

文章编号:1007-2780(2024)08-1116-12

基于融合特征分布学习与图像重建的异常检测

朱思宇, 朱磊*, 王文武, 乐华钢

(武汉科技大学 信息科学与工程学院, 湖北 武汉 430070)

摘要: 无监督学习是当前工业产品缺陷检测领域的主流研究方向, 目前主要分为基于重建和基于特征的两类方法。前者构建基于内容感知的映射以将异常区域映射为正常区域并通过残差图像来检测缺陷, 注重于图像整体的表现。后者利用高层语义特征以实现定位异常, 更加注重图像细节呈现。根据两种方法的优缺点, 本文提出一种基于特征与重建融合的方法, 有效结合两者优点互补其不足并实现统一的端到端的学习与推理。首先训练一个重建模型, 然后采用归一化流模型以充分学习输入正常样本的高可能性数据概率分布, 使其与重建模型相融合, 有效地提高重建模型缺陷检测以及缺陷定位的准确率。在广泛使用的 MVTec AD 数据集上, 提出的融合模型的平均图像级 AUROC 达到了 98.7%, 平均像素级 AUROC 达到了 94.2%, 特别地, 相比单一重建模型提升了 3.3%。提出的特征与重建网络融合模型显著提高了重建网络部分对于缺陷定位的不足, 使结果更为精确。

关键词: 异常检测; 重建; 内存模块; 流模型; 融合算法

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/CJLCD.2023-0304

Abnormity detection based on fusion feature distribution learning and image reconstruction

ZHU Siyu, ZHU Lei*, WANG Wenwu, YUE Huagang

(School of Information Science and Engineering, Wuhan University of Science and Technology,
Wuhan 430070, China)

Abstract: Unsupervised learning is the main research direction in the field of industrial product defect detection at present, and it is mainly divided into two types of methods: reconstruction based and feature based methods. The former constructs content-aware mappings to map abnormal regions into normal regions and detect defects through residual images, focusing on overall performance of the images. The latter uses high-level semantic features to achieve positioning exceptions and pay more attention to image detail presentation. According to the advantages and disadvantages of the two methods, a method is proposed based on the fusion of characteristics and reconstruction, which effectively combines advantages of the two methods to complement their shortcomings and realize unified end-to-end learning and reasoning. A reconstructed model is trained firstly, then a normalized flow model is adopted to fully learn the probability distribution of high probability data of input normal samples, and it is integrated with the reconstructed model

收稿日期: 2023-09-19; 修订日期: 2023-10-12.

基金项目: 国家自然科学基金 (No.62173262)

Supported by National Natural Science Foundation of China (No.62173262)

*通信联系人, E-mail: zhulei@wust.edu.cn

to effectively improve the accuracy of defect detection and defect positioning of the reconstructed model. On the widely used MVTec AD data set, the average image level AUROC of the proposed fusion model reaches 98.7%, the average pixel-level AUROC reaches 94.2%, in particular, an increase of 3.3% compared to a single reconstruction model. The convergence model of characteristics and proposed reconstruction network has significantly improved the shortcomings of defect positioning in the reconstruction network part, which makes experimental results more accurate.

Key words: abnormal detection; rebuild; memory modules; flow model; fusion algorithm

1 引言

在实际生活中,异常检测模型被广泛应用于工业生产检测、医学图像分析和视频监控等领域^[1-2]。一般的异常检测任务往往被视为单类分类问题,即判断一个物体有异常还是无异常。以监督的方式去进行异常检测任务往往需要足够的异常样本作为标记。然而,由于异常在当前成熟生产工艺条件下较少出现,且出现后不易观察、范围小、形式各种各样,极大地增加了标记异常样本的人工成本。因此现有的工作常常将异常检测任务定义为无监督或者半监督任务,即仅利用正常的样本去学习模型。在推理阶段,将异常值判定为异常样本。常见的异常检测方法有两类,一种是基于重建的方法^[3],另一种是基于特征的方法^[4]。

基于重建的方法将异常检测作为一项内容感知的图像重建任务。其主要思想是假设仅用正态样本训练的模型无法对异常区域进行合理的特征表达,从而在测试时通过重构误差来检测异常。现有基于重建的异常检测方法通常采用自编码器(AutoEncoder, AE)^[5]结构,其输入为没有异常的正常样本,通过由多层“卷积-降采样”构成的编码器提取样本中具有代表性的特征,随后通过由“卷积-上采样”的解码器将特征恢复为输出图像,通过计算输入图像与重建结果之间的差值图以直观地显示缺陷的存在。Sakurada等人^[6]首次将自编码器应用于异常检测,发现自编码器可以很好地编解码正常样本数据以及区分正常样本和异常样本。此后, Nicolau等人^[7]提出将自编码器和密度估计相结合,利用隐藏层的低密度来指示异常。为了进一步提高编码器提取正常样本主要特征的能力,文献[8]引入了生成对抗网络以提供由鉴别器生成的额外监督。

基于特征的方法通常提取正常样本的特征,并将其表示为可测量的分布。而正负样本的判断则是通过计算异常分数来实现的,例如,计算输出样本的分布与正常样本分布之间的距离来判断正负样本。由于基于特征的方法通过 ImageNet 预训练网络对图像特征进行提取,对于具有类似结构的图像能够得到较稳定的特征描述,最近的一些研究如 SPADE^[9]、PaDiM^[10]便利用在 ImageNet 上预训练的特征提取器作为提取特征方法。具体来说,通过在 ImageNet 数据集上预训练分类模型以得到有效的特征提取模型。然后对正常样本提取的特征建模,比较正常样本和异常样本之间特征分布的差异,从而计算异常分数来判别异常是否存在。另外一类是以流模型为基础的特征方法^[11],其可以实现两种不同分布之间的转换。此类方法同样利用在 ImageNet 上预训练的网络作为特征提取器,然后将复杂的正常样本分布映射为简单的高斯分布。由于在训练时利用正常样本进行训练,所以在判别时能够将正常样本映射为高可能性,异常样本映射为低可能性,然后通过得到的异常分数划分正常至异常区间,从而对异常的存在进行判别。

基于重建的方法具有异常检测原理直观、结构设计灵活等优点,但存在固有的难以平衡重建质量和异常发现能力的缺点。即较高的重建质量需要编码器减少降采样层级以尽可能地保留图像细节信息。相反,网络需具有较深的结构以提取足以描述图像内容的特征来保证对于正常与异常区域的鉴别能力。虽然此类方法采用在编码器和解码器之间加入跳转连接以缓解细节退化问题,但又引入了对于异常区域进行一致性映射的风险。基于特征的方法通过 ImageNet 预训练网络对图像特征的提取具有无需在目标数据集上进行自监督训练的优点,对于具有类似结

构的图像能够得到较为稳定的特征描述,即使对于一些包含较小缺陷的异常样本也能成功地进行检测。然而,在大规模数据集上进行预训练具有较大的代价和硬件需求,限制了特征提取网络设计的灵活性。此外,特征提取网络最终输出的特征图具有较小的分辨率,因此难以输出高像素级精度的异常检测结果。

本文在分析了上述两种方法优缺点的基础上提出了一种基于重建与特征网络融合的检测模型。该方法一方面利用两者的优点形成合力以提升检测效果,另一方面针对两者固有的问题

提出了对应的解决方案。

2 基本原理

本文所提出的基于无监督异常检测的模型框架如图1所示。该模型由两部分组成:上半部分为重建模型,主要由记忆模块、DWT小波正/逆变换为基础的编解码器组成;下半部分为特征模型,以重建网络中编码器输出的高层特征作为输入,并采用流模型范式以学习正常样本的全局分布。最终通过融合检测方法生成差值图像用于缺陷鉴别。

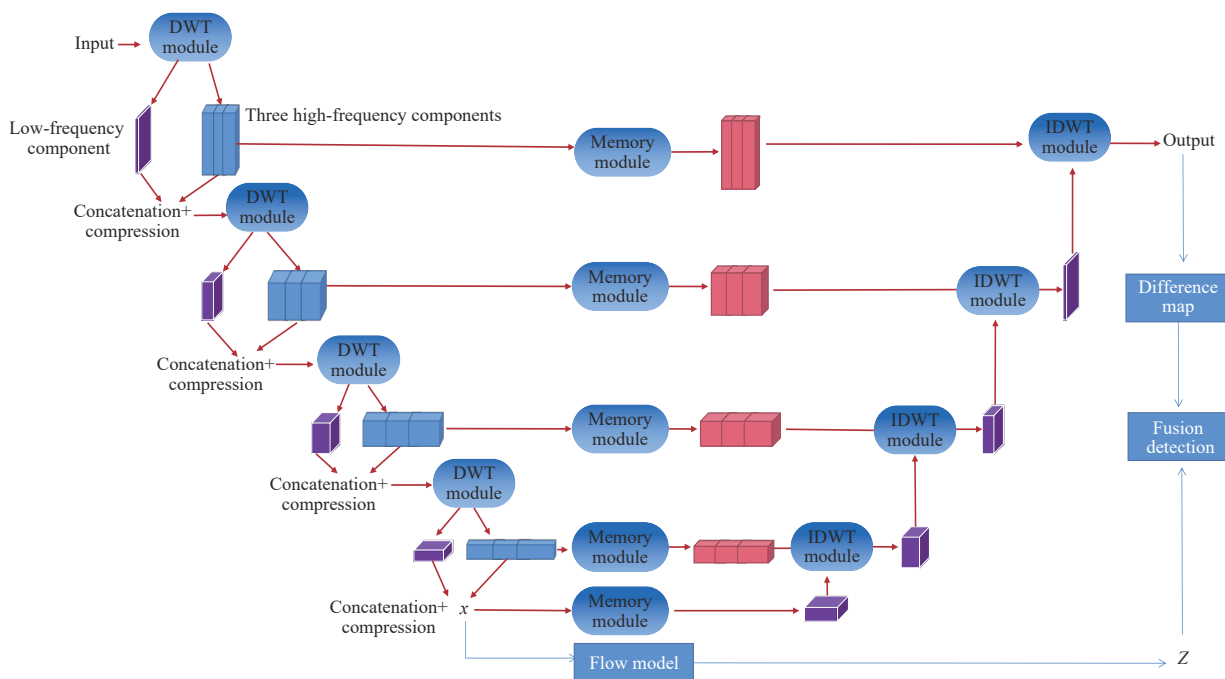


图1 模型的整体结构

Fig. 1 Overall structure of the model

2.1 编解码器

传统编码器通常采用多级“卷积组+池化”模块以提取多尺度深度特征。池化操作会稀释特征中用于描述细节的信息,导致重建图像模糊。为此研究者多采用跳转连接直接将浅层特征输出至解码器以缓解该问题。然而,浅层特征通常缺乏足够的高层语义以鉴别缺陷区域,易导致重建网络对缺陷进行恒等映射。为解决该问题,受文献[12]启发,本文引入离散小波变换(DWT)搭建所提出重建网络的编码器和解码器模块。

DWT是一种可逆的无损变换,当对一张图

片做离散小波变换时可以得到4张分辨率下降4倍的子图,它们分别储存原图片高频信息和低频信息。其中高频信息代表图像的细节信息,低频信息代表图像的基本信息。由于DWT天然具有降分辨率的特性,用离散小波变换替代下采样在理论上是合理的。同时,为了避免重建网络能够学习到正常样本中的主体特征以及对无关细节信息进行过滤,本文构建了记忆模块用于编码DWT模块输出的高频部分。如图1所示,网络整体架构采用U型结构,共分为4个层级,每一层将上一级的高频和经过拼接后的低频信息拼接后再按通道方向进行压缩,作为下一级的输入,

这样能够保留足够完整的图像信息,使网络能更好地进行图像重建。

2.2 记忆模块

一个标准的记忆模块在应用中分为两个阶段:(1)在训练阶段仅使用正常样本训练,利用正常样本的特征来更新记忆项目;(2)在推理阶段,识别编码的图像特征与缓存记忆项之间的归一化相似度,并利用该相似度重构编码特征。由于在训练时仅采用正常样本来构建记忆模块,因此通过使用存储的正常样本记忆项进行重构,可以有效地提高重建质量。

本文设计的记忆模块如图2所示,记忆模块中存储器设计为矩阵 $M \in \mathbb{R}^{N \times C}$,其中 N 为存储器最大存储容量, C 为存储的特性向量维度。 M 被设计为记录在训练过程中的正常样本,本文定义该存储器为内容可寻址内存,其寻址方案基于内存项 m 和查询 z 的相似性计算注意力权重 w ,如公式(1)所示:

$$w_i = \frac{\exp(d(z, m_i))}{\sum_{j=1}^N \exp(d(z, m_j))}, \quad (1)$$

其中: m_i 是内存项中第 i 个内存; w_i 为相对应的权重; $d(*, *)$ 定义为余弦相似度,其表达式如公式(2)所示:

$$d(z, m_i) = \frac{zm_i^T}{\|z\| \|m_i\|}. \quad (2)$$

可以看到,越是与存储的内存项接近,其权重 w_i 越大。值得注意的是,表征异常的特征可能通过 w_i 的线性组合被错误地恒等映射,为了缓解这个问题,本文采用硬收缩操作来保证 w_i 的稀疏性,如公式(3)所示:

$$\hat{w}_i = \frac{\max(w_i - \lambda, 0) \times w_i}{|w_i - \lambda| + \epsilon}, \quad (3)$$

其中, λ 表示收缩阈值, ϵ 是一个很小的量。最后,重建的特征可通过公式(4)得到。

$$\hat{z} = \hat{w}_i M = \sum_{i=1}^N \hat{w}_i m_i. \quad (4)$$

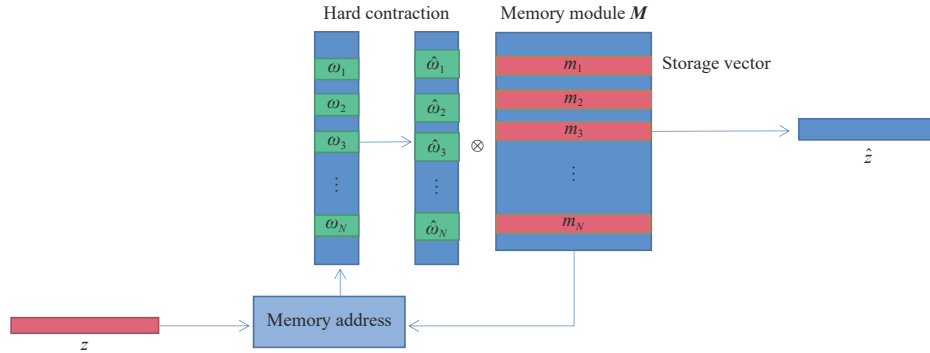


图2 记忆存储模型示意图

Fig. 2 Schematic diagram of memory storage model

2.3 归一化流模型

归一化流模型^[13]是一种生成模型,用于建模和处理高维数据的概率分布。它基于可逆的转换函数,将输入数据从一个分布映射到另一个分布,通常是将一个简单的先验分布映射到更复杂的后验分布。通过这样的转换,流模型可以从先验分布中生成新样本,并对后验分布进行推断和采样。流模型有一个很好的特殊性质,便是两种分布之间的映射是双向性的,并且可以在两个方向上求值。

本文设计的流模型结构如图3所示。本文将

重建网络中编码器的最高层输出特征作为特征网络的输入。如图1所示,最高层的特征 x 作为输入,其输入大小为 $512 \times 16 \times 16$ 。该特征首先沿通道方向进行随机排列以充分混叠特征。为了应用缩放和移位操作,特征图会沿着通道方向被均匀地划分为两部分,然后按照公式(5)、(6)进行组合:

$$y_{out,2} = y_{in,2} \odot e^{s_1(y_{in,1})} + t_1(y_{in,1}), \quad (5)$$

$$y_{out,1} = y_{in,1} \odot e^{s_2(y_{out,2})} + t_2(y_{out,2}), \quad (6)$$

其中: $y_{in,1}$ 和 $y_{in,2}$ 分别为输入时均匀划分的两部分, e 为指数函数, s, t 为任意的函数, $\odot$ 为元素

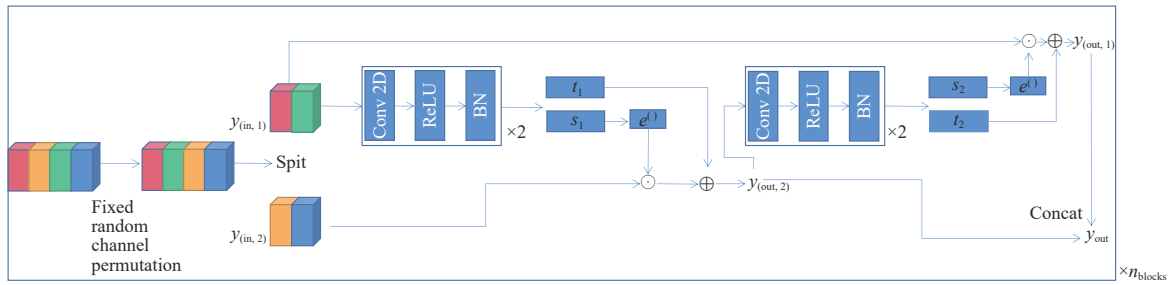


图3 流模型内部示意图

Fig. 3 Internal schematic diagram of the flow model

积。在本文设计的流模型中采用文献[11]中建议的利用正负样本可能性的不同区分正负样本的方法。值得注意的是,本文方法与其他流模型相比具有无需预训练的优势,即无需利用 ImageNet 网络进行参数初始化而直接利用重建网络中编码器进行特征提取。其优点在于:(1)流模型参数能够被更加贴合数据集的数据特征所优化;(2)可以不被特征提取器结构所限制,使前端重建网络设计更为灵活。

本文将重建网络中编码器最顶层的输出作为流模型的输入,并通过流模型训练得到输出 z 。最终通过 z 的结果,与重建网络的差值图相融合,进一步提高重建网络的效果。

2.4 损失函数

本文提出的模型的损失函数有两部分。第一部分为重建模型的损失函数。重建网络一般会用平均绝对误差(MAE)或者平均平方误差(MSE)损失作为重建指标,在这里本文利用了MSE损失作为主要训练损失。

同时,为了使重建的图像与正常样本的图像更具有结构相似性,提高重建的质量,本文额外增加了结构相似指标(SSIM)。SSIM损失是一种用于评估图像质量的指标,常用于图像重建中,其损失的计算基于对两幅图像结构信息的比较。具体而言,它首先将原始图像和重建图像分成多个重叠的图像块,然后计算每个图像块的亮度、对比度和结构信息,如公式(7)和(8)所示:

$$\mathcal{L}_{MSE} = \|x - \hat{x}\|_2, \quad (7)$$

$$L_{SSIM}(x, \hat{x}) = \frac{1}{N_p} \sum_{i=1}^H \sum_{j=1}^W 1 - SSIM(x, \hat{x})_{i,j}, \quad (8)$$

其中: x 为输入的图像, $\hat{x}$ 为重建图像, N_p 为图像像素总个数。

提出模型的损失函数的第二部分为流模型的

损失函数。在流模型中,训练的目的是尽可能使 z 被映射到高可能性,即 $z=0$ 处,如公式(9)所示:

$$P_X(x) = P_Z(z) \left| \det \frac{\partial z}{\partial x} \right|. \quad (9)$$

对其进行如公式(10)所示的最大似然处理:

$$\text{Log} P_X(x) = \log P_Z(z) + \log \left| \det \frac{\partial z}{\partial x} \right|, \quad (10)$$

其中, $\left| \det \frac{\partial z}{\partial x} \right|$ 为雅可比行列式。为了简化问题,本文认为 $P_Z(z)$ 为标准正态分布,这里,使用负对数似然损失 $\mathcal{L}(x)$ 得到最小化问题,如公式(11)所示:

$$\mathcal{L}(x) = \frac{\|z\|_2^2}{2} - \log \left| \det \frac{\partial z}{\partial x} \right|. \quad (11)$$

3 实验与分析

为了验证本文提出方法的性能,本文将提出方法与当前先进的基于重建和基于特征模型的方法进行了定性和定量的对比。所有实验在当前广泛使用的公共异常数据集 MVTEC AD 上进行。此外,本文还独立评价了本文提出方法中重建模型和特征模型对整体性能的贡献。最后,通过消融实验分析本文提出方法中重要的超参数对最终性能的影响。

3.1 图像数据集

本文所使用的数据集为工业异常检测数据集 MVTEC AD^[14]。MVTEC AD 共有 5 354 张图像,包含了 15 个不同的类别,其中有 10 种类别为对象类别,另外 5 种类别为纹理类别。在每个类别中,通常利用训练集中的正常样本进行训练。在测试时,利用测试集中包含的正常样本和异常样本来评估。最后,利用该数据集包含的真实缺陷二值图来定量地评估算法性能。

3.2 实验过程

本文提出方法的训练过程采用端到端的方案。首先冻结特征网络参数并对重建网络进行若干轮的训练以获得稳定的编码器特征,然后再整体训练两者以获得最终的模型参数。

3.2.1 重建网络部分实验

在所有实验中,MVTec AD数据集的每张输入图像的大小被缩放至 256×256 ,并将像素值归一化到 $[-1, 1]$ 的范围。本文使用加噪处理来增加图片的多样性,同时也能在一定程度上模拟异常,减少训练过程的过拟合。

本文构建的DWT模块具有良好的信息分解能力。如图4所示,DWT模块将原图分解为4个部分,分别为低频分量、水平分量、垂直分量和对角线分量,可以无损地代表原图像信息。在编码器的设计中,本文将DWT模块分解得到的3个高频分量按照通道方向进行拼接并送入记忆单元进行存储,同时,拼接好的高频分量与低频分量再次按通道方向进行拼接,然后进行维度压缩,作为下一级的输入。

我们发现在记忆模块中使存储量 $N=1\ 000$ 时效果最佳,并且设置收缩阈值 λ 为 0.002 , ε 为 $1E-12$ 。这样在模型收缩过程中,记忆模块相似

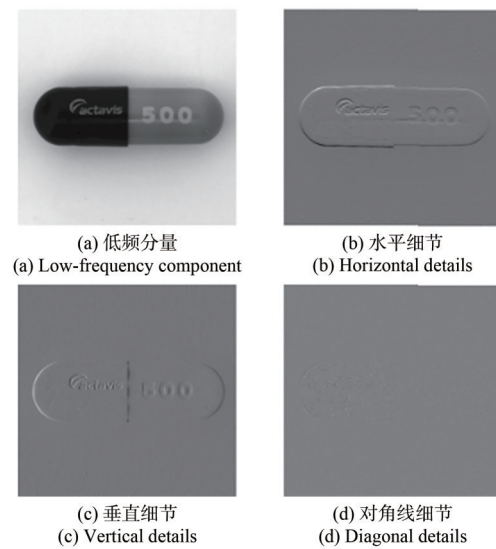


图4 DWT分解的4个频率分量示意图

Fig. 4 Schematic diagram of the four frequency components decomposed by DWT

性效果最佳。结果如表1所示,给出了训练过程输入的特征向量和输出的特征向量维度、输入输出特征向量的差值均值的绝对值(相似度)以及推理阶段正常样本和异常样本的相似度。其中 B 、 H 、 W 、 C 分别为图像的数据样数、高度、宽度以及通道数。

表1 记忆模块相似度比较图

Tab. 1 Comparison chart of memory module similarity

阶段	输入的特征向量维度 ($B \times H \times W, C$)	记忆模块的维度 (C, N)	输出的特征向量维度 ($B \times H \times W, C$)	输入输出的差值均值的绝对值(相似度)
训练阶段	($1 \times 128 \times 128, 192$)	($192, 1\ 000$)	($1 \times 128 \times 128, 192$)	$1.00E-05$
	($1 \times 64 \times 64, 384$)	($384, 1\ 000$)	($1 \times 64 \times 64, 384$)	$3.20E-05$
	($1 \times 32 \times 32, 768$)	($768, 1\ 000$)	($1 \times 32 \times 32, 768$)	$4.30E-05$
	($1 \times 16 \times 16, 1\ 536$)	($1\ 536, 1\ 000$)	($1 \times 16 \times 16, 1\ 536$)	$8.40E-05$
推理阶段(正常样本)	($1 \times 128 \times 128, 192$)	($192, 1\ 000$)	($1 \times 128 \times 128, 192$)	$2.10E-05$
	($1 \times 64 \times 64, 384$)	($384, 1\ 000$)	($1 \times 64 \times 64, 384$)	$3.40E-05$
	($1 \times 32 \times 32, 768$)	($768, 1\ 000$)	($1 \times 32 \times 32, 768$)	$5.26E-05$
	($1 \times 16 \times 16, 1\ 536$)	($1\ 536, 1\ 000$)	($1 \times 16 \times 16, 1\ 536$)	$8.84E-05$
推理阶段(异常样本)	($1 \times 128 \times 128, 192$)	($192, 1\ 000$)	($1 \times 128 \times 128, 192$)	$1.00E-03$
	($1 \times 64 \times 64, 384$)	($384, 1\ 000$)	($1 \times 64 \times 64, 384$)	$5.20E-03$
	($1 \times 32 \times 32, 768$)	($768, 1\ 000$)	($1 \times 32 \times 32, 768$)	$2.00E-04$
	($1 \times 16 \times 16, 1\ 536$)	($1\ 536, 1\ 000$)	($1 \times 16 \times 16, 1\ 536$)	$5.90E-03$

本次实验中共分离了4个层次。为了使图片更好地重建,本文采用跳连接操作以更好地获得

图片细节信息。图5给出了输入图片和重建图片的效果图、差值图和人工标注结果。

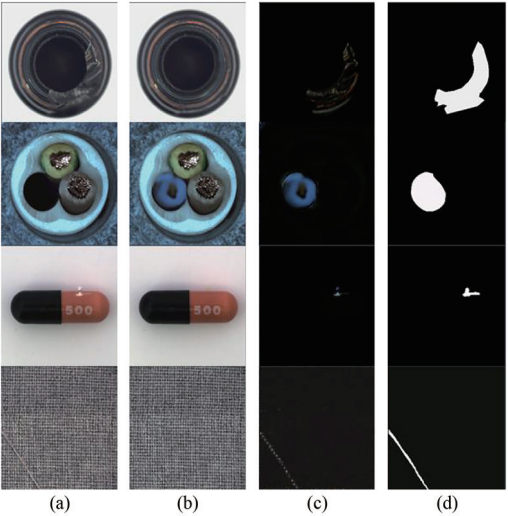


图5 重建结果图,分别为异常样本的原始图像(a)、重建结果(b)、差值图(c)和人工标注结果(d)。
Fig. 5 Reconstruction results, including the original image (a), reconstruction result (b), difference map (c) and manual annotation result (d) of the abnormal sample.

在重建网络的训练过程中,本文使用Adam^[15]方法进行网络优化,固定学习率为1E-4,其余

参数为默认值。
3.2.2 特征网络部分实验
本文直接把编码部分作为特征网络的特征提取器,并将重建网络中编码器最顶层的输出作为特征网络部分的输入。
本次实验中特征网络使用固定的最终输入大小为16×16。与文献[11]不同的是,本文特征网络的输入不是特征向量而是特征图张量,这样更有利于保留图像的空间信息。特征网络中归一化流由3个耦合块组成,利用全连接网络作为内部函数 s 和 t ,大小为1 024个神经元。模型中包含2个隐藏的密集层,每层利用Conv、BN、和ReLU激活层进行特征提取,卷积层在内部网络中对前2个耦合块使用3×3卷积核,对最后一个块使用5×5卷积核。在特征网络的训练过程中,本文使用Adam方法进行网络优化,学习率为1E-4,权重衰减为1E-5,动量值为 $\beta_1=0.8$ 和 $\beta_2=0.8$ 。

3.2.3 结果比较

图6给出了本文提出方法性能的定性评价结

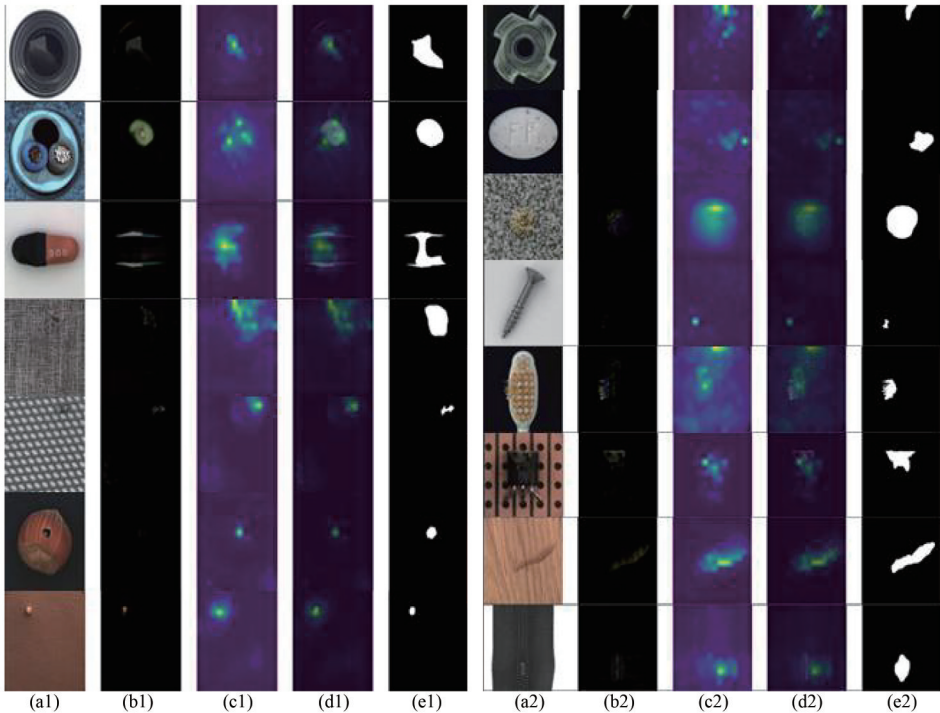


图6 本文实验结果图。图像共有两组,分别为原始图像(a1,a2)、基于重建的差值图(b1,b2)、基于特征的热力图(c1,c2)、融合特征网络后的缺陷热力图(d1,d2)以及人工标注结果(e1,e2)。
Fig. 6 Experimental results of this article. There are two sets of images, including original image (a1,a2), reconstruction-based difference map (b1,b2), feature-based heatmap (c1,c2), defect heatmap after fusion feature network (d1,d2) and manual annotation results (e1,e2).

果。从左至右分别显示了原始图像、基于重建的差值图、基于特征的热力图、融合特征网络后的缺陷热力图以及人工标注结果。本文采用的评价指标可以通过设置阈值来检测受试者工作特性下的面积 AUROC (Area Under the Receiver Operating Characteristic Curve)。AUROC 是目前无监督缺陷检测中最为通用的评价指标,其取值在 0~1 之间,值越大性能越好。实验中计算图像级和像素级两个指标,可以看到,重建网络输出的差值图能够定位缺陷位置,融合特征网络后,缺陷区域的定位更为准确。

如表 2 所示,本文将提出的方法与最近几年基于重建的方法,如 DAAD^[16]、DRAEM^[17]、SSM^[18]; 基于特征方法,如 DifferNet^[11]、CutPaste^[19]、CFLOW^[20]进行了定量比较。以图像级别 AUROC 为例,单独的重建方法并没有达到最优,在融合特征网络后,超过了最佳的 CFLOW 方法,达到了 98.7%。值得注意的是,由于本文提出的特征网络和重建网络共用一个主干网络,并未像独立特征方法引入额外的预训练数据,因此更能体现提出方法的有效性,且使主干网络的构建更具灵活性。

表 2 图像级别 AUROC 的比较

Tab. 2 Comparison of image levels AUROC

分类	样本	基于重建的方法			基于特征的方法			本文方法	
		DAAD/%	DRAEM/%	SSM/%	DifferNet/%	CutPaste/%	CFLOW/%	基于重建的方法/%	融合方法/%
物体类	bottle	97.6	99.2	99.9	99.0	99.9	100.0	100.0	100.0
	cable	84.4	91.8	77.3	95.9	100.0	100.0	97.7	98.2
	capsule	76.7	98.5	91.4	86.9	98.6	99.3	93.7	94.3
	hazelnut	92.1	100.0	91.5	99.3	93.3	96.8	98.8	99.8
	metal nut	75.8	98.7	88.7	96.1	86.6	91.9	95.7	96.6
	pill	90.0	98.9	89.1	88.8	99.8	99.9	97.9	98.3
	screw	98.7	93.9	85.0	96.3	90.7	99.7	100.0	100.0
	toothbursh	99.2	100.0	100.0	98.6	97.5	95.2	100.0	100.0
	transistor	87.6	93.1	91.0	91.1	99.8	99.1	98.6	99.2
	zipper	85.9	100.0	99.9	95.1	99.9	98.5	99.4	99.7
物体类均值		88.8	97.4	91.4	94.7	96.6	98.0	98.2	98.6
纹理类	carpet	86.6	97.0	76.3	92.9	100.0	100.0	90.8	95.3
	grid	95.7	99.9	100.0	84.0	96.2	97.6	100.0	100.0
	leather	86.2	100.0	99.9	97.1	95.4	97.7	96.5	98.9
	tile	88.2	99.6	94.4	99.4	100.0	98.7	100.0	100.0
	wood	98.2	99.1	95.9	99.8	99.1	99.6	100.0	100.0
纹理类均值		91.0	99.1	93.3	94.6	98.1	98.7	97.5	98.8
总均值		89.5	98.0	92.0	94.7	97.1	98.3	97.9	98.7

表 3 以像素级别的 AUROC 评价了本文提出方法的性能。包括重建网络的像素级别的 AUROC 结果以及当特征网络耦合块为 3 时,提出方法的像素级别的 AUROC 结果。可以看到,单独使用重建网络得到的像素级 AUROC 为 90.9%;融合特征网络后,像素级 AUROC 为

94.2%,提高了 3.3%,特别地,部分类别提升超过了 10%。综合图像级别和像素级别 AUROC 的实验表明了本文提出方法的有效性。值得注意的是实验进一步证明了基于特征的方法能够有效地对重建方法提供补充,且这种补充并非来自额外预训练数据的支持。

表 3 加入特征融合后的网络与原重建网络的比较

Tab. 3 Comparison of the network after adding feature fusion with the original reconstructed network

分类	样本	重建网络的像素级 AUROC/%	特征融合后网络的像素级 AUROC/%	差值/%
物体类	bottle	97.5	98.5	1.0
	cable	94.4	96.1	1.7
	capsule	95.5	97.1	1.6
	hazelnut	92.9	97.1	4.2
	metal nut	95.7	98.0	2.3
	pill	93.5	95.4	1.9
	screw	77.6	83.0	5.4
	toothbursh	98.5	98.6	0.1
	transistor	80.7	94.3	13.6
	zipper	94.6	95.1	0.5
物体类均值		92.1	95.3	3.2
纹理类	carpet	74.0	78.8	4.8
	grid	97.6	98.5	0.9
	leather	91.7	94.5	2.8
	tile	89.2	95.1	5.9
	wood	89.9	93.2	3.3
纹理类均值		88.5	92.0	3.5
总均值		90.9	94.2	3.3

3.3 消融实验

为了检验各模块及参数对提出方法性能的影响,本文开展了针对 SSIM 损失、流模型的耦合块个数、内置神经元个数等关键参数的消融实验。

首先我们去掉了 SSIM 损失函数之后,仅使用 MSE 损失作为重建网络的损失函数,得到如表 4 所示的结果。

表 4 不同损失对重建结果的影响

Tab. 4 Impact of different losses on reconstruction results

重建网络结果	图像级结果/%	像素级结果/%
MSE 损失和 SSIM 损失	97.9	90.9
MSE 损失	97.4	90.3

由表 4 可知,SSIM 损失对于网络重建质量的提升有一定的作用。本文认为,因为 SSIM 损失衡量的是两幅图像的相似度,且从亮度、对比度和结构 3 个方面比较两幅图片的差异,符合重建的思想,对重建的质量有了一定的提升。

此外,本文在确定了损失之后进一步研究了流模型耦合块个数对重建网络效果的提升。本文将耦合块个数从 2 增加到 7,表 5 为耦合块从 2 块增加到 7 块后网络性能的表现。实验结果表明,当耦合块数量为 3 时,提升的效果最佳。本文认为,过多的耦合块数量可能会影响最高层特征图的映射,导致信息的丢失,造成结果不准确。

最后,本文研究了内置神经元个数对融合网络效果提升的影响。本文分别在耦合块数量为 2、3、4 时,将内置神经元个数从 1 024 增加到 2 048,结果如表 6 所示。可以看出,当神经元数量为 1 024 时,耦合块数量为 3 时的图像级和像素级结果最好。当神经元数量为 2 048 时,图像级结果随着耦合块数量的增加而变好,但整体比不过神经元数量为 1 024 时的结果。像素级结果在耦合块为 3 时结果最好。本文认为,过多的神经元数量可能增加了流模型变换程中的复杂度,从而影响了特征图的映射,导致效果略微降低。综合表 6 数据,当耦合块数量为 3、神经元数量为 1 024 的结果表现最佳。

表 5 耦合块个数对重建网络质量的提升

Tab. 5 Improvement of reconstructed network quality by the number of coupling blocks

耦合块 数量	重建网络的图像级 AUROC 总均值/%	融合网络的图像级 AUROC 总均值/%	差值/%	重建网络的像素级 AUROC 总均值/%	融合网络的像素级 AUROC 总均值/%	差值/%
2	97.9	98.5	0.6	90.9	93.5	2.6
3	97.9	98.7	0.8	90.9	94.2	3.3
4	97.9	98.5	0.6	90.9	93.5	2.6
5	97.9	98.5	0.6	90.9	93.6	2.7
6	97.9	98.6	0.7	90.9	93.7	2.8
7	97.9	98.5	0.6	90.9	93.6	2.7

表 6 当耦合块为 2、3、4 时,神经元数量对网络质量的提升

Tab. 6 Improvement of the network quality by neurons number when the coupling blocks are 2, 3 and 4

耦合块 数量	神经元 数量	重建网络的图像级 AUROC 总均值/%	融合网络的图像级 AUROC 总均值/%	差值/%	重建网络的像素级 AUROC 总均值/%	融合网络的像素级 AUROC 总均值/%	差值/%
2	1 024	97.9	98.5	0.6	90.9	93.5	2.6
	2 048	97.9	98.1	0.2	90.9	93.7	2.8
3	1 024	97.9	98.7	0.8	90.9	94.2	3.3
	2 048	97.9	98.2	0.3	90.9	93.9	3.0
4	1 024	97.9	98.5	0.6	90.9	93.5	2.6
	2 048	97.9	98.4	0.5	90.9	93.8	2.9

4 结 论

本文提出了一个基于特征分布学习与图像重建网络相融合的工业产品缺陷检测模型。该工作证明了基于特征方法和重建方法这两种主流无监督缺陷检测方法能够相互形成有效的补充,且性能的提升并非依靠特征方法采用的预训练数据。在重建网络中,本文提出了一种基于DWT/IDWT的网络特征分解和降采样的编码器-解码器方案,用于缓解传统编码器-解码器中存在的难以平衡细节重建效果和缺陷检测能力的问题。此外,本文提出了特征网络与重建网络的融合方案,该方案利用特征定位准确的特性,加入了基于流模型的融合检测模型,融合了重建

网络重建特性与流模型的优点。更为重要的是,本文提出的特征网络以重建网络的编码器作为特征提取器,因此在学习过程中无需在大数据集上预训练,一方面去除了在大规模数据集上预训练的工作,另一方面使编码器的设计更为灵活。本文的重建网络算法检测准确率为:图像级:97.9%,像素级:90.9%;融合网络算法检测准确率为:图像级:98.7%,像素级:94.2%。公用数据库上的大量实验表明,所提出的方法能够有效地针对工业产品缺陷执行有效的检测。

但是本文结果在像素级别准确率方面仍有不足之处,对于最新方法,仍有提高的空间。在未来,希望能继续改进模型,如使用更好的存储模块以及更高效的特征映射方法。

参 考 文 献:

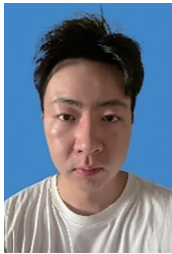
[1] NAYAK R, PATI U C, DAS S K. A comprehensive review on deep learning-based methods for video anomaly detection [J]. *Image and Vision Computing*, 2021, 106: 104078.

[2] PATCHA A, PARK J M. An overview of anomaly detection techniques: existing solutions and latest technological trends [J]. *Computer Networks*, 2007, 51(12): 3448-3470.

- [3] GONG D, LIU L Q, LE V, *et al.* Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection [C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea: IEEE, 2019: 1705-1714.
- [4] BERGMANN P, FAUSER M, SATTLEGGGER D, *et al.* Uninformed Students: student-teacher anomaly detection with discriminative latent embeddings [C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 4182-4191.
- [5] KINGMA D P, WELLING M. Auto-encoding variational Bayes [C]//*Proceedings of the 2nd International Conference on Learning Representations*. Banff, AB, Canada: Curran Associates Inc. , 2014.
- [6] SAKURADA M, YAIRI T. Anomaly detection using autoencoders with nonlinear dimensionality reduction [C]//*Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*. Gold Coast, Australia QLD, Australia: Association for Computing Machinery, 2014: 4-11.
- [7] CAO V L, NICOLAU M, MCDERMOTT J. A hybrid autoencoder and density estimation model for anomaly detection [C]//*Proceedings of the 14th International Conference on Parallel Problem Solving from Nature*. Edinburgh, UK: Springer, 2016: 717-726.
- [8] SABOKROU M, KHALOOEIM, FATHY M, *et al.* Adversarially learned one-class classifier for novelty detection [C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA: IEEE, 2018: 3379-3388.
- [9] COHEN N, HOSHEN Y. Sub-image anomaly detection with deep pyramid correspondences [J/OL]. *arXiv*, 2020: 2005.02357.
- [10] DEFARD T, SETKOV A, LOESCH A, *et al.* PaDiM: a patch distribution modeling framework for anomaly detection and localization [C]//*Proceedings of the ICPR International Workshops and Challenges Pattern Recognition*. Online: Springer, 2021: 475-489.
- [11] RUDOLPH M, WANDT B, ROSENHAHN B. Same but DifferNet: semi-supervised defect detection with normalizing flows [C]//*Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE, 2021: 1906-1915.
- [12] LIU P J, ZHANG H Z, ZHANG K, *et al.* Multi-level wavelet-CNN for image restoration [C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR)*. Salt Lake City, UT, USA: IEEE, 2018: 773-782.
- [13] REZENDE D J, MOHAMED S. Variational inference with normalizing flows [C]//*Proceedings of the 32nd International Conference on Machine Learning*. Lille, France: JMLR.org, 2015: 1530-1538.
- [14] BERGMANN P, FAUSER M, SATTLEGGGER D, *et al.* MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection [C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA: IEEE, 2019: 9584-9592.
- [15] KINGMA D P, BA J. Adam: a method for stochastic optimization [C]//*Proceedings of the 3rd International Conference on Learning Representations*. San Diego, CA, USA: Curran Associates Inc. , 2015.
- [16] HOU J L, ZHANG Y Y, ZHONG Q Y, *et al.* Divide-and-assemble: learning block-wise memory for unsupervised anomaly detection [C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, 2021: 8771-8780.
- [17] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DRAEM—A discriminatively trained reconstruction embedding for surface anomaly detection [C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, 2021: 8310-8319.
- [18] HUANG C Q, XU Q W, WANG Y F, *et al.* Self-supervised masking for unsupervised anomaly detection and localization [J]. *IEEE Transactions on Multimedia*, 2023, 25: 4426-4438.
- [19] LI C L, SOHN K, YOON J, *et al.* CutPaste: self-supervised learning for anomaly detection and localization [C]//*Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE, 2021: 9659-9669.

- [20] GUDOVSKIY D, ISHIZAKA S, KOZUKA K. CFLOW-AD: real-time unsupervised anomaly detection with localization *via* conditional normalizing flows [C]//*Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE, 2022: 1819-1828.

作者简介:



朱思宇,男,硕士研究生,2019年于武汉科技大学获得学士学位,主要研究方向为计算机视觉和缺陷检测。
E-mail:756831285@qq.com



朱磊,男,博士,副教授,2015年于华中科技大学获得博士学位,主要研究方向为目标检测与识别、语义分割、场景解析等。E-mail:zhulei@wust.edu.cn