiaoguang Wei, and Dong Zhan. Defgan: Defect detection gans with latent space pitting for high-speed railway insulator. *TIM*, 70:1–10, 2020. 2