

文章编号 1004-924X(2026)02-0296-13

高频引导的多尺度空间感知异常检测网络

张陈涛¹, 邹庆林¹, 刘 洋², 王亚飞², 王彩云³, 徐周毅¹, 郑高峰^{1*}

(1. 厦门大学 萨本栋微米纳米科学技术研究院, 福建 厦门 361000;

2. 内蒙古伊利实业集团股份有限公司, 内蒙古 呼和浩特 010000;

3. 国家乳业技术创新中心, 内蒙古 呼和浩特 010000)

摘要:针对现有工业异常检测算法对小尺寸缺陷检测精度低、多尺度特征提取能力弱、异常分割精度低等问题,提出一种结合高频残差引导和多尺度注意力特征融合的工业异常检测网络。首先,针对传统全频处理导致高频细节丢失的问题,设计了频域分离策略,利用高斯核滤波提取高频残差特征,强化网络对微小异常的检测能力;其次,针对常规卷积网络对复杂纹理的表征能力不足、异常与背景区分度不高的问题,在判别网络的编码器阶段嵌入全局增强的多尺度注意力模块 GEMA,通过并行双路径捕获水平与垂直方向的多尺度局部信息,强化不同空间位置的显著特征,提升复杂纹理背景下的特征判别性;最后,在判别网络的解码器阶段集成坐标注意力模块 CoordAtt,通过分解坐标轴动态调制特征权重,实现异常区域的精准空间定位。实验表明,在 MVTec AD 公开数据集上,所改进模型的图像级平均 AUROC 为 98.6%,像素级别的平均 AUROC 和 AP 分别为 97.6% 和 73.2%,有效提高了工业异常检测的效果。

关键词:工业异常检测;高频分量引导;多尺度空间感知;注意力机制

中图分类号:TH39;TP391.4 **文献标识码:**A

doi:10.37188/OPE.20263402.0296

CSTR:32169.14.OPE.20263402.0296

Multi-scale spatial sensing anomaly detection network guided by high frequency

ZHANG Chentao¹, ZOU Qinglin¹, LIU Yang², WANG Yafei², WANG Caiyu³,

XU Zhouyi¹, ZHENG Gaofeng^{1*}

(1. Pen-Tung Sah Institute of Micro-Nano Science and Technology, Xiamen University,
Xiamen 361000, China;

2. Inner Mongolia Yili Industrial Group Limited by Share Ltd, Hohhot 010000, China;

3. National Center of Technology Innovation for Dairy, Hohhot 010000, China)

* Corresponding author, E-mail: zheng_gf@xmu.edu.cn

Abstract: Aiming at the problems of low detection accuracy for small-sized defects, weak multi-scale feature extraction ability and low anomaly segmentation accuracy of existing industrial anomaly detection algorithms, an industrial anomaly detection network combining high-frequency residual guidance and multi-scale attention feature fusion was proposed. Firstly, aiming at the problem of high-frequency detail loss

收稿日期:2025-09-17; **修订日期:**2025-10-19.

基金项目:国家乳业技术创新中心项目(No. 2024-JSGG-008);内蒙古自治区中央引导地方科技发展资金项目(No. 2024ZY0074)

caused by traditional full-frequency processing, a frequency-domain separation strategy was designed. Gaussian kernel filtering was utilized to extract high-frequency residual features, enhancing the network's detection ability for minor anomalies. Secondly, aiming at the problems of insufficient representation ability of complex textures and low discrimination between anomalies and backgrounds in conventional convolutional networks, a globally enhanced multi-scale attention module GEMA is embedded in the encoder stage of the discriminative network. It captures multi-scale local information in the horizontal and vertical directions through parallel dual-path, enhancing the salient features at different spatial positions. Improve the feature discriminability in complex texture backgrounds; Finally, in the decoder stage of the discriminant network, the coordinate attention module CoordAtt is integrated. By decomposing the coordinate axes and dynamically modulating the feature weights, precise spatial positioning of abnormal areas is achieved. Experiments show that on the MVTec AD public dataset, the average AUROC at the image level of the improved model is 98.6%, and the average AUROC and AP at the pixel level are 97.6% and 73.2% respectively, effectively improving the effect of industrial anomaly detection.

Key words: industrial anomaly detection; high-frequency component guidance; multi-scale spatial perception; attention mechanism

1 引言

工业产品表面异常不仅会影响产品外观与性能,还可能诱发安全隐患,产品表面质量已成为生产质量管控的核心环节^[1-2],以及衡量产品市场竞争力的重要指标^[3]。早期的工业品异常检测主要靠人工进行^[4],但传统人工质量检测存在效率低^[5]、检测质量受作业人员状态波动影响、人力管理成本攀升等问题,难以满足现代工业高效质检要求。依赖人工设计特征的传统视觉算法,在应对金属拉丝、皮革纹路等复杂纹理背景时,其检测效果、稳定性易受光照变化等环境因素干扰^[6-7]。相较之下,基于深度学习的检测方法能够自动学习人工难以设计的缺陷特征^[8-9],在处理背景复杂、缺陷微小的工业图像时展现出突破传统方法的检测能力^[10-11],已经在缺陷检测领域得到了广泛应用^[12]。当前基于深度学习的工业异常检测方法主要分为有监督学习和无监督学习两类^[13]。无监督学习通过挖掘数据内在分布规律自主识别异常来判断样本是否存在异常。而在有监督学习中,网络利用已标记的异常信息进行训练,准确地识别图像中的缺陷位置、类别和缺陷大小。

在监督式工业异常检测领域中,以卷积神经网络(Convolutional Neural Network, CNN)为核心的深度学习框架已成为最常见的算法模式^[14]。

典型研究如李良福等^[15]构建的CNN-DDBC双通道分类器架构,通过特征空间映射实现桥梁裂纹的精准识别。Chu和Kitani^[16]提出了一个半监督的框架,用于在数据显著不平衡的情况下检测异常。Zeng等^[17]将异常检测视为目标检测的子集,并设计了一个基于参考的缺陷检测网络RDDN,使用模板参考和上下文参考来检测异常。Wan等^[18]为了有效利用少量的异常信息,提出了用于训练不平衡数据分布的logit诱导损失和用于表征异常特征的异常捕获模块ACM。以上监督式的工业异常检测方法需依赖大量精确标注的缺陷数据集进行参数优化,面对工业检测场景中普遍存在的数据长尾分布、缺陷模式未知及缺陷形态不规则等挑战,构建符合模型训练要求的大规模人工标注数据集较为困难,限制了监督式方法在工业异常检测中的应用。

相比于有监督算法,无监督深度学习算法模型设计更加简单、高效,无需标签数据,可以解决有监督方法需要大量标注数据的问题。Bergmann等^[19]率先构建基于自编码器的重建模型,通过结构相似性损失强化异常敏感度,但存在异常区域误重建的固有缺陷。为此,Collin等^[20]引入跳跃连接与污渍噪声注入,显著提升了模型鲁棒性。在数据增强方向,Zavrtanik等^[21]提出的DRAEM框架通过物理渲染合成异常样本,与Li等^[22]的CutPaste自监督策略共同推动了异常合

成策略的进步,但合成数据的真实性仍有待提升。Zavrtanik 等^[23]的 DSR 网络在特征空间构建异常子空间,标志着检测范式从图像级向特征级的转变。郭艳等^[24]在 YOLO 框架中引入多维注意力机制,增强对微小缺陷的特征提取能力,但对多尺度缺陷的提取能力依旧较弱。王素琴等^[25]融合 RGB 与频域特征检测相似背景下的细微缺陷,但并未利用频域特征对微小缺陷的检测进行改善。尽管上述方法取得显著进展,但当前无监督工业图像异常检测领域仍存在亟待突破的技术瓶颈:算法的多尺度特征提取能力弱、对小尺寸缺陷检测敏感度低、异常分割精度不高三个方面。

针对上述问题,本文设计了一个高频特征引导的无监督工业品表面异常检测算法。该算法利用大量正常工业品样本图像进行模型训练,构建图像的低维表征空间以实现异常图像的精准重构,继而通过对比分析重构图像与原始图像的差异特征实现精准的异常识别。为提升细微异常检测的鲁棒性,设计了频域分离模块,该模块采用高斯核卷积实现图像频域解耦,从而构建高频残差特征通道以强化细微异常表征能力。此外,为提升网络对局部细节和全局信息的融合能力,设计了多尺度增强模块(Global-Enhanced Multi-scale Augmentation module, GEMA),该模块通过扩展卷积与全局上下文融合的双路径架构,提升局部特征分辨力与全局语义关联性。在此基础上,为进一步提升异常区域的空间定位精度,设计了坐标注意力模块(Coordinate Atten-

tion Module, CoordAtt),该模块基于空间维度动态权重分配机制,建立坐标轴分解的注意力聚焦模式,最终实现异常区域的高精度空间定位。

2 本文方法

异常检测网络的总体结构如图 1 所示,由一个重建子网络和一个判别子网络组成。对重建子网络进行训练,在保持输入图像非异常区域不变的情况下,将图像中的异常区域重建为正常图像。同时,判别子网络学习联合重建-异常的输入,并从级联的重建和异常图像中输出正确的异常分割图。网络通过高频重建子网络与多尺度判别子网络的异构特征交互实现工业异常精准定位。首先采用频域解耦策略对输入图像进行物理先验引导的特征分离,利用固定参数的高斯核卷积主动剥离低频成分,将蕴含表面细节的高频特征单独输入重建子网络进行重建训练,有效强化对微小缺陷的敏感度。判别子网络在五级下采样编码结构中嵌入 GEMA 模块实施通道分组与双路径特征优化,通过自适应方向池化捕获空间长距离依赖,并结合深度可分离卷积提取局部上下文,在特征图分组归一化后通过动态权重融合实现多尺度感知增强。解码阶段引入坐标注意力机制 CoordAtt,通过分解 H/W 方向的特征向量生成空间注意力权重,在逐级上采样过程中与编码器多级特征进行跳跃连接,形成"全局粗定位-局部精修"的渐进式检测路径。网络通过特征级联实现信息互补,最终经卷积层融合生成像素级异常热力图。

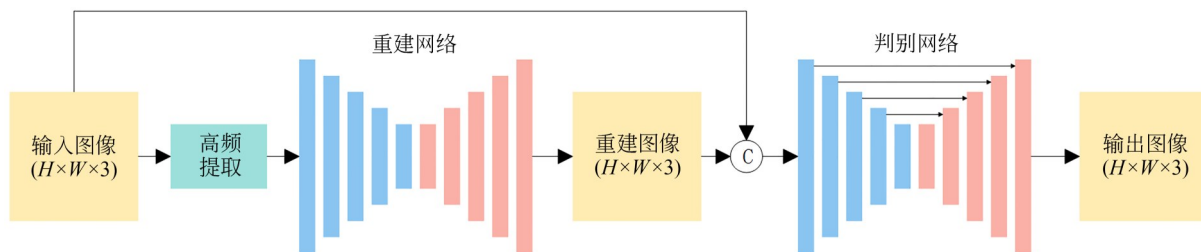


图1 网络总体结构

Fig. 1 Overall structure of the network

本文采用异常合成的方式进行数据增强,由于网络训练时异常样本不需要真实地反映目标

域中的异常外观,只需要偏离正常分布的外观,允许学习合适的距离函数,通过偏离正态性来识

别异常^[21]。本文采用的异常模拟方法遵循这一范式,训练时由Perlin噪声发生器^[26]生成噪声图像,捕捉多种异常形状,并通过随机均匀采样的阈值二值化为异常掩膜。发生器参数设置如下:频率设为0.01~0.1,这个范围是通过在不同值下生成纹理并进行视觉评估确定的,低频对应大面积异常、高频对应细节异常;振幅为0.2~0.8,通过在相同频率下调整振幅并评估视觉效果确定:小于0.2导致纹理过于平滑,大于0.8导致不自然的过度细节;八度固定3~5层,少于3层缺乏纹理层次,超过5层计算开销增加但视觉改善不明显。掩膜从与输入图像分布无关的异常外观数据集中采样异常,并将其叠加于无异常图像上,生成异常图像用于训练。融合叠加的公式如下:

$$I_a = \bar{M}_a \otimes I + \beta(M_a \otimes A), \quad (1)$$

其中: A 为异常外观, I 为原始正常图像, M_a 为掩膜, β 为不透明度参数,随机采样于 $[0.1, 1.0]$,下限0.1确保异常具有足够视觉辨识度,上限1.0覆盖明显缺陷的情况, $\otimes$ 为逐元素相乘。

异常外观来源于DTD数据集^[27],其包含47类纹理属性如条纹、斑点、网格等,共5 640张图像。

2.1 高频特征提取模块

高频特征提取模块的核心在于通过物理先验引导的频域解耦策略,将工业图像中的高频细节信息显式分离,从而强化网络对微小缺陷的敏感度。模块结构如图2所示,输入图像特征首先通过高斯模糊层分离出低频分量,高频分量由原始输入减去低频分量获得。

该模块以高斯低通滤波为基础,构建频域自

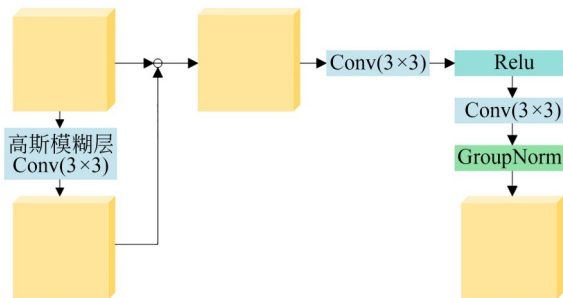


图2 高频特征提取模块示意图

Fig.2 Schematic diagram of the high-frequency feature extraction module

适应的特征分解机制。给定输入图像 $I \in \mathbb{R}^{H \times W \times C}$,首先通过固定参数的二维高斯核卷积提取低频分量:

$$L = I * G, \quad (2)$$

$$G(x, y) = \frac{1}{2\pi f^2} e^{-\frac{x^2 + y^2}{2f^2}}, \quad (3)$$

其中: $f=1.0$ 为控制滤波器的截止频率, f 过低会导致滤波器截止频率过高,会过度保留高频噪声,降低信噪比, f 过高会过度平滑微小缺陷的边缘特征,导致关键纹理信息丢失;核尺寸 $k=3$ 确保对常见工业缺陷特征的有效捕获,相比于 $k=5, k=7$ 等大核卷积,其计算效率更高,符合实时性要求。

公式(2)的操作在频域等价于理想低通滤波的近似实现:

$$F(L) = F(I) \otimes F(G), \quad (4)$$

其中: F 表示傅里叶变换, $\otimes$ 为逐元素相乘。

高频残差分量通过空间域残差计算获得:

$$H_{\text{high}} = I - L. \quad (5)$$

该残差在频域对应高通滤波结果:

$$F(H_{\text{high}}) = F(I) \otimes F(1 - F(G)). \quad (6)$$

其梯度场可分解为:

$$\nabla H_{\text{high}} = \nabla I - \nabla L. \quad (7)$$

这一特性使得模块能够保留原始图像中由缺陷引起的边缘突变,如裂纹的锐利梯度,同时抑制低频背景,如均匀材质表面的平滑区域的干扰。

为增强高频特征的模型兼容性,对 H_{high} 进行标准化处理:

$$\hat{H}_{\text{high}} = \frac{H_{\text{high}} - \mu(H_{\text{high}})}{\eta(H_{\text{high}})}, \quad (8)$$

其中, μ, η 分别表示计算特征的均值与标准差。公式(8)将高频分量的数值分布约束至零均值单位方差区间,避免因其他因素导致的幅值波动影响重建网络的稳定性。

与传统可学习滤波器相比,固定高斯核参数具有明确的物理意义,核尺寸 k 与缺陷最小可检测尺寸 d_{min} 满足关系 $d_{\text{min}} \propto k\sigma$,这使模块在保持端到端训练优势的同时具备可解释性。由于低频分量 L 的梯度 ∇L 在平滑区域趋近于零,高频残差的梯度 ∇H_{high} 在缺陷边界处近似等于 ∇I ,使得重建网络在反向传播时能够通过梯度幅值强

化异常区域的权重更新。固定滤波策略避免了可学习卷积核在训练初期可能导致的频域泄露问题,其计算复杂度稳定为 $O(k^2CHW)$ 。

传统边缘检测算子虽能提取高频特征,但其固定的方向敏感性与频带选择缺乏自适应性。高频特征提取模块通过残差计算实现宽频带覆盖,且后续可通过学习动态调整频带权重。

2.2 GEMA 模块

提出了一种全局增强多尺度注意力模块,该模块在经典 EMA 架构^[28]的基础上,通过引入动态感受野扩展与全局上下文感知机制,显著提升了工业缺陷检测任务中对多尺度异常特征的捕获能力。GEMA 构建了双路径特征交互体系:其一通过扩展卷积核实现自适应感受野调整,其二通过全局上下文注入增强跨区域语义关联。

GEMA 模块结构如图 3 所示,在特征处理路径上,首先将输入特征 $X \in R^{B \times C \times H \times W}$ 沿通道维度划分为 G 组子特征 $\{X_g\}_G$,每组特征分别进行双分支处理。

第一分支采用扩展率为 d 的空洞卷积进行特征变换得到的特征图上各点值为:

$$F_d^{(g)} = \sum_{k=-r}^r \sum_{l=-r}^r W_d(k, l) \cdot X_g(i + d \cdot k, j + d \cdot l), \quad (9)$$

其中: $r=1$ 为卷积核半径,该设置能够在保持较小计算复杂度的前提下,与动态调整的空洞率 d 协同工作, $d \in \{1, 2, 4, 8\}$ 为动态调整的空洞率,确保从局部细节到全局上下文的完整覆盖, X_g 为输入特征, W_d 为权重矩阵。

该机制实现了感受野的动态自适应调节。每个特征组通过独立的可学习参数控制空洞率大小,在训练过程中根据特征尺度特性自动优化。具体而言,模块将不同空洞率的特征图进行加权融合,并通过专门的尺度感知损失函数约束感受野调整过程,确保网络能够根据缺陷尺寸自动选择最优感知范围。尺度感知损失函数如下:

$$L_{\text{scale}} = \sum_{g=1}^G \left\| \nabla_d F_d^{(g)} \right\|^2. \quad (10)$$

该设计使得半径为 1 卷积核的实际感受野扩展至 $(2rd+1)(2rd+1)$,在保持参数量不变的前提下,显著增强了对大尺度连续性缺陷的捕获能力。

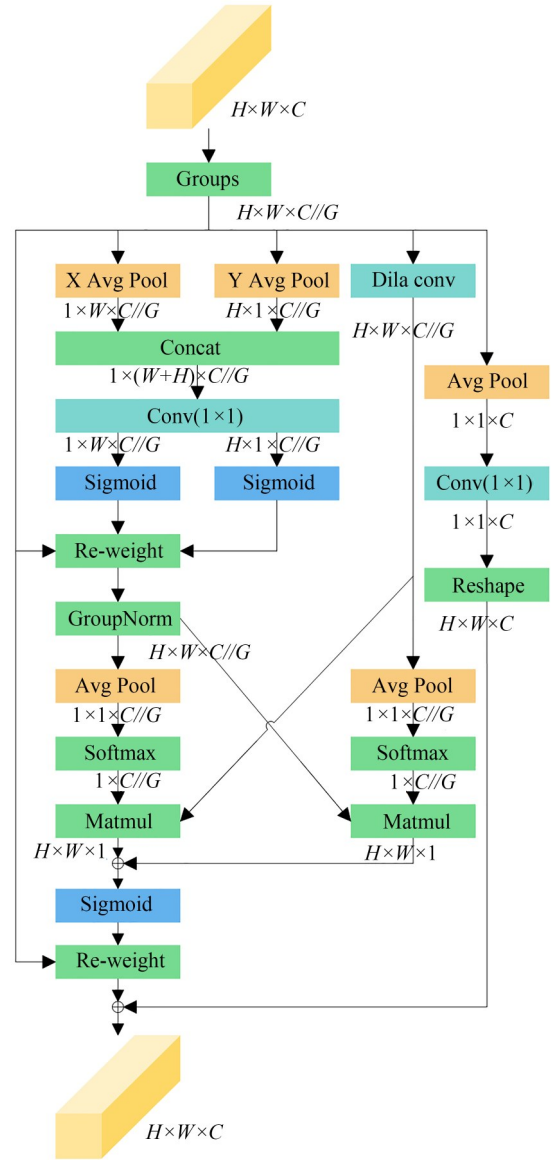


图 3 GEMA 模块示意图

Fig. 3 Schematic diagram of GEMA module

第二分支通过水平与垂直维度的非对称池化操作生成方向敏感注意力图:

$$A_h^{(g)} = \frac{1}{W} \sum_{j=1}^W X_g(:, j), \quad (11)$$

$$A_v^{(g)} = \frac{1}{H} \sum_{i=1}^H X_g(i, :), \quad (12)$$

其中: X_g 为输入特征, A 为注意力权重矩阵。

二者经 1×1 卷积与 Sigmoid 激活后形成空间注意力权重 $W_s^{(g)} \in [0, 1]^{H \times W}$,该权重矩阵可表达为:

$$W_s^{(g)} = \sigma \left(W_c * \left[A_h^{(g)}; A_v^{(g)} \right] \right), \quad (13)$$

其中: σ 表示 Sigmoid 函数,*代表卷积操作。

为克服局部注意力对全局上下文感知的局限性,GEMA引入全局上下文卷积分支,通过全局平均池化提取特征图的全局统计量:

$$G = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X(:, :, i, j). \quad (14)$$

该全局特征经 1×1 卷积映射后与局部特征进行加权融合,最终输出可表示为:

$$X_{out} = \sum_{g=1}^G (F_d^{(g)} \otimes W_s^{(g)}) + \phi(G) \otimes X, \quad (15)$$

其中: $\phi(\cdot)$ 表示上下文特征变换函数, $\otimes$ 为逐元素相乘。该融合策略在数学上保证了局部细节特征与全局语义信息的协同优化,其梯度传播可分解为:

$$\frac{\partial L}{\partial X} = \sum_{g=1}^G \left(\frac{\partial L}{\partial F_d^{(g)}} \cdot W_s^{(g)} \right) + \phi'(G) \cdot \frac{\partial L}{\partial X_{out}}. \quad (16)$$

微分特性使得模块在反向传播过程中,能够通过局部注意力梯度与全局上下文梯度优化特征表达。

2.3 CoordAtt 模块

CoordAtt 模块结构如图4所示,其将传统二维空间注意力分解为两个正交的一维操作,通过坐标轴解耦理论实现高效且精准的空间定位。给定输入特征 $F \in \mathbb{R}^{B \times H \times W \times C}$,模块首先沿水平与垂直方向分别进行坐标嵌入:

$$z_h(c, i) = \frac{1}{W} \sum_{j=1}^W F(c, i, j), \quad (17)$$

$$z_w(c, j) = \frac{1}{H} \sum_{i=1}^H F(c, i, j). \quad (18)$$

公式(17)、公式(18)的操作可视为对空间位置信息的显式编码,其中 z_h 捕获每行像素的统计特性, z_w 编码每列像素的分布规律。随后,通过跨通道交互对拼接后的特征进行压缩与重构:

$$z = \text{Conv}_{1 \times 1} \left(\text{BN} \left([z_h; z_w] \right) \right), z \in \mathbb{R}^{B \times (H+W) \times (C/r)}, \quad (19)$$

其中: r 为压缩率,BN表示批归一化层。利用 h-swish 激活函数 $\delta(x) = x \cdot \text{Sigmoid}(\beta x)$, β 为可学习参数,引入非线性变换,增强特征的表达能。

重构阶段通过分离卷积生成方向注意力权重:

$$a_h = \sigma(\text{Conv}_{1 \times 1}^h(z_h)), \quad (20)$$

$$a_w = \sigma(\text{Conv}_{1 \times 1}^w(z_w)), \quad (21)$$

其中: σ 为 Sigmoid 函数, $\text{Conv}_{1 \times 1}^h$ 和 $\text{Conv}_{1 \times 1}^w$ 分别为水平与垂直方向的卷积,对应池化后的特征。

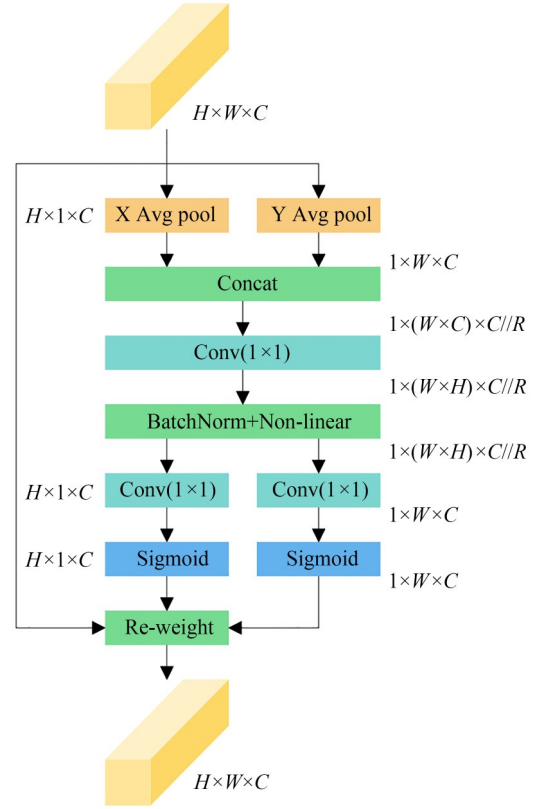


图4 CA模块示意图

Fig. 4 Schematic diagram of CA module

最终特征通过外积运算实现空间调制:

$$F'_{c,i,j} = F_{c,i,j} \cdot a_h(c, i) \cdot a_w(c, j). \quad (22)$$

该设计在数学上等价于对原始特征的逐位置缩放,其梯度传播过程可分解为:

$$\frac{\partial L}{\partial F_{c,i,j}} = \frac{\partial L}{\partial F'_{c,i,j}} \cdot a_h(c, i) a_w(c, j) + \sum_{k,l} \frac{\partial L}{\partial a_h(c, k)} \frac{\partial a_h(c, k)}{\partial F_{c,i,j}} + \sum_{m,n} \frac{\partial L}{\partial a_w(c, n)} \frac{\partial a_w(c, n)}{\partial F_{c,i,j}}. \quad (23)$$

第二、三项在反向传播时强化了异常边界区域的参数更新,使模块能够精确捕捉边缘模糊的缺陷形态。相较于传统空间注意力机制,CoordAtt 的参数量从 $O(C^2 + CHW)$ 降低至 $O(C^2/r + C(H+W))$,在保持定位精度的同时显著提升了计算效率,尤其适用于高分辨率工业

图像的实时处理需求。解码器中多级 CoordAtt 层的堆叠可逐步细化异常区域的空间边界,形成从粗到精的定位优化路径。

3 实 验

3.1 实验数据集

算法主要在 MVTec AD 数据集上进行实验验证,MVTec AD 是工业异常检测领域广泛使用的基准数据集,由德国 MVTec 公司于 2019 年发布。该数据集包含 15 个类别,其中包括 5 种纹理和 10 类物体共计 5 354 张高分辨率图像,训练集包含 3 629 张无缺陷的正常样本,测试集则包含 1 725 张涵盖 70 余种工业常见缺陷的异常样本,包括划痕、凹痕、结构变形、表面污染等多种人工模拟的真实工业缺陷类型,并提供像素级标注以支持无监督异常定位任务。所有图像都是分辨率为 $700 \times 700 \sim 1\,024 \times 1\,024$ 的三通道彩色图像。此外,为了验证算法的泛化性,还在 VisA 数据集^[29]上做实验验证。该数据集由 10 821 张高分辨率 RGB 图像组成,包括 9 621 张正常样本和 1 200 张异常样本,分为 12 种工业品类别,异常类型多样。上述两个数据集的一个关键设计理念是训练集仅包含正常样本,这种设定模拟了实际工业场景中缺陷样本稀缺的现实情况。

3.2 评价指标

采用受试者工作特性曲线下面积(Area Under The Receiver Operating Characteristic Curve, AUROC)作为评价指标。AUROC 是目前无监督工业异常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。在实验中计算图像级 AUROC 和像素级 AUROC 两个指标。图像级 AUROC 用于异常分类,像素级 AUROC 用于评估异常定位。然而,在只有一小部分像素异常的表面异常检测场景中,像素级 AUROC 并不能很好地反映定位精度。原因是假阳性率由非常高的非异常像素数量所主导,因此尽管有假阳性检测,假阳性率仍保持在较低水平。所以,额外加上了像素级的平均精度度量 AP 用于检测精度评估,它在表面异常检测中更适用于高度不平衡的类的检测精度判别。上述指标的定义如式(24)~式(28):

$$TPR = \frac{TP}{TP + FN}, \quad (24)$$

$$FPR = \frac{FP}{FP + TN}, \quad (25)$$

$$Image - AUROC = \int_0^1 TPR(FPR^{-1}(r))dr, \quad (26)$$

$$Pixel - AUROC = \sum_{i=1}^{N-1} (FPR_{i+1} - FPR_i) \times (TPR_i + TPR_{i+1})/2, \quad (27)$$

$$AP = \sum_{k=1}^K (R_k - R_{k-1}) \cdot P_k, \quad (28)$$

其中: TPR 为真阳率, FPR 为假阳率, TP 表示真阳性样本, TN 表示真阴性样本, FP 表示假阳性样本, FN 表示假阴性样本, N 为像素总数, P_k 为第 k 个阈值点的精确率, R_k 为对应召回率。

3.3 实验设置

所有实验均在搭载 Ubuntu 18.04 LTS 操作系统的计算平台上进行。模型的训练过程使用两块显存为 11 GB 的 NVIDIA GeForce RTX 2080 Ti 显卡。针对数据集中的每一个工业品类型,模型均独立训练至 700 个完整轮次。选用 Adam 优化器进行参数优化,其初始学习率设置为 0.001。为了提高模型的收敛效率并防止过拟合,本文采用了学习率衰减策略:在训练进程分别达到第 490, 560 和 630 个轮次时,学习率衰减为当前值的 0.2 倍,确保模型在训练后期能够更精细地调整参数,能够更好地收敛。训练的批量大小设置为 16。损失函数由 L2 损失和结构相似性损失 (Structural Similarity Index Measure, SSIM) 共同组成,通过反向传播驱动网络更新。

3.4 实验结果与分析

为了全面评估所提出方法的性能,在工业异常检测数据集 MVTec AD 和 VisA 上,与当前主流的工业异常检测方法进行对比实验,包括 DRAEM^[21], DSR^[23], RIAD^[30], PaDim^[31], US^[32], MMR^[33], SuperSimpleNet^[34], SPADE^[35] 和 DiffAD^[36], 实验结果如表 1~表 3 所示。由表 1 和表 2 可知,本文所提出的方法在 MVTec AD 数据集上检测的图像级平均 AUROC 达到了 98.6%, 像素级别的平均 AUROC 和 AP 分别达到了 97.6% 和 73.2%, 优于其他的异常检测算法。除了平均指标外,木板、金属网、瓷砖等工业品类别

的各项指标均高于其他的主流异常检测算法。其中,皮革、瓷砖、拉链等类别的图像级 AUROC 均达到了 100%;此外,金属网类别的像素级 AUROC 也达到了 100%,充分证明了其在特定场景下的检测有效性。由表 3 可知,本文提出的方法

在 VisA 数据集上检测的图像级平均 AUROC 达到了 91.2%,优于主流的异常检测算法。由表 4 可知本文所提出的方法在参数量以及运行效率方面与主流算法处于同一水平。

图 5 展示了本文模型对 MVTec AD 数据集

表 1 各主流模型在 MVTec AD 数据集上的图像级 AUROC 结果

Tab. 1 Image-level AUROC results of each mainstream model on the MVTec AD dataset

方法类别	DRAEM ^[21]	RIAD ^[30]	PaDim ^[31]	US ^[32]	MMR ^[33]	DSR ^[23]	SuperSimple Net ^[34]	本文方法
胶囊(Capsule)	98.5	88.4	91.3	86.1	96.9	98.3	98.7	97.2
瓶子(Bottle)	99.2	99.9	99.9	99.0	100	99.6	100	96.7
地毯(Carpet)	97.0	84.2	99.8	91.6	99.6	99.6	98.4	98.1
皮革(Leather)	100	100	100	88.2	100	99.3	100	100
药片(Pill)	98.9	83.8	93.3	87.9	98.2	98.9	98.1	98.5
晶体管(Transistor)	93.1	90.9	97.4	81.8	95.1	96.3	99.9	95.7
瓷砖(Tile)	99.6	98.7	98.1	99.1	98.7	100	99.7	100
电缆(Cable)	91.8	81.9	92.7	86.2	97.8	95.3	98.1	95.1
拉链(Zipper)	100	98.1	90.3	91.9	97.6	98.5	99.6	100
牙刷(Toothbrush)	100	100	96.1	95.3	100	100	92.2	100
金属螺母(Metal nut)	98.7	88.5	98.7	82.0	99.9	99.1	99.5	99.8
榛子(Hazelnut)	100	83.3	92.0	93.1	100	97.7	99.8	100
螺钉(Screw)	93.9	84.5	85.8	54.9	92.5	95.5	92.9	97.8
金属网(Grid)	99.9	99.6	96.7	81.0	100	100	99.3	100
木板(Wood)	99.1	93.0	99.2	97.7	99.1	94.7	99.3	100
平均(Average)	98.0	91.7	95.5	87.7	98.4	98.2	98.4	98.6

表 2 MVTec AD 数据集上像素级 AUROC 和 AP 的结果

Tab. 2 Results of pixel-level AUROC and AP on the MVTec AD dataset

方法类别	DRAEM ^[21]	RIAD ^[30]	PaDim ^[31]	US ^[32]	MMR ^[33]	DSR ^[23]	本文方法
胶囊(Capsule)	94.3/49.4	98.2/38.2	98.6/46.7	96.8/25.9	98.0/—	91.0/—	90.9/51.2
瓶子(Bottle)	99.1/86.5	98.4/76.4	98.2/77.3	97.8/74.2	98.3/—	98.8/—	99.2/90.2
地毯(Carpet)	95.5/53.5	96.3/61.4	99.0/60.7	93.5/52.2	98.8/—	96.0/—	97.3/71.7
皮革(Leather)	98.6/75.3	99.4/49.1	99.0/53.5	97.8/40.9	99.2/—	99.5/—	99.1/70.2
药片(Pill)	97.6/48.5	95.7/51.6	95.7/61.2	96.5/62.0	98.4/—	94.2/—	98.1/52.2
晶体管(Transistor)	90.9/50.7	87.7/39.2	97.6/72.0	73.7/27.1	90.2/—	80.3/—	90.8/57.3
瓷砖(Tile)	99.2/92.3	89.1/52.6	94.1/52.4	92.5/65.3	95.6/—	98.6/—	99.4/98.1
电缆(Cable)	94.7/52.4	84.2/24.4	96.7/45.4	91.9/48.2	95.4/—	97.7/—	96.7/65.6
拉链(Zipper)	98.8/81.5	97.8/63.4	98.4/58.2	95.6/36.1	98.0/—	98.4/—	98.0/76.1
牙刷(Toothbrush)	98.1/44.7	98.9/50.6	98.8/54.7	97.9/37.7	98.4/—	99.5/—	98.5/61.2
金属螺母(Metal nut)	99.5/96.3	92.5/64.3	97.3/77.4	97.2/83.5	95.9/—	94.1/—	99.1/92.6
榛子(Hazelnut)	99.7/92.9	96.1/33.8	98.1/61.1	98.2/57.8	98.5/—	99.1/—	99.6/91.8
螺钉(Screw)	97.6/58.2	98.8/43.9	98.4/21.7	97.4/7.8	99.5/—	98.1/—	99.4/65.6
金属网(Grid)	99.7/65.7	98.8/36.4	97.1/35.7	89.9/10.1	99.0/—	99.6/—	100/71.4
木板(Wood)	96.4/77.7	85.8/38.2	94.1/46.3	92.1/53.3	94.8/—	91.5/—	97.6/82.2
平均(Average)	97.3/68.4	94.2/48.2	97.4/55.0	93.9/45.5	97.2/—	95.8/—	97.6/73.2

表 3 VisA 数据集上图像级 AUROC 的结果

Tab. 3 Results of image-level AUROC on the VisA dataset

Algorithm Category	DRAEM ^[21]	SPADE ^[35]	PaDiM ^[31]	DiffAD ^[36]	SimpleNet ^[34]	本文方法
Candle	91.8	91.0	91.6	90.4	95.6	93.8
Capsules	74.7	61.4	70.7	87.6	76.7	82.1
Cashew	95.1	97.8	93.0	81.4	91.7	88.8
Chewing gum	94.8	85.8	98.8	94.0	99.1	92.2
Fryum	97.4	88.6	88.6	87.1	95.3	96.3
Macaroni1	97.2	95.2	87.0	87.6	90.8	90.1
Macaroni2	85.0	87.9	70.5	90.7	65.2	89.7
PCB1	47.6	72.1	94.7	75.0	60.1	77.3
PCB2	89.8	50.7	88.5	94.6	93.3	92.3
PCB3	92.0	90.5	91.0	94.7	94.9	95.3
PCB4	98.6	83.1	97.5	97.7	98.2	97.1
Pipe fryum	100	81.1	97.0	92.7	93.3	99.6
Average	88.7	82.1	89.1	89.5	87.9	91.2

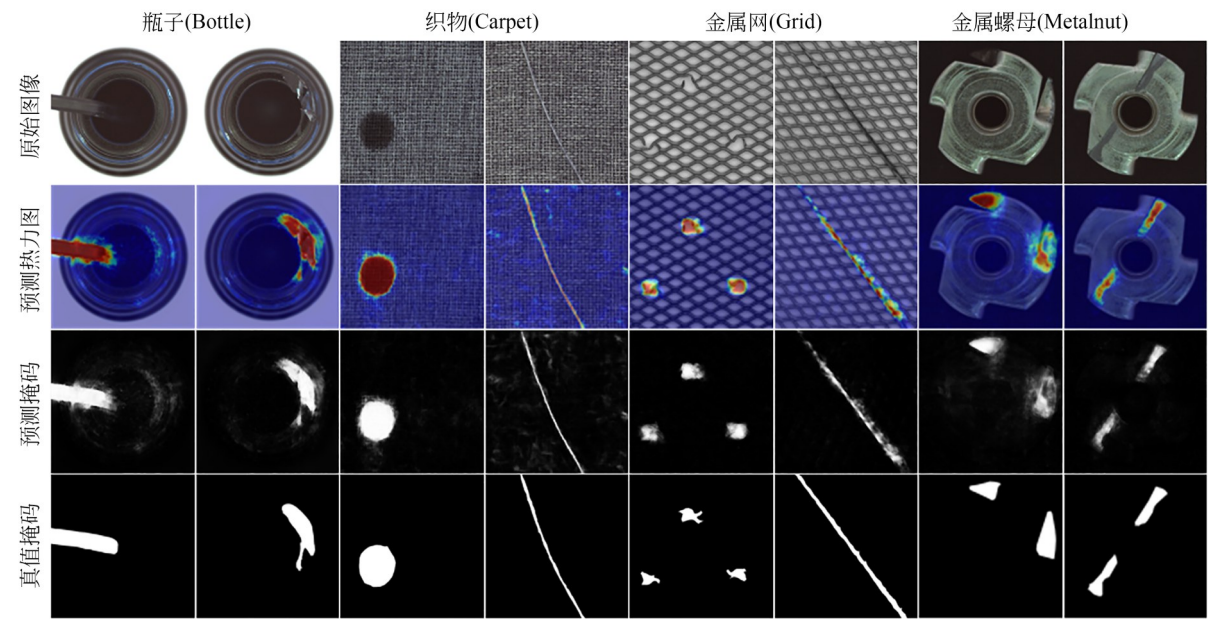
表 4 网络运行效率及参数量对比

Tab. 4 Comparison of network operation efficiency and parameter quantity

方法指标	DRAEM ^[21]	RIAD ^[30]	PaDim ^[31]	US ^[32]	MMR ^[33]	DSR ^[23]	本文方法
FPS	45	—	25	—	35	43	40
参数量/MB	93	—	68.9	—	102	38.4	97.6

中部分工业品进行异常检测效果。图中第一行为原始带异常的图像,第二行为本文模型检测输出的热力图,第三行为本文模型检测输出的预测图,第四行为工业品图像的异常真值图

(Ground Truth Mask)。这些可视化结果有力地证实了所提模型不仅能够有效检测出各类异常,更能实现高精度的定位,与前述定量分析结论高度一致。



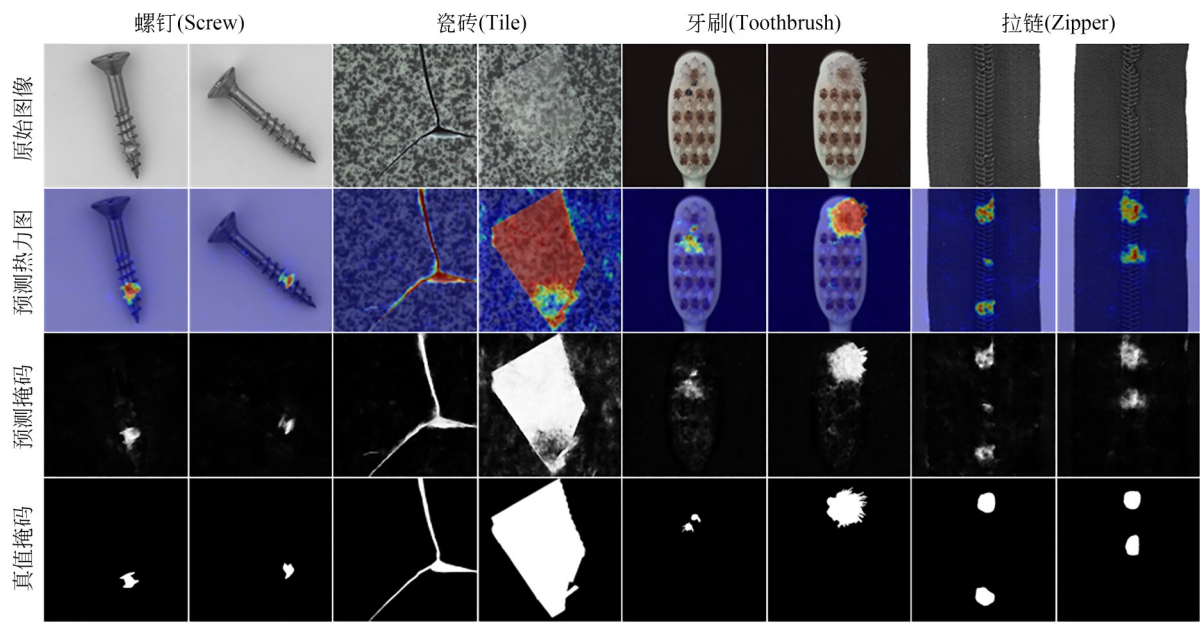


图5 部分检测结果图
Fig. 5 Partial test result chart

3.5 消融实验

为了系统评估所提出工业异常检测算法中各个核心组件的独立贡献及其有效性,进行了消融实验。所提出的工业异常检测算法主要包括以下几个关键部分:高频特征提取模块;全局增强多尺度注意力模块GEMA;交叉注意力模块CA,消融实验结果见表5。

表5 消融实验结果						
Tab. 5 Results of ablation experiments						
Baseline	高频	GEMA	CA	Auc-I	Auc-P	Ap-P
✓				97.7	97.0	69.6
✓	✓			98.1	97.0	70.1
✓		✓		97.9	97.4	71.8
✓			✓	97.8	97.0	72.0
✓	✓	✓		98.3	97.4	72.1
✓	✓		✓	98.2	97.1	72.2
✓		✓	✓	98.3	97.4	72.9
✓	✓	✓	✓	98.6	97.6	73.2

当在基线模型上加入高频特征提取模块时,图像级 AUROC 提高了 0.4%,像素级 AP 提升了 0.5%;加入 GEMA 模块后的模型,相较于基线模型图像级 AUROC 提升了 0.2%,像素级

AUROC 提升了 0.4%,像素级 AR 提升了 2.2%;当加入 CA 模块后,模型的图像级 AUROC 提升了 0.1%,像素级 AP 提升了 2.4%。这些实验结果数据充分证明,所提出的高频特征提取模块、GEMA 模块以及 CA 模块均对增强模型的工业异常检测能力起到了不可或缺且有效的促进作用。

高频特征提取模块首先通过高斯低通滤波分离图像的高频残差分量,有效抑制均匀背景干扰的同时保留缺陷引起的边缘梯度突变,为后续处理提供富含细节的特征输入;高效多尺度注意力模块在此基础上,通过动态分组策略和双路径注意力机制,实现对多尺度特征的自适应融合,其空洞卷积路径扩展感受野以捕获大范围异常模式,而方向敏感注意力路径则强化局部纹理判别力,使网络能够同时响应微观划痕和宏观结构缺陷;坐标注意力模块在接管特征优化任务,通过坐标轴分解的空间权重调制,将之前阶段提取的语义信息转化为精准的空间定位信号,其水平与垂直注意力机制能够有效矫正异常区域的边界偏移,最终通过残差连接实现特征重构。这种协同架构使得高频模块的细节保持能力、多尺度模块的语义抽象能力与坐标模块的空间感知能力形成互补。

4 结 论

针对现有工业异常检测方法中高频细节缺失及微小缺陷感知能力不足的局限性,提出一种基于频域特征解耦的工业异常检测框架。首先设计频域感知预处理模块,通过高斯核函数构建低频基准图像,进而提取高频残差分量,利用双域特征解耦机制强化高频通道对微小异常的响应能力。在特征编码阶段,构建全局增强的多尺度注意力机制 GEMA,通过可变形卷积建立跨尺度特征关联,并引入自适应全局上下文建模单元,有效捕获跨区域长程依赖关系,提升复杂纹理背景下异常特征的鉴别能力。同时采用混合感受野卷积策略,通过扩展卷积与常规卷积的并行架构,实现局部细节与全局语义的多粒度特征融合。在特征解码阶段,引入坐标轴分解的动态注意力机制 CoordAtt,通过对特征图进行空间轴向的注意力权重分配,建立异常区域的几何位置感知,显著提升像素级缺陷定位的边界精度。本文在 MVTec AD 数据集上的实验表明,所改进模型的图像级平均 AUROC 为 98.6%,像素级别的平均 AUROC 和 AP 分别为 97.6% 和 73.2%,说明所提出的异常检测方法能有效提高检测精度。

dAtt,通过对特征图进行空间轴向的注意力权重分配,建立异常区域的几何位置感知,显著提升像素级缺陷定位的边界精度。本文在 MVTec AD 数据集上的实验表明,所改进模型的图像级平均 AUROC 为 98.6%,像素级别的平均 AUROC 和 AP 分别为 97.6% 和 73.2%,说明所提出的异常检测方法能有效提高检测精度。

作者贡献声明:

张陈涛:方法论的提出、项目管理、指导;
邹庆林:实验设计、论文构思和撰写;
刘洋:参与实验与数据分析;
王亚飞:提供资源、数据整理和分析;
王彩云:调查研究;
徐周毅:论文校稿;
郑高峰:论文审核与编辑写作。

参考文献:

- [1] 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.
LV C K, SHEN F, ZHANG Z T, *et al.* Review of image anomaly detection[J]. *Acta Automatica Sinica*, 2022, 48(6): 1402-1428. (in Chinese)
- [2] 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 S R. Research on application of deep learning in field of industrial surface defects detection[J]. *Computing Technology and Automation*, 2022, 41(1): 59-65. (in Chinese)
- [3] 于瀚文,吴一全. 基于机器视觉的锂电池缺陷检测研究进展[J]. 仪器仪表学报, 2024, 45(9): 1-23.
YU H W, WU Y Q. Research progress on defect detection of lithium battery based on machine vision[J]. *Chinese Journal of Scientific Instrument*, 2024, 45(9): 1-23. (in Chinese)
- [4] 林珊玲,彭雪玲,王栋,等. 多尺度增强特征融合的钢表面缺陷目标检测[J]. 光学精密工程, 2024, 32(7): 1075-1086.
LIN S L, PENG X L, WANG D, *et al.* Object detection of steel surface defect based on multi-scale enhanced feature fusion[J]. *Opt. Precision Eng.*, 2024, 32(7): 1075-1086. (in Chinese)
- [5] 蔡引娣,张殿鹏,孙梓盟,等. 基于改进 YOLOv8 模型的增材制造微小气孔缺陷检测及其尺寸测量[J]. 光学精密工程, 2024, 32(21): 3222-3230.
CAI Y D, ZHANG D P, SUN Z M, *et al.* YOLOv8 model-based additive manufacturing micro porosity defect detection and its dimension measurement[J]. *Opt. Precision Eng.*, 2024, 32(21): 3222-3230. (in Chinese)
- [6] LIU K, WANG H Y, CHEN H Y, *et al.* Steel surface defect detection using a new haar-weibull-variance model in unsupervised manner[J]. *IEEE Transactions on Instrumentation and Measurement*, 2017, 66(10): 2585-2596.
- [7] 王冬云,唐楚,鄂世举,等. 基于导向滤波 Retinex 和自适应 Canny 的图像边缘检测[J]. 光学精密工程, 2021, 29(2): 443-451.
WANG D Y, TANG C, ES J, *et al.* Image edge detection based on guided filter Retinex and adaptive Canny[J]. *Opt. Precision Eng.*, 2021, 29(2): 443-451. (in Chinese)
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [9] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network[J]. *Chinese Journal of Com-*

- puters, 2017, 40(6): 1229-1251. (in Chinese)
- [10] 罗东亮, 蔡雨萱, 杨子豪, 等. 工业缺陷检测深度学习的方法综述[J]. 中国科学(信息科学), 2022, 52(6): 1002-1039.
- LUO D L, CAI Y X, YANG Z H, *et al.* Survey on industrial defect detection with deep learning [J]. *Science in China (Information Sciences)*, 2022, 52(6): 1002-1039. (in Chinese)
- [11] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- ZHANG S, GONG Y H, WANG J J. The development of deep convolution neural network and its applications on computer vision [J]. *Chinese Journal of Computers*, 2019, 42(3): 453-482. (in Chinese)
- [12] 牛小明, 曾理, 杨飞, 等. 零部件光学影像精准定位的轻量化深度学习网络[J]. 光学精密工程, 2023, 31(17): 2611-2625.
- NIU X M, ZENG L, YANG F, *et al.* Lightweight deep learning network for accurate localization of optical image components [J]. *Opt. Precision Eng.*, 2023, 31(17): 2611-2625. (in Chinese)
- [13] ZHU J W, DING C B, TIAN Y, *et al.* Anomaly heterogeneity learning for open-set supervised anomaly detection[C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 17616-17626.
- [14] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47(5): 1017-1034.
- TAO X, HOU W, XU D. A survey of surface defect detection methods based on deep learning [J]. *Acta Automatica Sinica*, 2021, 47(5): 1017-1034. (in Chinese)
- [15] 李良福, 马卫飞, 李丽, 等. 基于深度学习的桥梁裂缝检测算法研究[J]. 自动化学报, 2019, 45(9): 1727-1742.
- LI L F, MA W F, LI L, *et al.* Research on detection algorithm for bridge cracks based on deep learning [J]. *Acta Automatica Sinica*, 2019, 45(9): 1727-1742. (in Chinese)
- [16] CHU W H, KITANI K M. Neural batch sampling with reinforcement learning for semi-supervised anomaly detection [C]. *Computer Vision-ECCV 2020*. Cham: Springer, 2020: 751-766.
- [17] ZENG Z Y, LIU B, FU J L, *et al.* Reference-based defect detection network [J]. *IEEE Transactions on Image Processing*, 2021, 30: 6637-6647.
- [18] WAN Q, GAO L, LI X Y. Logit inducing with abnormality capturing for semi-supervised image anomaly detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 3523412.
- [19] BERGMANN P, LÖWE S, FAUSER M, *et al.* Improving unsupervised defect segmentation by applying structural similarity to autoencoders [C]. *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. February 25-27, 2019. Prague, Czech Republic. SCITEPRESS-Science and Technology Publications, 2019: 372-380.
- [20] COLLIN A S, DE VLEESCHOUWER C. Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise [C]. 2020 *25th International Conference on Pattern Recognition (ICPR)*. January 10-15, 2021, Milan, Italy. IEEE, 2021: 7915-7922.
- [21] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DRÆM-A discriminatively trained reconstruction embedding for surface anomaly detection [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021. Montreal, QC, Canada. IEEE, 2021: 8310-8319.
- [22] LI C L, SOHN K, YOON J, *et al.* CutPaste: self-supervised learning for anomaly detection and localization [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021. Nashville, TN, USA. IEEE, 2021: 9659-9669.
- [23] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DSR-a dual subspace re-projection network for surface anomaly detection [C]. *Computer Vision-EC-CV 2022*. Cham: Springer, 2022: 539-554.
- [24] 郭艳, 王智文, 赵润星. YOLO-POD: 基于多维注意力机制的高精度PCB微小缺陷检测算法 [J]. 电子学报, 2024, 52(7): 2515-2528.
- GUO Y, WANG Z W, ZHAO R X. YOLO-POD: high-precision PCB tiny-defect detection algorithm based on multi-dimensional attention mechanism [J]. *Acta Electronica Sinica*, 2024, 52(7):

- 2515-2528. (in Chinese)
- [25] 王素琴, 程成, 石敏, 等. 结合频率和 ViT 的工业产品表面相似特征缺陷检测方法[J]. 中国图象图形学报, 2024, 29(10): 3074-3089.
- WANG S Q, CHENG C, SHI M, *et al.* Defect detection method for industrial product surfaces with similar features by combining frequency and ViT[J]. *Journal of Image and Graphics*, 2024, 29(10): 3074-3089. (in Chinese)
- [26] PERLIN K. An image synthesizer[J]. *ACM SIGGRAPH Computer Graphics*, 1985, 19(3): 287-296.
- [27] SHARAN L, ROSENHOLTZ R, ADELSON E H. Accuracy and speed of material categorization in real-world images[J]. *Journal of Vision*, 2014, 14(9): 12.
- [28] OUYANG D L, HE S, ZHANG G Z, *et al.* Efficient multi-scale attention module with cross-spatial learning[C]. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. June 4-10, 2023, Rhodes Island, Greece. IEEE, 2023: 1-5.
- [29] ZOU Y, JEONG J, PEMULA L, *et al.* SPot-the-difference self-supervised pre-training for anomaly detection and segmentation[C]. *Computer Vision-ECCV 2022*. Cham: Springer, 2022: 392-408.
- [30] ZAVRTANIK V, KRISTAN M, SKOČAJ D. Reconstruction by inpainting for visual anomaly detection [J]. *Pattern Recognition*, 2021, 112: 107706.
- [31] 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*. Cham: Springer, 2021: 475-489.
- [32] BERGMANN P, FAUSER M, SATTLEGER D, *et al.* Uninformed students: student-teacher anomaly detection with discriminative latent embeddings[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020. Seattle, WA, USA. IEEE, 2020: 4182-4191.
- [33] ZHANG Z L, ZHAO Z B, ZHANG X W, *et al.* Industrial anomaly detection with domain shift: a real-world dataset and masked multi-scale reconstruction [J]. *Computers in Industry*, 2023, 151: 103990.
- [34] ROLIH B, FUČKA M, SKOČAJ D. SuperSimpleNet: unifying unsupervised and supervised learning for fast and reliable surface defect detection [C]. *Pattern Recognition*. Cham: Springer, 2025: 47-65.
- [35] COHEN N, HOSHEN Y. Sub-image anomaly Detection with Deep Pyramid Correspondences [EB/OL]. 2020: *arXiv*: 2005.02357. <https://arxiv.org/abs/2005.02357>
- [36] ZHANG X Y, LI N Q, LI J W, *et al.* Unsupervised surface anomaly detection with diffusion probabilistic model[C]. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. October 1-6, 2023, Paris, France. IEEE, 2024: 6759-6768.

作者简介:



张陈涛(1986—),男,福建漳平人,2010年于哈尔滨工业大学获得学士学位,2017年于厦门大学获得博士学位,现为厦门大学助理教授,主要研究方向为机器视觉与环境感知、智能机器人。E-mail:zhangct@xmu.edu.cn

通讯作者:



郑高峰(1984—),男,福建泉州人,博士,教授,博士生导师,2006年于华中科技大学获得学士学位,2011年于厦门大学获得博士学位,主要从事智能控制与微纳喷印的研究。E-mail:zheng_gf@xmu.edu.cn