

DOI:10.19651/j.cnki.emt.2518323

基于特征重构的工业图像无监督异常检测^{*}

陆畅 李文举 王旭彬 杨康

(上海应用技术大学智能技术学部 上海 201418)

摘要: 异常检测是现代工业制造中的一项重要任务,由于异常样本的稀缺性,仅需正常样本训练的无监督检测引起了广泛关注,其中,重构式检测因其简洁、通用的框架得到普遍应用。然而,现有算法多基于图像进行重构,异常和正常区域之间的区分度不够,同时,由于工业图像中异常位置、大小的未知性强,现有算法无法很好的捕获样本的整体结构特征。针对以上问题,本文提出了一种基于特征重构的工业图像异常检测算法。首先,利用预训练模型提取多尺度特征来作为重构对象,避免了像素空间重构对异常鉴别力不够的状况;其次,设计了一种全局特征提取模块来增强重构模型对全局特征的感知能力;最后,设计一种特征重组策略来联合训练重构模型,以进一步增强模型对样本整体结构的理解,从而提升重构的效果。在 MVTec-AD 上进行的大量实验表明,所提算法在样本级异常检测上实现了 98.7% 的 AUROC 分数,在像素级异常定位上实现了 98.3% 的 AUROC 分数,均达到了最先进的性能。

关键词: 异常检测;无监督学习;特征重构;自监督学习

中图分类号: TN06;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Unsupervised industrial anomaly detection based on feature reconstruction

Lu Chang Li Wenju Wang Xubin Yang Kang

(Faculty of Intelligent Technology, Shanghai Institute of Technology, Shanghai 201418, China)

Abstract: Anomaly detection is an important task in modern industrial manufacturing. Due to the scarcity of abnormal samples, unsupervised detection that only requires normal sample training has attracted widespread attention. Among them, reconstruction based detection has been widely applied due to its concise and universal framework. However, existing algorithms are mostly based on image reconstruction, thus the discrimination between abnormal and normal regions is insufficient. At the same time, due to the strong uncertainty of abnormal positions and sizes in industrial images, existing algorithms cannot capture the overall structural features of samples well. This article proposed an industrial image anomaly detection algorithm based on feature reconstruction to address the above issues. Firstly, the use of pre trained models to extract multi-scale features as reconstruction objects avoids the situation where pixel space reconstruction has insufficient ability to distinguish anomalies; secondly, a global feature extraction module was designed to enhance the perception ability of the reconstruction model towards global features; finally, design a feature recombination strategy to jointly train the reconstruction model, in order to further enhance the model's understanding of the overall structure of the samples and improve the reconstruction effect. A large number of experiments conducted on MVTec AD have shown that the proposed algorithm achieves an AUROC score of 98.7% in sample level anomaly detection and 98.3% in pixel level anomaly localization, both of which have reached state-of-the-art performance.

Keywords: anomaly detection; unsupervised learning; feature reconstruction; self-supervised learning

0 引言

图像异常检测是一种识别与正常情况或预期不符的对象或区域的技术。在现实场景中,异常数据往往是隐患的标志,因此及时发现异常数据非常重要。其广泛应用于工

业视觉检测^[1]、医学影像诊断^[2]以及视频监控^[3]等领域,是学术界以及工业界的一个重要课题。在工业生产中,这些异常情况可能会导致产品质量下降、生产效率低下等问题,而异常检测有助于解决这些问题。然而,异常样本非常稀少,并且实际生产中可能不断产生新型的缺陷,因此异常样

收稿日期:2025-03-11

* 基金项目:国家自然科学基金(52408343)项目资助

本的数量和种类都难以满足训练的需求。在这些情况下,由于目标类别样本的缺乏,传统的目标检测和图像分割的算法已不再适用^[4]。因此,基于无监督学习的异常检测,即仅使用正常样本进行训练模型是很有实际意义的。

图像的异常检测最早是基于传统图像处理的算法,如模板匹配^[5]和统计模型^[6]等。然而,这些传统算法需要手动设计和提取特征,这对于复杂样本来说是一项艰巨的任务。并且,传统算法通常是基于特定的任务和样本进行设计的,对不同类型的样本和缺陷类型适应性较差。近年来,深度学习在计算机视觉中的各个领域内都得到了长足的发展。相比于传统的算法,深度学习凭借其算法的便利性和通用性等优点,已被广泛应用于图像异常检测任务中。当前基于深度学习的无监督异常检测算法大致可以分为以下几类:基于重构的算法^[7-18]、基于预训练特征的算法^[19-24]、基于蒸馏的算法^[25-28]、基于一类分类的算法^[29-33]。其中,基于图像重构的算法原理简单有效,是无监督异常检测中一类重要的分支。该类算法基于一个核心假设:重构网络仅学习正常样本的特征,因此不能准确地重构异常区域。原始图像与重构图像之间的逐像素差异被用于估计异常程度。

目前针对重构算法的改进已经有许多不同的研究。一些算法使用生成模型来增强重构网络的性能,包括差分自编码器(variational autoencoder, VAE)^[7]和生成对抗网络(generative adversarial networks, GAN)^[8-9]。也有一些算法试图改善网络结构来增强网络的重构能力,如在网络中添加跳跃连接^[10]、在多尺度进行重构^[11-12]等。随着重构网络的过拟合,有时也能过于精准地还原异常区域,导致异常检测失效。Gong 等^[13]提出基于记忆的正则化来抑制重构网络过拟合,该算法将编码器提取的隐空间特征用记忆项替换,从而限制了解码器的表达,后续的工作,如文献^[14]继承了这一算法。也有一些算法通过代理任务重新解释重构任务,如文献^[15]通过合成异常将重构任务转换为自监督分割任务,文献^[16]将重构任务替换为图像补绘任务,这些算法不直接训练网络的重构能力,缓解了重构网络的过拟合现象。

但目前重构范式仍然存在一些问题:首先,先前的重构范式通常在像素空间中进行,逐像素的均方误差难以充分反映重构图像和原始图像之间的真实差异,因此网络更难进行准确地重构。其次,以往的重构范式往往忽略对样本整体结构信息的利用,导致网络依赖局部模式进行重构,因而更容易过拟合。

针对重构范式的存在第一个问题,本文尝试将重构在特征空间进行。最近,基于预训练特征的异常检测算法,取得了优秀的成果,这些算法在训练中通过预训练网络提取正常样本的深度特征,然后对所提取的正常特征图进行数学建模。如 Defard 等^[21]通过多元高斯分布建模特征图的每一个位置,在测试时,通过马氏距离度量测试特征图和正

常特征高斯分布的距离得到异常分数。而 Roth 等^[22]将提取的特征图保存以建立特征库,在测试时在特征库中检索测试特征的最近邻,并通过欧氏距离计算异常分数。尽管这些算法往往对所提取的特征图的每一个向量单独地建模,没有充分利用样本不同区域间的关系,但这些算法在捕捉样本的复杂特征方面表现更加出色。因此,本文尝试通过预训练模型来提取信息丰富的特征以改善重构范式的异常检测性能。针对重构范式存在的第二个问题,本文在重构网络中加入全局特征提取模块来增强网络对全局特征的捕获。并且,本文提出一种基于特征重组的自监督学习策略来联合训练重构模型,通过将待重构特征按照网格区域打乱,避免了模型简单地依赖局部特征进行重构,从而进一步增强了模型对样本整体的感知。

综上,本文针对异常检测中重构范式的普遍问题进行改进,提出一种基于特征重构的异常检测算法(feature reconstruction anomaly detection, FRAD)。

首先,本文提出一种基于预训练网络的特征提取器来获取语义信息丰富的特征,将重构目标由像素空间嵌入特征空间以获得更准确的异常检测结果。

其次,本文提出一种全局特征提取模块(global feature extraction, GFE)和基于特征重组的自监督学习策略来增强重构网络对样本整体结构信息的感知,从而抑制重构网络的过拟合。

最后,本文所提模型在工业异常检测数据集 MVTEC AD^[34]上进行了实验,实现了最先进的异常检测和异常定位效果。

1 算 法

本文提出了一种针对异常检测重构范式的改进,如图 1 所示,本文重构网络基于基本的卷积自编码器(convolutional autoencoder, CAE)架构搭建,包含 1 个编码器网络和一个解码器网络。编码器网络包含 3 个编码器模块,基于类似 ResNet 的瓶颈模块(bottleneck, BN)搭建,每个编码器模块包含 3 个 BN 层。每个 BN 层由 3 个卷积层构成,其中,第一个 1×1 卷积用于降维,以减少后续 3×3 卷积的计算量。中间的 3×3 卷积用于特征提取。BN 层中最后一个 1×1 卷积用于恢复维度,以保持输入输出维度一致,以便进行残差连接。解码器网络结构与编码器网络对称,中间由跳跃连接与编码器连接。当上采样时,BN 层中间的 3×3 卷积被替换为一个步长为 2 的 2×2 反卷积。来自编码器网络各层的特征通过跳跃连接直接送往解码器对应模块,和前一级解码器模块的输出在通道维度进行拼接,接着,通过一个卷积核为 1 的卷积层进行降维和融合以减少通道数并保留重要信息,再送入下一层解码器。

为了增强网络对异常的鉴别力,并抑制网络过拟合,本文通过所提改进对重构范式进行了改进,所提改进包含一个多级特征提取器,一种全局特征提取模块和一种自监督

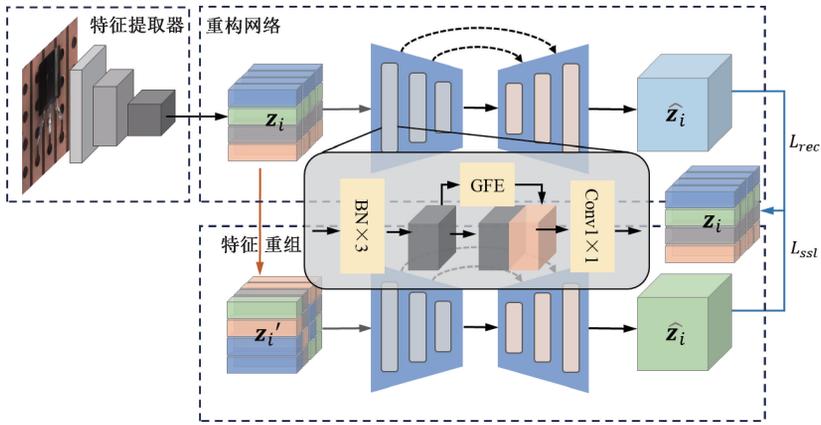


图 1 FRAD 的网络结构和训练设计

Fig. 1 Network structure and training design of FRAD

学习策略。如图 1 所示,在训练中,图像先由特征提取器提取多层次特征,并融合为特征图 z_i 。接着,特征图 z_i 被送入重构网络进行重建。同时,为了增强模型对目标整体结构的感知,本文将 z_i 按照网格划分区域打乱,并重组为 z_i' 。 z_i' 也送入重构网络进行重构(图 1 中下方分支)。

1.1 多级特征提取器

在训练过程中,样本图片首先被送入一个特征提取器来获得图像的特征图。特征图的质量直接决定了异常检测模型的性能:如果提取的特征缺乏区分性,则异常样本和正常样本在特征空间中可能会重叠,导致模型难以将它们准确区分。研究表明,在大型分类数据集上预训练的深度学习神经网络能够提取具有高度鉴别力的特征,这些特征在异常检测任务中显著提升了模型的性能^[19]。本文引入一种基于预训练网络的简单有效的多级特征融合算法,在下方简述其流程。

在训练阶段,目标是从预训练网络中提取多层特征,并将这些特征进行融合。如图 2 所示,特征提取器的输入是任意训练样本 $x_i \in X_{train} = \{x_1, x_2, \dots, x_N\}$,其中 X_{train} 表示一系列训练样本的集合, N 表示训练样本的总数。首先, x_i 被送入预训练网络来获取多层次特征,本文用 ϕ 表示该网络,用 L 表示本文选取的特征层的集合。对于其中每一层 $l \in L, \phi_l(x_i)$ 表示样本经过网络所得到的第 l 层输出特征。接下来,本文将所选取的每一层特征图通过双线性插值调整到一个相同的目标分辨率,这一步可以表示为:

$$\varphi_{i,l} = \text{resize}(\phi_l(x_i)) \in R^{C_l \times H \times W} \quad (1)$$

其中, $\varphi_{i,l}$ 表示分辨率调整过的第 l 层特征图, C_l 表示特征图 $\phi_l(x_i)$ 的通道数, H 和 W 为多层次特征图 $\{\phi_l(x_i) | l \in L\}$ 中分辨率最大的特征图的宽和高。

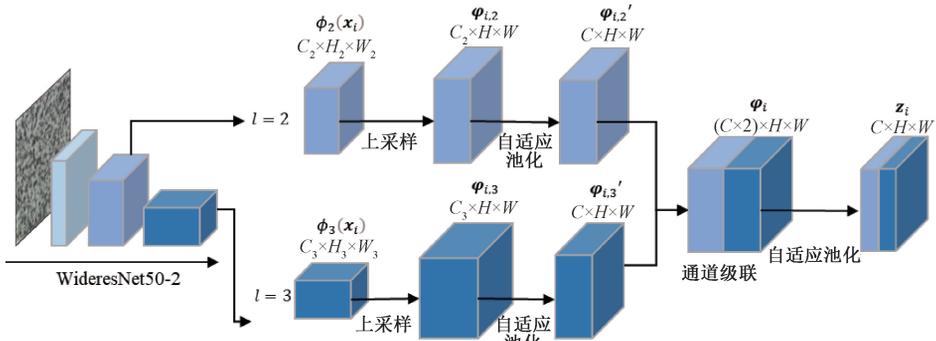


图 2 特征提取流程

Fig. 2 Feature extraction

用 $\varphi_{i,l,h,w} \in R^{C_l}$ 表示特征图 $\varphi_{i,l}$ 上位置 (h, w) 处的 C_l 维向量。本文通过 1 维自适应平均池化调整每一个向量 $\varphi_{i,l,h,w}$ 的通道数,得到:

$$\varphi_{i,l}' = \left\{ \text{pool}_{1d}(\varphi_{i,l,h,w}) \mid \begin{array}{l} h \in \{1, 2, \dots, H\}, \\ w \in \{1, 2, \dots, W\} \end{array} \right\} \in R^{C \times H \times W} \quad (2)$$

接着,对于本文选取的所有特征层 L ,在通道维度拼接多级特征 $\varphi_{i,l}'$,得到:

$$\varphi_i = \text{concat}(\{\varphi_{i,l}' | l \in L\}) \quad (3)$$

最后,对特征图 φ_i 每像素位置的特征向量再次应用 1 维自适应平均池化调整通道维度,得到特征图 $z_i \in R^{C \times H \times W}$ 。

之所以在拼接多级特征之前先将所有特征图的通道调整到同样维度,是因为本文认为来自不同深度的特征在异常鉴别中具有同样重要的作用,应当在多级特征中占据同样的比例。使用 1 维自适应平均池化调整通道维度的好处是其不仅能够控制特征图的维度,还能对特征图进行平滑操作,从而减少系统噪声带来的干扰。

1.2 特征重组策略

针对重构网络对样本整体特征缺乏掌握的问题,本文提出一种自监督学习策略来促进模型学习样本结构,这一步将待重构的特征图划分为多个网格区域,将这些子区域打乱后再由重构网络重构出原始特征,通过将打乱的特征图恢复,重构模型能够在此过程中学习到图像的整体结构。具体地,在通过上节提及的多级特征提取器得到多级特征表示 $z_i \in R^{C \times H \times W}$ 后,如图 3 所示,将这些特征在空间维度上按照正方形网格划分为 $K \times K$ 个子区域。本文的标准设置中, $K = 4$, 因此整个特征图将被划分为 4×4 共 16 个子区域。将划分后的每一个子区域 $z_{i,j} \in R^{C \times \frac{H}{K} \times \frac{W}{K}}$ 按顺序依次排列,得到一个有序的特征列表,记作: $S = [z_{i,j}]$ 。这里 i 是样本的序号, $j \in \{1, 2, \dots, K \times K\}$ 。

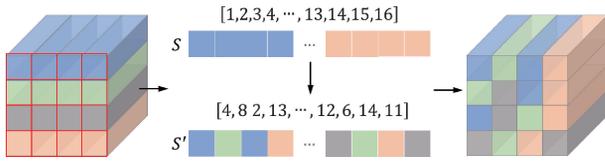


图 3 特征重组示意
Fig. 3 Feature shuffle

接着,对特征列表 S 中子区域特征的顺序进行标记,可以得到一个索引序列,当 $K = 4$ 时,这个索引为 $[1, 2, 3, \dots, 12, 13, 14, 15]$ 。接下来,本文将这个索引序列随机打乱,以生成一个新的索引序列,如图 3 所示,假设新的索引为 $[4, 8, 2, 13, 5, 10, 1, 16, 3, 9, 7, 15, 12, 6, 14, 11]$ 。根据这个重新排列后的索引,本文将子区域特征重新排列,从而得到了一个顺序打乱的列表 S' 。最后,根据列表 S' ,可以重组出打乱的特征图: $z'_i \in R^{C \times H \times W}$ 。如图 1 中所示,重组特征图 z'_i 也被与未重组特征图 z_i 一同送入重构模型进行重构,以联合优化重构模型。

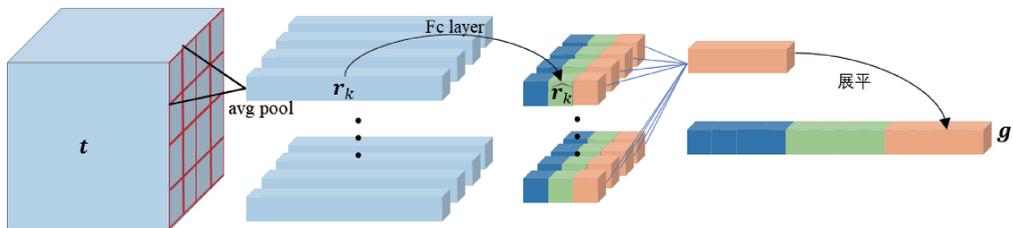


图 4 全局特征提取模块(GFE)
Fig. 4 Global feature extraction (GFE)

1.3 全局特征提取

卷积能通过局部感受野来提取特征,虽然这种机制能有效捕捉纹理和边缘等局部特征,但难以直接建模整个图像的上下文信息。尽管在整个卷积网络中随着层数的增加,卷积层通过叠加能够获得更大范围的感受野,但由于卷积操作的本质是局部计算,卷积网络中信息主要是通过相邻区域的局部特征进行传播,这导致模型在整合全局上下文信息时效率较低。在重构任务中,这会使模型过度依赖于局部的特征进行重构,从而导致模型过拟合。针对这一问题,本文提出一种全局特征提取模块 GFE,旨在显式捕捉图像的全局特征以提升重构网络的泛化性。

设输入 GFE 模块的特征图 $t \in R^{C_i \times H_i \times W_i}$, 获得全局特征的示意如图 4 所示,本文将其均等地划分为 $K \times K$ 的网格区域,每区域的特征可以表示为 $t_k \in R^{C_i \times \frac{H_i}{K} \times \frac{W_i}{K}}$, $k \in \{1, 2, \dots, K \times K\}$ 。本文用 $P = \frac{H_i \times W_i}{K \times K}$ 来表示每区域内包含的特征向量总数,从而可以将每区域的特征表示为:

$$t_k = \{t_{k,p} \mid p \in \{1, 2, \dots, P\}\} \quad (4)$$

其中, $t_{k,p} \in R^{C_i}$ 表示每区域内的每个特征向量。对所有区域特征 t_k , 本文计算其特征向量的平均值,得到区域表示特征:

$$r_k = \frac{1}{P} \sum_{p=1}^P t_{k,p}, r_k \in R^{C_i} \quad (5)$$

接着,对每一个区域表示特征 r_k , 本文通过全连接对其进行压缩,从而获得更紧凑的表示:

$$\hat{r}_k = fc(r_k) \in R^{\hat{C}_i} \quad (6)$$

然后,本文将所有 $K \times K$ 个向量 \hat{r}_k 所构成的特征图在空间维度展平,从而获得一个全局特征向量 $g \in R^{\hat{C}_i \times K \times K}$ 。这个全局特征向量包含了原特征每个网格区域内的紧凑表示信息。本文中令 $\hat{C}_i = \frac{C_i}{K \times K}$, 从而使全局特征向量 g 的通道维度 $\hat{C}_i \times K \times K = C_i$ 。为了将全局特征 g 和特征图 t 融合,本文将全局特征向量 g 在 $H_i \times W_i$ 维度上进行重复并将所得到的特征图和特征图 t 拼接得到融合特征。

1.4 损失函数

为了训练重构网络,损失函数需要评估原始特征 \mathbf{z}_i 和重构特征 $\hat{\mathbf{z}}_i$ 之间的相似度。由于本文的重构在高维特征空间下进行,重构损失通常会使用的欧式距离容易受到维度灾难的影响,即高维空间中所有的点对间的欧氏距离将趋于相似。因此,本文使用余弦相似度来度量特征向量之间的相似性,因为其受维度增加的影响更小。如图 5 所示,余弦相似度不同于欧式距离,仅关注两向量之间的方向夹角,而不关心向量的模长。重构网络的输入和输出的每个特征向量之间的余弦相似度可以表示为:

$$\cos(\mathbf{z}_{i,h,w}, \hat{\mathbf{z}}_{i,h,w}) = \frac{\mathbf{z}_{i,h,w}^T \cdot \hat{\mathbf{z}}_{i,h,w}}{\|\mathbf{z}_{i,h,w}\|_2 \cdot \|\hat{\mathbf{z}}_{i,h,w}\|_2} \quad (7)$$

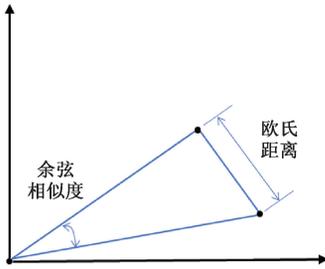


图 5 余弦相似度和欧式距离的几何示意

Fig. 5 Geometric schematic of cosine similarity and Euclidean distance

为了便于损失函数的优化,本文将其转化为余弦距离,得到:

$$D_{\cos}(\mathbf{z}_{i,h,w}, \hat{\mathbf{z}}_{i,h,w}) = 1 - \frac{\mathbf{z}_{i,h,w}^T \cdot \hat{\mathbf{z}}_{i,h,w}}{\|\mathbf{z}_{i,h,w}\|_2 \cdot \|\hat{\mathbf{z}}_{i,h,w}\|_2} \quad (8)$$

当两向量越相似时,余弦距离越接近 0。当两向量完全无关,即相互垂直时,余弦距离为 1。当两向量反向相关时,余弦距离为 2。

对特征图上的所有特征向量,本文计算余弦距离并求均值得到重构损失:

$$L_{rec} = \frac{1}{HW} \sum_{w=0}^W \sum_{h=0}^H D_{\cos}(\mathbf{z}_{i,h,w}, \hat{\mathbf{z}}_{i,h,w}) \quad (9)$$

本文将重组后的特征 \mathbf{z}'_i 和原始特征 \mathbf{z}_i 同时送入重构网络进行重构,对重组后的特征 \mathbf{z}'_i , 重构网络试图将其恢复为未重组特征 \mathbf{z}_i 的相似特征图,即对重组特征 \mathbf{z}'_i 的重构 $\hat{\mathbf{z}}'_i$, 本文计算它和未重组特征图 \mathbf{z}_i 的余弦距离损失,从而得到模型的自监督损失:

$$L_{ssl} = \frac{1}{HW} \sum_{w=0}^W \sum_{h=0}^H D_{\cos}(\mathbf{z}_{i,h,w}, \hat{\mathbf{z}}'_{i,h,w}) \quad (10)$$

本文联合两损失一同对重构模型进行优化,模型的总体损失函数为:

$$L_{total} = L_{rec} + \lambda L_{ssl} \quad (11)$$

其中, λ 为权值,用以平衡两特征 \mathbf{z}_i 和 \mathbf{z}'_i 的重构任务对模型的影响程度。由于特征图的布局被打乱,模型在尝试重构原特征时无法依赖局部特征进行简单的推理,从而

迫使模型转为学习样本的整体特征。通过加强对样本整体特征的学习,提升了重构的精度,并抑制了模型过拟合。

1.5 推理流程

推理时,如图 6 所示,本文仅保留特征提取器和重构网络,而不进行特征重组。每个测试样本 \mathbf{y}_i 经过多级特征提取器得到融合特征 \mathbf{z}_i , 送入重构网络,经过模型重构得到重构特征 $\hat{\mathbf{z}}_i$ 。式(8)也被用于计算 \mathbf{z}_i 和 $\hat{\mathbf{z}}_i$ 每特征向量间的余弦距离以生成像素级的异常分数图,该图像中较大的分数说明该位置存在异常。为了得到精确的定位图像,本文通过双线性插值将异常分数图上采样到样本图像 \mathbf{y}_i 的尺寸,再使用 $\sigma = 4$ 的高斯滤波器平滑图像,最终得到每像素异常分数 A_{pix} 。 A_{pix} 中的最大值作为样本级异常分数 A_{sam} 。

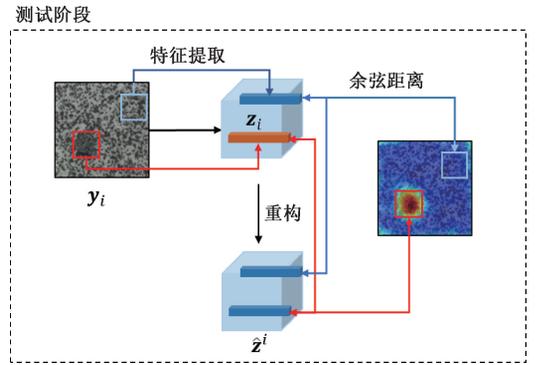


图 6 FRAD 的推理流程

Fig. 6 The reasoning process of FRAD

2 实 验

2.1 数据集

MvTec AD 数据集是一个用于无监督工业视觉异常检测的基准数据集,旨在改善和评估异常检测算法的性能。MvTec AD 共包含 15 个类别,包括 5 种纹理类别和 10 种物体类别,总计 3 629 张图片,涵盖多种常见的工业对象。它的训练集为来自真实世界的无标注正常样本,展示了该类别产品的正常状态。测试集中每一类都包含正常样本和具有特定缺陷的异常样本,如划痕、孔洞或变形等,并且每个异常样本图像都有详细的像素级标注。

2.2 评价指标

无监督工业异常检测任务中,评价指标主要使用接收者工作特征曲线下面积(area under receiver operating curve, AUROC)分数,其由 Bergmann 等^[25]在 MVTEC AD 论文中提出,既可以用于样本级异常检测任务,也可以用于像素级异常定位任务。对于异常定位任务, Bergmann 等^[25]提出使用每区域重叠分数(per-region overlap score, PRO-score)作为评价指标。下面对两种指标进行简要介绍:

1) AUROC

本文使用 MVTEC AD 论文中建议的接收者工作特征

曲线下面积,即 AUROC 作为评估所提算法性能的主要标准。AUROC 的计算涉及到两个基础指标,假阳性率(fake positive rate, FPR)和真阳性率(true positive rate, TPR)。TPR 表示正样本中被正确分类为正类的比例,也称召回率:

$$TPR = \frac{TP}{TP + FN} \quad (12)$$

其中, TP 表示被正确分类的正样本数量, FN 表示被错误分为负类的正样本数量。

假阳性率,也称误报率表示负类样本中被错误分类为正类的比例,计算公式为:

$$FPR = \frac{FP}{FP + TN} \quad (13)$$

其中, FP 表示被误分为正类的负样本数量, TN 表示被正确分为负类的负样本数量。将 FPR 定义为横轴, TPR 定义为纵轴,通过计算模型不同分类阈值下的 TPR 和 FPR 可以作出接收者工作特征曲线(receiver operating curve, ROC),这个曲线下的面积就是 AUROC。因此, AUROC 不依赖于特定的阈值,有助于理解模型在各类阈值下的表现。AUROC 取值在 0~1 之间,接近 1 说明模型在任意分类阈值取值下的召回率都较高并且误报率低。

另外, AUROC 计算时需要同时考虑 TPR 和 FPR,因此其分数不受正负样本比例影响。这一点在判断异常检测模型的性能方面非常实用,因为在工业环境中经常出现类别失衡。对异常检测任务,本文对每个样本的异常分数计算样本级 AUROC。对异常定位任务,本文通过每个样本的像素级异常分数图计算像素级 AUROC。

2) PRO

此外,本文还使用每区域重叠分数,即 PRO 分数来评估本文提出的算法的异常定位质量。PRO 分数的计算公式如下:

$$PRO = \frac{1}{N} \sum_{i=1}^N \max \left(\frac{G_i \cap P_j}{G_i \cup P_j} \right) \quad (14)$$

其中, N 表示单连通真实区域的总数量。对于图像中的每一个单连通的真实区域,算法会遍历所有预测区域,计算该真实区域与每个预测区域之间的交并比,并从中选取最大值作为该真实区域的匹配得分。最终,将所有真实区域的匹配得分取平均值,得到全局的 PRO 分数。

PRO 分数平等地考虑不同大小的异常区域的定位,若模型漏检某个小异常区域,则该区域的 PRO 分数贡献为 0,直接降低总分。如图 7 所示,图中有两个单连通的真实标注区域 G_1 和 G_2 , 以及它们对应的预测区域 P_1 和 P_2 。假设 $G_1 \cap P_1$ 的值是 80%, 而可以从图中看到 $G_2 \cap P_2$ 是 0%, 整个图像的 PRO 分数可以得到是 $(80\% + 0\%) / 2 = 40\%$ 。相比 AUROC 分数, PRO 分数能够更精准地评估模型对任意大小的每个独立异常区域的捕捉能力。

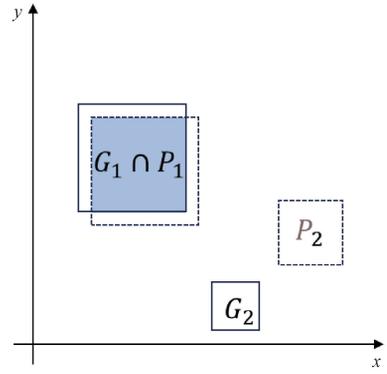


图 7 PRO 分数

Fig. 7 PRO score

2.3 设 置

实验中训练和测试图像被缩放并裁剪到 256×256 分辨率。本文使用在 ImageNet^[35] 上预训练的 WideResnet50-2^[36] 作为特征提取网络 ϕ 。本文选取更加通用的中间层特征,即 $l \in L = \{2, 3\}$ 。特征提取器输出的多级特征通道数 $C = 512$ 。本文采用 $\beta = (0.5, 0.999)$ 的 Adam 优化器来训练特征重构网络,学习率设置为 0.001。本文将 batch 大小设置为 16, 在一张 RTX 4090 GPU 上训练了 100 个 epoch。式(10)的超参数 $\lambda = 0.1$ 。对于 GFE 和自监督策略中特征图网格的划分,本文都设置为 $K = 4$ 。

2.4 异常检测结果

在 MVTec AD 上的异常检测 AUROC 分数如表 1(左侧数据)所示,分数由百分比表示。本文与各类不同范式的异常检测算法都进行了对比,包括基于重构的算法: DR Δ EM^[15]、RIAD^[16]、OmniAL^[17]和 DMU-TransNet^[18]; 基于预训练特征的算法: Padim^[21]; 基于蒸馏的算法: RD^[26]; 基于一类分类的算法: CutPaste^[31]。表中将每一类的所有算法间最优的结果加粗,并将第二名下划线。

结果显示,本文的算法实现了最高的平均图像异常检测性能,平均分数比第 2 名高了 0.2%。并且,本文的模型在纹理和物体类别上的平均 AUROC 分别达到了 99.6% 和 98.2%, 在两种样本大类上表现都很好。另外值得注意的是,在多个类别的异常检测中本文的模型的 AUROC 都达到了 100% 或接近 100% 的水平,这表示本文的模型对各种类型样本的异常检测都具有通用性。

2.5 异常定位结果

本文也在 MVTec AD 上检测了所提模型的异常定位性能,像素级 AUROC 分数如表 1(右侧数据)所示, PRO 分数如表 2 所示。分数均由百分比表示,表中将最优的结果加粗,并将第 2 名下划线。对于像素级 AUROC 指标,本文的算法的平均性能和 OmniAL 并列第 1, 比第 2 名高了 0.4%, 并且,相比 OmniAL 本文算法对物体样本的异常检测能力更好。而 PRO 分数的平均得分中,本文的算法以 94.6% 的成绩超过了第 2 名。较高的 PRO 分数说明本

表 1 MVTec AD 上的 AUROC 分数(样本级%/像素级%)
Table 1 AUROC scores on MVTec AD(sampel-level%/pixel-level%)

Category	DRÆM ^[15]	RIAD ^[16]	OmniAL ^[17]	DMU-TransNet ^[18]	Padim ^[21]	RD ^[26]	CutPaste ^[31]	Ours
carpet	97/95.5	84.2/96.3	98.7/ 99.4	97.6/98.4	99.8/99.1	98.9/98.9	93.9/98.3	100/99.2
grid	99.9/ 99.7	99.6/98.8	99.9/99.4	100 /97.9	96.7/97.3	100 /99.3	100 /98.8	100 /99.3
leather	100 /98.6	100/99.4	99/99.3	99.9/99.3	100 /99.2	100/99.4	100/99.4	100/99.4
tile	99.6/ 99.2	98.7/89.1	99.6 /99	99/94.1	98.1/94.1	99.3/95.6	94.6/89.1	99.2/95.5
wood	99.1/ <u>96.4</u>	93/85.8	93.2/ 97.4	99.8 /94.5	99.2/94.9	99.2/95.3	99.1/85.8	99/95.4
Average	99.1/ <u>97.9</u>	95.1/93.9	98.1/ 98.9	99.2/96.8	98.8/96.9	99.5/97.7	97.5/94.3	99.6 /97.8
bottle	99.2/ <u>99.1</u>	99.9/98.4	100/99.2	100 /96.2	99.9/98.3	100 /98.7	98.2/98.4	<u>100</u> /98.9
cable	91.8/94.7	81.9/84.2	98.2/97.3	<u>98.7</u> / <u>97.5</u>	92.7/96.7	95/97.4	81.2/84.2	99.2/98.3
capsure	98.5 /94.3	88.4/92.8	95.2/96.9	96.8/ 98.9	91.3/98.5	96.3/ <u>98.7</u>	98.2/92.8	96.1/98.5
hazel_nut	100/99.7	83.3/96.1	95.6/98.4	100 /99	92/98.2	99.9/98.9	98.3/96.1	100 / <u>99.2</u>
metal_nut	98.7/ 99.5	88.5/92.5	99.2/99.1	99.4/96.7	98.7/97.2	100 /97.3	99.9/92.5	100 /97.1
pill	98.9 /97.6	83.8/95.7	97.2/ 98.9	97.8/96.3	93.3/95.7	96.6/ <u>98.2</u>	94.9/95.7	<u>98</u> /97.7
screw	93.9/97.6	84.5/98.8	88/98	95.4/96.2	85.8/98.5	97/99.6	88.7/98.8	<u>96.7</u> / <u>99.5</u>
toothbrush	100 /98.1	100 /98.9	100/99.4	87.2/ <u>99.1</u>	96.1/98.8	99.5/ <u>99.1</u>	99.4/98.9	93.3/99.1
transistor	93.1/90.9	90.9/87.7	93.8/93.3	97.1/95.4	<u>97.4</u> / <u>97.5</u>	96.7/92.5	96.1/87.7	100/98.3
zipper	100 /98.8	98.1/97.8	100/99.5	99.7/ <u>98.9</u>	90.3/98.5	98.5/98.2	99.9/97.8	98.5/98.4
Average	97.4/97	89.9/94.3	96.7/ <u>98</u>	97.2/97.4	93.8/97.8	<u>98</u> /97.9	95.5/94.3	98.2/98.5
Total Average	98/97.3	91.7/94.2	97.2/ 98.3	97.9/97.2	95.4/97.5	<u>98.5</u> / <u>97.8</u>	96.2/94.3	98.7/98.3

表 2 MVTec AD 上的异常定位 PRO 分数

Table 2 PRO scores for anomaly localization on MVTec AD

Category	Padim ^[21]	US ^[25]	RD ^[26]	Ours
carpet	96.2	87.9	<u>97</u>	97.5
grid	94.6	<u>95.2</u>	97.6	97.6
leather	<u>97.8</u>	94.5	99.1	99.1
tile	86	94.6	<u>90.6</u>	90.2
wood	<u>91.1</u>	<u>91.1</u>	90.9	92.5
Average	93.1	92.7	<u>95</u>	95.4
bottle	94.8	93.1	<u>96.6</u>	96.8
cable	88.8	81.8	<u>91</u>	93.5
capsure	93.5	96.8	95.8	<u>96.2</u>
hazel_nut	92.6	96.5	95.5	<u>95.7</u>
metal_nut	85.6	94.2	92.3	88
pill	92.7	<u>96.1</u>	96.4	95.9
screw	94.4	94.2	98.2	<u>97.7</u>
toothbrush	93.1	93.3	94.5	<u>94.2</u>
transistor	<u>84.5</u>	66.6	78	87.8
zipper	<u>95.9</u>	95.1	96.4	95.7
Average	91.6	90.8	<u>93.5</u>	94.2
Total Average	92.1	91.4	<u>94</u>	94.6

文提出的算法可以定位不同尺寸的异常。无论是整个物体还是细微缺陷。

2.6 消融实验

本文进行了消融研究,以验证使用本文提出的模型的组成部分的效果;GFE 模块和特征重组策略。此外,本文还研究了特征提取器的不同特征层的选择对模型性能的影响。

表 3 中展示了所提的 GFE 模块和特征重组策略对模型的异常检测性能的影响,分别展示了 AUROC 分数和 PRO 分数。本文将不包括 GFE 模块并且不使用特征重组的特征重构网络作为基线。如表 3 所示,GFE 模块主要增强全局特征表征能力,对整体分类任务更有效,但对模型异常定位能力的提升有限。特征重组策略提升了 0.4% 的像素级 AUROC 和 0.3% 的 PRO,显著改善了模型像素级定位能力,但是也造成了样本级异常检测能力的轻微下降。当两者同时使用时,GFE 能提供可靠的上下文信息,而特征重组防止局部过拟合,从而二者结合实现全局和局部特征平衡。

表 3 所提出的模块在 MVTec AD 上的消融实验

Table 3 The ablation of proposed modules on MVTec AD		样本-AUROC	像素-AUROC	PRO
		97.9	97.6	93.8
✓		98.1	97.7	93.9
	✓	97.6	98.0	94.1
✓	✓	98.7	98.3	94.6

表 4 给出了使用不同的特征层进行重构的异常检测效果,分别展示了 AUROC 分数和 PRO 分数。结果显示,单层特征中,第 2 层具有最好的效果,可以获得接近最佳配置的性能。但第 4 行结果(标准配置)显示,第 2 层特征仍然能够从第 3 层中获得额外的信息,从而实现更加准确的异常检测效果。从单层特征来看,第 1 层效果最差,其包含最少的全局信息,因此对一些结构较为复杂的物体类别识别不够准确。

征提供的鉴别力,Padim 和本文算法都实现了对所有异常区域的定位,并且本文算法具备更准确的定位能力。

表 4 使用不同特征层在 MVTEC AD 上的消融实验

Table 4 The ablation of using different level features on MVTEC AD

1	2	3	样本-AUROC	像素-AUROC	PRO
✓			88.1	95.7	89.1
	✓		98.4	97.9	93.6
		✓	97.6	97.2	90.7
	✓	✓	98.7	98.3	94.6
✓	✓	✓	95.2	97.4	92.8

2.7 可视化定性结果

图 8 中展示了本文所提算法和其他一些异常检测算法的可视化效果对比,图中展示部分类别的异常样本结果。本文算法结合预训练特征改进了重构范式,因此,本文对比了一种基于重构的异常检测算法 DREAM,和一种基于预训练特征的异常检测算法 Padim。根据图 8,像素空间下的重构算法 DRÆM 具有更多细节的定位效果,但是未能准确的检测出所有的异常区域。而通过预训练特

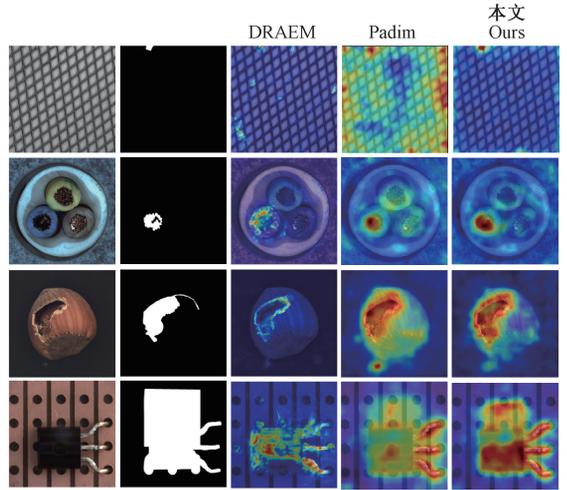
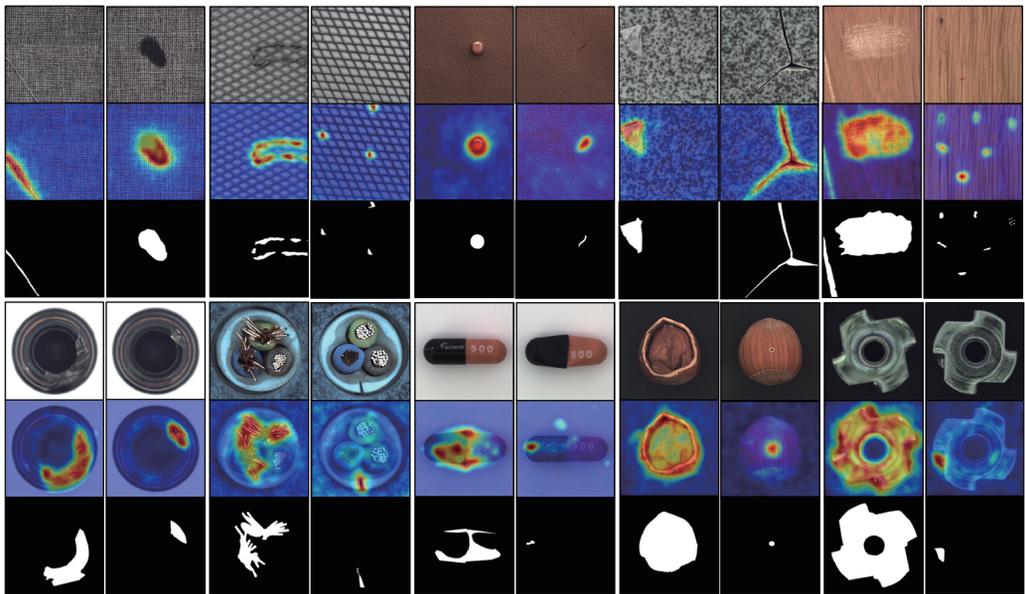


图 8 不同算法在 MVTEC AD 上的可视化结果对比
Fig. 8 Visualizations of different methods on MVTEC AD

所提算法在 MVTEC AD 上全部类别的异常定位可视化定性结果如图 9 所示,展示了包括纹理和物体类别的 15 类共 30 个样本(每类 2 个)。每一类从上到下的 3 行分别是测试图像、热度图和真值掩码,本文对每种类型的数据分别展示大面积缺陷或结构异常和微小的缺陷。能够看到,所提算法对各种类型的异常均能产生很好的定位效果。尤其对于物体类别的全局异常,所提模型具有格外好的定位表现,如图 9 中第 7 列的三极管缺失异常的样本,FRAD 能够完整地检测到缺失的三极管。



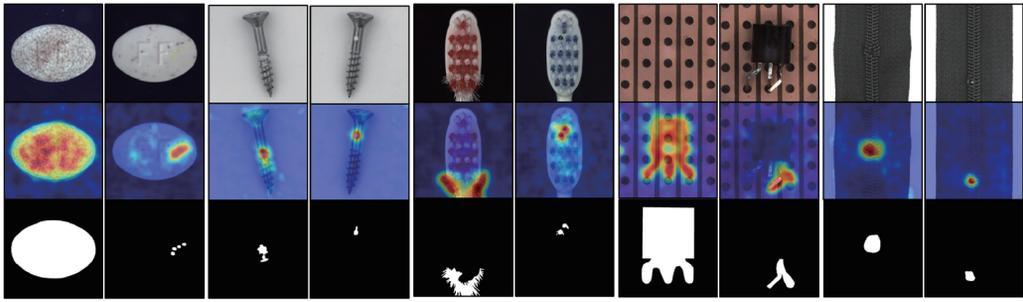


图 9 来自 MVTec AD 测试集的可视化结果

Fig. 9 Visualizations on MVTec AD testset

3 结 论

以往的重构范式通常在像素级重构,缺乏对复杂异常的鉴别力。同时,以往的重构范式没有关注样本的整体结构,导致重构过程依赖于局部特征。为了解决这些问题,本文提出一种基于特征重构的异常检测模型 FRAD,可以用于各种类型的异常检测。该模型通过预训练的网络将样本嵌入特征空间,提升对异常的鉴别力。所提出的全局特征提取模块能在不引入额外大量复杂度的情况下融合全局信息来增强重构网络。并且,FRAD 引入了一种自监督学习策略特征重组,通过重新排列特征图的布局来促进模型学习样本的整体结构。为了验证所提模型,本文在工业图像异常检测基准数据集 MVTec AD 上进行了实验,图像异常检测 AUROC 达到 98.7%,定位 AUROC 达到 98.3%,达到了最先进的异常检测和异常定位性能。

参考文献

- [1] 葛路,何仕荣. 深度学习在工业表面缺陷检测领域的应用研究[J]. 计算技术与自动化, 2022, 41(1): 59-65.
GE L, HE SH R. Research on applications of deep learning in industrial surface defect detection [J]. Computing Technology and Automation, 2022, 41(1): 59-65.
- [2] KERMANY D S, GOLDBAUM M, CAI W J, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning[J]. Cell, 2018, 172(5): 1122-1131.
- [3] LIU W, LUO W X, LIAN D Z, et al. Future frame prediction for anomaly detection-a new baseline[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 6536-6545.
- [4] 吕承侃,沈飞,张正涛,等. 图像异常检测研究现状综述[J]. 自动化学报, 2022, 48(6): 1402-1428.
LYU CH K, SHEN F, ZHANG ZH T, et al. Review of image anomaly detection [J]. Acta Automatica Sinica, 2022, 48(6): 1402-1428.
- [5] CHEN W J, HO J H, MUSTAPHA K B, et al. A

vision based system for anomaly detection and classification in additive manufacturing [C]. 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET). Penang, Malaysia: IEEE, 2019:87-92.

- [6] ZHANG H, JIN X T, WU Q M J, et al. Automatic visual detection system of railway surface defects with curvature filter and improved Gaussian mixture model[J]. IEEE Transactions on Instrumentation and Measurement, 2018, 67(7): 1593-1608.
- [7] BAUR C, WIESTLER B, ALBARQOUNI S, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images[J]. ArXiv preprint arXiv:1804.04488, 2018.
- [8] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]. International Conference on Information Processing in Medical Imaging. Springer, 2017: 146-157.
- [9] 周士金,邢杰. 基于记忆增强 GAN 的异常检测[J]. 计算机科学, 2022, 49(S2): 261-269.
ZHOU SH J, XING H J. Memory-augmented GAN-based anomaly detection[J]. Computer Science, 2022, 49(S2): 261-269.
- [10] AKÇAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection[C]. 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-8.
- [11] MISHRA P, PICIARELLI C, FORESTI G L. Image anomaly detection by aggregating deep pyramidal representations [J]. ArXiv preprint arXiv: 2011.06288, 2020.
- [12] 余文勇,张阳,姚海明,等. 基于轻量化重构网络的表面缺陷视觉检测[J]. 自动化学报, 2022, 48(9): 2175-2186.
YU W Y, ZHANG Y, YAO H M, et al. Visual

- inspection of surface defects based on lightweight reconstruction network[J]. *Acta Automatica Sinica*, 2022, 48(9): 2175-2186.
- [13] GONG D, LIU L Q, LE V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C]. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019:1705-1714.
- [14] 鄢宁,李岳阳,罗海驰. 基于块金字塔记忆模块的无监督异常检测[J]. *计算机工程*, 2023, 49(12): 304-310. YAN N, LI Y Y, LUO H CH. Unsupervised anomaly detection based on block pyramid memory module[J]. *Computer Engineering*, 2023, 49(12): 304-310.
- [15] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DRÆM-A discriminatively trained reconstruction embedding for surface anomaly detection[C]. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021: 8330-8339.
- [16] ZAVRTANIK V, KRISTAN M, SKOČAJ D. Reconstruction by inpainting for visual anomaly detection[J]. *Pattern Recognition*, 2021, 112: 107706.
- [17] ZHAO Y. OmniAL: A unified CNN framework for unsupervised anomaly localization[C]. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 3924-3933.
- [18] ZHOU W, WU SH H, WANG Y Y, et al. DMU-TransNet: Dense multi-scale u-shape transformer network for anomaly detection [J]. *Measurement*, 2024, 229: 114216.
- [19] COHEN N, HOSHEN Y. Sub-image anomaly detection with deep pyramid correspondences [J]. *ArXiv preprint arXiv:2005.02357*, 2020.
- [20] LEE S W, LEE S H, SONG B C. CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization [J]. *ArXiv preprint arXiv:2206.04325*, 2022.
- [21] DEFARD T, SETKOV A, LOESCH A, et al. PaDiM: A patch distribution modeling framework for anomaly detection and localization [C]. *Pattern Recognition, ICPR International Workshops and Challenges: Virtual Event, Part IV*. Springer, 2021: 475-489.
- [22] ROTH K, PEMULA L, ZEPEDA J, et al. Towards total recall in industrial anomaly detection[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 14318-14328.
- [23] BAE J, LEE J H, KIM S. PNI: Industrial anomaly detection using position and neighborhood information[C]. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023: 6373-6383.
- [24] HYUN J, KIM S, JEON G, et al. ReConPatch: Contrastive patch representation learning for industrial anomaly detection[C]. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024: 2052-2061.
- [25] BERGMANN P, FAUSER M, SATTLEGGER D, et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 4183-4192.
- [26] DENG H, LI X. Anomaly detection via reverse distillation from one-class embedding[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 9737-9746.
- [27] 陈亚当,陈柳任,余文斌,等. 多尺度特征融合的知识蒸馏异常检测算法[J]. *计算机辅助设计与图形学学报*, 2022, 34(10):1543-1549. CHEN Y D, CHEN L R, YU W B, et al. Knowledge distillation anomaly detection with multi-scale feature fusion [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2022, 34(10):1543-1549.
- [28] 邢鹏,蒋鑫,潘永华,等. 基于特征约束蒸馏学习的视觉异常检测[J]. *软件学报*, 2023, 34(9): 4378-4391. XING P, JIANG X, PAN Y H, et al. Feature constrained restricted distillation learning for visual anomaly detection [J]. *Journal of Software*, 2023, 34(9): 4378-4391.
- [29] RUFF L, VANDERMEULEN R A, GOERNITZ N, et al. Deep one-class classification[C]. *International Conference on Machine Learning (ICML)*, 2018.
- [30] LIU ZH K, ZHOU Y M, XU Y SH, et al. SimpleNet: A simple network for image anomaly detection and localization[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 20402-20411.
- [31] LI C L, SOHN K, YOON J, et al. CutPaste: Self-supervised learning for anomaly detection and localization[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 9664-9674.
- [32] 邓诗卓,滕达,李晓红,等. 基于球面正则化的支持向量描述视觉异常检测[J]. *仪器仪表学报*, 2024, 45(3): 315-325. DENG SH ZH, TENG D, LI X H, et al. Visual anomaly detection via spherical regularized support

- vector data description [J]. Chinese Journal of Scientific Instrument, 2024, 45(3): 315-325.
- [33] 王飞州,程凡永,张明艳,等. 自监督瓷砖表面异常检测与定位[J]. 电子测量技术, 2023, 46(17): 180-188.
WANG F ZH, CHENG F Y, ZHANG M Y, et al. Self-supervised tile surface anomaly detection and localization[J]. Electronic Measurement Technology, 2023, 46(17): 180-188.
- [34] BERGMANN P, FAUSER M, SATTLAGER D, STEGER C. MVTec AD-a comprehensive real-world dataset for unsupervised anomaly detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 9592-9600.
- [35] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2009:248-255.
- [36] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[J]. ArXiv preprint arXiv: 1605.07146, 2017.

作者简介

陆畅, 硕士研究生, 主要研究方向为计算机视觉、异常检测。

E-mail: 2500428614@qq.com

李文举(通信作者), 博士, 教授, 主要研究方向为计算机视觉、模式识别、智能检测。

E-mail: wjli@sit.edu.cn

王旭彬, 硕士研究生, 主要研究方向为机器视觉、人工智能。

E-mail: highg767@qq.com

杨康, 博士, 讲师, 主要研究方向为基于机器学习的桥梁检测与基础设施智慧运维。

E-mail: yangkang@sit.edu.cn