

面向工业图像异常检测的连续密集标准化流模型

张邹铨, 张辉, 吴天月, 陈天才

引用本文

张邹铨, 张辉, 吴天月, 陈天才. 面向工业图像异常检测的连续密集标准化流模型[J]. 计算机科学, 2023, 50(12): 212-220.

ZHANG Zouquan, ZHANG Hui, WU Tianyue, CHEN Tiancai. [Continuous Dense Normalized Flow Model for Anomaly Detection in Industrial Images](#) [J]. Computer Science, 2023, 50(12): 212-220.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[使用RAP生成可传输的对抗网络流量](#)

Generate Transferable Adversarial Network Traffic Using Reversible Adversarial Padding
计算机科学, 2023, 50(12): 359-367. <https://doi.org/10.11896/jsjcx.221000155>

[基于迭代非对称盲点网络的低剂量CT重建算法](#)

Low-dose CT Reconstruction Algorithm Based on Iterative Asymmetric Blind Spot Network
计算机科学, 2023, 50(12): 221-228. <https://doi.org/10.11896/jsjcx.230300014>

[基于双空间共轭自编码器的多时相高光谱异常变化检测](#)

Multi-temporal Hyperspectral Anomaly Change Detection Based on Dual Space Conjugate Autoencoder
计算机科学, 2023, 50(12): 175-184. <https://doi.org/10.11896/jsjcx.221100092>

[基于特征融合与边界修正显著性目标检测](#)

Feature Fusion and Boundary Correction Network for Salient Object Detection
计算机科学, 2023, 50(12): 166-174. <https://doi.org/10.11896/jsjcx.221100203>

[基于Transformer特征融合的时间序列分类网络](#)

Transformer Feature Fusion Network for Time Series Classification
计算机科学, 2023, 50(12): 97-103. <https://doi.org/10.11896/jsjcx.221100112>

面向工业图像异常检测的连续密集标准化流模型

张邹铨¹ 张辉² 吴天月¹ 陈天才³

1 长沙理工大学电气与信息工程学院 长沙 410000

2 湖南大学机器人学院 长沙 410000

3 湖南大学电气与信息工程学院 长沙 410000

(zouquan_zhang@163.com)

摘要 工业产品表面异常检测是生产制造中不可或缺的环节。在实际工业生产中,普遍存在异常样本所占比例低且未知异常复杂多变等现象,进而造成在小样本数据集上过拟合、泛化能力不佳等一系列负面影响。近年来,标准化流思想为基于深度学习的工业图像异常检测带来了新途径,但标准化流的固有架构易导致模型表达能力不足。针对上述难点,提出了一种面向工业图像异常检测的连续密集标准化流模型。首先,设计一种基于对比学习的特征提取网络预训练策略,将模拟异常数据和少量真实异常数据加入对比学习任务中,并训练特征骨干网络 AlexNet 拉近或拉远特定样本间的距离;其次,设计连续密集标准化流模型,采用可逆变换的复合架构来构造密集流模块,增强生成式模型对分布的拟合能力。在 MVTec AD 和 Magnetic Tile Defects 以及自制的工业布匹数据集上的实验结果表明,与其他的异常检测模型相比,所提方法在 3 个数据集上的检测性能达到了最优或次优的水平。

关键词: 工业图像检测;异常检测;深度学习;标准化流;对比学习

中图分类号 TP391

Continuous Dense Normalized Flow Model for Anomaly Detection in Industrial Images

ZHANG Zouquan¹, ZHANG Hui², WU Tianyue¹ and CHEN Tiancai³

1 School of Electrical & Information Engineering, Changsha University of Science and Technology, Changsha 410000, China

2 School of Robotics, Hunan University, Changsha 410000, China

3 School of Electrical & Information Engineering, Hunan University, Changsha 410000, China

Abstract Anomaly detection on the surface of industrial products is an indispensable link in manufacturing. In actual industrial production, there are common phenomena such as low proportion of abnormal samples and complex and changeable unknown abnormal, which in turn cause a series of negative effects such as overfitting and poor generalization ability on few-shot datasets. In recent years, the idea of normalized flow has brought a new approach to the field of industrial image anomaly detection based on deep learning, but the inherent architecture of normalized flow easily leads to insufficient model expressiveness. Aiming at the above difficulties, a continuous dense normalized flow model for industrial image anomaly detection is proposed. First, a feature extraction network pre-training strategy based on contrastive learning is designed, which involves simulated abnormal data and a small amount of real abnormal data in the contrastive learning task, and trains the feature backbone network AlexNet to narrow or widen the distance between specific samples. Secondly, a continuous dense normalized flow model is designed, and it uses a composite architecture of reversible transformation to construct a dense flow module to enhance the fitting ability of the generative model to the distribution. The experimental datasets include MVTec AD, Magnetic Tile Defects and self-made industrial cloth datasets. Compared with other anomaly detection models, our method achieves optimal or sub-optimal detection performance on the three datasets, respectively.

Keywords Industrial image detection, Anomaly detection, Deep learning, Normalized flow, Contrastive learning

到稿日期:2022-10-23 返修日期:2023-03-17

基金项目:国家重点研发计划(2021ZD0114503);国家自然科学基金(61971071,62027810);湖南省杰出青年科学基金(2021JJ10025);湖南省研究生科研创新项目(CX20210797)

This work was supported by the National Key R&D Program of China(2021ZD0114503), National Natural Science Foundation of China(61971071,62027810), National Science Fund for Distinguished Young Scholars of Hunan Province, China(2021JJ10025) and Postgraduate Scientific Research Innovation Project of Hunan Province(CX20210797).

通信作者:张辉(zhanghuihy@126.com)

1 引言

工业异常指与预期结果差异大,不同于常规的不良数据实例。为了保证工业的生产安全和产品质量,异常检测在工业界和学术界受到广泛关注。工业图像异常检测一直是制造业中的重要环节,是高端制造业智能化发展的关键任务。而随着科技日新月异的进步,工业领域中引入了诸多优质的先进制造技术,衍生出种类繁多、特点各异的工业生产手段,这使得异常变得更为复杂多样。与正常数据相比,工业异常出现频率较低^[1]。

在实际生产中,大多数场景依旧使用传统的人工检测技术作为工业解决方案,该方法要求检测人员有丰富的生产经验积累。但其在复杂、高强度的工作环境下存在异常类别难评定、检测效果不稳定等诸多问题,制约了传统方法在高度智能化工业异常检测任务上的应用范围和实际价值。

近年来,深度学习的跨越式发展应用于各种大规模的工业检测场景,提高了工业生产效率和生产的自动化程度,同时减少了人力成本。但以往基于深度学习的工业图像异常检测方法多采用有监督的学习方式,异常样本被视作训练目标^[2]。这种常规的有监督深度学习算法不可避免地需要大量手动标记的异常数据,与工业异常情况不可预知、难于归纳的实际问题相悖,导致面向特定缺陷所设计的检测方法对未知异常失效。

本文围绕基于对比学习的生成式异常检测模型展开研究,并提出了一种面向工业图像异常检测的连续密集标准化流模型。第2章介绍了相关工作,着重从自动编码器、生成对抗网络以及标准化流这3种不同的生成式模型进行阐述;第3章对连续密集标准化流模型所涉及的相关知识进行说明;第4章介绍了本文实验设置及步骤,分析对比实验结果,并通过消融实验验证模型中各模块有效性;最后总结全文并展望未来。

针对实际工业检测应用中存在的异常样本少、标注成本高,流模型在通用的工业异常检测任务中表达能力受限,以及正常与异常图像数据在位置与形态上相似度高的问题,提出了一种基于特征增强对比的密集流模型,本文的主要工作概括如下:

(1)对于异常数据,尽管工业界已知的缺陷样本不足的现象十分常见,但依旧能让模型“浏览”几张异常,使得模型对异常的定义更清晰。因此提出针对特征提取模块的基于对比学习的预训练策略,仅使用模拟异常样本和少量的真实异常样本来检测所有的未知异常,提高模型的判别能力。

(2)引入标准化流的思想,通过标准化流模型获得的概率分布差异来检测异常,构建密集仿射耦合模块,以增强流模型的表情能力。

(3)该模型在公共的MVTec AD数据集^[3]、MTD(Magnetic Tile Defects)数据集^[4]以及自制的布匹数据集上检测性能较高,验证了本文方法的有效性与可靠性。

2 相关工作

近年来,国内外学者结合深度学习算法对工业界不同

产品的检测提供了许多无监督与弱监督下的检测方法。研究的范围主要集中在基于相似度的方法与生成式方法。基于相似度的异常检测方法利用深度卷积神经网络以获取样本深度高维特征正态信息,并采用基于统计学模型来鉴别样本分布^[5]。然而该方法在测试阶段十分缓慢。

现有的诸多异常检测方法都是基于生成式模型,譬如自动编码器。该模型包含编码及解码两个模块,编码层将输入的高维数据降维成低维特征,并将其映射到潜在空间分布,从而提取到最有效的特征,而解码器则是将低维特征进行图像重构还原^[6-7]。在异常检测任务中,自动编码器对输入的正常图像进行编解码,通过约束神经网络重构输出拟合原图片,以此学习正常图像分布模式。Yu等^[8]提出了一种全卷积自动编码器重构网络,通过设计结合结构性损失和平均绝对误差(Mean Absolute Error, MAE)损失的函数作为重构网络的目标函数,解决了传统自动编码器对不规则纹理表面缺陷检测效果较差的难题。Zavrtanik等^[9]运用重构子网络将模拟异常样本重构为正常样本,其后衔接一个判别子网络输出异常样本的异常分数。Kim等^[10]提出采用渐进式自动编码器和对比学习的思想来缓解异常样本未知而导致泛化性能差的问题。同时,经过与CutPaste^[11]增强图片对比后的异常检测模型可以更有效地重建正常图像。Ulger等^[12]采用变分推断的思想来训练神经网络,通过 β -VAE获得更好的解纠缠表示。但变分自动编码器及其变体往往在训练时学习近似概率模型的后验分布,无法直接优化边缘似然函数以形成闭式解。

另一种基于生成对抗网络(Generative Adversarial Network, GAN)的异常检测方法通常利用GAN能生成逼真图像的特点。在训练过程中,网络中的判别器与生成器分别试图最大化和最小化目标函数的最终取值,因此能重构出比自动编码器更清晰的图像。AnoGAN模型^[13]首次将GAN思想引入图像异常检测任务中,该模型在非异常数据上进行训练以学习流形。同一般的自动编码器一样,AnoGAN通过对比生成图像与原图像的差异来确定异常。然而该方法计算性能低下,时间开支较大,限制了其在工业场景中的适用性。Xing等^[14]在传统对抗自动编码器中引入了双对抗机制,即使用了两个具有不同功能的判别器,并进行两组对抗学习,使得模型具备更优的重构能力。然而,大多数生成对抗网络的训练过程较困难,不同的超参数选择都会对模型能力产生不同程度的影响。同时,对于概率生成模型而言,GAN的训练不考虑样本分布,更易出现模型崩解。

标准化流(Normalized Flows, NF)是一种基于可逆变换的概率生成模型。与上述生成模型不同的是,标准化流以概率分布作为基底,尝试将标准概率密度拟合到未知且复杂的概率密度上^[15]。因此该模型具备直接估计样本参数分布的能力,并允许两者相互映射及评估。NICE(Non-linear Independent Components Estimation)模型^[16]作为最原始的NF模型,为捕获未知的数据分布提供了一般解式,奠定了坚实的理论基础。为了扩张潜在语义空间,Huang等^[17]将多层耦合后的数据映射到增强状态空间中,缓解了可逆性所施加的架构限制而导致的表达能力弱的问题。但这种方法也增加了模型

复杂性,面对高分辨率图像及面向工业的实时监测任务无能为力。标准化流在异常检测领域具有很好的效果,Rudolph等^[18]提出了DifferNet,采用Real-NVP标准化流体系统构建异常检测的概率密度估计模块,该方法在MVTec AD数据集^[3]上平均接受者操作特征曲线下面积(Area Under the Receiver Operating Characteristic, AUROC)达到93.25%,但其使用的

是较为简单的耦合块结构,并未考虑标准化流模型在表达能力上的缺陷。

3 基于连续密集标准化流的工业图像异常检测模型

本文提出的连续密集标准化流模型主要分为特征提取模块及标准化流模块,具体网络结构如图1所示。

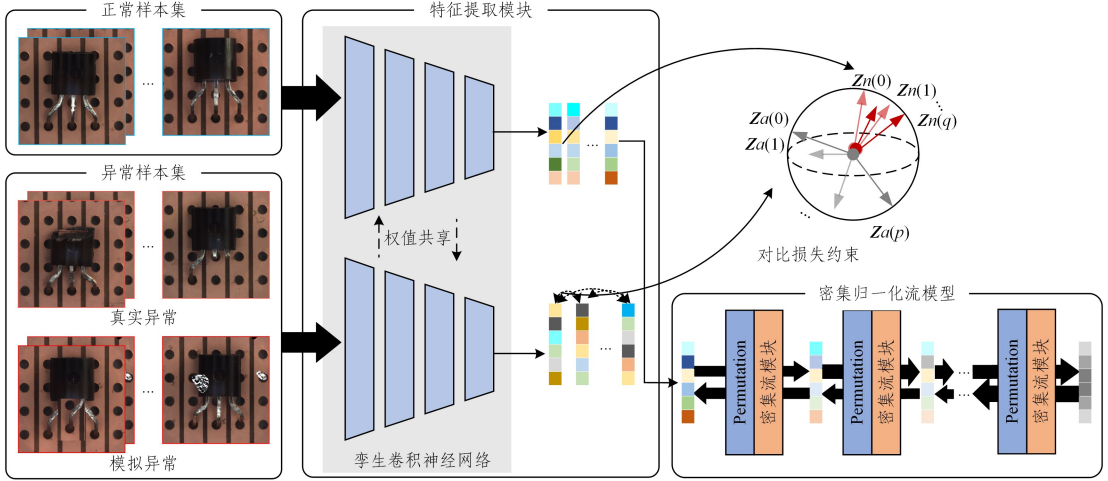


图1 网络总体框架图

Fig. 1 Overall network framework of the proposed model

特征提取模块作为预训练阶段的主体,将正常样本及混合异常样本输入到孪生网络中,经过特征级增广模块后执行对比学习任务。在训练阶段,将正常样本输入到特征提取模块的主干网络以获得多尺度特征,经过池化及维度拼接后输入到标准化流模块。模型通过连续有限的信息混合及密集耦合子模块,实现样本潜在空间分布和基础分布间概率密度的可逆转换。

3.1 基于对比学习的特征提取网络预训练策略

特征提取网络预训练策略可以为下游任务提供正样本的最佳特征表示,在特征空间上明确正负例间的决策边界,缓解工业场景下因异常复杂不可知而出现检测模型失效的问题。本文通过生成模拟异常数据来扩展异常数据,将工业正常样本、模拟异常样本及少量真实异常样本作为孪生网络的输入;对异常样本支路的高维向量进行特征级增广,打破模拟异常及真实异常间的壁垒;通过最小化对比损失函数更新网络参数,使得正样本在潜在空间中的距离尽可能缩小,同时拉远正负特征对的距离。因此,特征提取模块能够关注到更具区分性的深度语义信息,从而提高下游任务中异常检测的精度。

本文采用裁剪粘贴的方法对正常样本进行数据增强。具体而言,预先截取样本部分区域,改变该部分的对比度和亮度等信息,最后将其覆盖于原图像的随机位置,由此获得一类模拟异常。采用DRAEM模型^[9]中的生成伪异常方法,运用柏林噪声构造随机的像素级标签,将DTD数据集(Describable Textures Dataset)^[19]作为异常来源,对原正常样本的特定区域进行覆盖,获得第二种模拟异常。

定义两个分别包含工业正常样本及真假混合的异常样本数据集 $I_n = \{I_i^a\}_{i=1}^{N_n}$ 与 $I_a = \{I_j^a\}_{j=1}^{N_a}$, N_n 和 N_a 分别为正常和异常

样本图像总数, I_i^a 为第 i 张正常样本, I_j^a 为第 j 张异常图像。其中数据集 I_a 包含少量真实异常 I_{real} 以及模拟异常 I_{fake} ,将两者分别输入到孪生网络 F_n 和 F_a 中,将AlexNet^[20]作为基础的孪生骨干网络。为了扩充负样本,并使得混合异常数据在特征级上创造更难以鉴别的负例,对 F_a 中最后一个全连接层输出的高维数据进行特征增广,即:

$$\hat{z}^a = \tau * z_i^a + (1 - \tau) * z_j^a \quad (1)$$

其中, \hat{z}^a 为特征增广后的新负例,其由两个不同的原负特征 z_i^a 和 z_j^a 插值得来; τ 是一个 $[0, 1]$ 的随机标量。为了防止模型对正常数据丧失泛化能力,由此生成的特征数目不到原负特征数目的一半。

本文使用三元组损失变体作为对比学习任务中的目标函数,其将高维特征的欧氏距离映射到指数核空间中,以增加困难负样本的权重,对比损失函数 L_t 如式(3)所示:

$$d(\mathbf{u}, \mathbf{v}) = \left\| \frac{\mathbf{u}}{\|\mathbf{u}\|_2} - \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\|_2 \quad (2)$$

$$L_t = \frac{1}{C_N * K} \sum_{i=1, i \neq j}^N \sum_{j=1}^N \sum_{k=1}^K \max(e^{-\frac{d(z_i^a, z_j^a)}{\sigma}} - e^{-\frac{d(z_i^a, z_k^a)}{\sigma}} + m, 0) \quad (3)$$

其中, $d(\mathbf{u}, \mathbf{v})$ 为两个归一化特征向量之间的欧氏距离, σ 是调节指数核强度的有关超参数, m 为设定的距离阈值。

3.2 连续密集标准化流模型

3.2.1 基本原理

标准化流通过构造一种处处可逆的变换,将任意且未知的数据分布映射到具有明确定义分布的潜在空间中。由于标准化流的特性,两个分布得以等价变换,相关公式如下:

$$\mathbf{z} \sim p(\mathbf{z}) \quad (4)$$

$$\mathbf{x} = g(\mathbf{z}) = f^{-1}(\mathbf{z}) \quad (5)$$

其中, \mathbf{x} 为隐含向量表示; \mathbf{z} 为潜在表示; 而 $p(\mathbf{z})$ 则代表已知的潜在分布; g 为归一化流模型中的可逆变换函数, 由 k 个微分同胚映射器组合而成, 使得流模型能够不断地对概率分布进行微调, 同时弥补效果不佳的组件带来的负增益。在图 1 的密集归一化流模型中, 从左至右为标准化方向: 从目标分布中的样本 x 开始, 使用 $\mathbf{z}=f(\mathbf{x})$ 从 $p(\mathbf{z})$ 中获取样本, 从而得以计算 \mathbf{z} 的似然性。从右到左是构造目标概率密度方向: 从先验分布 $p(\mathbf{z})$ 中对 \mathbf{z} 进行采样, 利用函数 g 逆变换为期望样本。通过变量替换的思想, 使得未知的复杂分布与已知的标准分布之间的概率密度函数彼此关联, 给定数据点 x 处的密度 $p_x(x)$ 的相关推导如式(6)和式(7)所示:

$$|p_x(\mathbf{x})d\mathbf{x}| = |p_z(\mathbf{z})d\mathbf{z}| \quad (6)$$

$$p_x(\mathbf{x}) = p_z(\mathbf{z}) \left| \frac{d\mathbf{z}}{d\mathbf{x}} \right| = p_z(\mathbf{z}) \left| \frac{d\mathbf{z}}{d\mathbf{h}_k} \frac{d\mathbf{h}_k}{d\mathbf{h}_{k-1}} \dots \frac{d\mathbf{h}_1}{d\mathbf{x}} \right|$$

$$= p_z(\mathbf{z}) |\det J_f(\mathbf{z})| \quad (7)$$

式(7)由变量 \mathbf{x} 与 \mathbf{z} 的概率密度在所属空间中的积分为 1

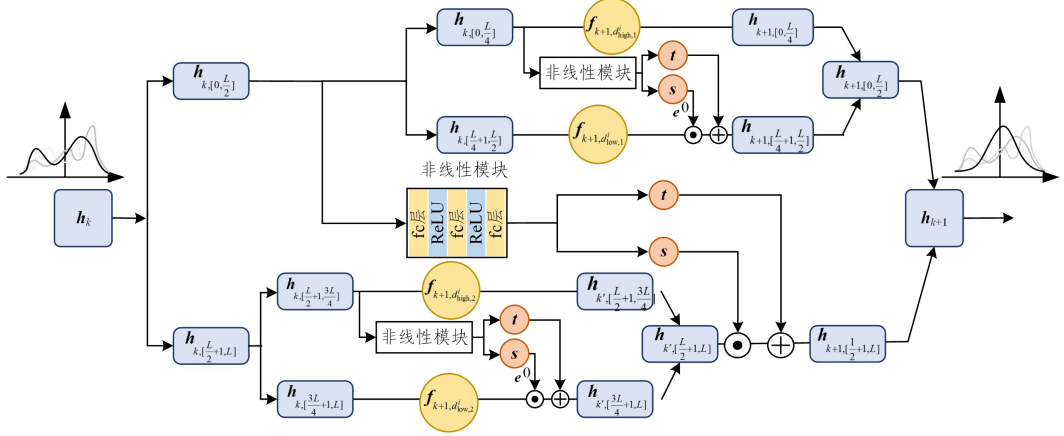


图 2 密集标准化流模块

Fig. 2 Dense normalized flow module

输入向量经过随机排列后被分割成两个部分, $\mathbf{h}_{k+1, [\frac{D}{2}+1, D]}$ 的生成依赖于与 $\mathbf{h}_k, [0, \frac{D}{2}]$ 相关的变换因子, 在 $f_{k+1, d_{high}}$ 与 $f_{k+1, d_{low}}$ 中, 输入向量被不断排列及分割, 重复进行向量间的相关性变换, 直到填满整个模块的下三角雅可比矩阵, 从而避免有效信息的缺失。每个子模块中的仿射层数目都会变为原来的两倍, 但其通道复杂程度将大幅度下降, 其中, 任意的 $s(h)$ 与 $t(h)$ 均相互独立。为防止指数的爆炸式增长而引发的模型崩溃, 对 $s(h)$ 采用非线性函数对该项进行软夹紧, 该函数如式(8)所示:

$$s_{clamp} = \frac{2\alpha}{\pi} \tan^{-1} \frac{s}{\alpha} \quad (8)$$

其中, α 的经验值为 2。

每一个填充雅可比行列式的小块均采用 Real-NVP^[22] 耦合层结构搭建, 一个密集流模块的数学定义如下:

$$\mathbf{h}_{k+1} = f_{k+1}(\mathbf{h}_k)$$

$$= \begin{bmatrix} f_{k+1, d_{high}}(\mathbf{h}_{k, [0, \frac{L}{2}]}) \\ Q(f_{k+1, d_{low}}(\mathbf{h}_{k, [\frac{L}{2}+1, L]}) | \mathbf{h}_{k, [0, \frac{L}{2}]}, \theta) \end{bmatrix} \quad (9)$$

其中, 函数 $Q(\cdot)$ 为可微变换函数, 旨在将分割后的两个特征向量进行关联耦合。其逆变换如下:

推导而来, 其中 $\left| \frac{d\mathbf{z}}{d\mathbf{x}} \right|$ 表示概率密度相关空间缩放大小。对于多变量输入而言, 雅可比矩阵 $J_f(\mathbf{z})$ 量化了概率变换后相对变化的大小, 由行列式的物理意义可知, 若 $J_f(\mathbf{z})$ 的绝对值为将 \mathbf{z} 映射到 \mathbf{x} 时的值, 空间大小缩放的倍数即为概率密度缩放倍数的倒数。

3.2.2 密集流模块

简单的耦合层结构所得到的雅可比行列式仅考虑全局的变量替换, 限制了概率分布空间的表达能力, 增加了未知分布精细化映射到标准分布过程的难度, 导致模型需要过多且不必要的重复块去拟合。受文献[21]的启发, 本文构造了一种连续密集标准化流模型作为密度估计器以进行工业异常检测, 在密集标准化流子模块中, 通过不断缩减概率分布空间, 与相关同维度特征空间相互作用, 来提升模型的表达性。其中一个密集标准化流子模块结构如图 2 所示。

$$\mathbf{h}_{k+1}^{-1} = f_{k+1}^{-1}(\mathbf{h}_k)$$

$$= \begin{bmatrix} f_{k+1, d_{high}}^{-1}(\mathbf{h}_{k, [0, \frac{L}{2}]}) \\ Q^{-1}(f_{k+1, d_{low}}^{-1}(\mathbf{h}_{k, [\frac{L}{2}+1, L]}) | \mathbf{h}_{k, [0, \frac{L}{2}]}, \theta) \end{bmatrix} \quad (10)$$

使用最大似然估计训练模型, 让似然函数最大化以寻求最优的模型参数。本文将目标函数表示为负对数似然损失, 目标函数如式(11)所示:

$$L_{nf}(\mathbf{x}) = \frac{\|\mathbf{z}\|_2^2}{2} - \log \left| \det \frac{d\mathbf{z}}{d\mathbf{x}} \right| \quad (11)$$

3.2.3 分类与定位

与 DifferNet^[18] 相同, 连续密集标准化流模型将沿用负对数似然的方法判别输入的数据是否为异常。为获得更完备的上下文信息和位置信息, 本文传递各个图像的多个不同的旋转及亮度变换视图, 提升了模型检测性能。本文模型的预测阶段计算式如式(12)所示: :

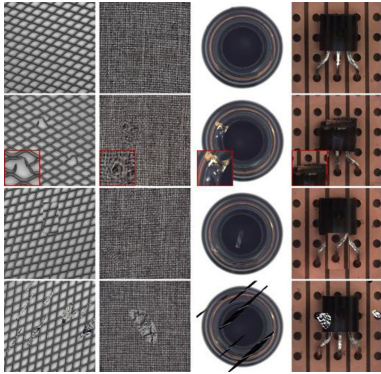
$$\tau(\mathbf{I}_i^{\rho}) = E_{P_i \in \rho} [-\log p_Z(f_{cd}(f_{ex}(P_i(\mathbf{I}^{\rho}))))] \quad (12)$$

其中, ρ 表示第 i 张测试图像的多个变换, 而 P_i 是多个变换中第 j 个变换操作; τ 为最终通过模型后所得到的异常分数; f_{ex} 与 f_{cd} 分别为特征主干网络及连续密集流模块两部分。当 $\tau(\mathbf{I}^{\rho})$ 大于所设阈值 θ 时, 预测输入图片为异常。

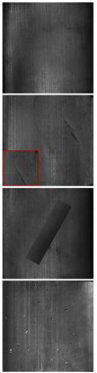
4 实验结果与分析

4.1 数据集

本文以 MVTec AD 数据集^[3]作为主要实验数据集,并将实验扩展至 MTD 数据集^[4]以及自制的工业布匹数据集上进行验证,以此证明本文方法的通用性。MVTec AD 是一个专门用于对工业异常检测方法进行基准测试的数据集,其分为 15 个不同的对象和纹理类别,每个类别包括一组无缺陷的训练图像和一组具有各种异常的图像以及没有缺陷的图像的测试集。在该数据集中,正常图像与异常图像间仅在部分区域存在偏差,这使得正常分布和离群的异常分布之间仅存在细微差异。该数据集共包含超 5000 张高分辨率图像,原始图像分辨率在 700×700 像素到 1024×1024 像素之间。在本文中,各类训练正样本数量为 3629 张,混合异常样本数量为 477 张,其中参与训练的真实异常为各类对象和纹理数据集的 1~2 种已知异常,数量为 310 张。对于地毯、瓷砖这两类纹理,采用数据扩充的方法生成的模拟异常在视觉上与正常样本并无差别,因此这两类纹理不采用模拟数据参与对比学习。



(a) MVTec AD 数据集部分示例



(b) MTD 数据集示例



(c) 自制布匹数据集示例

图 3 工业数据集示例

Fig. 3 Example of industrial datasets

MTD 数据集共有 1344 张磁瓦图像,其中包含无缺陷磁瓦图像以及 5 类磁瓦缺陷数据。本文在 952 张无缺陷图像按 80% 的比例随机选取 762 张作为训练集,其余 190 张图像作为测试数据集,混合异常样本数量为 127 张。

自制布匹数据集包含 764 张印有衣物标签的布匹图像,其中正常图像有 339 张,包含刮痕、脏污等两类异常的图像共 425 张。由于布匹图像长宽比例大,于是对其进行裁剪,裁剪后正常样本共计 2696 张,参与训练的正常样本量为 2157 张,

混合异常样本数量为 190 张。

本文以 DifferNet^[18]作为基线,数据集在训练阶段均被缩放成 448×448 , 224×224 与 112×112 这 3 个尺度,从而显著丰富样本空间,并对所有的训练数据采用随机旋转等图像增强方法。由于在预训练过程中,参与训练的数据集较小,从头开始训练会使后续模型效果不佳,泛化能力弱,同时会浪费大量算力。因此,我们使用 ImageNet 预训练权重对特征提取模块进行初始化,对本文数据集进行迁移学习。本文对所有数据集采用相同的基本参数设置:初始学习率为 0.0002, epoch 数为 280, batch size 设为 96, 使用 Adam 优化器优化。实验室服务器的操作系统为 Ubuntu18.04.4; CPU 配置为 Intel Core i7-10700K CPU @ 3.80GHz, GPU 型号为 NVIDIA GeForce GTX 3070, 深度学习框架为 Pytorch 1.7.1。

4.2 实验结果与分析

本文使用了在公开数据集中广泛采用的评估指标 AUROC 来评价所提模型的性能。AUROC 计算未知的不同阈值下真阳性率与假阳性率所构建的曲线下的面积, AUROC 值越大,表示模型分类效果越好。

4.2.1 消融实验

为验证基于对比学习的特征提取网络预训练策略及连续密集流的有效性,本文利用上述指标对各模块进行评估与分析,消融实验结果如表 1 所列。其中加粗字体为每行最优值,基线+(1)表示添加了基于对比学习的特征提取网络预训练策略的 DifferNet 模型,基线+(2)表示采用了连续密集流模型的 DifferNet。在基线上采用基于对比学习的特征提取网络预训练策略后,纹理类 AUROC 指标为 95.61%, 提高了 3.85%, 总体指标较基线提高了 2.99%。在基线上采用连续密集流模型后, tile 和 toothbrush 等类别的 AUROC 指标大幅提升,与基线相比提高了 0.76%。同时引入预训练策略和连续密集流模型的 AUROC 指标达 5.22%, 纹理类提升了 0.74%, 对象类别提升了 4.13%。

表 1 MVTec AD 数据集下模型引入各模块的实验结果

Table 1 Ablation experiment results of model adding each module

on MVTec AD dataset

(单位: %)

类别	基线	基线+(1)	基线+(2)	本文方法
Carpet	91.09	80.18	92.50	98.39
纹理类	73.43	98.48	73.68	99.31
Leather	96.84	100.00	97.83	100.00
Tile	98.02	99.41	99.71	99.21
Wood	99.40	100.00	99.55	100.00
纹理类均值	91.76	95.61	92.65	99.38
对象类	99.05	100.00	98.97	100.00
Bottle	94.00	98.46	95.71	99.81
Cable	88.07	98.72	87.31	97.22
Capsule	99.86	99.64	99.89	100.00
Hazelnut	94.18	100.00	95.45	97.90
Metal_nut	87.45	99.10	88.68	92.12
Pill	95.90	83.50	94.38	96.71
Screw	96.11	100.00	99.72	100.00
Toothbrush	90.67	86.33	91.96	96.66
Transistor	94.64	99.75	94.77	99.84
Zipper	93.99	96.55	94.68	98.03
对象类均值	93.25	96.24	94.01	98.47
平均值				

综上,本文提出的基于对比学习的特征提取网络预训练策略对检测未知异常十分有效,真实异常成本低,因此,模型对未知的异常的抽象定义更具象化,模拟异常扩充了真实异常数据,使得正常与异常在潜在分布中的决策边界更明确,同时也是十分经济的数据扩充方法。在对比学习任务下,随着模型参数不断被训练优化,大部分区域相似的异类在空间中被拉远,同类样本互相聚合,从而提升了后续标准化流模型的辨别能力;在连续密集流模型中,以正常样本作为输入将更容易被收敛,在每个密集流中,大部分的信息得以被充分利用与表达,使得变换后的潜在分布在最后一层与标准正态分布精准对齐。

4.2.2 模型超参数分析

在连续密集流模型中主要是耦合层数以及全连接层大小这两种超参数影响模型效果。表2中,AUROC指标表现形式为“纹理类别平均值/对象类别平均值/总平均值”,加粗字体为对应部分的最优值。从表2可以看出,AUROC指标随着耦合层数目的增加而升高,密集流块数量为4时,不同大小全连接层下的指标较密集流块数量为2时提高了0.09%~0.58%;随着全连接层的增加,标准化流模块数目带来的影响减小。对于不同大小全连接层,在全连接层维度为768时指标趋于平稳,最佳的密集流块数量为3,指标为98.47%。在超参数设计的指标中,最低指标为97.72%,其仍旧比基线高3.1%,且流模块数从8缩减为2。

表3 在MVTec AD数据集下不同异常检测模型的检测结果

Table 3 Detection results of different anomaly detection models on MVTec AD dataset

		(单位:%)								
类别		GANomaly	CAVGA-R _w	DevNet	PaDiM	DRA	CutPaste	DRAEM	本文方法	
纹 理 类	Carpet	82.10	82.00	86.70	99.90	94.00	100.00	97.00	98.39	
	Grid	74.30	81.00	96.70	95.70	98.70	99.10	99.90	99.31	
	Leather	80.80	84.00	99.90	100.00	100.00	100.00	100.00	100.00	
	Tile	72.00	86.00	98.70	97.40	99.40	99.80	99.60	99.21	
	Wood	92.00	89.00	99.90	98.80	99.80	99.80	99.10	100.00	
纹理类均值		80.24	84.40	96.38	98.36	98.38	99.74	99.12	99.38	
对 象 类	Bottle	79.40	96.00	99.30	99.80	100.00	100.00	99.20	100.00	
	Cable	71.10	97.00	89.20	92.20	90.90	96.20	91.80	99.81	
	Capsule	72.10	93.00	86.50	91.50	93.50	95.40	98.50	97.22	
	Hazelnut	87.40	92.00	100.00	93.30	100.00	99.90	100.00	100.00	
	Metal_nut	69.40	88.00	99.10	99.20	99.70	98.60	98.70	97.90	
	Pill	67.10	97.00	86.60	94.40	90.40	93.30	98.90	92.12	
	Screw	100.00	79.00	97.00	84.40	97.70	86.60	93.90	96.71	
	Toothbrush	70.00	99.00	86.00	97.20	82.60	90.70	100.00	100.00	
	Transistor	80.80	89.00	92.40	97.80	91.50	97.50	93.10	96.66	
	Zipper	74.40	96.00	99.00	90.90	100.00	99.90	100.00	99.84	
	对象类均值		77.17	92.60	93.51	94.07	94.63	95.81	97.41	98.03
	平均值		78.20	89.87	94.47	95.50	95.88	97.12	97.98	98.47

表4 MTD与布匹数据集下不同异常检测模型的检测结果

Table 4 Detection results of different anomaly detection models on

MTD and homemade cloth datasets

模型	MTD	布匹
GANomaly	47.15	30.80
DevNet	83.37	98.68
PaDiM	94.07	95.11
DRA	71.12	96.38
CutPaste	85.85	91.31
DRAEM	91.62	77.80
本文	94.12	97.20

表2 不同流模块数与不同全连接层维度下的检测结果

Table 2 Detection results with different number of stream modules and different fully connected layer dimensions

(单位:%)			
超参数	耦合层数=2	耦合层数=3	耦合层数=4
全连接层	98.57/97.29/	98.96/97.88/	98.92/97.89/
维度=352	97.72	98.24	98.23
全连接层	99.04/97.16/	98.93/ 98.76 /	99.12/97.99/
维度=448	97.79	98.15	98.37
全连接层	99.00/97.44/	98.63/97.86/	98.81/98.09/
维度=512	97.96	98.11	98.33
全连接层	99.00/98.00/	99.17 /98.12/	98.72/98.28/
维度=768	98.33	98.47	98.42

4.2.3 不同异常检测模型实验对比

为了进一步验证本文模型的有效性,我们将该模型与其他异常检测模型进行对比实验,表3列出了不同模型在MVTec AD数据集上的检测结果。其中,PaDiM^[23]和GANomaly^[24]分别采用了引言中所述的基于生成对抗式与基于统计学的方法,而DRAEM^[9]和Cutpaste^[11]这两种模型则是生成了模拟异常参与训练。DevNet^[25],DRA^[26]与CAVGA-R_w^[27]是同为弱监督模式下的异常检测模型。本文方法在检测精度上比其他方法都更优秀,在6个类别上均达到最优值,并在5个类别上指标高达100%。表4列出了模型在MTD数据集^[4]与自制的布匹数据集上的检测结果,可以看出,该模型对不同种类及对象的工业数据集具有通用性,同时对正负样本具备较好的决策能力,达到了较高的分类精度。通过基于对比学习任务的训练策略和高表达能力的标准化流模型,本文方法得以学习到更广义的正态和异常表示。

4.2.4 实验结果可视化

在测试阶段,不同类别的正常与异常样本在模型推理中所得到的异常得分具有显著差别。图4为MVTec AD数据集中15种类别数据的异常分数统计图,其中横坐标为模型测试所得分数,纵坐标为样本个数,蓝色柱状图代表正常,黄色柱状图代表异常。可以直观地看出,在多个设定阈值下,异常样本都能轻易地被判别出来。图5给出了不同方法在MVTec AD数据集中各类别的ROC曲线。其中红线代表本文方法,可以看出,所提方法大幅降低了误检和漏检的可能性。



图 4 MVTEC AD数据集各类别的异常分数统计图(电子版为彩图)

Fig. 4 Anomaly score statistics by category on MVTEC AD

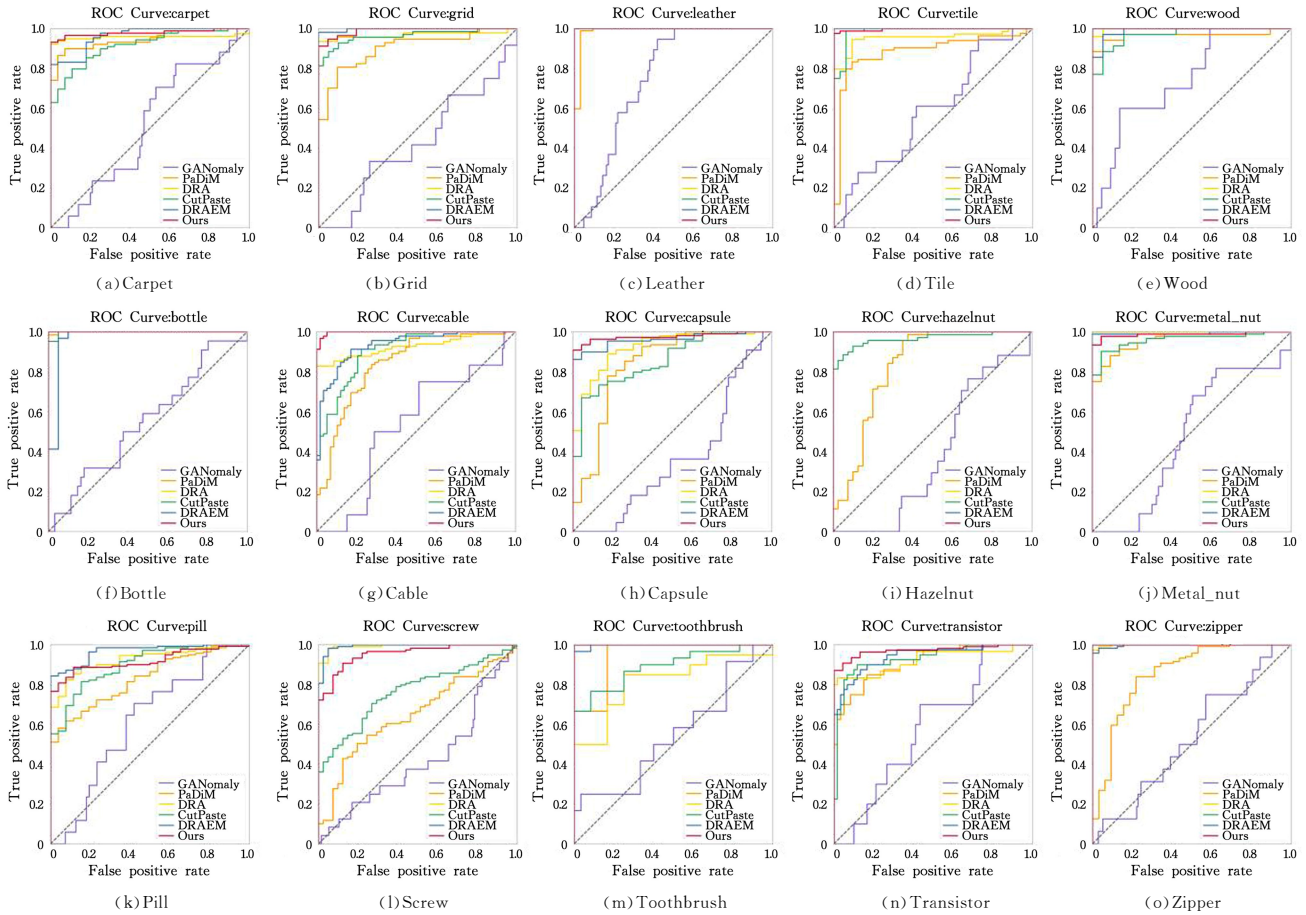


图 5 不同方法在 MVTEC AD 数据集中各类别的 ROC 曲线(电子版为彩图)

Fig. 5 ROC curves of different categories on MVTEC AD datasets by different methods

鉴于特征提取网络深度相对较浅以及标准化流模型的双射性,由负最大似然函数所得到的异常分数能与输入图片产生直接联系。具体来说,预测过程中具有非零梯度的 ε 值对应的原始图像将更趋向于是异常区域。这使得模型更具可解释性。梯度图示例如图6所示,在所生成的梯度图中,斑点代表该处可能为异常。而梯度图中的光斑能够捕捉到异常区域的大致位置,因此本文方法具备一定的异常定位能力。

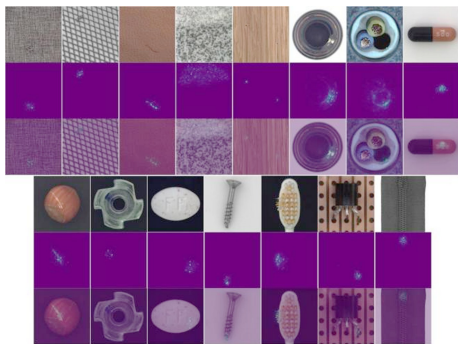


图6 MVTEC AD数据集各类别的异常梯度示意图

Fig.6 Schematic diagram of abnormal gradients of various categories on MVTEC AD dataset

结束语 为缓解工业实际场景中异常样本数据匮乏以及提高检测模型的表达能力,本文提出了一种面向工业图像异常检测的连续密集标准化流模型。将正常样本和由模拟异常样本与少量异常样本混合而成的数据集输入到孪生网络中,提取到具有判别性的深度特征,以异常数据为输入的网络分支末端采用特征级增广,从而获得丰富且更难以鉴别的高维异常特征。其后,构造三元组损失和对比损失来最大化正常样本间的向量相似度,拉远异类样本间的距离。使用Real-NVP^[22]耦合层体系结构,将其作为搭建密集流模块的基本架构,对非线性部分应用软夹紧,保持模型的稳定性,通过多个层级的可逆变换,提高模型对正常样本分布的拟合能力。

本文在弱监督模式下在公共数据集以及自制数据集上有着较高的检测性能,体现了本文方法的通用性。然而在测试阶段,对单张图像采用48次旋转及尺寸变换来获取数据更为详尽的信息,虽然极大地提高了异常检测能力,但同时也降低了测试效率。此外,一些异常检测任务面临着严格无监督的环境,在后续工作中,将进一步研究仅包含正常样本作为训练数据的快速、高精度、高效率的工业异常检测方法。

参考文献

- [1] CHALAPATHY R, CHAWLA S. Deep Learning for Anomaly Detection: A Survey [EB/OL]. (2019-01-23) [2022-10-10]. <https://arxiv.org/pdf/1901.03407.pdf>.
- [2] LV C K, SHEN F, ZHANG Z T, et al. Review of Image Anomaly Detection [J]. Acta Automatica Sinica, 2022, 48(6): 1402-1428.
- [3] BERGMANN P, FAUSER M, SATTLEGGER D, et al. MVTEC AD-A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE Press, 2019: 9592-9600.
- [4] HUANG Y, QIU C, YUAN K. Surface Defect Saliency of Magnetic Tile [J]. The Visual Computer, 2020, 36(1): 85-96.
- [5] XU T, TIAN C Y, LIU C H. Deep Learning for Abnormal Crowd Behavior Detection: A Review [J]. Computer Science, 2021, 48(9): 125-134.
- [6] WANG Y S, YAO H X, SUN X S, et al. Representation Ability Research of Autoencoders in Deep Learning [J]. Computer Science, 2015, 42(9): 56-60, 65.
- [7] BERGMANN P, LÖWE S, FAUSER M, et al. Improving Unsupervised Defect Segmentation by Applying Structural Similarity to Autoencoders[EB/OL]. (2019-02-01) [2022-10-10]. <https://arxiv.org/pdf/1807.02011.pdf>.
- [8] 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.
- [9] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DRAEM: A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Montreal: IEEE Press, 2021: 8330-8339.
- [10] KIM D, JEONG D, KIM H, et al. Spatial Contrastive Learning for Anomaly Detection and Localization [J]. IEEE Access, 2022, 10: 17366-17376.
- [11] LI C L, SOHN K, YOON J, et al. Cutpaste: Self-Supervised Learning for Anomaly Detection and Localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Nashville: IEEE Press, 2021: 9664-9674.
- [12] ULGER F, YUKSELS E, YILMAZ A. Anomaly Detection for Solder Joints using β -VAE[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2021, 11(12): 2214-2221.
- [13] 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.
- [14] XING H J, HAO Z. Novelty Detection Method Based on Global and Local Discriminative Adversarial Autoencoder [J]. Computer Science, 2021, 48(6): 202-209.
- [15] KOBYZEV I, PRINCE S J D, BRUBAKER M A. Normalizing Flows: An Introduction and Review of Current Methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(11): 3964-3979.
- [16] DINH L, KRUEGER D, BENGIO Y. NICE: Non-Linear Independent Components Estimation[EB/OL]. (2015-04-10) [2022-10-10]. <https://arxiv.org/pdf/1410.8516.pdf>.
- [17] HUANG C W, DINH L, COURVILLE A. Augmented Normalizing Flows: Bridging the Gap between Generative Flows and Latent Variable Models[EB/OL]. (2020-02-17) [2022-10-10]. <https://arxiv.org/pdf/2002.07101.pdf>.
- [18] RUDOLPH M, WANDT B, ROSENHAHN B. Same Same but Different: Semi-Supervised Defect Detection with Normalizing

- Flows[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Waikoloa; IEEE Press, 2021: 1907-1916.
- [19] CIMPOI M, MAJI S, KOKKINOS I, et al. Describing Textures in the Wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus; IEEE Press, 2014: 3606-3613.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [21] KRUSE J, DETOMMASO G, KÖTHE U, et al. HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. 2021: 8191-8199.
- [22] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density Estimation using Real NVP[EB/OL]. (2017-02-27)[2022-07-08]. <https://arxiv.org/pdf/1605.08803.pdf>.
- [23] 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. Springer, 2021: 475-489.
- [24] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Ganomaly: Semi-Supervised Anomaly Detection via Adversarial Training[C]//Asian Conference on Computer Vision. Springer, 2018: 622-637.
- [25] PANG G, DING C, SHEN C. Explainable Deep Few-Shot Anomaly Detection with Deviation Networks[EB/OL]. (2021-08-01)[2022-10-10]. <https://arxiv.org/pdf/2108.00462.pdf>.
- [26] DING C, PANG G, SHEN C. Catching Both Gray and Black Swans: Open-Set Supervised Anomaly Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2022: 7388-7398.
- [27] VENKATARAMANAN S, PENG K C, SINGH R V, et al. Attention Guided Anomaly Localization in Images[C]//European Conference on Computer Vision. 2020: 485-503.



ZHANG Zouquan, born in 1998, master. His main research interests include image processing, deep learning and so on.



ZHANG Hui, born in 1983, Ph.D, professor. His main research interests include machine vision and sparse representation.

(责任编辑:何杨)