Prototypical Residual Networks for Anomaly Detection and Localization

Hui Zhang[1,2] Zuxuan Wu[1,2] Zheng Wang[3] Zhineng Chen[1,2*] Yu-Gang Jiang[1,2]
[1]Shanghai Key Lab of Intell. Info. Processing, School of CS, Fudan University
[2]Shanghai Collaborative Innovation Center of Intelligent Visual Computing
[3]School of Computer Science, Zhejiang University of Technology

Abstract

Anomaly detection and localization are widely used in industrial manufacturing for its efficiency and effectiveness. Anomalies are rare and hard to collect and supervised models easily over-fit to these seen anomalies with a handful of abnormal samples, producing unsatisfactory performance. On the other hand, anomalies are typically subtle, hard to discern, and of various appearance, making it difficult to detect anomalies and let alone locate anomalous regions. To address these issues, we propose a framework called Prototypical Residual Network (PRN), which learns feature residuals of varying scales and sizes between anomalous and normal patterns to accurately reconstruct the segmentation maps of anomalous regions. PRN mainly consists of two parts: multi-scale prototypes that explicitly represent the residual features of anomalies to normal patterns; a multi-size self-attention mechanism that enables variable-sized anomalous feature learning. Besides, we present a variety of anomaly generation strategies that consider both seen and unseen appearance variance to enlarge and diversify anomalies. Extensive experiments on the challenging and widely used MVTec AD benchmark show that PRN outperforms current state-of-the-art unsupervised and supervised methods. We further report SOTA results on three additional datasets to demonstrate the effectiveness and generalizability of PRN.

1. Introduction

The human cognition and visual system has an inherent ability to perceive anomalies [53]. Not only can humans distinguish between defective and non-defective images, but they can also point to the location of anomalies even if they have seen none or only a limited number of anomalies. Anomaly detection (image-level binary classification) and anomaly localization (pixel-level binary classification) are introduced for the same purpose, and have been widely used in various scenarios due to their efficiency and remarkable accuracy, including industrial defect detection [4, 7, 34, 61], medical image analysis [52] and video surveillance [32].

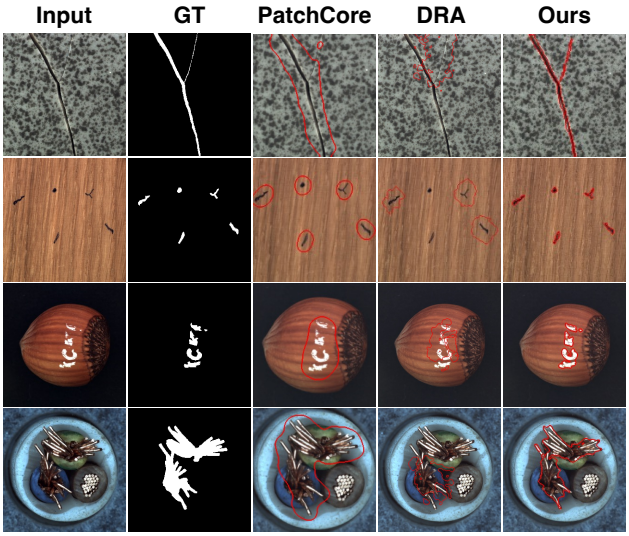

Figure 1. Anomaly detection and localization examples on MVTec [4]. Compared with the unsupervised method PatchCore [41] and the supervised method DRA [13], the proposed PRN is able to locate the anomalous regions more accurately.

Given its importance, a significant amount of work has been devoted to anomaly detection and anomaly localization, but few have addressed both detection and localization problems well at the same time. We argue that real-world anomalous data weaken these models mainly in three aspects: I) the amount of abnormal samples is limited and significant fewer than normal samples, producing data distributions that lead to a naturally **imbalanced learning** problem; II) anomalies are typically subtle and hard to discern, since normal patterns still dominate the anomalous image; **identifying abnormal regions** out of the whole image is the key to anomaly detection and localization; III) the appearance of anomalies varies significantly, *i.e.*, abnormal regions can take on a variety of sizes, shapes and numbers, and such **appearance variations**

* Corresponding author.

用于异常检测和定位的原型残差网络

张辉[1,2] 吴祖煊[1,2] 王铮[3] 陈志能[1,2*] 蒋宇凡[1,2] [1]复旦大学计算机科学技术学院 上海市智能信息处理重点实验室 [2]上海智能视觉计算协同创新中心 [3]浙江工业大学计算机科学与技术学院

摘要

Anomaly detection and localization are widely used in industrial manufacturing for its efficiency and effectiveness. Anomalies are rare and hard to collect and supervised models easily over-fit to these seen anomalies with a handful of abnormal samples, producing unsatisfactory performance. On the other hand, anomalies are typically subtle, hard to discern, and of various appearance, making it difficult to detect anomalies and let alone locate anomalous regions. To address these issues, we propose a framework called Prototypical Residual Network (PRN), which learns feature residuals of varying scales and sizes between anomalous and normal patterns to accurately reconstruct the segmentation maps of anomalous regions. PRN mainly consists of two parts: multi-scale prototypes that explicitly represent the residual features of anomalies to normal patterns; a multi-size self-attention mechanism that enables variable-sized anomalous feature learning. Besides, we present a variety of anomaly generation strategies that consider both seen and unseen appearance variance to enlarge and diversify anomalies. Extensive experiments on the challenging and widely used MVTec AD benchmark show that PRN outperforms current state-of-the-art unsupervised and supervised methods. We further report SOTA results on three additional datasets to demonstrate the effectiveness and generalizability of PRN.

1. 引言

人类认知与视觉系统天生具备感知异常的能力[53]。人类不仅能够区分缺陷图像与正常图像，甚至在从未见过或仅接触过少量异常样本的情况下，仍能精准指出异常位置。异常检测（图像级二元分类）与异常定位（像素级二元分类）正是基于这一目的被提出，并已获得广泛应用。

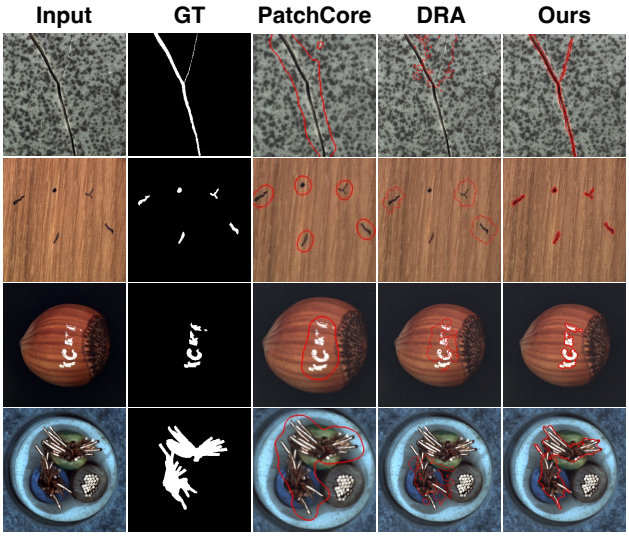

图1. MVTec [4]上的异常检测与定位示例。相较于无监督方法PatchCore [41]和有监督方法DRA [13]，所提出的PRN能够更精确地定位异常区域。

由于其高效性和卓越的准确性，它们被应用于多种场景，包括工业缺陷检测[4, 7, 34, 61]、医学图像分析[52]以及视频监控[32]。

鉴于其重要性，已有大量工作致力于异常检测与异常定位，但少有研究能同时妥善解决检测与定位问题。我们认为，现实世界中的异常数据主要在三个方面削弱了这些模型的性能：I) 异常样本数量有限且远少于正常样本，导致数据分布不均衡，形成天然的不平衡学习问题；II) 异常通常细微难辨，因为正常模式仍主导异常图像；从整幅图像中识别异常区域是异常检测与定位的关键；III) 异常的外观差异显著，*i.e*，异常区域可能呈现多种尺寸、形状和数量，此类外观变化

* Corresponding author.

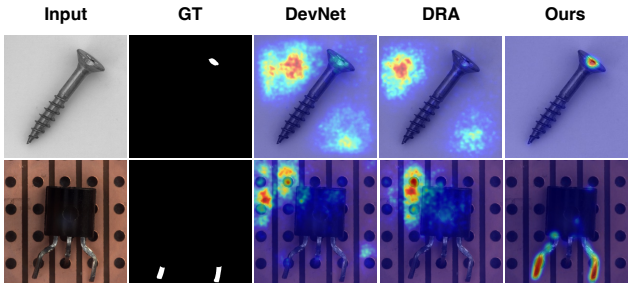

Figure 2. Indecipherable problem of supervised methods DevNet [35] and DRA [13]. Both images are detected as anomalous. Other methods mistakenly highlight normal regions rather than defect regions, whereas PRN correctly pinpoints the defect regions.

make it challenging to well-localizing all the anomalies.

Without adequate anomalies for training, unsupervised models become the de facto dominant approaches, which get rid of the imbalance problem by learning the distribution of normal samples [5, 9, 10, 12, 18, 25, 41–43, 47, 67] or generating sufficient synthetic anomalies [26, 28, 50, 63, 68]. However, these methods are opaque to genuine anomalies, resulting in implicit decisions that may induce many false negatives and false positives. Besides, unsupervised methods rely heavily on the quality of normal samples, and thus are not robust enough and perform poorly on uncalibrated or noisy datasets [19]. As shown in Fig. 1, unsupervised models predict broad regions around the anomaly. We attribute this problem to less discriminative abilities of these methods.

Recently, several supervised methods [13, 35, 46] are introduced. DeepSAD [46] enlarges the margin between the anomaly and the one-class center in the latent space to obtain more compact one-class descriptors by limit seen anomalies. DRA [13] and DevNet [35] formulate anomaly detection as a multi-instance learning (MIL) problem, scoring an image as anomaly if any image patch is a defect region. MIL-based methods enforce the learning at fine-grained image patch level, which effectively reduces the interference of normal patches in the anomalous images. Yet, these approaches typically struggle to accurately locate all anomalous regions with image-level supervision, as shown in Fig. 1. In particular, when the anomalous regions only occupy a tiny part of image patches, image-level representation may be dominated by the normal regions and disregards tiny anomalous, which may cause inconsistent image-level and pixel level performance as shown in Table 1. Furthermore, as shown in Fig. 2, these methods also encounter uninterpretable problems when making decisions.

In this paper, we propose a framework called Prototypical Residual Network (PRN) as an effective remedy for aforesaid issues on anomaly detection and localization. First, we propose multi-scale prototypes to represent normal patterns. In contrast to previous methods for constructing normal patterns from concatenated feature memory [41] or random sampled feature maps [63], we construct normal patterns with prototypes of intermediate feature maps of different scales, thereby preserving the spatial information and providing precise and representative normal patterns. Further, we obtain the feature map residuals via the deviation between the anomalous image and the closest prototype at each scale, and we add multi-scale fusion blocks to exchange information across different scales. Second, since the appearance of anomaly regions varies a lot, it is necessary to learn relationships among patches from multiple receptive fields. Thus, we introduce a multi-size self-attention [33, 55, 58, 59] mechanism, which operates on patches of different receptive fields to detect patch-level inconsistencies at different sizes. Finally, unlike previous methods [13, 35] that use image-level supervision for training, our model learns to reconstruct the anomaly segmentation map with pixel-level supervision, which focuses more on the anomalous regions and preserves better generalization. Besides, we put forward a variety of anomaly generation strategies that efficiently mitigate the impact of data imbalance and enrich the anomaly appearance. With the proposed modules, our method achieves more accurate localization than previous unsupervised and supervised methods, as shown in Fig. 1 and Fig. 2.

The main contributions of this paper are summarized as follows:

- We propose a novel Prototypical Residual Networks for anomaly detection and localization. Equipped with multi-scale prototypes and the multi-size self-attention mechanism, PRN learns residual representations among multi-scale feature maps and within the multi-size receptive fields at each scale.
- We present a variety of anomaly generation strategies that considering both seen and unseen appearance variance to enlarge and diversify anomalies.
- We perform extensive experiments on four datasets to show that our approach achieves new SOTA anomaly detection performance and outperforms current SOTA in anomaly localization performance by a large margin.

2. Related Work

Unsupervised Approaches. Unsupervised paradigm assumes that only normal data is available during training [36, 44, 53]. Auto-Encoder based methods [3, 6, 15, 65] rely on the hypothesis that the model is trained to reconstruct normal regions well but fails for abnormal regions. Although localization results based on the difference between the input and the reconstructed image are often intuitive and interpretable, their performance is limited. Generative models are introduced to obtain better reconstruction performance. However, the generation effect of VAE [11, 11, 30]

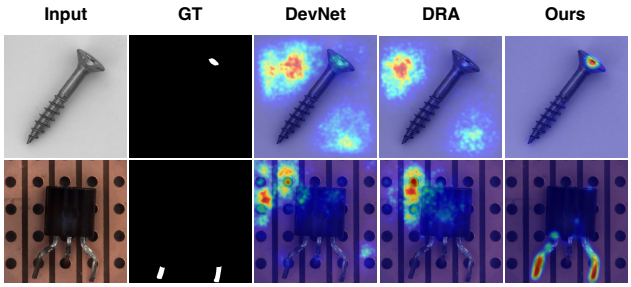

图2. 监督方法DevNet [35]和DRA [13]的不可解问题。两幅图像均被检测为异常。其他方法错误地高亮了正常区域而非缺陷区域，而PRN则准确定位了缺陷区域。

使得对所有异常进行精确定位变得具有挑战性。

在缺乏足够异常样本进行训练的情况下，无监督模型成为事实上的主导方法，它们通过学习正常样本的分布[5, 9, 10, 12, 18, 25, 41–43, 47, 67]或生成足够的合成异常[26, 28, 50, 63, 68]来规避数据不平衡问题。然而，这些方法对真实异常缺乏可解释性，导致隐含的决策可能产生大量漏报和误报。此外，无监督方法极度依赖正常样本的质量，因此鲁棒性不足，在校准不足或含噪声的数据集上表现较差[19]。如图1所示，无监督模型在异常周围预测出宽泛区域。我们将此问题归因于这些方法判别能力的不足。

最近，几种监督方法[13, 35, 46]被提出。DeepSAD[46]通过在潜在空间中扩大异常与单类中心的间隔，并利用有限可见异常来获得更紧凑的单类描述符。DRA[13]和DevNet[35]将异常检测构建为多示例学习（MIL）问题，若图像中任意图像块属于缺陷区域，则将该图像判定为异常。基于MIL的方法在细粒度图像块层面进行学习，有效减少了异常图像中正常区域的干扰。然而，如图1所示，这些方法通常难以在图像级监督下准确定位所有异常区域。特别是当异常区域仅占据图像块的极小部分时，图像级表征可能被正常区域主导而忽略微小异常，这可能导致图像级与像素级性能不一致，如表1所示。此外，如图2所示，这些方法在决策时还会遇到可解释性不足的问题。

本文提出了一种名为原型残差网络（PRN）的框架，作为解决前述异常检测与定位问题的有效方案。首先，我们提出多尺度原型来表示正常模式。与以往构建正常模式的方法不同——与从拼接特征记忆[41]或随机采样特征图[63]中构建模式不同，我们通过不同尺度中间特征图的原型来构建正常模式，从而保留空间信息并提供精确且具有代表性的正常模式。进一步，我们通过异常图像与各尺度最接近原型之间的偏差获取特征图残差，并加入多尺度融合模块以实现跨尺度信息交换。其次，由于异常区域的外观差异较大，需要从多个感受野学习图像块间的关系。因此，我们引入了多尺寸自注意力机制[33,55,58,59]，该机制在不同感受野的图像块上操作，以检测不同尺寸的图像块级不一致性。最后，与先前使用图像级监督进行训练的方法[13,35]不同，我们的模型通过像素级监督学习重建异常分割图，这更关注异常区域并保持更好的泛化能力。此外，我们提出了多种异常生成策略，有效缓解了数据不平衡的影响并丰富了异常外观。通过所提出的模块，我们的方法实现了比以往无监督和有监督方法更精确的定位，如图1和图2所示。

本文的主要贡献总结如下：

- 我们提出了一种新颖的原型残差网络，用于异常检测与定位。该网络配备多尺度原型和多尺寸自注意力机制，能够学习多尺度特征图之间以及每个尺度内多尺寸感受野之间的残差表示。
- 我们提出了多种异常生成策略，这些策略同时考虑已见和未见的外观变化，以扩大并丰富异常情况的多样性。
- 我们在四个数据集上进行了大量实验，结果表明，我们的方法在异常检测性能上达到了新的SOTA水平，并在异常定位性能上大幅超越了当前的SOTA。

2. 相关工作

无监督方法。无监督范式假设训练期间仅能获取正常数据[36, 44, 53]。基于自编码器的方法[3, 6, 15, 65]依赖以下假设：模型被训练以良好重建正常区域，但对异常区域表现不佳。虽然基于输入图像与重建图像差异的定位结果通常直观且可解释，但其性能有限。为获得更好的重建性能，研究者引入了生成模型。然而，VAE[11, 11, 30]的生成效果

or GAN [2, 22, 27, 48, 49, 66] over normal areas in the image is poor, leading to coarse reconstruction and false detection. Normalizing flows based methods [18, 42, 43, 67] learn bijective transformations between data distributions and well-defined densities, however the computational cost of these approaches is significant. Knowledge distillation-based methods [5, 12, 47, 56] transform the anomaly detection task into a feature comparison between teacher and student networks. Deep feature modeling-based methods [9, 10, 23, 25, 34, 41, 70] bulid a feature space for input images and then detect and localize anomalies by comparing the features. Self-supervised learning-based methods [38, 39, 64] designed proxy tasks such as predicting or recovering hidden regions or properties in input images [53]. One-class classification based methods [45, 51, 64] aim to map training images or patches to a small hypersphere in the feature space. These approaches address the imbalance problem by being opaque to anomalous samples, but suffer from implicit decisions that result in subpar performance on subtle and challenging anomalies.

Supervised Approaches. A recent emerging trend focuses on supervised anomaly detection by leverging seen anomalies to increase the differentiation between anomalous and normal samples. Some existing works [16, 31, 46] are learned with a minority of anomaly based on one-class classification metric. Some anomaly-focused deviation losses proposed in [35, 69] mitigate the bias derived from the seen anomalies. A multi-head model is introduced in [13] to learn disentangled anomaly representations, where each head is dedicated to capturing a specific type of anomaly. Due to the imbalanced learning problem, these methods are prone to over-fitting to seen anomalies and fail to generalize to unseen anomalies, resulting in poor anomaly detection performance. In addition, image-level representations may be dominated by normal regions while disregarding the representations of subtle anomaly regions, resulting in an inability to accurately localize anomalies that come in a variety of sizes, shapes and numbers.

3. Method

Together with the proposed anomaly generation strategies (Appendix A.1) that retain a balanced data distribution, we propose the Prototypical Residual Network (PRN) to reconstruct the segmentation map for anomaly detection and localization. Overall, we adopt a U-Net [40]-like architecture as shown in Fig. 3. The encoder is a pre-trained ResNet-18 [21], and the decoder consists of upsampling and convolution blocks. The skip-connection branches of PRN are equipped with the proposed Multi-scale Prototypes (MP, Sec. 3.1), Multi-scale Fusion blocks (MF, Sec. 3.2) and a Multi-size Self-Attention mechanism (MSA, Sec. 3.3). In the following, we will concretely describe each part.

3.1. Multi-scale Prototypes

Prototype Initialization. We define $\mathcal{X}_N$ to be the set of all normal samples during training ($\forall x \in \mathcal{X}_N : y_x = 0$). y_x denotes that if an image x is normal (0) or abnormal (1). Following [10,41], we use a network pre-trained on ImageNet to obtain feature maps of the input image at different scales. We use $\mathcal{F}_{i,j} = \mathcal{F}_j(x_i)$ ($j \in \{1, 2, 3, 4\}$) to denote the j-th block output of input x_i from a ResNet-like architecture such as ResNet-18 [21]. Assume the feature map $\mathcal{F}_{i,j} \in \mathbb{R}^{c^j \times h^j \times w^j}$ to be a tensor of depth c^j, height h^j and width w^j. Firstly, the j-th scale prototypes $\mathcal{P}_j \in \mathbb{R}^{K \times c^j \times h^j \times w^j}$ are K feature maps randomly sampled from $\mathcal{F}_j(\mathcal{X}_N)$, and are updated by k-means clustering [20]. L2 distance is used to calculate the distance between two feature maps. As the number of normal samples in different datasets varies considerably, to have a suitable amount of prototypes, we set the number of prototypes to a certain ratio of the number of normal samples. As a result, the value of K varies by datasets. The ablation on the proportion number is detailed in Sec. 4.3, and is typically 10%. Three scales of prototypes are employed ($j \in \{1, 2, 3\}$). Model parameters are frozen during clustering. After clustering, the prototypes $\mathcal{P}_j \in \mathbb{R}^{K \times c^j \times h^j \times w^j}$ at each scale remain unchanged during subsequent model training.

Residual Representation. Given the i-th input image and its corresponding feature map $\mathcal{F}_{i,j}$ at j-th block, we can find the closest prototype $\mathcal{P}_j^*$ at j-th scale by calculating the L2 distance between $\mathcal{F}_{i,j}$ and each of the prototypes $\mathcal{P}_j$. We define the anomalous residual representation of $\mathcal{F}_{i,j}$ to its closest prototype as

$$\begin{aligned} \mathcal{D}_{i,j} =& D(\mathcal{F}_{i,j} - \mathcal{P}_j^*), \\ \text{s.t. } \mathcal{P}_j^* =& \underset{\mathcal{P}_j^k \subset \mathcal{P}_j}{\arg\min} \|\mathcal{F}_{i,j} - \mathcal{P}_j^k\|_2 \end{aligned} \tag{1}$$

where $D(\cdot, \cdot)$ implements the element-wise Euclidean distance between two tensors, $\mathcal{D}_{i,j} \in \mathbb{R}^{c^j \times h^j \times w^j}$ is the residual from the nearest cluster prototype $\mathcal{P}_j^*$. Note that the input sample can match distinct prototypes at different scales, as the prototypes are learned independently at each scale.

3.2. Multi-scale Fusion

To enable information exchanging across multi-scale representations, we propose to use Multi-scale Fusion blocks (MF) inspired by [17, 57]. As shown in Fig. 4, the fused output feature map is the sum of the transformed representations of three input feature maps. The feature map $\mathcal{F}_{i,j}^*$ is fused with others as follows:

$$\mathcal{F}_{i,j}^* = f_{1j}(\mathcal{F}_{i,1}) + f_{2j}(\mathcal{F}_{i,2}) + f_{3j}(\mathcal{F}_{i,3}) \tag{2}$$

The choice of the transform function $f_{rj}(\cdot)$ depends on the input feature map index r and the output feature map index

或GAN [2, 22, 27, 48, 49, 66] 在图像正常区域上的表现较差，导致重建粗糙和误检测。基于归一化流的方法 [18, 42, 43, 67] 学习数据分布与明确定义密度之间的双射变换，但这些方法的计算成本较高。基于知识蒸馏的方法 [5, 12, 47, 56] 将异常检测任务转化为教师网络与学生网络之间的特征比较。基于深度特征建模的方法 [9, 10, 23, 25, 34, 41, 70] 为输入图像构建特征空间，然后通过比较特征来检测和定位异常。基于自监督学习的方法 [38,39,64] 设计了代理任务，例如预测或恢复输入图像中的隐藏区域或属性 [53]。基于单类分类的方法 [45,51,64] 旨在将训练图像或图像块映射到特征空间中的一个小型超球体内。这些方法通过对异常样本不敏感来解决不平衡问题，但存在隐含决策的缺陷，导致在细微和具有挑战性的异常上表现不佳。

监督式方法。近期的一个新兴趋势侧重于通过利用已知异常来增强异常与正常样本之间的区分度，从而实现监督式异常检测。现有的一些研究[16, 31, 46]基于单类分类指标，利用少量异常样本进行学习。文献[35, 69]提出的异常聚焦偏差损失函数缓解了已知异常带来的偏差。文献[13]引入了一种多头模型来学习解耦的异常表示，其中每个头专门用于捕捉特定类型的异常。由于存在不平衡学习问题，这些方法容易对已知异常过拟合，难以泛化到未知异常，导致异常检测性能不佳。此外，图像级表示可能被正常区域主导，而忽略细微异常区域的表示，导致无法准确定位各种尺寸、形状和数量的异常。

3. 方法

结合所提出的保持数据分布平衡的异常生成策略（附录A.1），我们提出了原型残差网络（PRN）来重建分割图以进行异常检测与定位。整体上，我们采用如图3所示的类U-Net[40]架构：编码器采用预训练的ResNet-18[21]，解码器由上采样和卷积块组成。PRN的跳跃连接分支配备了提出的多尺度原型模块（MP，第3.1节）、多尺度融合块（MF，第3.2节）以及多尺寸自注意力机制（MSA，第3.3节）。下文将具体阐述各部分设计。

3.1. 多尺度原型

原型初始化。我们定义$\mathcal{X}_N$为训练期间所有正常样本的集合（$\forall x \in \mathcal{X}_N$：$y_x = 0$）。$y_x$表示图像$x$是正常（0）还是异常（1）。遵循[10,41]的方法，我们使用在ImageNet上预训练的网络来获取输入图像在不同尺度下的特征图。我们使用$\mathcal{F}_{i,j} = \mathcal{F}_j(x_i)\,(j \in \{1,2,3,4\})$表示来自类ResNet架构（如ResNet-18 [21]）的输入x_i的第j个块输出。假设特征图$\mathcal{F}_{i,j} \in \mathbb{R}^{c^j \times h^j \times w^j}$是一个深度为$c^j$、高度为$h^j$、宽度为$w^j$的张量。首先，第$j$个尺度的原型$\mathcal{P}_j \in \mathbb{R}^{K \times c^j \times h^j \times w^j}$是从$\mathcal{F}_j(\mathcal{X}_N)$中随机采样的$K$个特征图，并通过k-means聚类[20]进行更新。使用L2距离计算两个特征图之间的距离。由于不同数据集中正常样本的数量差异很大，为了获得合适数量的原型，我们将原型数量设置为正常样本数量的特定比例。因此，K的值因数据集而异。比例数的消融实验详见第4.3节，通常设置为10%。我们采用三个尺度的原型（$j \in \{1,2,3\}$）。聚类过程中模型参数被冻结。聚类后，每个尺度的原型$\mathcal{P}_j \in \mathbb{R}^{K \times c^j \times h^j \times w^j}$在后续模型训练中保持不变。

残差表示。给定第i个输入图像及其在第j个块对应的特征图$\mathcal{F}_{i,j}$，我们可以通过计算$\mathcal{F}_{i,j}$与每个原型$\mathcal{P}_j$之间的L2距离，找到第j个尺度上最接近的原型$\mathcal{P}_j^*$。我们将$\mathcal{F}_{i,j}$到其最近原型的异常残差表示定义为

$$\begin{aligned} \mathcal{D}_{i,j} =& D(\mathcal{F}_{i,j} - \mathcal{P}_j^*), \\ \text{s.t. } \mathcal{P}_j^* =& \underset{\mathcal{P}_j^k \subset \mathcal{P}_j}{\arg\min} \|\mathcal{F}_{i,j} - \mathcal{P}_j^k\|_2 \end{aligned} \tag{1}$$

其中$D(\cdot,\cdot)$实现了两个张量间的逐元素欧几里得距离，$\mathcal{D}_{i,j} \in \mathbb{R}^{c^j \times h^j \times w^j}$表示与最近聚类原型$\mathcal{P}_j^*$的残差。需要注意的是，由于原型在不同尺度上独立学习，输入样本可以在不同尺度上匹配不同的原型。

3.2. 多尺度融合

为了实现多尺度表征间的信息交换，我们受[17, 57]启发，提出采用多尺度融合模块（MF）。如图4所示，融合后的输出特征图是三个输入特征图经变换后的表征之和。特征图{v*}与其他特征的融合方式如下：

$$\mathcal{F}_{i,j}^* = f_{1j}(\mathcal{F}_{i,1}) + f_{2j}(\mathcal{F}_{i,2}) + f_{3j}(\mathcal{F}_{i,3}) \tag{2}$$

变换函数$f_{rj}(\cdot)$的选择取决于输入特征图索引r和输出特征图索引

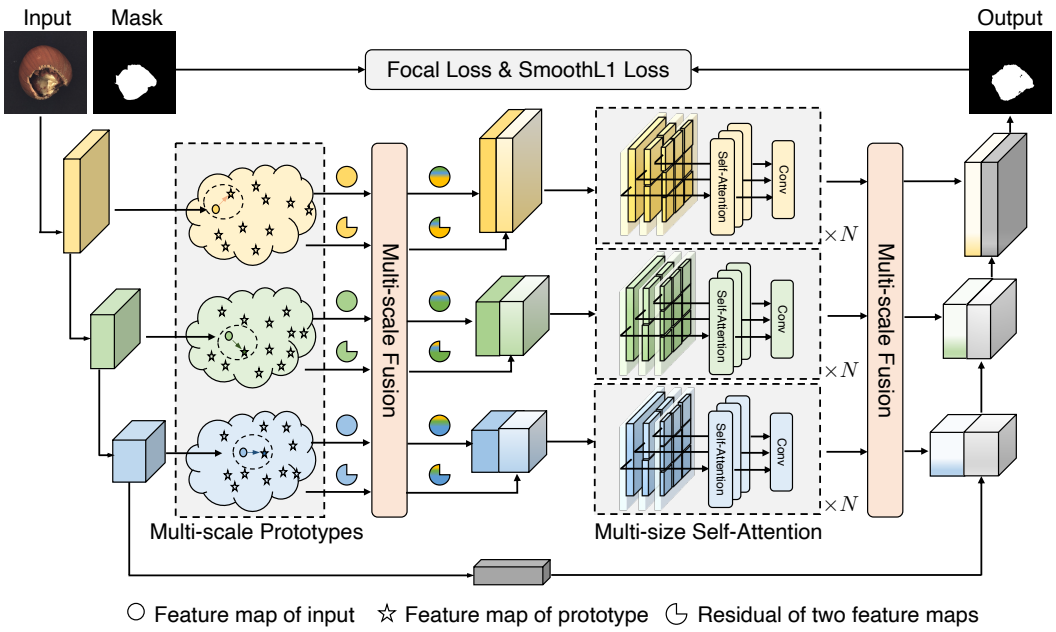

Figure 3. An overview of the proposed Prototypical Residual Network. Anomalous feature residuals of inputs are obtained via Multi-scale Prototypes for each scale feature map from the nearest cluster prototype. Feature maps and residuals at different scales are separately fused by Multi-scale Fusion blocks. Multi-size Self-Attention learns feature residuals on patches of different sizes at each scale, which are further enhanced by another Multi-scale Fusion block. Please see text for details.

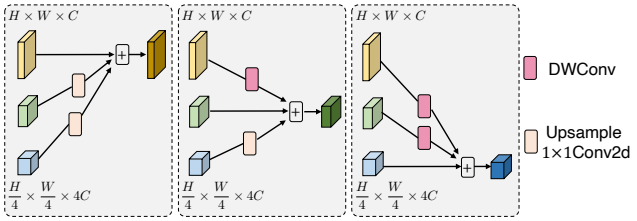

Figure 4. Three feature maps of different scales are fused by a multi-scale fusion block.

j ($r, j \in \{1, 2, 3\}$). If $r = j$, $f_{rj}(\mathcal{F}_{i,r}) = \mathcal{F}_{i,r}$. If $r < j$, $f_{rj}(\mathcal{F}_{i,r})$ down-samples the input feature map $\mathcal{F}_{i,r}$ through depth-wise separable convolutions with a stride of 2^{j-r}, a kernel size of $2^{j-r}+1$ and a padding of 2^{j-r-1}. If $r > j$, $f_{rj}(\mathcal{F}_{i,r})$ up-samples the input feature map $\mathcal{F}_{i,r}$ through a bilinear up-sampling followed by a 1×1 convolution. The anomalous residual representation $\mathcal{D}_{i,j}$ also follows the fusion paradigm in Fig. 4 and is concated with $\mathcal{F}^*_{i,j}$ along the depth dimension to obtain $\mathcal{C}^*_{i,j} \in \mathbb{R}^{2c^j \times h^j \times w^j}$.

3.3. Multi-size Self-Attention

As the anomalous regions vary in magnitude, to further detect local inconsistencies in the concatenated feature maps $\mathcal{C}^*_{i,j}$, we introduce a Multi-size Self-attention (MSA) mechanism. MSA splits $\mathcal{C}^*_{i,j}$ into patches of different sizes $p_s \in \{h^j, h^j/2, h^j/4, h^j/8\}$ and computes patch-wise self-attention [54,55,58,60,62] in different heads. Different heads correspond to different patch sizes, as shown in Fig. 3. To be specific, we first extract patches of shape $2c^j \times p_s \times p_s$ from $\mathcal{C}^*_{i,j}$, and flatten them into 1-dimension vectors for the s-th head. And then we use fully-connected layers to embed the flattened vectors into query embeddings $\mathcal{Q}^s_{i,j} \in \mathbb{R}^{\mathcal{N} \times c^s}$, where $\mathcal{N} = \left(h^j/p_s\right) \times \left(w^j/p_s\right)$ and $c^s = 2c^j \times p_s \times p_s$. We obtain key embeddings $\mathcal{K}^s_{i,j}$ and value embeddings $\mathcal{V}^s_{i,j}$ with the similar operations. The attention matrix is calculated by the following process:

$$\mathcal{A}^s_{i,j} = \text{softmax}\left(\frac{\mathcal{Q}^s_{i,j}\left(\mathcal{K}^s_{i,j}\right)^T}{c^s}\right)\mathcal{V}^s_{i,j} \qquad (3)$$

After that, $\mathcal{A}^s_{i,j}$ is reshaped to the original spatial resolution. Similar operations are implemented to obtain features from heads of different patch sizes. Finally, these features are concatenated and passed through a 2D residual block to obtain the output $\mathcal{T}_{i,j} \in \mathbb{R}^{2c^j \times h^j \times w^j}$. We stack the MSA for N times (N = 3 in this paper). To further fuse multiple scales of $\mathcal{T}_{i,j}$, we use another MF block to obtain $\mathcal{T}^*_{i,j}$, which is the output of the skip-connection as shown in Fig. 3.

3.4. Anomaly Generation Strategies

To alleviate the data imbalance problem, we propose two kinds of online anomaly generation strategies that can generate various types of anomalies. One strategy is to create

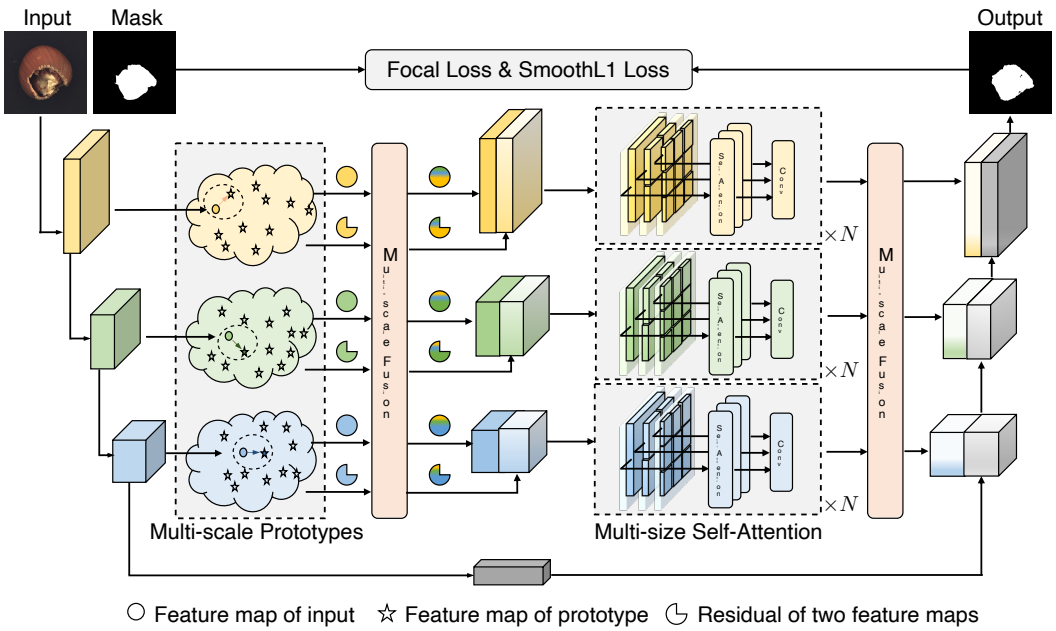

图3. 所提出的原型残差网络概览。通过多尺度原型从最近聚类原型获取各尺度特征图的输入异常特征残差。不同尺度的特征图与残差分别通过多尺度融合块进行融合。多尺寸自注意力机制在各尺度上学习不同尺寸图像块的特征残差，并通过另一多尺度融合块进一步增强。详见正文说明。

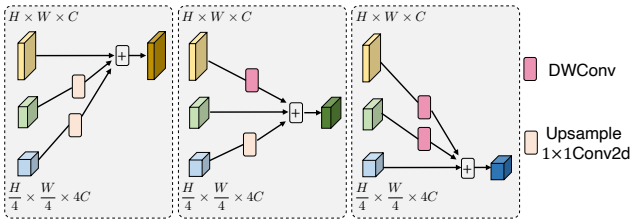

图4. 通过多尺度融合模块融合了三种不同尺度的特征图。

j ($r, j \in \{1, 2, 3\}$)。若 $r = j$，则 $f_{rj}(\mathcal{F}_{i,r}) = \mathcal{F}_{i,r}$。若 $r < j$，则 $f_{rj}(\mathcal{F}_{i,r})$ 通过步长为 2^{j-r}、卷积核大小为 $2^{j-r}+1$、填充为 2^{j-r-1} 的深度可分离卷积对输入特征图 $\mathcal{F}_{i,r}$ 进行下采样。若 $r > j$，则 $f_{rj}(\mathcal{F}_{i,r})$ 通过双线性上采样接一个 1×1 卷积对输入特征图 $\mathcal{F}_{i,r}$ 进行上采样。异常残差表示 $\mathcal{D}_{i,j}$ 同样遵循图 4 中的融合范式，并与 $\mathcal{F}^*_{i,j}$ 沿深度维度拼接以得到 $\mathcal{C}^*_{i,j} \in \mathbb{R}^{2c^j \times h^j \times w^j}$。

3.3. 多尺度自注意力

由于异常区域在幅度上存在差异，为了进一步检测拼接特征图$\mathcal{C}^*_{i,j}$中的局部不一致性，我们引入了多尺度自注意力（MSA）机制。MSA将$\mathcal{C}^*_{i,j}$分割成不同尺寸的块$p_s \in \{h^j, h^j/2, h^j/4, h^j/8\}$，并计算块间的自注意力。

注意力[54,55,58,60,62]在不同头部中。不同头部对应不同的补丁尺寸，如图3所示。具体而言，我们首先从$\mathcal{C}^*_{i,j}$中提取形状为$2c^j \times p_s \times p_s$的补丁，并将其展平为1维向量用于第$s$个头部。随后，我们使用全连接层将展平后的向量嵌入为查询嵌入$\mathcal{Q}^s_{i,j} \in \mathbb{R}^{\mathcal{N} \times c^s}$，其中$\mathcal{N} = (h^j/p_s) \times (w^j/p_s)$和$c^s = 2c^j \times p_s \times p_s$。我们通过类似操作获得键嵌入$\mathcal{K}^s_{i,j}$和值嵌入$\mathcal{V}^s_{i,j}$。注意力矩阵通过以下过程计算：

$$\mathcal{A}^s_{i,j} = \text{softmax}\left(\frac{\mathcal{Q}^s_{i,j}\left(\mathcal{K}^s_{i,j}\right)^T}{c^s}\right)\mathcal{V}^s_{i,j} \tag{3}$$

随后，$\mathcal{A}^s_{i,j}$被重塑至原始空间分辨率。类似的操作被应用于从不同补丁大小的头部获取特征。最终，这些特征被拼接并通过一个2D残差块以获得输出$\mathcal{T}_{i,j} \in \mathbb{R}^{2c^j \times h^j \times w^j}$。我们将MSA堆叠N次（本文中N=3）。为进一步融合$\mathcal{T}_{i,j}$的多尺度信息，我们使用另一个MF块来获得$\mathcal{T}^*_{i,j}$，即如图3所示的跳跃连接的输出。

3.4. 异常生成策略

为了缓解数据不平衡问题，我们提出了两种在线异常生成策略，能够生成多种类型的异常。一种策略是创建

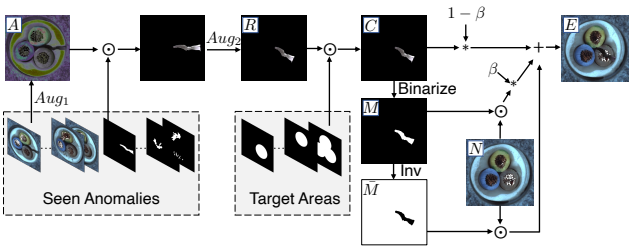

Figure 5. Generating extended anomalies. The anomalous region is augmented by random augmentation and placed on a target area of the normal sample to generate various anomalies online.

in-distribution anomalies by placing augmented anomalous regions from seen anomalies on normal samples, and these generated anomalies are named extended anomalies (EA). EA enlarge the amount of anomalies and mitigate the data imbalance problem. Another strategy is to create out-of-distribution anomalies [68] using normal samples without knowledge of the seen anomalies. These generated anomalies are named simulated anomalies (SA) , which supplement potential unseen anomalies.

Extended Anomalies. Instead of simply augmenting the entire image from the seen anomalies, we augment the specific anomalous regions of the seen anomalies and place them at any possible position within the normal sample. First, augmentations (Fig. 5, Aug_1) are applied to a randomly selected anomaly from the seen anomalies in order to generate color varieties (Fig. 5, A). Aug_1 takes two random operations from { equalize, solarize, posterize, sharpness, autocontrast, invert, gamma-contrast }. After that, we argument the selected anomaly with random spatial transformations as { rotate, shear, shift } to obtain position and shape varieties (Fig. 5, R). Since the extended anomalies should be as realistic and reasonable as possible, we propose a soft position constrain to place R in the foreground. More specifically, Target Areas (TA) is used to refer to areas where anomalies can be placed. We crop R, using a randomly sampled target area, to obtain clipped anomaly region (Fig. 5, C). If R has no overlap with the target region, we perform Aug_2 again until R has overlap with the target region. We binarized C to obtain the ground truth mask (Fig. 5, M). The proposed extended anomalies (Fig. 5, E) is therefore defined as:

$$E = \bar{M} \odot N + (1 - \beta)C + \beta\,(M \odot N) \qquad (4)$$

where $\bar{M}$ is the inverse of M, $\odot$ is the element-wise multiplication operation, β is the opacity parameter [68] for better combination of abnormal and normal parts. For object datasets and texture datasets, the target areas are part of the foreground of the object and part of the whole image, respectively. The shapes of the target are the set of geometries: {circle, rectangular, polygonal}.

Simulated Anomalies. Similar to DRAEM [68], we multiply Perlin [37] noise with random textures from the DTD [8] dataset and apply these augmented textures to normal images. As these anomalies significantly differs from the seen anomalies, we refer to these out-of-distribution anomalies as heterologous anomalies (HEA). To further expand the diversity of simulated anomalies, we introduce homologous anomalies (HOA), in which anomalies multiplied by the Perlin noise are augmented normal images. Note that the TA mentioned above is also applied to the generation of simulated anomalies. More details about HEA and HOA are presented in the supplementary materials.

3.5. Training and Inference

The decoder of PRN outputs an anomaly score map $\mathcal{M}_o$, which is of the same shape as the ground truth mask $\mathcal{M}$. Inspired by [68] and [63] , a focal loss [29] and a smooth L1 loss [14] are applied to increase the robustness toward accurate segmentation of hard examples and reduce the oversensitivity to outliers, respectively. Thus, the total loss $\mathcal{L}_{total}$ used for training PRN is defined as

$$\mathcal{L}_{total} = \text{Smooth}_{\mathcal{L}1}\left(\mathcal{M}_o, \mathcal{M}\right) + \lambda \mathcal{L}_{focal}\left(\mathcal{M}_o, \mathcal{M}\right) \qquad (5)$$

When the predicted $\mathcal{M}_o$ is accurate and sufficiently close to $\mathcal{M}$, $\mathcal{M}_o$ can be interpreted not only as the pixel-level anomaly localization result, but also as an image-level anomaly estimation for anomaly detection. Specifically, we take the average of the top-K anomalous pixels as the image-level anomaly score for anomaly detection. In a preliminary study, we trained a classification network based on $\mathcal{M}_o$ for image-level anomaly detection, but did not observe an improvement over top-K estimation.

4. Experiments

4.1. Experimental Details

Datasets. To validate the effectiveness and generalizability of our approach, we perform experiments on various datasets, *i.e.*, MVTec Anomaly Detection (MVTec AD [4]), DAGM [61], BeanTech anomaly detection dataset (BTAD [34]), and KolektorSDD2 [7]. There are 10 object sub-datasets and 5 texture sub-datasets in MVTec AD. Each sub-dataset presents a diverse set of anomalies, which enables a general evaluation of surface anomaly detection methods. DAGM contains 10 textured objects with small abnormal regions that are visually very similar to the background. BTAD includes three categories of real-world industrial products showcasing different body and surface defects. KolektorSDD2 is a dataset of surface defects that vary in shape, size, and color, from small scratches and spots to large surface defects. We adopt the general supervised setting [13,35], where the training set of each sub-dataset contains only 10 abnormal samples. More details will be provided in the supplementary materials.

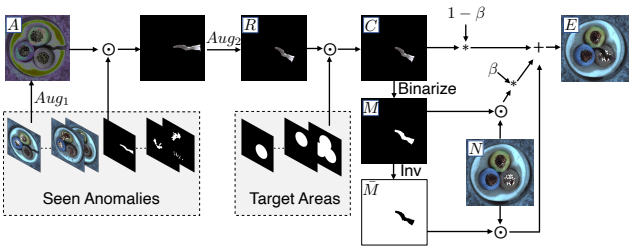

图5. 生成扩展异常。通过随机增强对异常区域进行扩充，并将其置于正常样本的目标区域上，以在线生成多种异常。

通过将已见过异常中的增强异常区域置于正常样本上，我们创建了分布内异常，这些生成的异常被称为扩展异常（EA）。EA增加了异常的数量，缓解了数据不平衡问题。另一种策略是利用正常样本，在未见过异常知识的情况下创建分布外异常[68]。这些生成的异常被称为模拟异常（SA），它们补充了潜在的未见异常。

扩展异常。我们并非简单地从已见异常中增强整个图像，而是增强已见异常的特定异常区域，并将它们放置在正常样本内的任何可能位置。首先，对从已见异常中随机选择的异常应用增强（图5，Aug_1），以生成颜色变体（图5，A）。Aug_1从{均衡化、曝光、色调分离、锐度、自动对比度、反转、伽马对比度}中随机选择两种操作。之后，我们通过随机空间变换（如{旋转、剪切、平移}）来增强所选异常，以获得位置和形状变体（图5，R）。由于扩展异常应尽可能真实合理，我们提出了一种软位置约束，将R放置在前景中。更具体地说，目标区域（TA）用于指代可以放置异常的区域。我们使用随机采样的目标区域裁剪R，以获得裁剪后的异常区域（图5，C）。如果R与目标区域没有重叠，我们再次执行Aug_2，直到R与目标区域有重叠。我们将C二值化以获得真实掩码（图5，M）。因此，提出的扩展异常（图5，E）定义为：

$$E = \bar{M} \odot N + (1-\beta)C + \beta\left(M \odot N\right) \tag{4}$$

其中$\bar{M}$是M的逆运算，$\odot$表示逐元素乘法操作，β为透明度参数[68]，用于更好地融合异常与正常部分。对于物体数据集和纹理数据集，目标区域分别是物体前景的一部分和整张图像的一部分。目标形状采用几何集合：{圆形、矩形、多边形}。

模拟异常。类似于DRAEM [68]，我们将Perlin [37]噪声与来自DTD [8]数据集的随机纹理相乘，并将这些增强后的纹理应用于正常图像。由于这些异常与已见异常存在显著差异，我们将这些分布外异常称为异源异常（HEA）。为进一步扩展模拟异常的多样性，我们引入了同源异常（HOA），其中通过Perlin噪声相乘的异常是经过增强的正常图像。请注意，上述TA方法也应用于模拟异常的生成。关于HEA和HOA的更多细节详见补充材料。

3.5. 训练与推理

PRN解码器输出一个与真实掩码$\mathcal{M}$形状相同的异常分数图$\mathcal{M}_o$。受[68]和[63]的启发，我们采用焦点损失[29]和平滑L1损失[14]，分别用于增强对困难样本精确分割的鲁棒性以及降低对异常值的过度敏感性。因此，用于训练PRN的总损失$\mathcal{L}_{total}$定义为

$$\mathcal{L}_{total} = \text{Smooth}_{\mathcal{L}1}\left(\mathcal{M}_o, \mathcal{M}\right) + \lambda\mathcal{L}_{focal}\left(\mathcal{M}_o, \mathcal{M}\right) \tag{5}$$

当预测的$\mathcal{M}_o$准确且足够接近$\mathcal{M}$时，$\mathcal{M}_o$不仅可以解释为像素级异常定位结果，也可作为图像级异常检测的异常估计值。具体而言，我们取前K个最异常像素的平均值作为图像级异常检测的评分。在初步研究中，我们基于$\mathcal{M}_o$训练了用于图像级异常检测的分类网络，但未观察到其相较于前K估计方法的性能提升。

4. 实验

4.1. 实验细节

数据集。为了验证我们方法的有效性和泛化能力，我们在多个数据集上进行了实验，*i.e*，包括MVTec异常检测数据集（MVTec AD [4]）、DAGM [61]、BeanTech异常检测数据集（BTAD [34]）以及KolektorSDD2 [7]。MVTec AD中包含10个物体子数据集和5个纹理子数据集。每个子数据集呈现了多样化的异常类型，从而能够全面评估表面异常检测方法。DAGM包含10个纹理物体，其异常区域较小且在视觉上与背景极为相似。BTAD涵盖了三大类真实工业产品，展示了不同的本体和表面缺陷。KolektorSDD2是一个表面缺陷数据集，缺陷在形状、大小和颜色上各不相同，从小划痕和斑点到大面积表面缺陷均有涵盖。我们采用通用的监督设置[13,35]，其中每个子数据集的训练集仅包含10个异常样本。更多细节将在补充材料中提供。

Category	Unsupervised							Supervised		
	KDAD [47]	CFLOW [18]	DRAEM [68]	SSPCAB [39]	CFA [25]	RD4AD [12]	PatchCore [41]	DevNet [35]	DRA [13]	Ours
Carpet	80.3/95.5	97.6/**99.2**	96.9/97.5	93.1/92.6	99.3/98.6	98.7/98.9	99.1/99.0	82.5/97.2	92.5/98.2	**99.7**/99.0
Grid	75.3/89.4	98.1/98.9	99.9/**99.7**	99.7/99.5	98.6/97.6	**100**/98.3	97.3/98.7	90.6/87.9	98.6/86.0	99.4/98.4
Leather	92.3/98.1	99.9/99.7	**100**/99.0	98.7/96.3	**100**/99.1	**100**/99.4	**100**/99.3	92.2/94.2	98.9/93.8	**100/99.7**
Tile	91.5/80.2	97.1/96.2	**100**/99.2	**100**/99.4	99.2/95.1	99.7/95.7	99.3/95.8	99.9/92.7	**100**/92.3	**100/99.6**
Wood	94.5/85.3	98.7/86.0	99.5/95.5	98.4/96.5	**100**/94.7	99.5/95.8	99.6/95.1	97.9/86.4	99.1/82.9	**100/97.8**
Bottle	99.2/95.7	99.9/97.2	98.0/99.1	95.6/99.2	**100**/98.6	**100**/98.8	**100**/98.6	99.7/93.9	**100**/91.3	**100/99.4**
Cable	90.3/80.2	97.6/97.8	90.9/95.2	92.7/95.1	**99.9/98.8**	96.1/97.2	**99.9**/98.5	98.7/88.8	94.2/86.6	98.9/**98.8**
Capsule	81.4/95.2	97.0/**99.1**	91.3/88.1	96.9/90.2	97.4/98.4	96.1/98.7	98.0/99.0	71.9/91.8	95.1/89.3	**98.0**/98.5
Hazelnut	98.8/95.0	**100**/98.8	**100**/99.7	**100**/99.7	**100**/98.6	**100**/99.0	**100**/98.7	99.7/91.1	**100**/89.6	**100/99.7**
Metal Nut	77.1/83.3	98.5/98.6	**100**/99.6	**100**/99.4	**100**/98.7	**100**/97.3	99.9/98.3	98.8/77.8	99.1/79.5	**100/99.7**
Pill	84.4/89.9	96.2/98.9	97.1/97.3	97.4/97.2	97.7/98.0	98.7/98.1	97.5/97.6	87.1/82.6	88.3/84.5	**99.3/99.5**
Screw	82.4/95.8	93.1/98.9	98.7/99.3	97.8/99.0	95.1/98.3	97.8/**99.7**	98.2/99.5	97.2/60.3	**99.5**/54.0	95.9/97.5
Toothbrush	97.1/95.5	98.8/99.0	**100**/97.3	97.9/97.3	**100**/98.8	**100**/99.1	**100**/98.6	79.2/84.6	87.5/75.5	**100/99.6**
Transistor	84.9/75.9	92.9/98.2	91.7/85.2	88.0/84.8	**100**/98.1	95.5/92.3	99.9/96.5	89.1/56.0	88.3/79.1	99.7/**98.4**
Zipper	93.7/95.3	97.1/**99.1**	**100/99.1**	**100**/98.4	99.5/98.6	97.9/98.3	99.5/98.9	99.1/93.7	99.7/96.9	99.7/98.8
Average	88.2/90.0	97.5/97.7	97.6/96.7	97.1/96.3	99.1/98.0	98.7/97.8	99.2/98.1	92.2/85.3	96.1/85.3	**99.4/99.0**

Table 1. Anomaly Detection and Localization on MVTec [4]. Best results on Image AUROC or Pixel AUROC are highlighted in bold.

Evaluation Metrics. Following previous work, we evaluate the results via the area under the receiver operating characteristic curve at the image level (Image-AUROC) and pixel level (Pixel-AUROC). However, anomalous regions typically only occupy a tiny fraction of the entire image. Thus, Pixel-AUROC can not accurately reflect the localization accuracy due to the fact that the false positive rate is dominated by the extremely high number of non-anomalous pixels and remains low despite false positive detection [53]. To comprehensively measure localization performance, we introduce Per Region Overlap (PRO) [5] score and pixel-level Average Precision (AP) [68]. The PRO score treats anomaly regions of varying sizes equally [12, 41], while AP is more appropriate for highly imbalanced classes, especially for industrial anomaly localization, where accuracy is critical [68].

Implementation Details. All images in four datasets are resized to 256×256. We use layer1, layer2 and layer3 of ResNet-18 [21] pre-trained on ImageNet to obtain feature maps with $64 \times 64 \times 64$, $128 \times 32 \times 32$ and $256 \times 16 \times 16$ scales respectively and frozen these blocks during training. The number of prototypes depends on the dataset and accounts for 10% of the total number of normal samples in the dataset. The maximum number of iterations of k-means for each scale is set to 300. We use Adam optimizer [24] for the parameter optimization, with an initial learning rate 10^{-4} and a weight decay of 10^{-2}. α and γ in focal loss is set to 0.5 and 4 respectively. λ in the total loss is set to 5. We train for 700 epochs with a batch size of 64 consisting of 32 normal samples, 16 extended anomalies and 16 simulated anomalies to ensure the diversity of anomalies. We take the average of the top 100 anomalous pixels as the image-level anomaly score. We compare PRN to seven unsupervised SOTA methods and two supervised SOTA methods. The results we report are based on the implementation provided by these methods. The backbone of PatchCore [41], RD4AD [12], CFLOW [18] and CFA [25] is WideResNet50. SSPCAB [1,39] replaces the penultimate convolutional layer of reconstructive encoder in DRAEM [68]. DevNet [35] proposes that the anomaly score given by the network can be further back-propagated to the original image pixels to infer which pixels are the major contributors to the anomaly for anomaly localiaztion. We use this approach to obtain the anomaly localization performance of DRA [13].

4.2. Anomaly Detection and Localization on MVTec

Anomaly detection and localization results on MVTec are shown in Table 1. Our method achieves the highest image AUROC (detection) and the highest pixel AUROC (localization) in 10 out of 15 classes. The average image AUROC results show that our method outperforms unsupervised SOTA by 0.2% and supervised SOTA by 3.3%. Meanwhile, for pixel AUROC, our method outperforms unsupervised SOTA by 0.9% and supervised SOTA by 13.7%.

For a comprehensive presentation of the capabilities on anomaly localization, two additional metric results, PRO and AP, are shown in Table 2. PRN outperforms the previous unsupervised SOTA by 2.2% and the previous supervised SOTA by 22.8% on the PRO metric. This confirms that PRN is more effective at simultaneously localizing anomalous regions of varying sizes. The more challenging AP metric further demonstrates the excellent anomaly localization capability of PRN. A better AP score is achieved in 12 out of 15 classes and is comparable to SOTA in other classes. In terms of overall AP, our approach even outperforms unsupervised SOTA by 10.5% and supervised SOTA by 52.6%. This significant improvement over AP goes a long way to demonstrate that PRN is more discriminative between normal and abnormal pixels. We further compare the pre-trained based approaches in terms of inference time per image (second) and performance, as shown in Table 3. All experiments were conducted on NVIDIA GeForce RTX 3090, using a uniform

Category	Unsupervised							Supervised		
	KDAD [47]	CFLOW [18]	DRAEM [68]	SSPCAB [39]	CFA [25]	RD4AD [12]	PatchCore [41]	DevNet [35]	DRA [13]	Ours
Carpet	80.3/95.5	97.6/**99.2**	96.9/97.5	93.1/92.6	99.3/98.6	98.7/98.9	99.1/99.0	82.5/97.2	92.5/98.2	**99.7**/99.0
Grid	75.3/89.4	98.1/98.9	99.9/**99.7**	99.7/99.5	98.6/97.6	**100**/98.3	97.3/98.7	90.6/87.9	98.6/86.0	99.4/98.4
Leather	92.3/98.1	99.9/99.7	**100**/99.0	98.7/96.3	**100**/99.1	**100**/99.4	**100**/99.3	92.2/94.2	98.9/93.8	**100/99.7**
Tile	91.5/80.2	97.1/96.2	**100**/99.2	**100**/99.4	99.2/95.1	99.7/95.7	99.3/95.8	99.9/92.7	**100**/92.3	**100/99.6**
Wood	94.5/85.3	98.7/86.0	99.5/95.5	98.4/96.5	**100**/94.7	99.5/95.8	99.6/95.1	97.9/86.4	99.1/82.9	**100/97.8**
Bottle	99.2/95.7	99.9/97.2	98.0/99.1	95.6/99.2	**100**/98.6	**100**/98.8	**100**/98.6	99.7/93.9	**100**/91.3	**100/99.4**
Cable	90.3/80.2	97.6/97.8	90.9/95.2	92.7/95.1	**99.9/98.8**	96.1/97.2	**99.9**/98.5	98.7/88.8	94.2/86.6	98.9/**98.8**
Capsule	81.4/95.2	97.0/**99.1**	91.3/88.1	96.9/90.2	97.4/98.4	96.1/98.7	98.0/99.0	71.9/91.8	95.1/89.3	**98.0**/98.5
Hazelnut	98.8/95.0	**100**/98.8	**100**/99.7	**100**/99.7	**100**/98.6	**100**/99.0	**100**/98.7	99.7/91.1	**100**/89.6	**100/99.7**
Metal Nut	77.1/83.3	98.5/98.6	**100**/99.6	**100**/99.4	**100**/98.7	**100**/97.3	99.9/98.3	98.8/77.8	99.1/79.5	**100/99.7**
Pill	84.4/89.9	96.2/98.9	97.1/97.3	97.4/97.2	97.7/98.0	98.7/98.1	97.5/97.6	87.1/82.6	88.3/84.5	**99.3/99.5**
Screw	82.4/95.8	93.1/98.9	98.7/99.3	97.8/99.0	95.1/98.3	97.8/**99.7**	98.2/99.5	97.2/60.3	**99.5**/54.0	95.9/97.5
Toothbrush	97.1/95.5	98.8/99.0	**100**/97.3	97.9/97.3	**100**/98.8	**100**/99.1	**100**/98.6	79.2/84.6	87.5/75.5	**100/99.6**
Transistor	84.9/75.9	92.9/98.2	91.7/85.2	88.0/84.8	**100**/98.1	95.5/92.3	99.9/96.5	89.1/56.0	88.3/79.1	99.7/**98.4**
Zipper	93.7/95.3	97.1/**99.1**	**100/99.1**	**100**/98.4	99.5/98.6	97.9/98.3	99.5/98.9	99.1/93.7	99.7/96.9	99.7/98.8
Average	88.2/90.0	97.5/97.7	97.6/96.7	97.1/96.3	99.1/98.0	98.7/97.8	99.2/98.1	92.2/85.3	96.1/85.3	**99.4/99.0**

表1. MVTec [4]上的异常检测与定位。图像AUROC或像素AUROC的最佳结果以**粗体**标出。

评估指标。遵循先前工作，我们通过图像级别（Image-AUROC）和像素级别（Pixel-AUROC）的接收者操作特征曲线下面积来评估结果。然而，异常区域通常仅占整个图像的极小部分。因此，由于误报率受大量非异常像素主导，即使存在误报检测，其值仍保持较低水平，Pixel-AUROC无法准确反映定位精度[53]。为全面衡量定位性能，我们引入区域重叠度（PRO）分数[5]和像素级平均精度（AP）[68]。PRO分数对大小不同的异常区域给予同等权重[12, 41]，而AP更适用于高度不平衡的类别，尤其在精度至关重要的工业异常定位任务中[68]。

实现细节。四个数据集中的所有图像均被调整为256×256的尺寸。我们使用在ImageNet上预训练的ResNet-18[21]的layer1、layer2和layer3层，分别获取尺度为64×64×64、128×32×32和256×16×16的特征图，并在训练过程中冻结这些模块。原型数量取决于数据集，占数据集中正常样本总数的10%。每个尺度的k-means最大迭代次数设为300。我们使用Adam优化器[24]进行参数优化，初始学习率为10^{-4}，权重衰减为10^{-5}。焦点损失中的α和γ分别设为0.5和4。总损失中的λ设为5。我们训练700个周期，批大小为64，其中包含32个正常样本、16个扩展异常样本和16个模拟异常样本，以确保异常样本的多样性。我们取前100个异常像素的平均值作为图像级异常分数。我们将PRN与七种无监督SOTA方法和两种有监督SOTA方法进行比较。我们报告的结果基于这些方法提供的实现。PatchCore[41]的骨干网络，RD4AD [12]、CFLOW [18]和CFA [25]均采用WideResNet50架构。SSPCAB [1,39]替换了DRAEM [68]重建编码器中倒数第二层卷积层。DevNet [35]提出，网络给出的异常分数可进一步反向传播至原始图像像素，以推断哪些像素是导致异常定位的主要贡献因素。我们采用该方法获取DRA [13]的异常定位性能。

4.2. MVTec数据集上的异常检测与定位

MVTec数据集上的异常检测与定位结果如表1所示。在15个类别中，我们的方法在10个类别上取得了最高的图像AUROC（检测）和最高的像素AUROC（定位）。平均图像AUROC结果表明，我们的方法比无监督SOTA方法高出0.2%，比有监督SOTA方法高出3.3%。同时，在像素AUROC方面，我们的方法比无监督SOTA方法高出0.9%，比有监督SOTA方法高出13.7%。

关于异常定位能力的全面展示，表2中列出了两项额外指标结果——PRO与AP。在PRO指标上，PRN较先前无监督SOTA方法提升2.2%，较有监督SOTA方法提升22.8%。这证实了PRN能更有效地同时定位不同尺寸的异常区域。更具挑战性的AP指标进一步彰显了PRN卓越的异常定位能力：在15个类别中有12类获得更高AP分数，其余类别表现与SOTA方法相当。就整体AP而言，我们的方法甚至超越无监督SOTA方法10.5%，超越有监督SOTA方法52.6%。AP指标的大幅提升充分证明PRN在正常与异常像素间具有更强的判别能力。我们进一步在单图推理时间（秒）与性能方面对基于预训练的方法进行比较，如表3所示。所有实验均在NVIDIA GeForce RTX 3090上采用统一配置进行。

Category	Unsupervised							Supervised		
	KDAD [47]	CFLOW [18]	DRAEM [68]	SSPCAB [39]	CFA [25]	RD4AD [12]	PatchCore [41]	DevNet [35]	DRA [13]	Ours
Carpet	92.5/45.6	**97.6**/68.3	92.9/65.1	86.4/48.6	93.6/57.2	95.4/56.5	95.5/62.2	85.8/45.7	92.2/52.3	97.0/**82.0**
Grid	72.9/7.3	96.0/41.2	**98.3/62.8**	98.0/57.9	92.9/25.8	94.2/15.8	94.0/24.5	79.8/25.5	71.5/26.8	95.9/45.7
Leather	97.5/26.8	99.2/64.5	97.4/**72.9**	94.0/60.7	95.4/48.5	98.2/47.6	96.9/45.3	88.5/8.1	84.0/5.6	**99.2**/69.7
Tile	74.3/27.7	89.1/60.1	**98.2**/95.2	98.1/96.1	83.3/55.9	85.6/54.1	91.3/56.2	78.9/52.3	81.5/57.6	**98.2/96.5**
Wood	76.5/24.3	82.8/29.0	90.3/74.6	92.8/78.9	85.9/49.0	91.4/48.3	87.1/49.3	75.4/25.1	69.7/22.7	**95.9/82.6**
Bottle	88.6/54.8	94.0/68.1	96.8/88.9	96.3/89.4	94.6/80.3	96.3/78.0	95.4/76.8	83.5/51.5	77.6/41.2	**97.0/92.3**
Cable	66.2/12.6	94.1/60.6	81.0/56.4	80.4/52.0	91.7/74.7	94.1/52.6	96.8/67.0	80.9/36.0	77.7/34.7	**97.2/78.9**
Capsule	90.1/10.1	94.0/48.8	82.7/39.6	92.5/46.4	93.0/48.3	**95.5**/47.2	93.4/46.0	83.6/15.5	79.1/11.7	92.5/**62.2**
Hazelnut	94.3/34.2	97.1/59.9	**98.5**/92.6	98.2/93.4	95.2/60.0	96.9/60.7	90.9/53.2	83.6/22.1	86.9/22.5	97.4/**93.8**
Metal Nut	76.9/34.1	91.5/88.0	97.0/97.0	**97.7**/94.7	91.4/92.2	94.9/78.6	92.6/86.6	76.9/35.6	76.7/29.9	95.8/**98.0**
Pill	86.4/20.9	95.2/82.0	88.4/47.6	89.6/48.3	95.4/81.9	96.7/76.5	94.5/75.7	69.2/14.6	77.0/21.6	**97.2/91.3**
Screw	85.2/6.1	95.8/43.9	95.0/**66.5**	95.2/61.7	93.5/28.7	**98.5**/52.1	97.5/34.7	31.1/1.4	30.1/5.0	92.4/44.9
Toothbrush	87.3/18.3	95.3/46.3	85.6/45.5	85.5/39.3	86.8/55.7	92.3/51.1	94.0/37.9	33.5/6.7	56.1/4.5	**95.6/78.1**
Transistor	68.1/25.8	82.5/67.5	70.4/39.0	62.5/38.1	**95.1**/76.2	83.3/54.1	92.3/66.9	39.1/6.4	49.0/11.0	94.8/**85.6**
Zipper	86.5/31.5	96.6/65.2	**96.8/77.6**	95.2/76.4	94.3/65.2	95.3/57.5	96.1/62.3	81.3/19.6	91.0/42.9	95.5/**77.6**
Total Average	82.9/25.34	93.4/59.6	91.3/68.1	90.8/65.5	92.1/60.0	93.9/55.4	93.9/56.3	71.4/24.4	73.3/26.0	**96.1/78.6**

Table 2. Results of the PRO and AP metrics for anomaly localization performance on MVTec [4].

	Backbone	I ↑	P ↑	O ↑	A ↑	T ↓
CFLOW	WResNet50	97.5	97.7	93.4	59.6	0.127
RD4AD		98.7	97.8	93.9	55.4	0.094
PatchCore		99.2	98.1	93.9	56.3	0.133
CFLOW	ResNet18	96.2	98.1	92.8	59.2	0.106
RD4AD		97.9	97.1	92.7	53.7	0.076
DRA		96.1	84.1	71.5	25.7	0.223
PRN(Ours)		**99.4**	**99.0**	**96.1**	**78.6**	**0.064**

Table 3. Comparison of pre-trained based approaches in terms of performance and inference time (second) on MVTec [4]. "I", "P", "O", "A" and "T" respectively refer to the five metrics of image auroc, pixel auroc, pixel pro, pixel ap, and inference time per image.

Module				Performance			
U-Net	MP	MSA	MF	I ↑	P ↑	O ↑	A ↑
✓				97.4	91.7	88.6	58.5
✓	✓		✓	98.9	98.5	95.3	77.0
✓		✓	✓	97.8	97.0	92.1	74.0
✓	✓	✓		98.7	98.5	95.4	78.1
✓	✓	✓	✓	**99.4**	**99.0**	**96.1**	**78.6**

Table 4. Ablations of different modules in PRN.

standard. Our approach not only gains the best performance, but also significantly reduces the inference time.

We qualitatively evaluate the performance of anomaly localization compared to state-of-the-art methods DRAEM [68] and PatchCore [41] by visualizing the results in Fig. 6. Our model accurately locates the anomalies and clearly focus on all anomalous regions, regardless of their sizes, shapes and numbers. Additional qualitative results are provided in the supplementary material.

4.3. Ablation Study

The importance of MP, MSA and MF. We investigate the importance of each modules in PRN and the results are reported in Table 4. We have the U-Net-like architecture without any module on the skip-connection branch as the baseline. Overall, PRN outperforms the baseline by a large margin, especially on the P, O, and A metrics. All metrics are significantly boosted by employing the MP that performs explicit residual representation. When applying the MSA which performs variable-sized anomalous feature learning, the performance is further improved. This confirms the effectiveness of information exchanging across multi-size receptive fields. Finally, removing the MF causing the degradation of performance, indicates that it is necessary to exchange information across different scales.

Effects of different anomaly generation strategies. We perform ablation studies to investigate the impact of the different components of the proposed anomaly generation strategies in Table 5. The proposed EA alleviates the problem of seen appearance variance, but does not adequately explore the underlying unseen anomalies. Table 5 indicates that the performance of the model increases with the variety of generated anomalies. We argue that the proposed SA consisting of both HEA and HOA can generate anomalies of various sizes, shapes and numbers, allowing our model to generalize to unseen anomalies. Besides, the proposed TA imposes soft constraints on the locations where anomaly regions are imposed, making the generated anomalies as realistic and reasonable as possible, thus significantly improving the performance of the model.

The effect of prototype proportion. The effect of the ratio of prototypes to total normal samples is compared in Table 6. Note that 100% means that no clustering is performed. Each feature map of a normal sample is regarded as a prototype and the number of prototypes is equal to the number of normal samples. The poor performance of the $\text{PRN}_{100\%}$ indicates that the residual representation obtained from the closest cluster prototype is more representative than that obtained from the single closest sample. Besides, too

Category	Unsupervised							Supervised		
	KDAD [47]	CFLOW [18]	DRAEM [68]	SSPCAB [39]	CFA [25]	RD4AD [12]	PatchCore [41]	DevNet [35]	DRA [13]	Ours
Carpet	92.5/45.6	**97.6**/68.3	92.9/65.1	86.4/48.6	93.6/57.2	95.4/56.5	95.5/62.2	85.8/45.7	92.2/52.3	97.0/**82.0**
Grid	72.9/7.3	96.0/41.2	**98.3/62.8**	98.0/57.9	92.9/25.8	94.2/15.8	94.0/24.5	79.8/25.5	71.5/26.8	95.9/45.7
Leather	97.5/26.8	99.2/64.5	97.4/**72.9**	94.0/60.7	95.4/48.5	98.2/47.6	96.9/45.3	88.5/8.1	84.0/5.6	**99.2**/69.7
Tile	74.3/27.7	89.1/60.1	**98.2**/95.2	98.1/96.1	83.3/55.9	85.6/54.1	91.3/56.2	78.9/52.3	81.5/57.6	**98.2/96.5**
Wood	76.5/24.3	82.8/29.0	90.3/74.6	92.8/78.9	85.9/49.0	91.4/48.3	87.1/49.3	75.4/25.1	69.7/22.7	**95.9/82.6**
Bottle	88.6/54.8	94.0/68.1	96.8/88.9	96.3/89.4	94.6/80.3	96.3/78.0	95.4/76.8	83.5/51.5	77.6/41.2	**97.0/92.3**
Cable	66.2/12.6	94.1/60.6	81.0/56.4	80.4/52.0	91.7/74.7	94.1/52.6	96.8/67.0	80.9/36.0	77.7/34.7	**97.2/78.9**
Capsule	90.1/10.1	94.0/48.8	82.7/39.6	92.5/46.4	93.0/48.3	**95.5**/47.2	93.4/46.0	83.6/15.5	79.1/11.7	92.5/**62.2**
Hazelnut	94.3/34.2	97.1/59.9	**98.5**/92.6	98.2/93.4	95.2/60.0	96.9/60.7	90.9/53.2	83.6/22.1	86.9/22.5	97.4/**93.8**
Metal Nut	76.9/34.1	91.5/88.0	97.0/97.0	**97.7**/94.7	91.4/92.2	94.9/78.6	92.6/86.6	76.9/35.6	76.7/29.9	95.8/**98.0**
Pill	86.4/20.9	95.2/82.0	88.4/47.6	89.6/48.3	95.4/81.9	96.7/76.5	94.5/75.7	69.2/14.6	77.0/21.6	**97.2/91.3**
Screw	85.2/6.1	95.8/43.9	95.0/**66.5**	95.2/61.7	93.5/28.7	**98.5**/52.1	97.5/34.7	31.1/1.4	30.1/5.0	92.4/44.9
Toothbrush	87.3/18.3	95.3/46.3	85.6/45.5	85.5/39.3	86.8/55.7	92.3/51.1	94.0/37.9	33.5/6.7	56.1/4.5	**95.6/78.1**
Transistor	68.1/25.8	82.5/67.5	70.4/39.0	62.5/38.1	**95.1**/76.2	83.3/54.1	92.3/66.9	39.1/6.4	49.0/11.0	94.8/**85.6**
Zipper	86.5/31.5	96.6/65.2	**96.8/77.6**	95.2/76.4	94.3/65.2	95.3/57.5	96.1/62.3	81.3/19.6	91.0/42.9	95.5/**77.6**
Total Average	82.9/25.34	93.4/59.6	91.3/68.1	90.8/65.5	92.1/60.0	93.9/55.4	93.9/56.3	71.4/24.4	73.3/26.0	**96.1/78.6**

表2. MVTec [4] 异常定位性能的PRO与AP指标结果。

	Backbone	I↑	P↑	O↑	A↑	T↓
CFLOW	WResNet50	97.5	97.7	93.4	59.6	0.127
RD4AD		98.7	97.8	93.9	55.4	0.094
PatchCore		99.2	98.1	93.9	56.3	0.133
CFLOW	ResNet18	96.2	98.1	92.8	59.2	0.106
RD4AD		97.9	97.1	92.7	53.7	0.076
DRA		96.1	84.1	71.5	25.7	0.223
PRN(Ours)		**99.4**	**99.0**	**96.1**	**78.6**	**0.064**

表3. 基于预训练的方法在MVTec [4]上的性能与推理时间（秒）对比。“I”、“P”、“O”、“A”和“T”分别指图像AUROC、像素AUROC、像素PRO、像素AP和单图推理时间这五项指标。

Module				Performance			
U-Net	MP	MSA	MF	I↑	P↑	O↑	A↑
✓				97.4	91.7	88.6	58.5
✓	✓		✓	98.9	98.5	95.3	77.0
✓		✓	✓	97.8	97.0	92.1	74.0
✓	✓	✓		98.7	98.5	95.4	78.1
✓	✓	✓	✓	**99.4**	**99.0**	**96.1**	**78.6**

表4. PRN中不同模块的消融实验。

标准。我们的方法不仅获得了最佳性能，还显著减少了推理时间。

我们通过在图6中可视化结果，对异常定位性能与最先进方法DRAEM[68]和PatchCore[41]进行了定性评估。我们的模型能准确定位异常，并清晰聚焦于所有异常区域，无论其大小、形状和数量如何。更多定性结果见补充材料。

4.3. 消融研究

MP、MSA与MF的重要性。我们研究了PRN中各个模块的重要性，结果如表4所示。我们以跳跃连接分支上不包含任何模块的类U-Net架构作为基线。总体而言，PRN大幅超越基线性能，尤其在P、O和A指标上。通过采用执行显式残差表示的MP模块，所有指标均得到显著提升。当引入执行可变尺寸异常特征学习的MSA模块后，性能得到进一步改善。这证实了跨多尺度感受野进行信息交换的有效性。最后，移除MF模块会导致性能下降的性能表明，有必要在不同尺度间交换信息。

不同异常生成策略的效果。我们进行了消融研究，以探讨表5中提出的异常生成策略各组成部分的影响。所提出的EA缓解了可见外观变化的问题，但未能充分探索潜在的未见异常。表5表明，模型的性能随着生成异常种类的增加而提升。我们认为，由HEA和HOA共同构成的SA能够生成不同大小、形状和数量的异常，使我们的模型能够泛化到未见异常。此外，所提出的TA对异常区域施加的位置施加了软约束，使生成的异常尽可能真实合理，从而显著提升了模型的性能。

原型比例的影响。在表6中比较了原型与总正常样本比例的影响。请注意，100%意味着未执行聚类操作，每个正常样本的特征图均被视为一个原型，且原型数量等于正常样本数量。$PRN_{100\%}$的较差性能表明，从最接近的聚类原型获得的残差表示比从单个最接近样本获得的表示更具代表性。此外，过

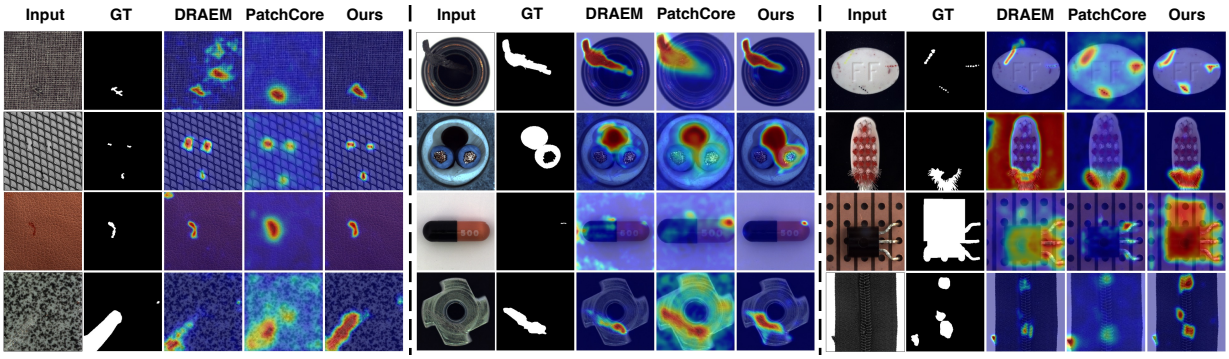

Figure 6. Qualitative examples on MVTec [4]. PRN achieves more accurate localization results for various types of anomalies.

Anomaly Generation				Performance			
EA	HEA	HOA	TA	I ↑	P ↑	O ↑	A ↑
✓			✓	98.6	97.2	93.4	75.7
✓	✓		✓	99.1	98.4	95.4	77.4
✓		✓	✓	98.6	98.4	95.7	75.2
	✓	✓	✓	98.7	98.2	95.1	73.4
✓	✓	✓		98.4	98.4	94.9	77.6
✓	✓	✓	✓	**99.4**	**99.0**	**96.1**	**78.6**

Table 5. Ablations of anomaly generation strategies.

	I ↑	P ↑	O ↑	A ↑	T ↓
$PRN_{5\%}$	99.2	98.6	95.4	78.1	**0.063**
$PRN_{10\%}$	**99.4**	**99.0**	**96.1**	**78.6**	0.064
$PRN_{20\%}$	99.2	98.8	95.7	77.3	0.066
$PRN_{100\%}$	86.2	91.4	75.4	49.9	0.074

Table 6. Ablations of the ratio of prototypes to total normal samples.

	DevNet [35]				DRA [13]				**PRN(Ours)**			
	I ↑	P ↑	O ↑	A ↑	I ↑	P ↑	O ↑	A ↑	I ↑	P ↑	O ↑	A ↑
1	79.6	75.3	51.0	16.5	88.9	78.8	58.2	19.1	98.8	98.3	95.4	74.7
5	86.7	83.7	66.9	22.7	93.5	82.8	68.6	21.9	99.2	98.6	95.6	76.4
10	92.2	85.3	71.4	24.4	96.1	85.3	73.3	26.0	**99.4**	**99.0**	**96.1**	**78.6**

Table 7. Impact of the number of seen anomalies used.

few prototypes lead to insufficient discrimination between prototypes, resulting in inferior performance. The results indicate that a proportion of 10% produces the optimum performance. In addition, using fewer prototypes can speed up inference.

Effects of the number of seen anomalies used. As shown in Table 7, we explore the impact of the number of anomalies used. Our approach significantly outperforms Devnet [35] and DRA [13] using different numbers of seen anomalies, which demonstrates the effectiveness of our proposed anomaly generation strategies and the robustness of PRN to datasets of different levels of imbalance.

4.4. Evaluation on other benchmarks

To further evaluate the anomaly detection and localization capabilities of PRN, we benchmark PRN on three additional widely used datasets, namely DAGM [61], BTAD [34] and KolektorSDD2 [7]. As shown in Table 8, PRN achieves new SOTA performance on all three datasets, proving its effectiveness and generalization. Results for more detailed comparisons and some qualitative examples are provided in the supplementary material.

	DAGM [61]				BTAD [34]				KolektorSDD2 [7]			
	I ↑	P ↑	O ↑	A ↑	I ↑	P ↑	O ↑	A ↑	I ↑	P ↑	O ↑	A ↑
DRAEM	91.1	83.4	70.5	35.6	89.0	87.1	61.6	19.2	81.1	85.6	67.9	39.1
CFLOW	91.2	95.1	87.6	45.2	90.5	96.1	71.6	**54.0**	95.2	97.4	93.8	46.0
SSPCAB	90.4	84.5	71.9	33.9	88.3	83.5	54.1	13.0	83.4	86.2	66.1	44.5
RD4AD	90.7	94.1	85.5	40.8	94.4	96.9	75.8	53.5	96.0	**97.6**	94.7	43.5
PatchCore	92.5	96.1	88.0	49.0	92.6	96.9	76.3	51.5	94.6	97.1	89.3	49.8
DRA	93.5	95.1	88.8	47.6	94.2	75.4	56.2	12.4	86.8	84.4	56.9	3.6
Ours	**98.2**	**96.6**	**93.8**	**49.4**	**94.7**	**97.1**	**78.0**	**54.0**	**96.4**	**97.6**	**94.9**	**72.5**

Table 8. Comparison of PRN with other approaches on DAGM, BTAD, and KolektorSDD2.

5. Conclusion

In this paper, we proposed a novel framework called Prototypical Residual Network for anomaly detection and localization. PRN learns residual resentations across multi-scale feature maps and within multi-size receptive fields at each scale, enabling accurate detection and localization of anomalous regions that come in a variety of sizes, shapes and numbers. In addition, we propose various anomaly generation strategies to expand and diversify the anomalies. We conduct in-depth experiments on four popular datasets to confirm the effectiveness and generalizability of our approach. PRN achieves new SOTA on anomaly detection and and significantly surpasses previous arts in anomaly localization performance.

Limitations. Our approach requires the dataset to provide accurate ground truth masks for anomalies. Using a single image-level anomaly average score for anomalous images with different defect sizes does not favor tiny defects. We leave this intriguing extension to future work.

Acknowledgement This project was supported by NSFC under Grant No. 62102092 and No. 62032006.

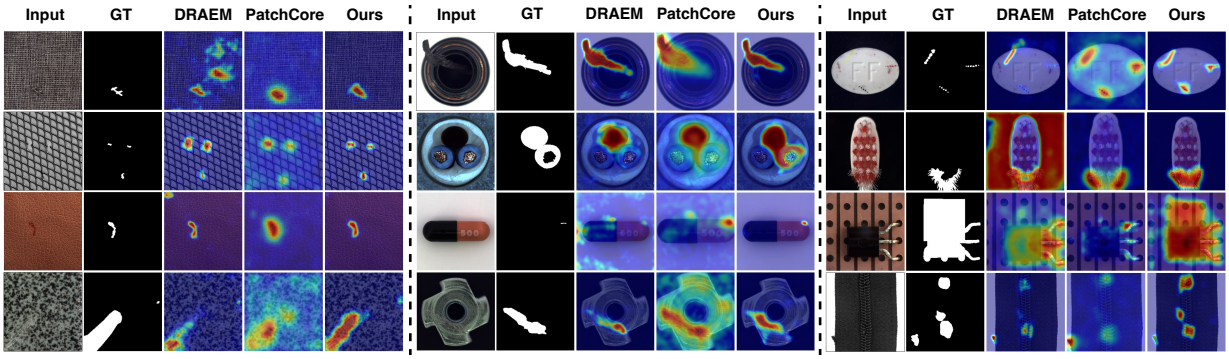

图6. MVTec [4]上的定性示例。PRN针对各类异常实现了更精准的定位结果。

Anomaly Generation				Performance			
EA	HEA	HOA	TA	I↑	P↑	O↑	A↑
✓			✓	98.6	97.2	93.4	75.7
✓	✓		✓	99.1	98.4	95.4	77.4
✓		✓	✓	98.6	98.4	95.7	75.2
	✓	✓	✓	98.7	98.2	95.1	73.4
✓	✓	✓		98.4	98.4	94.9	77.6
✓	✓	✓	✓	**99.4**	**99.0**	**96.1**	**78.6**

表5. 异常生成策略的消融实验。

	I↑	P↑	O↑	A↑	T↓
$PRN_{5\%}$	99.2	98.6	95.4	78.1	**0.063**
$PRN_{10\%}$	**99.4**	**99.0**	**96.1**	**78.6**	0.064
$PRN_{20\%}$	99.2	98.8	95.7	77.3	0.066
$PRN_{100\%}$	86.2	91.4	75.4	49.9	0.074

表6. 原型与总正常样本比例的消融实验。

	DevNet [35]				DRA [13]				**PRN(Ours)**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
1	79.6	75.3	51.0	16.5	88.9	78.8	58.2	19.1	98.8	98.3	95.4	74.7
5	86.7	83.7	66.9	22.7	93.5	82.8	68.6	21.9	99.2	98.6	95.6	76.4
10	92.2	85.3	71.4	24.4	96.1	85.3	73.3	26.0	**99.4**	**99.0**	**96.1**	**78.6**

表7. 使用已知异常数量{v*}的影响。

原型数量过少会导致原型间区分度不足，从而影响性能表现。结果表明，10%的比例能产生最佳性能。此外，使用更少的原型可以加快推理速度。

所用可见异常数量影响。如表7所示，我们探讨了所用异常数量的影响。在使用不同数量可见异常时，我们的方法显著优于Devnet [35]和DRA [13]，这证明了我们提出的异常生成策略的有效性以及PRN对不同不平衡程度数据集的鲁棒性。

4.4. 在其他基准测试上的评估

为了进一步评估PRN的异常检测与定位能力，我们在三个额外的

	DAGM [61]				BTAD [34]				KolektorSDD2 [7]			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
DRAEM	91.1	83.4	70.5	35.6	89.0	87.1	61.6	19.2	81.1	85.6	67.9	39.1
CFLOW	91.2	95.1	87.6	45.2	90.5	96.1	71.6	**54.0**	95.2	97.4	93.8	46.0
SSPCAB	90.4	84.5	71.9	33.9	88.3	83.5	54.1	13.0	83.4	86.2	66.1	44.5
RD4AD	90.7	94.1	85.5	40.8	94.4	96.9	75.8	53.5	96.0	**97.6**	94.7	43.5
PatchCore	92.5	96.1	88.0	49.0	92.6	96.9	76.3	51.5	94.6	97.1	89.3	49.8
DRA	93.5	95.1	88.8	47.6	94.2	75.4	56.2	12.4	86.8	84.4	56.9	3.6
Ours	**98.2**	**96.6**	**93.8**	**49.4**	**94.7**	**97.1**	**78.0**	**54.0**	**96.4**	**97.6**	**94.9**	**72.5**

表8. DAGM、BTAD和KolektorSDD2数据集上PRN与其他方法的比较。

广泛使用的数据集，即DAGM [61]、BTAD [34]和KolektorSDD2 [7]。如表8所示，PRN在所有三个数据集上均取得了新的SOTA性能，证明了其有效性和泛化能力。更详细的比较结果和一些定性示例见补充材料。

5. 结论

本文提出了一种名为原型残差网络的新框架，用于异常检测与定位。PRN通过在多尺度特征图中学习残差表示，并在每个尺度内结合多尺寸感受野，能够精准检测并定位不同尺寸、形状和数量的异常区域。此外，我们提出了多种异常生成策略以扩展并丰富异常样本的多样性。我们在四个常用数据集上进行了深入实验，验证了所提方法的有效性与泛化能力。PRN在异常检测任务中取得了当前最优性能，并在异常定位指标上显著超越了现有方法。

局限性。我们的方法要求数据集提供准确的异常真实掩码。对于具有不同缺陷大小的异常图像，使用单一图像级别的异常平均分数不利于微小缺陷的检测。我们将这一有趣的扩展留待未来工作。

致谢 本项目由国家自然科学基金资助，项目批准号：62102092 和 62032006。

A. Appendix

A.1. Anomaly Generation Strategies

This section details the generation of simulated anomalies, as shown in Fig. 7. A noise image is generated by a Perlin noise generator [37, 68] (Fig. 7,P), and then the noise parts within a target area are retained as the ground truth mask (Fig. 7,M). As the shape, size, and number of generated anomalous regions vary widely, we synthesize simulated anomalies (Fig. 7,S) as:

$$S = \bar{M} \odot N + (1 - \beta)(M \odot A) + \beta\,(M \odot N) \qquad (6)$$

where N is the normal sample, A is the source image of the anomaly, $\bar{M}$ is the inverse of M, $\odot$ is the element-wise multiplication operation, β is the opacity parameter for better combination of abnormal and normal regions. When A is an image randomly sampled from the DTD dataset [8] and is augmented (Aug_1, Fig. 5 in Section 3.4), we define S as a HEterologous Anomaly (HEA). Correspondingly, when A is an image randomly sampled from augmented normal samples, we define S as a HOmology Anomaly (HOA). In particular, the normal image is first augmented (Aug_1, Fig. 5 in Section 3.4), then is evenly divided into an 8×8 grid and randomly arranged before being reassembled [63].

Fig. 8 shows the anomalies generated by different strategies. In addition to increasing the number, extended anomalies (EA) increase the variety of seen anomalies. HEA and HOA supplement potential unseen anomalies with anomalies significantly different from seen anomalies.

A.2. Dataset Split

MVTec AD [4] is a widely used anomaly detection and localization benchmark with 15 classes, each containing one to several subclasses of anomalies. Following the general setting proposed by DRA [13], the 10 labeled anomaly samples are sampled from all possible anomaly classes in the test set per dataset. These sampled anomalies are then removed from the test data. Both BTAD [34] and KolektorSDD2 [7] are real-world industrial datasets containing three product types and one product type, respectively. The general setting used in BTAD and SDD2 is same to that used in MVTec. DAGM [61] contains 10 texture classes, and the original training set for each class consists of normal and abnormal samples. For each class, we first move all anomalous samples from the original training set to the original test set, and then randomly select ten anomalous samples from the test set as part of the new training set. These sampled anomalies are then removed from the test set.

A.3. More Detailed Comparison

Table 9 includes fine-grained anomaly detection and localization performance comparisons on all DAGM sub-datasets. We observe that PRN consistently performs well on all 10 sub-datasets and, in the average scenario, performs best across all four criteria. In particular, our approach outperforms previous methods by a large margin in two metrics, image auroc and pro.

We also compare the anomaly detection and location performance of each method in detail on the three BTAD products and report the numerical results in Table 10. It can be concluded that our method achieves consistently higher performance than the others on different categories .

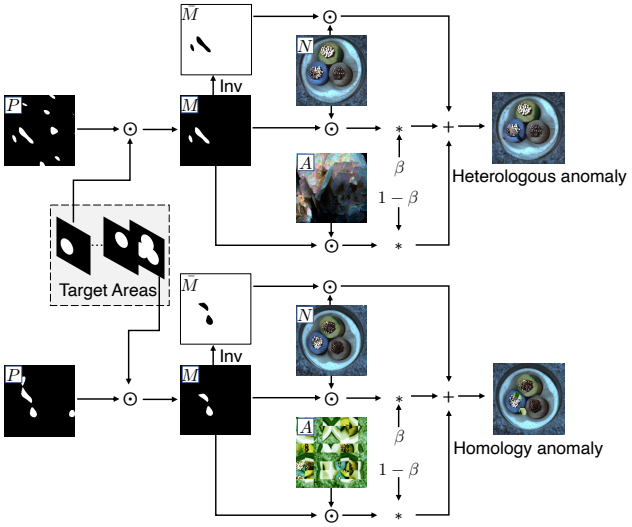

Figure 7. Generating simulated anomalies.

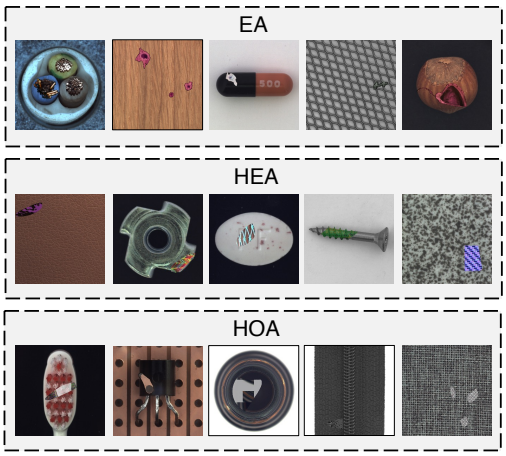

Figure 8. Examples of anomalies generated by different strategies.

A.4. More Qualitative Examples

We further qualitatively evaluate the performance of anomaly detection and location compared to state-of-the-art methods by introducing additional visualizations, as shown in Fig. 9, Fig. 10 and Fig. 11. Our method accurately detects and localizes anomalies in a wide range of sizes, shapes and

A. 附录

A.1. 异常生成策略

本节详细介绍了模拟异常生成的过程，如图7所示。首先通过柏林噪声生成器[37, 68]生成噪声图像（图7,P），随后保留目标区域内的噪声部分作为真实掩码（图7,M）。由于生成的异常区域在形状、大小和数量上存在较大差异，我们通过以下方式合成模拟异常（图7,S）：

$$S = \bar{M} \odot N + (1-\beta)(M \odot A) + \beta\,(M \odot N) \qquad (6)$$

其中 N 是正常样本，A 是异常源图像，$\bar{M}$ 是 M 的逆运算，$\odot$ 是逐元素乘法操作，β 是为更好融合异常与正常区域的不透明度参数。当 A 是从 DTD 数据集 [8] 中随机采样并经增强处理的图像（Aug_1，见第 3.4 节图 5）时，我们将 S 定义为异源异常（HEA）。相应地，当 A 是从增强后的正常样本中随机采样的图像时，我们将 S 定义为同源异常（HOA）。具体而言，正常图像首先经过增强处理（Aug_1，见第 3.4 节图 5），随后被均匀划分为 8×8 网格并随机排列，最后重新拼接 [63]。

图8展示了不同策略生成的异常。除了增加数量外，扩展异常（EA）还增加了已见异常的种类。HEA和HOA通过引入与已见异常显著不同的异常，补充了潜在的未见异常。

A.2. 数据集划分

MVTec AD [4] 是一个广泛使用的异常检测与定位基准数据集，包含15个类别，每个类别涵盖一种至多种异常子类。遵循 DRA [13] 提出的通用设置，我们从每个数据集的测试集中所有可能的异常类别中抽取10个带标签的异常样本。这些被抽样的异常样本随后会从测试数据中移除。BTAD [34] 和 KolektorSDD2 [7] 均为真实工业数据集，分别包含三种产品类型和一种产品类型。BTAD 和 SDD2 采用的通用设置与 MVTec 相同。DAGM [61] 包含10个纹理类别，每个类别的原始训练集同时包含正常样本和异常样本。对于每个类别，我们首先将所有异常样本从原始训练集移至原始测试集，然后从测试集中随机选择十个异常样本作为新训练集的一部分。这些被抽样的异常样本随后会从测试集中移除。

A.3. 更详细的比较

表9包含了所有DAGM子数据集上细粒度异常检测与定位的性能对比。我们观察到PRN在所有10个数据集上均表现稳定优异。

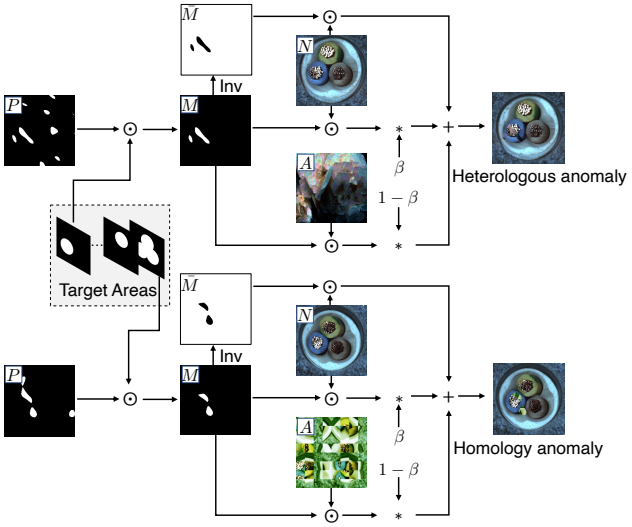

图7. 生成模拟异常。

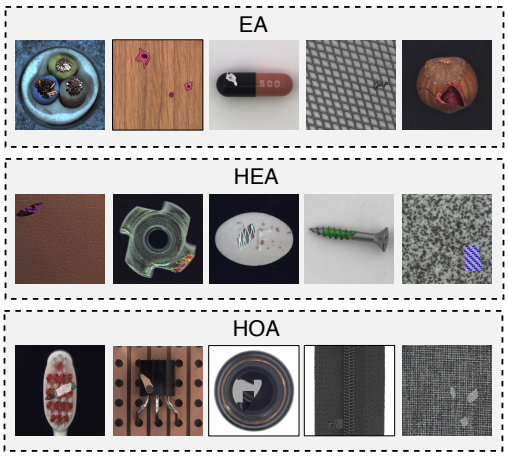

图8. 不同策略生成异常值的示例。

子数据集，并且在平均场景下，在所有四个标准中表现最佳。特别是，我们的方法在两个指标——图像AUROC和PRO上，以较大优势超越了先前的方法。

我们还在三种BTAD产品上详细比较了每种方法的异常检测与定位性能，并将数值结果呈现在表10中。可以得出结论：在不同类别上，我们的方法始终比其他方法实现了更高的性能。

A.4. 更多定性示例

我们通过引入额外的可视化（如图9、图10和图11所示），进一步定性评估了异常检测与定位性能，并与现有先进方法进行了对比。我们的方法能够准确检测并定位各种尺寸、形状和

Category	DRAEM [68]				CFLOW [18]				SSPCAB [39]				RD4AD [12]				PatchCore [41]				**Ours**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
Class1	86.9	75.4	56.3	20.9	91.6	**94.1**	84.1	33.4	95.3	78.9	61.2	29.9	95.2	92.8	83.0	41.6	84.4	89.6	72.8	13.1	**100**	92.7	**90.3**	**50.1**
Class2	85.8	83.7	66.1	18.2	98.2	99.6	98.2	50.2	93.9	92.0	80.4	18.3	99.7	**99.7**	99.1	**57.8**	**100**	**99.7**	**99.3**	55.5	96.0	97.1	95.6	44.8
Class3	98.0	90.3	78.2	32.9	88.3	93.7	86.0	32.9	**99.6**	90.3	79.2	31.7	81.2	93.9	85.2	31.8	94.0	**96.2**	**92.4**	**50.9**	99.2	94.2	91.3	32.4
Class4	99.3	98.6	95.5	62.4	**100**	**99.5**	**98.5**	65.1	99.9	99.1	97.6	74.7	99.9	99.1	97.7	64.6	100	99.4	98.4	**88.2**	99.7	98.2	96.7	67.2
Class5	**97.9**	56.4	39.9	21.9	86.3	94.3	84.5	**50.7**	81.1	53.6	35.9	15.5	74.1	86.7	64.3	31.2	90.6	**95.2**	77.3	29.6	96.9	94.9	**86.1**	30.2
Class6	**100**	96.0	89.3	71.5	96.5	96.1	87.9	46.9	**100**	95.4	88.3	70.0	92.0	88.3	68.9	30.3	99.4	98.1	93.5	71.2	**100**	**98.4**	**95.7**	**71.7**
Class7	**100**	96.7	90.8	58.1	98.9	96.0	91.8	61.4	100	94.8	87.0	51.1	99.8	95.2	91.4	65.7	99.9	**96.9**	**94.8**	**77.7**	**100**	95.1	91.3	51.3
Class8	**99.7**	92.9	90.4	34.2	56.7	79.9	51.0	3.2	96.4	91.1	88.9	23.2	65.2	86.2	67.6	7.0	60.6	86.4	56.5	7.8	93.4	**97.1**	**95.1**	**34.4**
Class9	50.2	49.7	13.3	0.1	99.9	**99.9**	**99.8**	**65.1**	50.9	60.4	26.1	0.1	**100**	99.8	99.4	26.5	96.4	99.4	95.7	45.9	97.1	98.7	96.8	46.4
Class10	92.7	94.2	85.4	35.7	95.7	98.0	94.4	42.9	86.5	89.1	74.7	24.4	99.6	99.0	97.9	51.1	**99.9**	**99.6**	**99.0**	49.6	**99.9**	**99.6**	**99.0**	**65.6**
Average	91.1	83.4	70.5	35.6	91.2	95.1	87.6	45.2	90.4	84.5	71.9	33.9	90.7	94.1	85.5	40.8	92.5	96.1	88.0	49.0	**98.2**	**96.6**	**93.8**	**49.4**

Table 9. Anomaly Detection and Localization on DAGM [61]. "I", "P", "O" and "A" respectively refer to the five metrics of image auroc, pixel auroc, pro and ap. The best results are highlighted in bold.

Category	DRAEM [68]				CFLOW [18]				SSPCAB [39]				RD4AD [12]				PatchCore [41]				**Ours**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
01	98.5	91.5	61.4	17.0	93.4	94.8	60.1	**39.6**	96.2	92.4	62.8	18.1	98.8	95.7	72.8	49.3	96.6	96.5	78.4	47.1	**100**	**96.6**	**81.4**	38.8
02	68.6	73.4	39.0	23.3	79.0	93.9	**56.9**	65.5	69.3	65.6	28.6	15.8	**84.9**	**96.0**	55.8	**66.1**	81.3	94.9	54.0	56.3	84.1	95.1	54.4	65.7
03	99.8	96.3	84.3	17.2	99.1	99.5	97.9	56.8	99.4	92.4	71.0	5.0	99.5	99.0	98.8	45.1	**99.9**	99.2	96.4	51.2	**99.9**	**99.6**	**98.3**	**57.4**
Average	89.0	87.1	61.6	19.2	90.5	96.1	71.6	**54.0**	88.3	83.5	54.1	13	94.4	96.9	75.8	53.5	92.6	96.9	76.3	51.5	**94.7**	**97.1**	**78.0**	**54.0**

Table 10. Anomaly Detection and Localization on BTAD [34].

numbers, as demonstrated by qualitative comparison results. Moreover, we argue that some of the localization errors can be attributed to inaccurate ground truth labels on anomalies. An example of this is shown in the second row of Fig. 11, where the ground truth does not label all anomalous regions. Another example is shown on the left in the fourth row of Fig. 10, where the ground truth labels a broad anomaly region, but our method correctly localizes the anomaly region. These imprecise annotations inevitably impact the anomaly localization scores of the evaluated methods.

References

[1] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh Ahuja, and Utku Genc. Anomalib: A deep learning library for anomaly detection. *arXiv preprint arXiv:2202.08341*, 2022. 6

[2] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *ACCV*, 2018. 3

[3] Samet Akçay, Amir Atapour-Abarghouei, and Toby P Breckon. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In *IJCNN*, 2019. 2

[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, 2019. 1, 5, 6, 7, 8, 9, 10

[5] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher

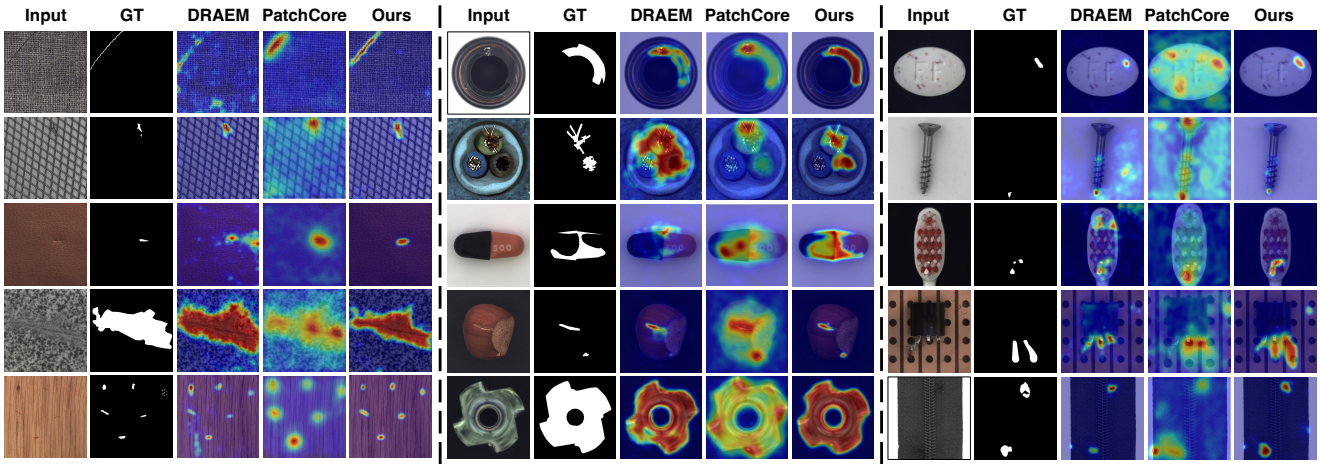

Figure 9. More qualitative examples on MVTec [4].

Category	DRAEM [68]				CFLOW [18]				SSPCAB [39]				RD4AD [12]				PatchCore [41]				**Ours**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
Class1	86.9	75.4	56.3	20.9	91.6	**94.1**	84.1	33.4	95.3	78.9	61.2	29.9	95.2	92.8	83.0	41.6	84.4	89.6	72.8	13.1	**100**	92.7	**90.3**	**50.1**
Class2	85.8	83.7	66.1	18.2	98.2	99.6	98.2	50.2	93.9	92.0	80.4	18.3	99.7	**99.7**	99.1	**57.8**	**100**	**99.7**	**99.3**	55.5	96.0	97.1	95.6	44.8
Class3	98.0	90.3	78.2	32.9	88.3	93.7	86.0	32.9	**99.6**	90.3	79.2	31.7	81.2	93.9	85.2	31.8	94.0	**96.2**	**92.4**	**50.9**	99.2	94.2	91.3	32.4
Class4	99.3	98.6	95.5	62.4	**100**	**99.5**	**98.5**	65.1	99.9	99.1	97.6	74.7	99.9	99.1	97.7	64.6	100	99.4	98.4	**88.2**	99.7	98.2	96.7	67.2
Class5	**97.9**	56.4	39.9	21.9	86.3	94.3	84.5	**50.7**	81.1	53.6	35.9	15.5	74.1	86.7	64.3	31.2	90.6	**95.2**	77.3	29.6	96.9	94.9	**86.1**	30.2
Class6	**100**	96.0	89.3	71.5	96.5	96.1	87.9	46.9	**100**	95.4	88.3	70.0	92.0	88.3	68.9	30.3	99.4	98.1	93.5	71.2	**100**	**98.4**	**95.7**	**71.7**
Class7	**100**	96.7	90.8	58.1	98.9	96.0	91.8	61.4	100	94.8	87.0	51.1	99.8	95.2	91.4	65.7	99.9	**96.9**	**94.8**	**77.7**	**100**	95.1	91.3	51.3
Class8	**99.7**	92.9	90.4	34.2	56.7	79.9	51.0	3.2	96.4	91.1	88.9	23.2	65.2	86.2	67.6	7.0	60.6	86.4	56.5	7.8	93.4	**97.1**	**95.1**	**34.4**
Class9	50.2	49.7	13.3	0.1	99.9	**99.9**	**99.8**	**65.1**	50.9	60.4	26.1	0.1	**100**	99.8	99.4	26.5	96.4	99.4	95.7	45.9	97.1	98.7	96.8	46.4
Class10	92.7	94.2	85.4	35.7	95.7	98.0	94.4	42.9	86.5	89.1	74.7	24.4	99.6	99.0	97.9	51.1	**99.9**	**99.6**	**99.0**	49.6	**99.9**	**99.6**	**99.0**	**65.6**
Average	91.1	83.4	70.5	35.6	91.2	95.1	87.6	45.2	90.4	84.5	71.9	33.9	90.7	94.1	85.5	40.8	92.5	96.1	88.0	49.0	**98.2**	**96.6**	**93.8**	**49.4**

表9. DAGM [61]上的异常检测与定位。“I”、“P”、“O”和“A”分别指图像auroc、像素auroc、pro和ap这五项指标。最佳结果以**粗体**标出。

Category	DRAEM [68]				CFLOW [18]				SSPCAB [39]				RD4AD [12]				PatchCore [41]				**Ours**			
	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑	I↑	P↑	O↑	A↑
01	98.5	91.5	61.4	17.0	93.4	94.8	60.1	**39.6**	96.2	92.4	62.8	18.1	98.8	95.7	72.8	49.3	96.6	96.5	78.4	47.1	**100**	**96.6**	**81.4**	38.8
02	68.6	73.4	39.0	23.3	79.0	93.9	**56.9**	65.5	69.3	65.6	28.6	15.8	**84.9**	**96.0**	55.8	**66.1**	81.3	94.9	54.0	56.3	84.1	95.1	54.4	65.7
03	99.8	96.3	84.3	17.2	99.1	99.5	97.9	56.8	99.4	92.4	71.0	5.0	99.5	99.0	98.8	45.1	**99.9**	99.2	96.4	51.2	**99.9**	**99.6**	**98.3**	**57.4**
Average	89.0	87.1	61.6	19.2	90.5	96.1	71.6	**54.0**	88.3	83.5	54.1	13	94.4	96.9	75.8	53.5	92.6	96.9	76.3	51.5	**94.7**	**97.1**	**78.0**	**54.0**

表10. BTAD [34]上的异常检测与定位。

如定性比较结果所示，数值上亦有所体现。此外，我们认为部分定位误差可归因于异常标注真值的不准确性。图11第二行展示了此类案例：标注真值未覆盖所有异常区域。另一案例如图10第四行左侧所示，标注真值划定了宽泛的异常区域，而我们的方法准确定位了异常区域。这些不精确的标注不可避免地影响了各评估方法的异常定位得分。

参考文献

[1] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh Ahuja, 以及 Utku Genc。Anomalib：一个用于异常检测的深度学习库。*arXiv preprint arXiv:2202.08341*，2022年6月

[2] Samet Akcay, Amir Atapour-Abarghouei 和 Toby P Breckon。Ganomaly：通过对抗性训练进行半监督异常检测。发表于 *ACCV*，2018年。3[3] Samet Akçay, Amir Atapour-Abarghouei 和 Toby P Breckon。Skip-ganomaly：跳跃连接与对抗训练的编码器-解码器异常检测。发表于 *IJCNN*，2019年。2[4] Paul Bergmann, Michael Fauser, David Sattlegger 和 Carsten Steger。Mvtec ad——一个用于无监督异常检测的综合真实世界数据集。发表于 *CVPR*，2019年。1, 5, 6, 7, 8, 9, 10[5] Paul Bergmann, Michael Fauser, David Sattlegger 和 Carsten Steger。无先验知识的学生：师生

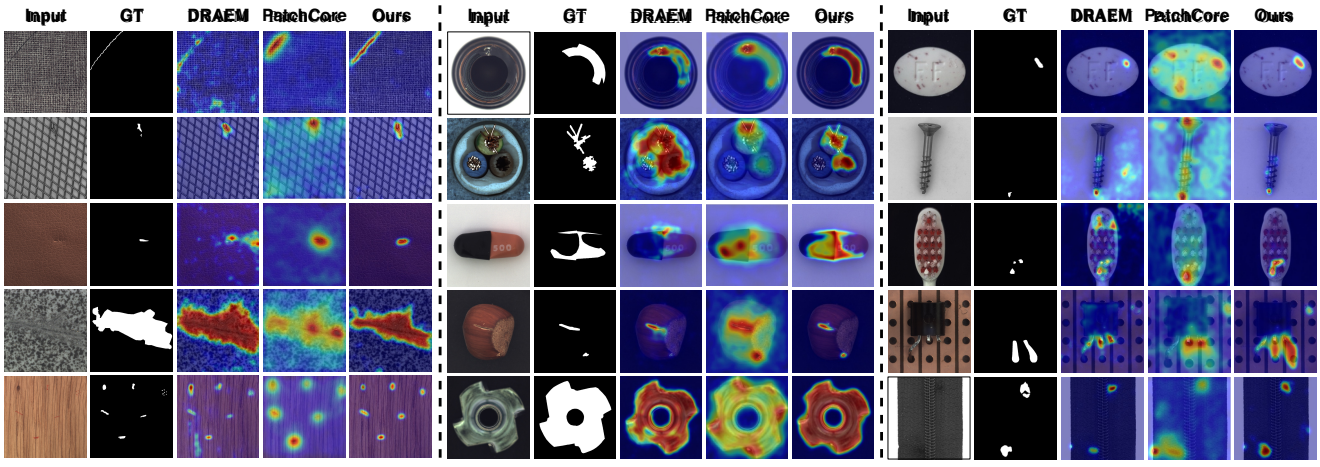

图9. MVTec [4]上的更多定性示例。

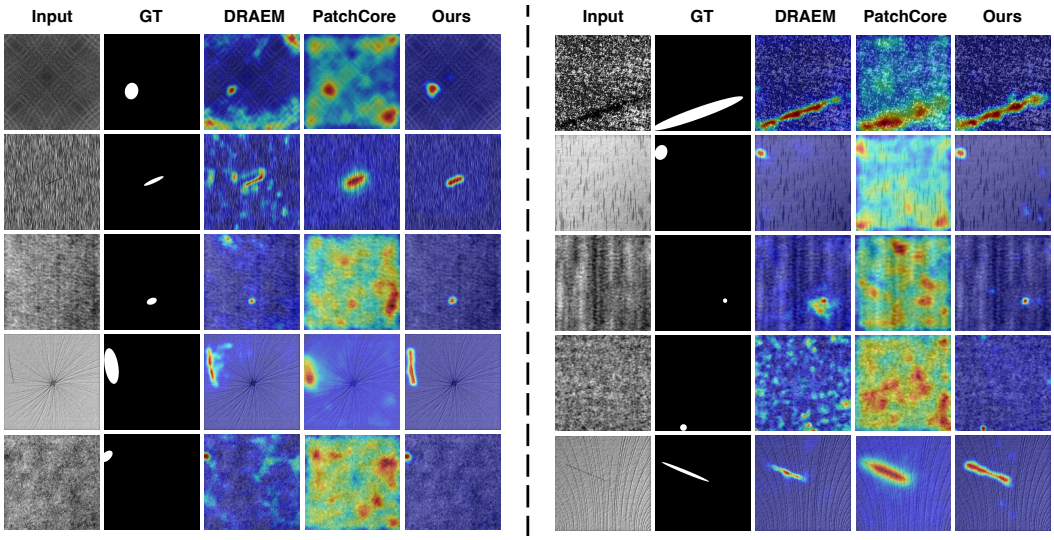

Figure 10. Qualitative examples on DAGM [61].

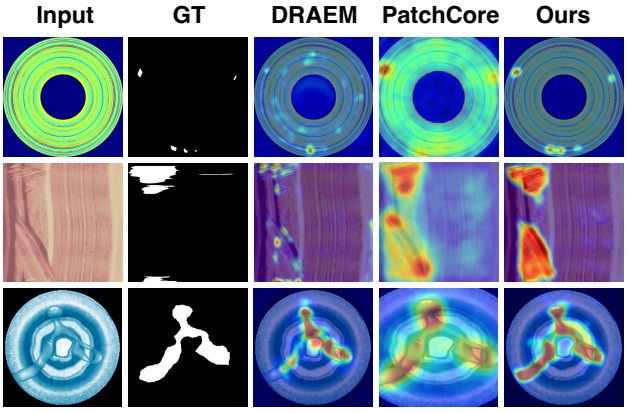

Figure 11. Qualitative examples on BTAD [34].

anomaly detection with discriminative latent embeddings. In *CVPR*, 2020. 2, 3, 6

[6] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*, 2018. 2

[7] Jakob Božič, Domen Tabernik, and Danijel Skočaj. Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Comput Ind*, 2021. 1, 5, 8, 9

[8] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *CVPR*, 2014. 5, 9

[9] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 2, 3

[10] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *ICPR*, 2021. 2, 3

[11] David Dehaene, Oriel Frigo, Sébastien Combrexelle, and Pierre Eline. Iterative energy-based projection on a normal data manifold for anomaly localization. *arXiv preprint arXiv:2002.03734*, 2020. 2

[12] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, 2022. 2, 3, 6, 7, 10

[13] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *CVPR*, 2022. 1, 2, 3, 5, 6, 7, 8, 9

[14] Ross Girshick. Fast r-cnn. In *ICCV*, 2015. 5

[15] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *ICCV*, 2019. 2

[16] Nico Görnitz, Marius Kloft, Konrad Rieck, and Ulf Brefeld. Toward supervised anomaly detection. *JAIR*, 2013. 3

[17] Jiaqi Gu, Hyoukjun Kwon, Dilin Wang, Wei Ye, Meng Li, Yu-Hsin Chen, Liangzhen Lai, Vikas Chandra, and David Z Pan. Multi-scale high-resolution vision transformer for semantic segmentation. In *CVPR*, 2022. 3

[18] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *WACV*, 2022. 2, 3, 6, 7, 10

[19] Songqiao Han, Xiyang Hu, Hailiang Huang, Mingqi Jiang, and Yue Zhao. Adbench: Anomaly detection benchmark. *arXiv preprint arXiv:2206.09426*, 2022. 2

[20] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *JSTOR*, 1979. 3

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 3, 6

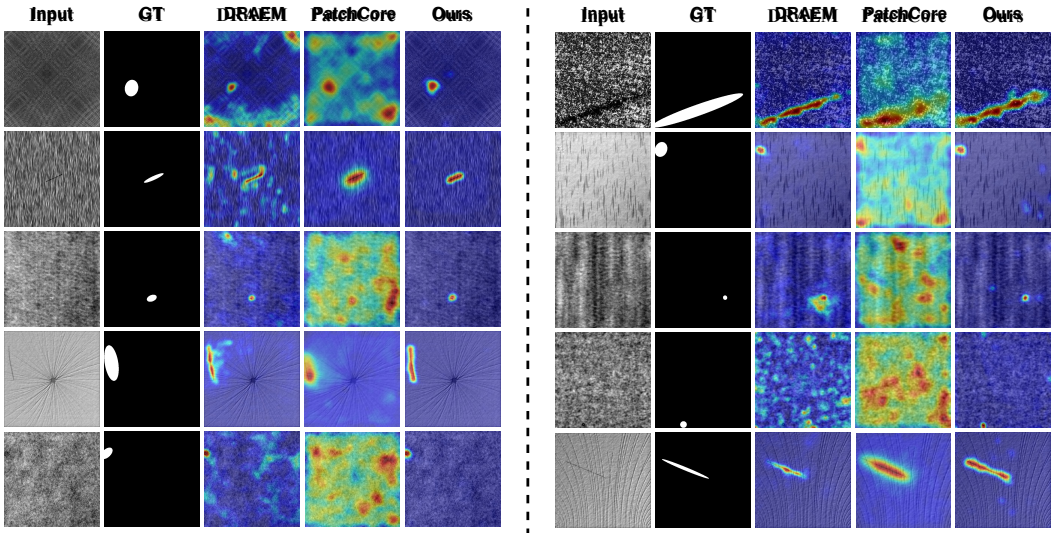

图10. DAGM [61]上的定性示例。

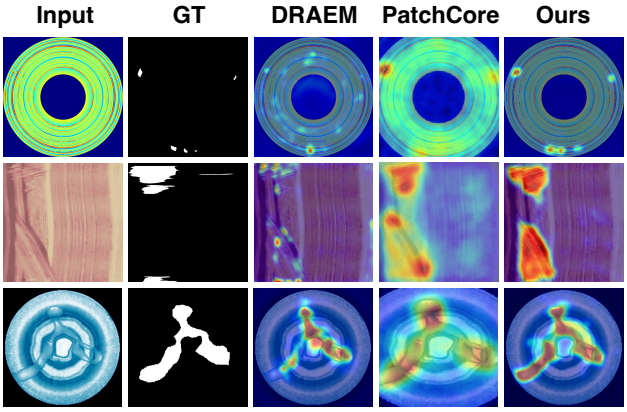

图 11. BTAD [34] 上的定性示例。

基于判别性潜在嵌入的异常检测。于*CVPR*，2020年。2, 3, 6 [6] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattleg- ger, 及 Carsten Steger。通过将结构相似性应用于自编码器改进无监督缺陷分割。*arXiv preprint arXiv:1807.02011*，2018年。2 [7] Jakob Božič, Domen Tabernik, 及 Danijel Skočaj。表面缺陷检测的混合监督：从弱监督到全监督学习。*Comput Ind*，2021年。1, 5, 8, 9 [8] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, 及 Andrea Vedaldi。自然场景中的纹理描述。于*CVPR*，2014年。5, 9 [9] Niv Cohen 与 Yedid Hoshen。基于深度金字塔对应关系的子图像异常检测。*arXiv preprint arXiv:2005.02357*，2020年。2, 3 [10] Thomas Defard, Aleksandr Setkov, Angelique Loesch, 及 Romaric Audigier。Padim：一种基于补丁分布建模的方法异常检测与定位框架。发表于*ICPR*，2021年。第2、3页 [11] David Dehaene, Oriel Frigo, Sébastien Combrexelle, Pierre Eline。基于迭代能量投影的正常数据流形异常定位方法。*arXiv preprint arXiv:2002.03734*，2020年。第2页 [12] 邓寒秋、李星宇。基于单类嵌入反向蒸馏的异常检测。发表于*CVPR*，2022年。第2、3、6、7、10页 [13] 丁丑波、庞冠松、沈春华。灰天鹅与黑天鹅双重捕获：开放集监督异常检测。发表于*CVPR*，2022年。第1、2、3、5、6、7、8、9页 [14] Ross Girshick。Fast R-CNN。发表于*ICCV*，2015年。第5页 [15] 董功、刘凌巧、Le Vuong、Budhaditya Saha、Moussa Reda Mansour、Svetha Venkatesh、Anton van den Hengel。记忆常态以检测异常：基于记忆增强深度自编码器的无监督异常检测。发表于*ICCV*，2019年。第2页 [16] Nico Görnitz, Marius Kloft, Konrad Rieck, Ulf Brefeld。面向监督式异常检测。*JAIR*，2013年。第3页 [17] 顾佳琪、Kwon Hyoukjun、王迪林、叶炜、李萌、陈宇昕、赖亮珍、Vikas Chandra、潘子平。用于语义分割的多尺度高分辨率视觉Transformer。发表于*CVPR*，2022年。第3页 [18] Denis Gudovskiy, Shun Ishizaka, Kazuki Kozuka。CFLOW-AD：基于条件归一化流的实时无监督异常检测与定位。发表于*WACV*，2022年。第2、3、6、7、10页 [19] 韩松乔、胡希阳、黄海亮、蒋明琪、赵越。AdBench：异常检测基准测试集。*arXiv preprint arXiv:2206.09426*，2022年。第2页 [20] John A Hartigan, Manchek A Wong。AS 136算法：K均值聚类算法。*JSTOR*，1979年。第3页 [21] 何恺明、张祥雨、任少卿、孙剑。深度残差学习在图像识别中的应用。发表于*CVPR*，2016年。第3、6页

[22] Jinlei Hou, Yingying Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, and Hong Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In *ICCV*, 2021. 3

[23] Jin-Hwa Kim, Do-Hyeong Kim, Saehoon Yi, and Taehoon Lee. Semi-orthogonal embedding for efficient unsupervised anomaly segmentation. *arXiv preprint arXiv:2105.14737*, 2021. 3

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[25] Sungwook Lee, Seunghyun Lee, and Byung Cheol Song. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *ACCESS*, 2022. 2, 3, 6, 7

[26] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, 2021. 2

[27] Jie Li, Xing Xu, Lianli Gao, Zheng Wang, and Jie Shao. Cognitive visual anomaly detection with constrained latent representations for industrial inspection robot. *Appl. Soft Comput.*, 2020. 3

[28] Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection. *arXiv preprint arXiv:2203.00259*, 2022. 2

[29] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017. 5

[30] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J Radke, and Octavia Camps. Towards visually explaining variational autoencoders. In *CVPR*, 2020. 2

[31] Wen Liu, Weixin Luo, Zhengxin Li, Peilin Zhao, Shenghua Gao, et al. Margin learning embedded prediction for video anomaly detection with a few anomalies. In *IJCAI*, 2019. 3

[32] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *CVPR*, 2018. 1

[33] Lingchen Meng, Hengduo Li, Bor-Chun Chen, Shiyi Lan, Zuxuan Wu, Yu-Gang Jiang, and Ser-Nam Lim. Adavit: Adaptive vision transformers for efficient image recognition. In *CVPR*, 2022. 2

[34] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *ISIE*, 2021. 1, 3, 5, 8, 9, 10, 11

[35] Guansong Pang, Choubo Ding, Chunhua Shen, and Anton van den Hengel. Explainable deep few-shot anomaly detection with deviation networks. *arXiv preprint arXiv:2108.00462*, 2021. 2, 3, 5, 6, 7, 8

[36] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *CSUR*, 2021. 2

[37] Ken Perlin. An image synthesizer. *ACM SIGGRAPH*, 1985. 5, 9

[38] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *ICIAP*, 2022. 3

[39] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In *CVPR*, 2022. 3, 6, 7, 10

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 3

[41] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, 2022. 1, 2, 3, 6, 7, 10

[42] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but differnet: Semi-supervised defect detection with normalizing flows. In *WACV*, 2021. 2, 3

[43] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *WACV*, 2022. 2, 3

[44] Lukas Ruff, Jacob R Kauffmann, Robert A Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G Dietterich, and Klaus-Robert Müller. A unifying review of deep and shallow anomaly detection. *IEEE*, 2021. 2

[45] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *PMLR*, 2018. 3

[46] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. *arXiv preprint arXiv:1906.02694*, 2019. 2, 3

[47] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *CVPR*, 2021. 2, 3, 6, 7

[48] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal*, 2019. 3

[49] 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 *IPMI*, 2017. 3

[50] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *ECCV*, 2022. 2

[51] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput*, 2001. 3

[52] Philipp Seeböck, Sebastian Waldstein, Sophie Klimscha, Bianca S Gerendas, René Donner, Thomas Schlegl, Ursula Schmidt-Erfurth, and Georg Langs. Identifying and categorizing anomalies in retinal imaging data. *arXiv preprint arXiv:1612.00686*, 2016. 1

[53] Xian Tao, Xinyi Gong, Xin Zhang, Shaohua Yan, and Chandranath Adak. Deep learning for unsupervised anomaly localization in industrial images: A survey. *TIM*, 2022. 1, 2, 3, 6

[22] 侯金磊、张莹莹、钟乔勇、谢迪、濮世亮、周虹。分而治之：学习块级记忆用于无监督异常检测。于*ICCV*，2021年。3[23] Jin-Hwa Kim、Do-Hyeong Kim、Saehoon Yi、Taehoon Lee。用于高效无监督异常分割的半正交嵌入。*arXiv preprint arXiv:2105.14737*，2021年。3[24] Diederik P Kingma、Jimmy Ba。Adam：一种随机优化方法。*arXiv preprint arXiv:1412.6980*，2014年。6[25] Sungwook Lee、Seunghyun Lee、Byung Cheol Song。CFA：基于耦合超球面的特征自适应用于目标导向的异常定位。*ACCESS*，2022年。2, 3, 6, 7[26] Chun-Liang Li、Kihyuk Sohn、Jinsung Yoon、Tomas Pfister。CutPaste：用于异常检测与定位的自监督学习。于*CVPR*，2021年。2[27] 李杰、徐兴、高连丽、王铮、邵杰。面向工业巡检机器人的带约束潜在表征认知视觉异常检测。*Appl. Soft Comput.*，2020年。3[28] 梁宇飞、张江宁、赵世伟、吴润泽、刘勇、潘书文。用于无监督异常检测的全频通道选择表征。*arXiv preprint arXiv:2203.00259*，2022年。2[29] 林腾仪、Priya Goyal、Ross Girshick、何恺明、Piotr Dollár。用于密集目标检测的焦点损失。于*ICCV*，2017年。5[30] 刘文倩、李润泽、郑萌、Srikrishna Karanam、吴子彦、Bir Bhanu、Richard J Radke、Octavia Camps。面向视觉解释变分自编码器。于*CVPR*，2020年。2[31] 刘文、罗伟新、李正新、赵培林、高升华等。嵌入边际学习的视频异常检测预测（含少量异常样本）。于*IJCAI*，2019年。3[32] 刘文、罗伟新、连东泽、高升华。未来帧预测用于异常检测——一个新的基线。于*CVPR*，2018年。1[33] 孟令琛、李恒铎、Bor-Chun Chen、蓝世一、吴祖玄、蒋宇刚、Ser-Nam Lim。AdaViT：用于高效图像识别的自适应视觉Transformer。于*CVPR*，2022年。2[34] Pankaj Mishra、Riccardo Verk、Daniele Fornasier、Claudio Piciarelli、Gian Luca Foresti。VT-ADL：一种用于图像异常检测与定位的视觉Transformer网络。于*ISIE*，2021年。1, 3, 5, 8, 9, 10, 11[35] 庞冠松、丁丑波、沈春华、Anton van den Hengel。基于偏差网络的可解释深度少样本异常检测。*arXiv preprint arXiv:2108.00462*，2021年。2, 3, 5, 6, 7, 8[36] 庞冠松、沈春华、曹龙冰、Anton Van Den Hengel。异常检测的深度学习综述。*CSUR*，2021年。2[37] Ken Perlin。一种图像合成器。*ACM SIGGRAPH*，1985年。5, 9[38] Jonathan Pirnay、Keng Chai。用于异常检测的修复Transformer。于*ICIAP*，2022年。3[39] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moes- lund, Mubarak Shah。用于异常检测的自监督预测卷积注意力块。发表于*CVPR*，2022年。3, 6, 7, 10[40] Olaf Ronneberger, Philipp Fischer, Thomas Brox。U-Net：用于生物医学图像分割的卷积网络。发表于*MICCAI*，2015年。3[41] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schōlkopf, Thomas Brox, Peter Gehler。迈向工业异常检测的完全召回。发表于*CVPR*，2022年。1, 2, 3, 6, 7, 10[42] Marco Rudolph, Bastian Wandt, Bodo Rosenhahn。相同但又不同：基于标准化流的半监督缺陷检测。发表于*WACV*，2021年。2, 3[43] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, Bastian Wandt。用于基于图像的缺陷检测的全卷积跨尺度流。发表于*WACV*，2022年。2, 3[44] Lukas Ruff, Jacob R Kauffmann, Robert A Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G Dietterich, Klaus-Robert Mūller。深度与浅层异常检测的统一综述。*IEEE*，2021年。2[45] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Mūller, Marius Kloft。深度单类分类。发表于*PMLR*，2018年。3[46] Lukas Ruff, Robert A Vandermeulen, Nico Gōrnitz, Alexander Binder, Emmanuel Mūller, Klaus-Robert Mūller, Marius Kloft。深度半监督异常检测。*arXiv preprint arXiv:1906.02694*，2019年。2, 3[47] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, Hamid R Rabiee。用于异常检测的多分辨率知识蒸馏。发表于*CVPR*，2021年。2, 3, 6, 7[48] Thomas Schlegl, Philipp Seebōck, Sebastian M Waldstein, Georg Langs, Ursula Schmidt-Erfurth。f-AnoGAN：基于生成对抗网络的快速无监督异常检测。*Med Image Anal*，2019年。3[49] Thomas Schlegl, Philipp Seebōck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, Georg Langs。利用生成对抗网络进行无监督异常检测以指导标记发现。发表于*IPMI*，2017年。3[50] Hannah M Schlūter, Jeremy Tan, Benjamin Hou, Bernhard Kainz。用于自监督异常检测与定位的自然合成异常。发表于*ECCV*，2022年。2[51] Bernhard Schōlkopf, John C Platt, John Shawe-Taylor, Alex J Smola, Robert C Williamson。估计高维分布的支持域。*Neural Comput*，2001年。3[52] Philipp Seebōck, Sebastian Waldstein, Sophie Klimscha, Bianca S Gerendas, René Donner, Thomas Schlegl, Ursula Schmidt-Erfurth, Georg Langs。识别与分类视网膜成像数据中的异常。*arXiv preprint arXiv:1612.00686*，2016年。1[53] Xian Tao, Xinyi Gong, Xin Zhang, Shaohua Yan, Chandranath Adak。工业图像中无监督异常定位的深度学习：综述。*TIM*，2022年。1, 2, 3, 6

[54] Rui Tian, Zuxuan Wu, Qi Dai, Han Hu, Yu Qiao, and Yu-Gang Jiang. Resformer: Scaling vits with multi-resolution training. In *CVPR*, 2023. 4

[55] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, 2017. 2, 4

[56] 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

[57] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *TPAMI*, 2020. 3

[58] Junke Wang, Zuxuan Wu, Wenhao Ouyang, Xintong Han, Jingjing Chen, Yu-Gang Jiang, and Ser-Nam Li. M2tr: Multi-modal multi-scale transformers for deepfake detection. In *ICMR*, 2022. 2, 4

[59] Rui Wang, Dongdong Chen, Zuxuan Wu, Yinpeng Chen, Xiyang Dai, Mengchen Liu, Yu-Gang Jiang, Luowei Zhou, and Lu Yuan. Bevt: Bert pretraining of video transformers. In *CVPR*, 2022. 2

[60] Zejia Weng, Xitong Yang, Ang Li, Zuxuan Wu, and Yu-Gang Jiang. Semi-supervised vision transformers. In *ECCV*, 2022. 4

[61] Matthias Wieler and Tobias Hahn. Weakly supervised learning for industrial optical inspection. 2007. 1, 5, 8, 9, 10, 11

[62] Zhen Xing, Qi Dai, Han Hu, Jingjing Chen, Zuxuan Wu, and Yu-Gang Jiang. Svformer: Semi-supervised video transformer for action recognition. In *CVPR*, 2023. 4

[63] Minghui Yang, Peng Wu, Jing Liu, and Hui Feng. Memseg: A semi-supervised method for image surface defect detection using differences and commonalities. *arXiv preprint arXiv:2205.00908*, 2022. 2, 5, 9

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

[65] Sanyapong Youkachen, Miti Ruchanurucks, Teera Phatrapomnant, and Hirohiko Kaneko. Defect segmentation of hot-rolled steel strip surface by using convolutional auto-encoder and conventional image processing. In *IC-ICTES*, 2019. 2

[66] Jongmin Yu, Du Yong Kim, Younkwan Lee, and Moongu Jeon. Unsupervised pixel-level road defect detection via adversarial image-to-frequency transform. In *IV*, 2020. 3

[67] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021. 2, 3

[68] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, 2021. 2, 5, 6, 7, 9, 10

[69] Jianpeng Zhang, Yutong Xie, Guansong Pang, Zhibin Liao, Johan Verjans, Wenxing Li, Zongji Sun, Jian He, Yi Li, Chunhua Shen, et al. Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection. *TMI*, 2020. 3

[70] Ye Zheng, Xiang Wang, Rui Deng, Tianpeng Bao, Rui Zhao, and Liwei Wu. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. In *ICME*, 2022. 3

[54] 田睿、吴祖煊、戴琦、胡瀚、乔宇、蒋宇刚。Resformer：通过多分辨率训练扩展视觉Transformer。发表于*CVPR*，2023年。4[55] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N. Gomez、 ukasz Kaiser、Illia Polosukhin。注意力机制就是你所需要的一切。发表于*NIPS*，2017年。2, 4[56] 王郭栋、韩淑敏、丁二锐、黄迪。基于师生特征金字塔匹配的无监督异常检测。*arXiv preprint arXiv:2103.04257*，2021年。3[57] 王京东、孙珂、程天恒、姜博瑞、邓超睿、赵阳、刘东、穆亚东、谭明奎、王兴刚等。面向视觉识别的深度高分辨率表示学习。*TPAMI*，2020年。3[58] 王俊科、吴祖煊、欧阳文浩、韩昕彤、陈晶晶、蒋宇刚、李思南。M2TR：用于深度伪造检测的多模态多尺度Transformer。发表于*ICMR*，2022年。2, 4[59] 王睿、陈冬冬、吴祖煊、陈胤鹏、戴熙洋、刘梦辰、蒋宇刚、周洛纬、袁路。BEVT：视频Transformer的BERT预训练。发表于*CVPR*，2022年。2[60] 翁泽佳、杨熙彤、李昂、吴祖煊、蒋宇刚。半监督视觉Transformer。发表于*ECCV*，2022年。4[61] Matthias Wieler、Tobias Hahn。面向工业光学检测的弱监督学习。2007年。1, 5, 8, 9, 10, 11[62] 邢震、戴琦、胡瀚、陈晶晶、吴祖煊、蒋宇刚。SVFormer：用于动作识别的半监督视频Transformer。发表于*CVPR*，2023年。4[63] 杨明辉、吴鹏、刘静、冯辉。MemSeg：一种利用差异性与共性进行图像表面缺陷检测的半监督方法。*arXiv preprint arXiv:2205.00908*，2022年。2, 5, 9[64] Jihun Yi、Sungroh Yoon。Patch SVDD：用于异常检测与分割的块级支持向量数据描述。发表于*ACCV*，2020年。3[65] Sanyapong Youkachen、Miti Ruchanurucks、Teera Phatrapomnant、Hirohiko Kaneko。基于卷积自编码器与传统图像处理的热轧钢带表面缺陷分割。发表于*IC-ICTES*，2019年。2[66] Jongmin Yu、Du Yong Kim、Younkwan Lee、Moongu Jeon。基于对抗性图像-频率变换的无监督像素级道路缺陷检测。发表于*IV*，2020年。3[67] 余佳伟、郑晔、王祥、李威、吴雨霜、赵瑞、吴立伟。FastFlow：基于二维标准化流的无监督异常检测与定位。*arXiv preprint arXiv:2111.07677*，2021年。2, 3[68] Vitjan Zavrtanik、Matej Kristan、Danijel Skočaj。DRAEM——一种判别性训练的重构嵌入表面异常检测方法。发表于*ICCV*，2021年。2, 5, 6, 7, 9, 10[69] 张建鹏、谢宇彤、庞冠松、廖志斌、Johan Verjans、李文星、孙宗稷、何健、李毅、沈春华等。基于置信度感知异常检测的胸部X光病毒性肺炎筛查。*TMI*，2020年。3[70] 郑晔、王祥、邓瑞、包天鹏、赵瑞、吴立伟。聚焦你的分布：从粗到细的非-异常检测与定位的对比学习。在*ICME*，2022年。3