

基于特征改进的自监督工业图像异常检测

许昌源, 贾可, 金治成, 李涵鑫, 王文润, 周记

(成都信息工程大学 计算机学院, 四川 成都 610000)

摘要: 工业图像的异常检测要求实现图像缺陷识别与定位, 一些异常检测算法使用 ImageNet 数据集得到的预训练模型进行缺陷识别, 未关注 ImageNet 与工业图像的域偏差对异常检测任务的影响。为此, 提出一种优化记忆库特征以提升异常检测性能的方法。首先, 通过粗粒度分割训练缓解域偏差问题; 其次, 引导特征学习对应区域块的关键信息; 最后, 改进损失函数以增强同类样本特征在特征空间中的聚集性, 拉大不同类样本特征间的类间距, 得到更优的记忆库特征。在 MVTec AD 数据集上, 所提方法在图像级、像素级的 AUROC 指标上达到 99.2%、97.6%, 相较于基准模型提升 0.7%、1.4%, 检测效果更优。

关键词: 迁移学习; 自监督学习; 记忆库特征; 粗粒度分割训练; 工业图像异常检测

DOI: 10.11907/rjdk.251192

扫描二维码阅读全文:



中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2025)010-0034-06

Self-Supervised Anomaly Detection in Industrial Images Based on Feature Improvement

XU Changyuan, JIA Ke, JIN Zhicheng, LI Hanxin, WANG Wenrun, ZHOU Ji

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610000, China)

Abstract: The anomaly detection of industrial images requires the recognition and localization of image defects. Some anomaly detection algorithms use pre trained models obtained from the ImageNet dataset for defect recognition, without paying attention to the impact of domain bias between ImageNet and industrial images on anomaly detection tasks. Therefore, a method is proposed to optimize the features of the memory bank to improve anomaly detection performance. Firstly, alleviate domain bias issues through coarse-grained segmentation training; Secondly, guide the feature learning to learn the key information of the corresponding region block; Finally, improving the loss function enhances the clustering of similar sample features in the feature space, widens the class spacing between different sample features, and obtains better memory features. On the MVTec AD dataset, the proposed method achieved 99.2% and 97.6% in AUROC metrics at the image and pixel levels, respectively. Compared to the benchmark model, it improved by 0.7% and 1.4%, indicating better detection performance.

Key Words: transfer learning; self-supervised learning; memory bank features; coarse-grained segmentation training; industrial image anomaly detection

0 引言

随着制造业的不断发展与工业自动化的不断普及, 工业生产活动对产品良品率的要求也日益提高。然而, 人工检测效率低下、成本较高, 难以满足大规模生产加工中进

行实时异常检测的要求。工业图像异常检测是指通过工业相机拍照, 识别出采集图像中的异常样本及异常区域, 通常包括各类产品表面的划痕、凸起、污渍、气泡等。

目前, 基于深度学习的图像异常检测方法精度高、速度快, 具有广泛的应用前景, 但主要挑战在于如何仅使用正常图像来建模, 并对未知的异常进行识别与定位。基于

收稿日期: 2025-03-19

基金项目: 四川省科技计划项目(2023YFG0305); 成都信息工程大学科研基金项目(KYTZ202156)

作者简介: 许昌源(2000-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为工业图像异常检测; 贾可(1983-), 男, 博士, 成都信息工程大学计算机学院副教授、硕士生导师, 研究方向为图像处理与工业异常检测; 金治成(1998-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为异常检测; 李涵鑫(1999-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为目标检测; 王文润(2001-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为异常检测; 周记(1999-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为行为分析。本文通讯作者: 贾可。

传统工业视觉的异常检测方法,比较正常图像与异常图像的结构差异进行异常判定,处理方法包含阈值分割、边缘检测、傅里叶变化等。Wang 等^[1]提出基于傅里叶图像重建的磁铁表面切割缺陷检测方法,对图像进行二维离散傅里叶变换得到频谱图像。其中,缺陷在频谱中表现为亮线,正常图像频谱无显著亮线。Prasitmeeboon 等^[2]提出基于支持向量机与颜色直方图的缺陷检测流程,但基于灰度或颜色特征的方法易受光照不均或阴影干扰,导致误检。Zhang 等^[3]提出一种结合自适应阈值分割与形态学重建的方法,检测铝合金轮毂 X 射线图像中的缺陷,并在不同的环境中对不同类型的缺陷进行分割,以验证该方法的鲁棒性。然而,传统异常检测方法依赖于手工设计的特征,需人工设计特征提取规则,不同场景需重新调整参数,难以适应复杂多变的工业环境。

在基于深度学习的工业图像异常检测中,AnoGAN、f-AnoGAN 等异常检测算法,基于生成对抗网络 (Generative Adversarial Networks, GAN) 模型学习图像的正常特征表示,但模型无法对缺陷部分进行重构来进行缺陷检测^[4,5]。Liang 等^[6]提出一种全频率信道选择重构网络 OCR-GAN,针对异常类图像的频率分布与正常类图像存在显著差异的问题,从频率角度处理异常检测任务,将重建过程建模为全频率图像恢复的组合。Yang 等^[7]使用多记忆增强模块生成的特征来重建正常区域,使用图像轮廓特征重建正常区域的轮廓。Rudolph 等^[8]基于标准化流的 DifferNet,在多个尺度上使用基于归一化流的图像特征密度估计来检测图像中的缺陷,用较少的训练样本也能产生良好的检测效果。张兰尧等^[9]提出 ValidFlow,通过可分离卷积减少网络计算量,使用卷积注意力模块优化检测性能。

在基于特征嵌入的方法中,通常使用的预训练模型包括 VGG、ResNet、EfficientNet、WideResNet、ViT 等^[10-14]。Defard 等^[15]提出的 PaDiM,利用预训练的卷积神经网络来提取图像特征,每个图像块的位置会被映射为一个嵌入向

量,正常类的特征分布通过高斯分布来建模。同时,采用随机降维方法简化嵌入向量维度,降低计算复杂度,提升检测速度。Roth 等^[16]提出基于记忆库的异常检测方法 PatchCore,将单类正常图像的 Patch 特征存储在记忆库中,比较输入图像的特征与记忆库中的特征得到异常得分,记忆库的特征是决定基于特征嵌入的图像异常检测算法性能优劣的关键之一。

然而,直接使用预训练模型提取正常图像的特征,既忽视了异常检测数据集和 ImageNet 数据集的分布差异,又忽视了自然图像分类与是否包含缺陷的二分类任务的不同之处,因此预训练模型提取的特征与图像异常检测所需的特征存在目标域偏差。为了解决上述问题,本文主要贡献为:①使用预训练模型在工业数据集上进行粗粒度分割训练,优化模型对工业异常检测任务的特征提取能力,同时通过区域块级别的缺陷预测,引导特征学习关键区域信息;②为了避免模型在正常类别特征与异常类别特征边界模糊的情况下,无法正确区分类别而引起的漏检、误检情况,引入附加角裕度损失函数 (Additive Angular Margin Loss, AAM-Softmax),在模型训练中对目标类别特征角度增加一个间隔值来减小相同类别特征的间距,以贴合异常检测任务的要求。

1 本文方法

本文的图像异常检测方法可分为以下两个阶段:第一阶段由特征提取器、特征适配器、分类器组成,通过粗粒度分割训练实现迁移学习,得到适应于工业图像异常检测的特征提取器;在第二阶段的异常检测训练阶段中,提取每个类别的正常图像块特征并保存在记忆库中,在推理阶段中,提取输入图像特征与记忆库中的最近邻特征,并计算得到异常分数与异常定位图。模型总体结构如图 1 所示。

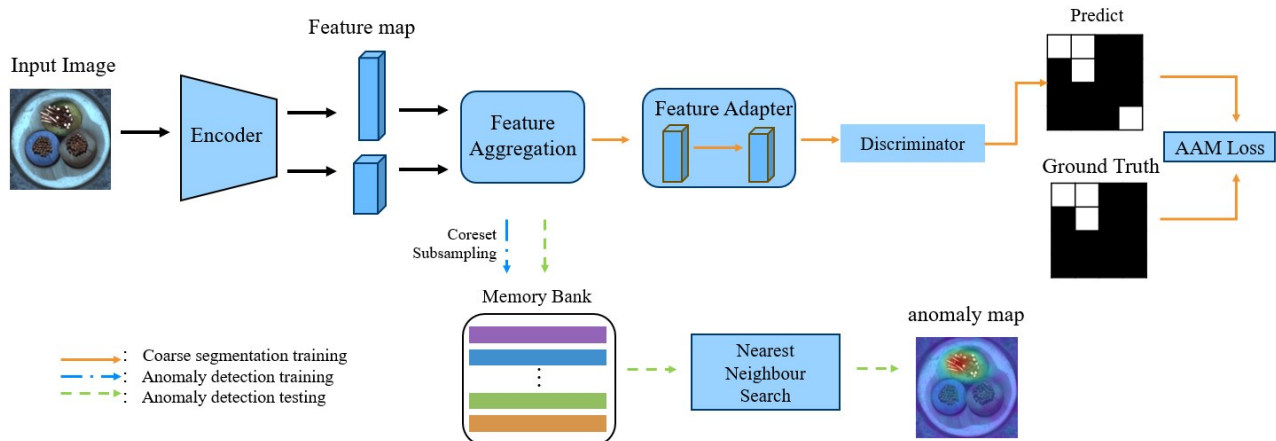


Fig. 1 Overall structure of the model

图 1 模型总体结构

1.1 特征提取器与特征融合

通常情况下,输入图像经过特征提取器,从预训练网络 ϕ 的不同层中输出特征。由于主干网络深层特征的语义信息偏向于原自然图像分类需求,与异常检测无关,浅层特征缺少整图的语义信息。然而,中间层特征保留了更多细节和空间信息,对微小变化更敏感,适合检测局部异常,因此在后续流程中使用主干网络的中间层特征进行输出。

对于输入图像 x_i ,定义在预训练网络的第 j 层输出的特征图为 $\phi_{i,j} \in R^{c \times h \times w}$, c, h, w 代表特征图的通道数、高度、宽度, $\phi_{i,j} = \phi_j(x_i)$ 。本文使用 $\phi_{i,j}^{h^*,w^*}$ 表示特征图中位于 (h^*, w^*) 处的特征,其中 $h^* \in \{1, \dots, h\}$, $w^* \in \{1, \dots, w\}$ 。为了增加感受野的大小和对小空间偏差的鲁棒性,对 $\phi_{i,j}^{h^*,w^*}$ 周围特征进行聚合,周围的特征向量如式(1)所示。

$$\mathcal{N}_p^{h^*,w^*} = \left\{ (m, n) \mid m \in [h^* - \lfloor p/2 \rfloor, \dots, h^* + \lfloor p/2 \rfloor], \right. \\ \left. n \in [w^* - \lfloor p/2 \rfloor, \dots, w^* + \lfloor p/2 \rfloor] \right\} \quad (1)$$

式中: p 代表Patch的大小。

式(2)为 $\phi_{i,j}^{h^*,w^*}$ 中的原始特征经过位于Patch区域的邻域特征聚合后的结果。

$$\phi_{i,j}(\mathcal{N}_p^{h^*,w^*}) = f_{agg} \left(\left\{ \phi_{i,j}^{m,n} \mid (m, n) \in \mathcal{N}_p^{h^*,w^*} \right\} \right) \quad (2)$$

式中: f_{agg} 代表特征聚合操作;特征提取器输出的各层特征图 $\phi_{i,j}$ 经过邻域特征聚合后可得到新的特征图。

后续,为了充分融合不同层特征的多尺度信息,使用双线性插值对齐特征图大小,再使用自适应平均池化操作进行特征融合得到特征向量。

1.2 特征适配器与分类器

本文使用特征提取器提取到输入图像相应的Patch特征后,在特征适配器中使用多层感知器来增强特征表达能力。当图像中的正常区域特征与异常区域特征相似,而在低维空间具有重叠或区分度不高时,通过多层感知器的非线性变换将Patch特征映射到更具区分性的空间,有助于得到更具判别性的特征,增强特征在高维空间的区分度。

为了更好地对异常图像中的局部缺陷进行分类训练。将真值图通过最大池化降采样至特征图大小,与特征提取器得到的Patch特征形成一一对应关系,每个Patch特征对应一个区域块的标签,本文称之为粗粒度的分割训练。相较于逐像素的分割训练,所提分割方法使模型对原图中每个区域的图像进行独立的二分类,专注于学习局部特征,以提升对局部缺陷的检测能力。

1.3 损失函数

Softmax Loss不特别强调类别之间的角度或距离,而是根据类别概率进行优化。因此,当待检测样本的特征处于决策边界附近时,模型可能难以学习到具有足够判别力的特征,从而导致漏检或误检。

$$L_1 = -\log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^N e^{W_j^T x_i + b_j}} \quad (3)$$

式中: W_{y_i} 代表 y_i 类的权重向量; W_j 代表 j 类的权重向量; x_i 代表样本 i 的特征向量; b_{y_i} 代表真实类别 y_i 的偏差; b_j 代表第 j 类的偏差。

AAM-Softmax(Additive Angular Margin loss)是一种基于角度的损失函数。首先,将偏差置0,由于 $W_j^T x_i = \|W_j\| \|x_i\| \cos \theta_j$, $W_j \in R^d$ 代表权重的第 j 列权重,通过L2-normalization将 $\|x_i\|$ 与 $\|W_j\|$ 归一化为1,并将 $\|x_i\|$ 缩放为 s ,使特征分布在半径为 s 的超球面上,如式(4)所示。

$$L_2 = -\log \frac{e^{s \cos \theta_{y_i}}}{e^{s \cos \theta_{y_i}} + \sum_{j=1, j \neq y_i}^N e^{s \cos \theta_j}} \quad (4)$$

同时,将角度 θ_{y_i} 加上一个角度间隔值 m ,进一步拉大不同类样本特征的间隔,如式(5)所示。综上,AAM-Softmax通过余弦相似度衡量特征间的相似性,并使特征分布在半径为 s 的球面上,给目标类别的角度增加一个固定的角度间隔值 m 来增强类的可分性。

$$L_3 = -\log \frac{e^{s \cos(\theta_{y_i} + m)}}{e^{s \cos(\theta_{y_i} + m)} + \sum_{j=1, j \neq y_i}^N e^{s \cos \theta_j}} \quad (5)$$

实验表明,AAM-Softmax能有效增强同类样本在特征空间中的聚集性,提升不同类别之间的分离度,从而在高维特征空间中形成更紧凑、明确的类别边界,以此提升图像异常检测的性能。

1.4 异常检测阶段

在异常检测训练阶段,先将每个类别的正常图像 X_i 经过特征提取器得到输出的特征 $\phi_{i,j}$,再将邻域特征聚合后得到的特征向量 $S_{i,j}, S_{i,j+1}$ 存储在记忆库 M 中,如式(6)所示。

$$M = \bigcup_{x_i \in X_i} f_{pooling}(S_{i,j}, f_{resize}(S_{i,j+1})) \quad (6)$$

为了减小存储内存并提升推理速度,在训练阶段对每类图像的Patch特征使用Mini Batch K-Means聚类,将簇内样本到中心的平均距离的倒数作为簇密度,每个簇的采样权重取簇密度的倒数,基于采样权重设置采样率,对每个簇进行贪婪核心集子采样。

在推理阶段,将输入图像的特征与子集中的特征通过欧氏距离计算得到异常得分,并将异常得分通过双线性插值匹配原图大小,通过高斯平滑后得到异常定位图。

2 实验结果与分析

2.1 数据集

本文使用工业检测领域的MVTec AD数据集,包含5类纹理类图像、10类物体类图像,如地毯、皮革、螺母、玻璃瓶等^[17]。其中,每个类别的训练集包含正常图片;测试集

包含缺陷图片、正常图片及缺陷图片的像素级标注;缺陷类别包括裂纹、孔洞、污渍、划痕等。

在粗粒度分割训练中,需要提供模拟的异常图片进行分类训练。本文参考 MemSeg 中的模拟异常图的生成方法,将训练集中正常图片的灰度图经过二值化后提取待检测物体区域,使用二维柏林噪声模拟零散分布的缺陷位置,使用纹理数据集的图像生成纹理缺陷。

首先,对原始图像进行包括调整对比度、亮度、锐度、色彩饱和度,反转图像颜色、旋转图像等多种变化,对增强后的图像划分网格,打乱及重新排列生成结构缺陷;其次,从两类缺陷制作方法中随机选取一类模拟的缺陷,叠加在待检测物体上得到模拟的异常图片。正常图像、模拟异常图像与真实的异常图像如图 2 所示。

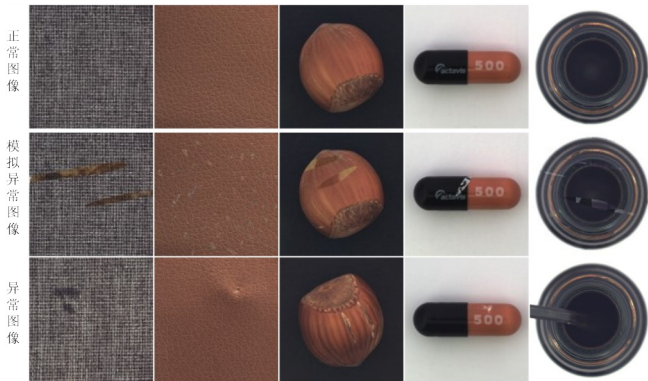


Fig. 2 Comparison results of normal images, simulated abnormal images, and abnormal images

图2 正常图像、模拟异常图像与异常图像的比较结果

2.2 模型参数与实验环境

本文在对预训练模型进行迁移训练时,选择 EfficientNet-B5 作为主干网络,其在 ImageNet 上的分类指标更好且各层的特征维度更低,有利于减少生成记忆库的大小。同时,选择 stage 4、stage 5 层进行特征输出,训练阶段使用 Adam 优化器,学习率为 1.0×10^{-4} ,AAM-Softmax 的学习率为 1.0×10^{-3} ,超参数 $s=30$ 、 $m=0.5$ 。

在数据预处理阶段,将所有图片缩放为 256×256 大小,使用中心裁剪法将图像裁剪为 224×224 ,对每个类训练 150 个 epoch,异常检测的测试阶段 BatchSize=2。实验环境配置如表 1 所示。

Table 1 Experimental environment configuration

表1 实验环境配置

实验环境	环境配置
CPU	Intel(R)Xeon(R)CPU E5-2678 v3 @ 2.50 GHz
GPU	NVIDIA GeForce8*GTX2080Ti
操作系统	Linux
深度学习框架	Pytorch 1.13.1
Python	3.9

2.3 评估指标

实验使用 AUROC (area under the receiver operating characteristic curve)作为评价指标,描述模型在二分类中不

同阈值下的分类表现,适用于图像异常检测中类别不平衡的情况。ROC 曲线中,FPR(False Positive Rate)为横坐标、TPR(True Positive Rate)为纵坐标,通过改变分类分数阈值,绘制 TPR、FPR 的散点图得到的曲线称为 ROC 曲线,如式(7)、式(8)所示。

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

式中:TP(True Positive)为真阳性,代表模型正确预测正类的样本数量;FP(False Positiv)为假阳性,代表模型错误预测正类的样本数量;TN(True Negative)为真阴性,代表模型正确预测负类的样本数量;FN(False Negativ)为假阴性,代表模型错误预测负类的样本数量。

AUROC 是指 ROC 曲线下的面积,取值范围为 0-1。当 AUROC=1 时,至少存在一个阈值使分类器完全准确地区分正类、负类样本。

2.4 比较方法

本文方法与主流的图像异常检测算法 DRAEM、KDA D、Cutpaste、PaDiM、PatchCore,在 MVTec AD 数据集上各类别的图片级、像素级 AUROC 值如表 2 所示^[18-21]。

Table 2 Detection results of different algorithms in the MVTec AD dataset

表2 不同算法在MVTec AD数据集中的检测结果 (%)

	DRAEM	KDA D	CutPaste	PaDim	PatchCore	Ours
Carpet	96.9/97.5	80.3/95.5	93.1/98.3	99.8/99.0	98.1/97.6	99.5/97.9
Grid	99.9/99.7	75.3/89.4	99.9/97.5	96.7/97.1	98.4/96.8	99.2/98.2
Leather	100/99.0	92.3/98.1	100/99.5	100/99.0	100/98.5	100/97.4
Tile	100/99.2	91.5/80.2	93.4/90.5	98.1/94.1	99.0/97.3	99.1/98.6
Wood	99.5/95.5	94.5/85.3	98.6/95.5	99.2/94.1	97.8/96.4	99.6/97.1
Bottle	98.0/ 99.1	99.2/95.7	98.3/97.6	99.9/98.2	100/97.4	100/97.8
Cable	90.9/95.2	90.3/80.2	80.6/90.0	92.7/96.7	99.0/ 98.9	99.5/97.4
Capsule	91.3/88.1	81.4/95.2	96.2/97.4	91.3/ 98.6	98.3/96.3	97.9/98.5
Hazelnut	100/99.7	98.8/95.0	97.3/97.3	92.0/98.1	100/94.5	100/97.8
Metal Nut	100/99.6	77.1/83.3	99.3/93.1	98.7/97.3	100/97.8	100/96.9
Pill	97.1/97.3	84.4/89.9	92.4/95.7	93.3/95.7	94.7/91.6	96.8/97.6
Screw	98.7/99.3	82.4/95.8	86.3/96.7	85.8/98.4	93.8/97.9	97.6/97.4
Toothbrush	100/97.3	97.1/95.5	98.3/98.1	96.1/ 98.8	99.7/93.8	99.4/97.1
Transistor	91.7/85.2	84.9/75.9	95.5/93.0	97.4/ 97.6	99.2/91.3	99.6/96.2
Zipper	100/99.1	93.7/95.3	99.4/ 99.3	90.3/98.4	99.5/96.7	99.3/98.3
Average	97.6/96.7	88.2/90.0	95.2/96.0	95.5/97.4	98.5/96.2	99.2/97.6

DRAEM 包含重建子网络与判别子网络,重建网络采用编码器—解码器结构,将模拟异常图像重建为正常图像来学习正常样本的特征;判别网络输入为原始图像与重建图像,通过 U-Net 网络直接输出像素级的异常热图。

KDA D 是一种基于多分辨率知识蒸馏的异常检测与定位方法,通过损失函数将输入梯度处理为热图,结合高斯滤波和形态学操作去除噪声,以准确定位异常区域。

Cutpaste 在进行异常检测任务时,通过自监督学习构建单类的分类器,使用 ImageNet 的预训练特征进行迁移学习。异常样本是在原图上裁剪不同大小的矩形区域,并随

机粘贴在原图上所构成。然而,该方法生成的缺陷纹理只来自原图且只能生成矩形缺陷,无法很好地模拟实际缺陷。

Padim 使用马氏距离计算图像异常分数,通过协方差矩阵对特征进行归一化。然而,该方法的计算复杂度较高,对样本量的要求较高。

PatchCore 使用预训练模型的中间层特征生成补丁特征,将其存储为记忆库。测试时,计算测试图像中每个补丁特征与内存库中最近邻的距离,并将其作为异常分数。

由表 2 可知,本文方法在图片级 AUROC 中的 7 个类别中取得了最高分,在像素级 AUROC 中的两个类别中取得了最高分,在平均分指标上最优,相较于使用 EfficientNet-B5 网络的 Patchcore 算法,在 I-AUROC、P-AUROC 上提升了 0.7%、1.4%。

本文算法通过第一阶段的粗粒度分割训练减小同类特征间的类间距,从而使第二阶段提取的正常图像特征更紧密,实验结果如表 3 所示。由此可知, PatchCore 算法对于 Carpet 类图像,生成的记忆库维度为 (21 952, 1 024)。为了减少内存占用并提升检测速度,根据聚类后的簇密度来设置不同簇的核心集来降采样率,选择 EfficientNet-B5 作为主干网络,最终生成的记忆库维度为 (17 829, 128),以降低记忆库中的特征条数和特征维度,减少内存占用情况,提升推理速度,使算法更贴合实际工业生产中异常检测的实时性要求。此外,所提模型的图像级 AUROC、像素级 AUROC 指标更优。

Table 3 Comparison results between the proposed model and the benchmark model (Carpet class)

表 3 本文模型与基准模型的比较结果 (Carpet 类)

	WideResNet50	ours
记忆库维度	21 952, 1 024	17 829, 128
推理速度/ms	58.4	35.7
I-AUROC/P-AUROC/%	98.1/97.6	99.5/97.9

图 3 为异常检测的缺陷定位图。本文以药丸、电缆、胶囊、木材、瓷砖类为例,每列分别代表输入图像、标签图、预训练模型的缺陷定位图、粗粒度分割训练后模型的缺陷定位图。由此可见,针对药丸类尺寸较小的颜色异常问题,本文模型能从同颜色的斑点中准确找出缺陷位置;在电缆类中待检物品结构多样,本文模型相较于基准模型的误检情况更少;模型能准确定位胶囊类的细微划痕和木材类水滴,说明该方法对不同尺度缺陷的检测泛化能力更强;对于瓷砖类的细长缺陷,本文模型的检测图在缺陷部分的红色更多,代表异常得分更高,缺陷区域的特征与正常区域的特征距离更远。

综上,经过微调后的模型在异常检测任务中对缺陷的定位更准确,对正常区域的误判情况更少。

2.5 消融实验

EfficientNet-B5 网络通过堆叠不同的 MB block 组成,对于原始预训练模型,浅层特征包含更多边缘和纹理信

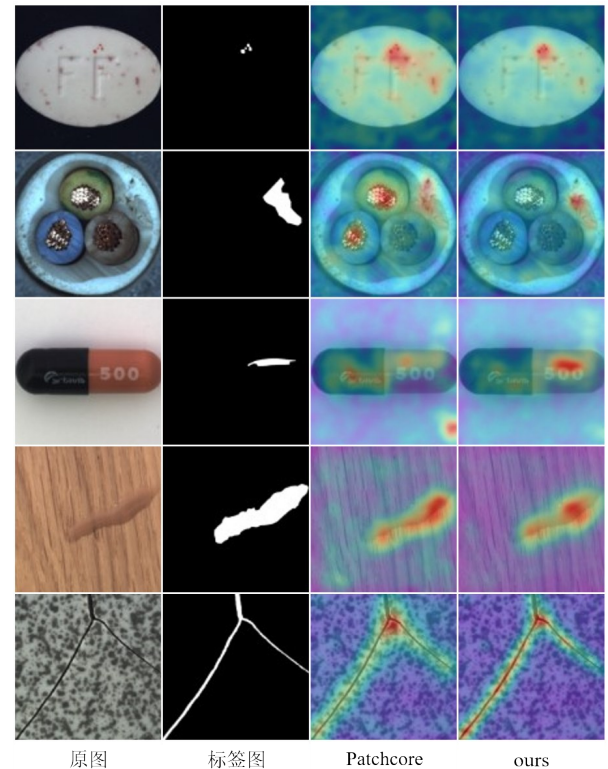


Fig. 3 Defect location map

图 3 缺陷定位图

息,深层特征偏向于原始的图像分类任务,具有更多的语义信息。为了保证特征通用性,遵循了 PatchCore 中的建议,使用中间层特征输出。

表 4 为对不同层特征输出进行粗粒度分割训练后得到的主干网络参数,在后续图像异常检测阶段的检测效果。由此可知,仅使用单层特征时特征信息单一,检测结果较差。尽管,使用 Layer 3、Layer 4、Layer 5 的多尺度融合特征能得到更好的像素级 AUROC 指标,但也会增加模型复杂度、记忆库大小,降低推理速度。为此,本文最终采用 Layer 4 与 Layer 5 层特征进行输出。

Table 4 Detection results using features from different layers

表 4 不同层特征的检测结果 (%)

	纹理类	结构类	所有类
Layer 4	99.0/97.6	98.7/96.9	98.9/97.4
Layer 5	98.6/97.4	98.3/96.6	98.4/97.1
Layer 4+5	99.5/97.9	99.0/97.5	99.2/97.6
Layer 3+4+5	99.2/98.2	98.6/97.9	98.7/98.1

本文使用 PatchCore 算法做为 baseline,采用与本文相同的主干网络的预训练模型进行比较,以突显公平性,消融实验结果如表 5 所示。其中,ILT (Image Classification Training) 代表使用训练集的图像进行分类训练来微调特征提取网络,训练采用交叉熵损失函数;CST (Coarse-grained Segmentation Training) 代表使用训练集的图像进行粗粒度分割训练来微调特征提取网络;CGS-FT (Coarse-grained Segmentation Fine-Tuning) 为在微调阶段使用 AAM-Softmax 作为损失函数的粗粒度分割训练微调特征

提取网络,在异常检测推理阶段通过聚类对每个簇设置不同的采样率进行核心集降采样,以形成记忆库进行异常检测。由此可知,图像级别的分类训练会导致模型检测指标退化,原因为模型注重于图像分类,特征不适用于异常检测任务。粗粒度分割训练通过引导模型进行 Patch 级别的缺陷分类,并使用 AAM-Softmax 减小同类特征的类间距,能在图像级、像素级 AUROC 上得到提升,证实了特征提取器的改善效果明显。

Table 5 Ablation experiment results
表 5 消融实验结果 (%)

模型	I-AUROC/P-AUROC/%
baseline	98.5/96.2
baseline+ILT	97.4/95.4
baseline+CST	98.9/96.9
CGS-FT	99.2/97.6

3 结语

在基于记忆库方法的异常检测中,预训练模型虽然具有良好的特征提取能力,但对工业图像异常检测任务而言还不够。为此,本文通过粗粒度分割训练方法有效消除记忆库中特征的域偏差,结合下游异常检测任务的要求拉大不同类特征的间距,缩小同类特征间距。在 MVTec AD 数据集上的实验表明,所提模型的图像级、像素级 AUROC 检测精度达到 99.2%、97.6%。

此外,相较于基准模型,本文将主干网络更换为更轻量的 EfficientNet-B5,并根据聚类后簇的密度设置采样率进行贪婪核心集降采样,降低生成的记忆库特征维度。例如,Carpet 类的推理阶段检测速度提升了约 40%,算法更符合工业场景下的实时异常检测需求。后续,可在多尺度的缺陷检测效果和更轻量的检测模型方面进行深入探索与改良。

参考文献:

[1] WANG F, ZUO B. Detection of surface cutting defect on magnet using fourier image reconstruction[J]. Journal of Central South University, 2016, 23(5): 1123-1131.

[2] PRASITMEEBOON P, YAU H. Defect detection of particleboards by visual analysis and machine learning[C]// 2019 5th International Conference on Engineering, Applied Sciences and Technology, 2019: 1-4.

[3] ZHANG J, GUO Z, JIAO T, et al. Defect detection of aluminum alloy wheels in radiography images using adaptive threshold and morphological reconstruction[J]. Applied Sciences, 2018, 8(12): 2365.

[4] 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, 2017: 146-157.

[5] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks[J]. Medical Image Analysis, 2019, 54: 30-44.

[6] LIANG Y, ZHANG J, ZHAO S, et al. Omni-frequency channel-select-

tion representations for unsupervised anomaly detection[J]. IEEE Transactions on Image Processing, 2023, 32: 4327-4340.

[7] YANG Y, XU X Y, ZHANG L N, et al. Industrial surface anomaly detection based on multi-memory enhancement module and image contour reconstruction[J/OL]. Computer Engineering and Applications, 2024: 1-17. <https://link.cnki.net/urlid/11.2127.TP.20241025.1428.002>.
杨曜,许湘云,张琳娜,等.基于多记忆增强模块及图像轮廓重建的工业表面异常检测[J/OL]. 计算机工程与应用, 2024: 1-17. <https://link.cnki.net/urlid/11.2127.TP.20241025.1428.002>.

[8] RUDOLPH M, WANDT B, ROSENHAHN B. Same same but different: semi-supervised defect detection with normalizing flows[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 1907-1916.

[9] ZHANG L Y, CHEN X L, ZHANG D M, et al. Validflow: unsupervised image defect detection based on normalized flow[J]. Data Acquisition and Processing, 2023, 38(6): 1445-1457.
张兰尧,陈晓玲,张达敏,等. ValidFlow: 基于标准化流的无监督图像缺陷检测[J]. 数据采集与处理, 2023, 38(6): 1445-1457.

[10] KAREN S, ANDREW Z. Very deep convolutional networks for large-scale image recognition[DB/OL]. <https://arxiv.org/abs/1409.1556>.

[11] HE K, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[12] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks[C]// International Conference on Machine Learning, 2019: 6105-6114.

[13] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[DB/OL]. <https://arxiv.org/abs/1605.07146>.

[14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. <https://iclr.cc/virtual/2021/oral/3458>.

[15] DEFARD T, SETKOV A, LOESCH A, et al. Padim: a patch distribution modeling framework for anomaly detection and localization[C]// International Conference on Pattern Recognition, 2021: 475-489.

[16] ROTH K, PEMULA L, ZEPEDA J, et al. Towards total recall in industrial anomaly detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 14318-14328.

[17] BERGMANN P, FAUSER M, SATTLEGGGER D, et al. MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9584-9592.

[18] ZAVRTANIK V, KRISTAN M, SKOČAJ D. Draem—a discriminatively trained reconstruction embedding for surface anomaly detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 8330-8339.

[19] SALEHI M, SADJADI N, BASELIZADEH S, et al. Multiresolution knowledge distillation for anomaly detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14902-14912.

[20] LI C L, SOHN K, YOON J, et al. Cutpaste: self-supervised learning for anomaly detection and localization[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 9664-9674.

[21] DEFARD T, SETKOV A, LOESCH A, et al. Padim: a patch distribution modeling framework for anomaly detection and localization[C]// International Conference on Pattern Recognition, 2021: 475-489.

(责任编辑:刘嘉文)