

基于变分特征解纠缠的锐度感知工业缺陷分类算法

钟依慧¹, 马瑛¹, 江少进², 杨丰玉¹

(1. 南昌航空大学软件学院, 江西 南昌 330000; 2. 江西昌河飞机工业集团有限公司, 江西 景德镇 333000)

摘要: 在工业生产领域, 缺陷分类对于保障产品质量和安全性起着至关重要的作用。但是, 工业缺陷数据集存在类内差异大、类间差异小的特性, 加上缺陷样本数量有限, 致使现有缺陷分类模型在实际工业环境中的表现欠佳。针对这一问题, 提出一种基于变分特征解纠缠(VFD)和锐度感知(SA)的工业缺陷分类算法 VFD-SA。首先, 引入变分自编码器(VAE)将缺陷特征解纠缠为类别判别特征和类内方差特征; 然后, 通过重采样策略增强类内方差特征, 与原始特征相结合以提升特征表示的区分度, VFD 使模型更加专注于缺陷的类别判别特征, 同时对与缺陷类别无关的细节和背景具有一定的容忍度, 进而提高模型的缺陷分类性能; 最后, 通过引入 SA 训练策略, 优化损失函数的几何形状, 进一步提升模型的泛化能力。在轧钢缺陷数据集 NEU-CLS、金属缺陷数据集 GC10-DET 和自制紧固件缺陷数据集上进行实验, 结果表明, VFD-SA 的准确率分别达到 100%、93.52% 和 99.48%, 显著优于现有缺陷分类算法, 能够满足各类工业场景下的缺陷分类需求。

关键词: 缺陷图像分类; 变分自编码器; 锐度感知; 小样本; 工业缺陷

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070667

Sharpness-Awareness Industrial Defect Classification Algorithm Based on Variational Feature Disentanglement

ZHONG Yihui¹, MA Ying¹, JIANG Shaojin², YANG Fengyu¹

(1. School of Software, Nanchang Hangkong University, Nanchang 330000, Jiangxi, China;

2. Jiangxi Changhe Aircraft Industry Group Co., Ltd., Jingdezhen 333000, Jiangxi, China)

【Abstract】 In industrial production, defect classification plays a crucial role in ensuring product quality and safety. However, industrial defect datasets are characterized by significant intra-class variations and minimal inter-class differences. Coupled with a limited number of defect samples, this results in subpar performance of existing defect classification models in real-world industrial settings. To address this issue, an industrial defect classification algorithm called VFD-SA is proposed, which is based on Variational Feature Disentanglement (VFD) and Sharpness-Aware (SA) training. First, a Variational Auto-Encoder (VAE) is introduced to disentangle the defect features into category-discriminative and intra-class variance features. Subsequently, a resampling strategy is employed to enhance the intra-class variance features, which are then combined with the original features to improve the discriminative power of the feature representations. The VFD enables the model to focus more on category-discriminative features of defects while exhibiting tolerance to details and backgrounds unrelated to defect categories, thereby enhancing the model's defect classification performance. Finally, by incorporating the SA training strategy, the geometric shape of the loss function is optimized to further improve the generalization capability of the model. Experimental results on the NEU-CLS dataset for steel rolling defects, GC10-DET dataset for metal defects, and a self-made dataset for fastener defects demonstrate that the VFD-SA achieves accuracy rates of 100%, 93.52%, and 99.48%, respectively, significantly outperforming existing defect classification algorithms and meeting defect classification requirements across various industrial scenarios.

【Key words】 defect image classification; Variational Auto-Encoder (VAE); Sharpness-Awareness (SA); few-shot; industrial defect

0 引言

在国家建设和工业生产中, 质量控制是确保产品可靠性和安全性的核心环节, 直接关系到基础设

施的安全^[1]。然而, 受现代生产技术的复杂性和各种不可控的外界因素的影响, 产品可能出现多种缺陷^[2], 这些缺陷涵盖表面缺陷(如钢铁表面的斑块、夹杂物、划痕等)以及结构缺陷(如形变、错位与缺损

基金项目: 江西省重点研发项目(20202BBEL53002)。

作者简介: 钟依慧, 女, 硕士, 主研方向为缺陷检测; 马瑛, 博士; 江少进、杨丰玉(通信作者), 硕士。

收稿日期: 2024-11-29

修回日期: 2025-02-25

E-mail: yangfengyu@nchu.edu.cn

等)。表面或结构缺陷均可能对工业材料、部件的耐磨性、抗腐蚀性和疲劳强度产生负面影响,损害产品完整性与功能,甚至引发安全事故。

工业缺陷分类是为了确定产品是否满足质量标准和技术要求而进行的测试或测量活动。传统缺陷分类主要依靠人工目视检查,存在检测速度慢、效率低、成本高以及因人为因素致使检测结果可信度低等问题^[3]。随着人工智能技术的发展,基于计算机视觉和深度学习的表面缺陷分类方法逐步取代传统人工检测方法。在早期,深度学习主要依靠大量标注数据,借助卷积神经网络提取图像特征进行缺陷分类或分割。然而,面对样本不足或缺陷类型未知的情况时,这些方法表现欠佳。因此,当前基于深度学习的缺陷分类工作基本聚焦于如何高效地提取图像的缺陷特征。文献[4]通过引入权重系数,使网络关注不同特征,并整合多尺度缺陷特征,提升网络对特征的感知能力。文献[5]采用新颖的多尺度并行卷积网络提取不同尺度特征,实现实时缺陷分类。不过,这些方法仅着眼于不同缺陷图像间的整体差异,忽略了缺陷特征间的内部关联,限制了模型提取缺陷特征的能力,未能从根本上提升缺陷分类性能。

在工业缺陷分类领域,数据集常呈现同类缺陷差异大、不同缺陷间相似度高特性^[6],致使现有缺陷分类模型在实际工业环境中表现不佳。为解决该问题,本文结合细粒度图像分类领域的特征提取思路^[7],提出一种融合变分特征解纠缠(VFD)和锐度感知(SA)的工业缺陷分类算法(VFD-SA)。一方面,通过引入变分自编码器(VAE),将缺陷特征解纠缠为类别判别特征和类内方差特征,再利用重采样策略增强类内方差特征,与原始特征相结合,提升特征表示的区分度,VFD让模型更专注于缺陷的判别性特征,同时对与缺陷类别无关的细节和背景有一定容忍度,从而提高模型的缺陷分类性能;另一方面,引入锐度感知训练策略^[8],通过优化损失函数的几何形状来提升模型的泛化能力。

1 相关工作

1.1 基于深度学习的工业缺陷分类

随着深度学习的迅速发展,众多研究人员将其应用于各类工业缺陷分类场景^[9],深度学习已成为当下最主流的缺陷分类方法。相较于传统方法,深度学习借助卷积、池化等操作,能够提取更深入、更抽象的图像特征,具有更强的表征能力^[10]。卷积神经网络通过卷积运算,从输入图像中捕获不同层次

的语义信息,有效学习大量样本的特征表示,赋予模型强大的泛化能力。

在近期研究中,文献[5]开发了轻量级ConCNN模型,利用多尺度图像输入,实现钢材表面缺陷的高效实时分类,大幅缩短了训练时间,降低了资源消耗,同时提升了分类准确性与响应速度。文献[11]针对钣金件表面缺陷快速检测的需求,提出基于卷积神经网络的少样本学习方法,结合数据增强技术与窗口滑动检测策略,满足了工业生产对实时性和准确性的双重要求。文献[12]采用基于伪标签的半监督学习策略,借助教师-学生模型架构与课程学习技术,显著提升了瓷砖表面缺陷分类在小样本情况下的精度。文献[13]改进VGG16网络,融合CBAM注意力机制与Inception网络结构,并运用数据增强技术,增强了缺陷图像分类的性能与鲁棒性。文献[14]提出融合特征金字塔和注意力机制的多阶段分类算法,通过对抗性弱化和特征融合技术,提高缺陷区域识别的准确性与鲁棒性。文献[15]提出结合传统机器学习和深度学习的无监督动态加权联合分类方法,经纹理特征聚类、深度特征提取和基于KL(Kullback-Leibler)散度的动态权重标注,有效提升了带钢表面缺陷分类的准确性。

1.2 变分自编码器

VAE^[16]是自编码器的变体,其核心思想是运用变分推断,从最大似然角度对数据分布进行建模。典型VAE模型结构如图1所示,输入图像经编码器的卷积层编码,输出均值向量与标准差向量,两者共同确定一个潜在概率分布,基于此分布采样得到潜在向量,该向量是输入图像在低维潜在空间的表征,蕴含其主要特征。解码器以潜在向量作为输入,重构出与原始输入极为相似的图像。如此,VAE实现了输入数据的压缩与重构。因为潜在向量从概率分布采样而来,模型具备生成新数据的能力,可通过改变潜在向量的值,生成与训练数据特征相似但不同的新图像。

VAE最初作为深度生成概率模型用于图像生成。文献[17]提出跨域解耦方法,借助基于生成对抗网络(GAN)和仅以成对图像为输入的跨域自编码器进行双向图像翻译,将内部表示解耦为共享与独有部分。该设计在多样化样本生成、跨域检索、特定域图像转移和插值等任务中展现出良好性能。

后续研究者发现,VAE在图像数据集上具备学习解纠缠表示的潜力,能提升模型的鲁棒性与泛化性。文献[18]提出通过预测旋转图像进行自监督语

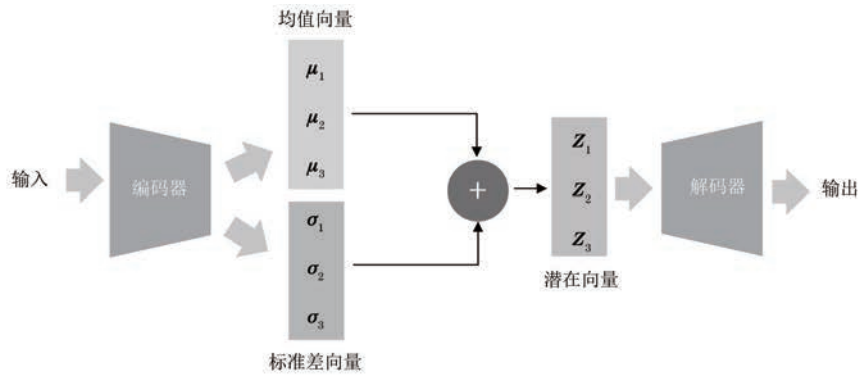


图 1 VAE 整体框架

Fig. 1 The overall framework of VAE

义特征学习的方法,利用 ConvNet 模型取得与监督方法相当的效果。受文献[18]研究启发,文献[19]通过联合训练图像旋转预测和实例区分,实现语义旋转相关与无关特征表示的解耦,有助于提升图像分类、检索、分割等任务的泛化能力。文献[7]提出新的解耦数据增强框架,对输入图像特征进行解耦,对判别特征进行采样,以实现数据增强,用于解决小样本分类问题。类似地,文献[20]将图像分解为类特异性和外观特异性特征,结合从基类学习提取的这两类特征,为新类别生成额外样本。

2 VFD-SA 算法

2.1 缺陷变分解纠缠

在工业缺陷分类领域,深度学习模型面临两大难题:

1) 缺陷特征不明显,数据集常呈现同类缺陷内部差异大、不同类缺陷间相似度高的特性,致使模型性能欠佳。如图 2 所示,同一缺陷内部也存在明显差异,如油点缺陷可能表现为点状、块状或丝状等不同形态,颜色和深浅程度也各不相同。而类间缺陷还存在缺陷类似的问题,如油点和铁皮灰。

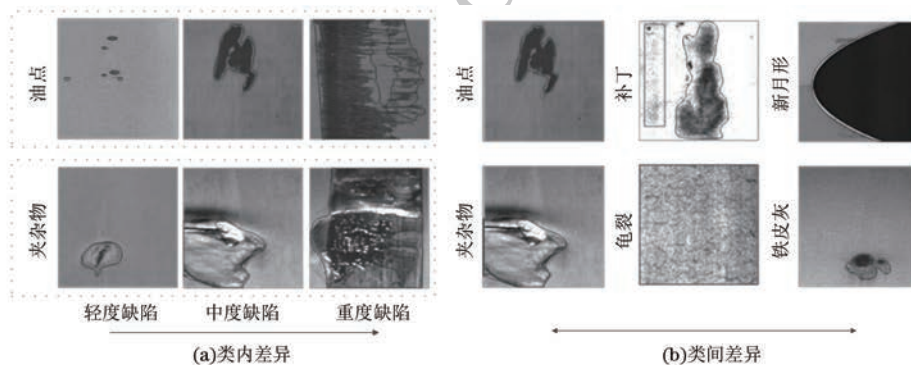


图 2 类内和类间缺陷的差异

Fig. 2 The differences between intra-class and inter-class defects

2) 缺陷样本数据集稀缺,模型可学习的训练样本极为有限,而训练完成的模型在真实工业场景应用时需面对海量未知数据,泛化性严重不足。

本文利用一种新颖的缺陷特征解纠缠方法,实现简单有效的数据增强。具体而言,将输入的表面缺陷图像特征表示分解为 2 个分量,即类别判别特征 z_l 和类内方差特征 z_v 。将类内方差特征 z_v 约束为遵循先验分布,随后从该分布中重复采样新的类内方差特征 z_v ,并将其与原始类别判别特征 z_l 相加,得到增强的特征表示。最终,利用增强特征与原始特征,共同训练表面缺陷分类器。缺陷变分解纠缠整体流程如图 3 所示。

给定一个输入图像 i ,使用特征提取模块将 i 映射为特征图 $X^{(i)}$ 。然后,根据 $X^{(i)}$ 计算类内方差特征 $z_v^{(i)}$ 和类别判别特征 $z_l^{(i)}$,此时输入图像的嵌入 $z^{(i)}$ 可以表示为:

$$z^{(i)} = z_l^{(i)} + z_v^{(i)} \quad (1)$$

类别判别特征 $z_l^{(i)}$ 可以通过最小化交叉熵损失来学习:

$$L_{\text{cls}}(X^{(i)}) = L_{\text{cross-entropy}}(W(z_l^{(i)}), y^{(i)}) \quad (2)$$

式中: W 是具有全连接层的分类器; $y^{(i)}$ 是第 i 个输入图像的真实类别标签; $L_{\text{cls}}(X^{(i)})$ 通过衡量模型预测结果与真实标签的差异,从而驱动模型准确依据类特定特征判断图像类别。

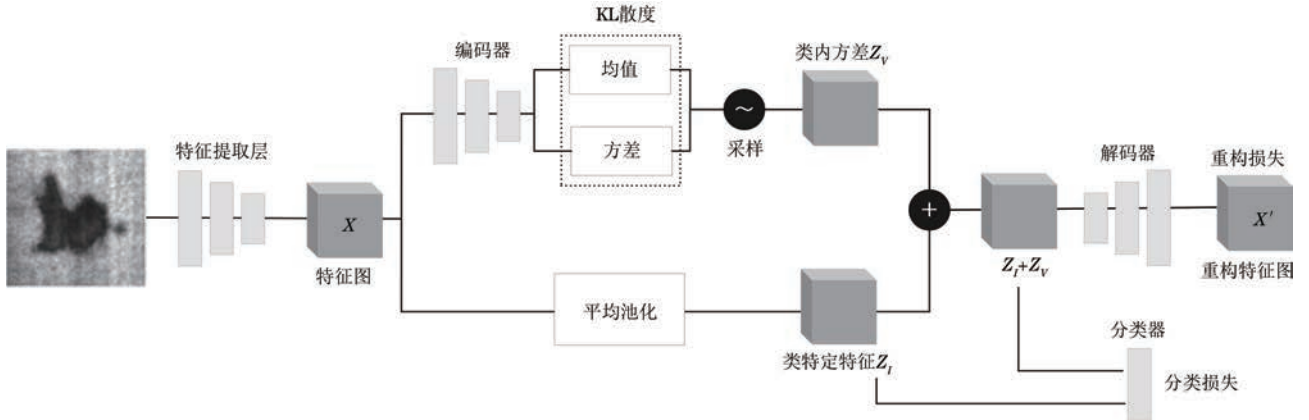


图 3 变分特征解纠缠整体流程

Fig.3 Overall process of variational feature disentanglement

在处理类内方差特征 z_v 时,由于真实后验分布 $p(z_v|X)$ 难以直接计算,本文采用变分推断对变量 z_v 的后验分布进行建模,使用分布 $q(z_v|X)$ 来近似真实后验分布 $p(z_v|X)$,两者之间的 KL 散度定义如下:

$$\text{KL}(q(z_v|X) \parallel p(z_v|X)) = \int_Z q(Z|X) \log_a \frac{q(Z|X)}{p(Z|X)} \quad (3)$$

式中: Z 是变量 z_v 的取值空间; $\text{KL}(q(z_v|X) \parallel p(z_v|X))$ 用于量化分布 $q(z_v|X)$ 与 $p(z_v|X)$ 之间的差异。最小化式(3),使近似分布更好地代表真实分布。

直接优化 KL 散度较为困难,因此,引入证据下界(ELBO)。变分推断表明,最大化边缘似然 $p(X^{(i)})$ 等价于最大化证据下界,定义如下:

$$\text{ELBO}^{(i)} = \mathbb{E}_{q(z_v^{(i)}|X^{(i)})} [\log_a p(X^{(i)}|z_v^{(i)})] - \text{KL}(q(z_v^{(i)}|X^{(i)}) \parallel p(z_v)) \quad (4)$$

式中: $\mathbb{E}_{q(z_v^{(i)}|X^{(i)})} [\log_a p(X^{(i)}|z_v^{(i)})]$ 表示在近似分布 $q(z_v^{(i)}|X^{(i)})$ 下模型生成数据(特征图 $X^{(i)}$)的对数似然的期望,反映模型对数据的拟合程度; $\text{KL}(q(z_v^{(i)}|X^{(i)}) \parallel p(z_v))$ 为近似分布与类内方差特征 z_v 的先验分布 $p(z_v)$ 之间的差异。最大化 ELBO,可同时优化近似分布 $q(z_v^{(i)}|X^{(i)})$ 的参数(如均值和标准差)以及模型其他参数,使模型更好地理解数据,促使近似分布接近真实后验分布。

通常情况下,类内变异分布可以用高斯分布来建模^[8],将 z_v 的先验分布设置为以原点为中心的各向同性多变量高斯分布 $p(z_v) = \mathcal{N}(0, I)$ 。对于后验分布,将其设置为具有对角线协方差的多元高斯分布:

$$q(z_v^{(i)}|X^{(i)}) = \mathcal{N}(\mu^{(i)}, \sigma^{(i)}) \quad (5)$$

式中: $\mu^{(i)}$ 和 $\sigma^{(i)}$ 由概率编码器计算得出,分别代表后验分布的均值和标准差。为从分布中采样类内方差特征 $z_v^{(i)}$,且解决直接采样在模型训练反向传播中难以求导的问题,采用重参数化技巧(Reparameterization trick),如下所示:

$$z_v^{(i)} = \mu^{(i)} + \sigma^{(i)} * \epsilon, \epsilon \sim \mathcal{N}(0, I) \quad (6)$$

式中: ϵ 是从标准正态分布中采样的随机噪声。

$z_l^{(i)}$ 由给定的 $X^{(i)}$ 所确定,因此,得出 $p(X^{(i)}|z_v^{(i)}) = p(X^{(i)}|z_v^{(i)}, z_l^{(i)}) = p(X^{(i)}|z_l^{(i)})$ 。为估计最大似然 $p(X^{(i)}|z_v^{(i)})$,使用解码器从 $z_l^{(i)}$ 中重构原始特征图,并最小化原始特征图与重构图之间的 L2 距离。根据式(4),进而推导出用于类内方差建模的损失函数:

$$L_{\text{intra}}(X^{(i)}) = \|X^{(i)} - \hat{X}^{(i)}\|^2 + \text{KL}(q(z_v^{(i)}|X^{(i)}) \parallel p(z_v)) \quad (7)$$

$L_{\text{intra}}(X^{(i)})$ 由重构误差和 KL 散度两部分组成: $\|X^{(i)} - \hat{X}^{(i)}\|^2$ 通过最小化重建特征图与原始特征图之间的 L2 距离,确保编码器从输入中提取有意义的信息;KL 散度则鼓励后验分布 $q(z_v^{(i)}|X^{(i)})$ 与先验分布 $p(z_v)$ 保持一致。两项协同驱动模型对类内方差进行有效建模。

本文并不直接最小化 KL 散度,而是如文献[21]中所述,将其分解为 3 个项:

$$\begin{aligned} \text{KL}(q(z_v|X) \parallel p(z_v)) = & \text{KL}(q(z_v|X) \parallel q(z_v)p(X)) + \\ & \text{KL}\left(q(z_v) \parallel \prod_j q(z_{v_j})\right) + \\ & \sum_j \text{KL}(q(z_{v_j}) \parallel p(z_{v_j})) \end{aligned} \quad (8)$$

式中: z_{v_j} 表示潜在变量的第 j 维。

式(8)的三项分别为索引-编码互信息、总相关性和维度 KL 散度。对索引-编码互信息和总相关

性项的惩罚会导致更加纠缠的表示^[7],而维度 KL 项能够确保潜在变量不会偏离先验分布太远。与文献[21]类似,本文使用权重 α 对总相关性进行惩罚,并将 L_{intra} 重写如下:

$$L_{\text{intra}}(X^{(i)}) = \|X^{(i)} - \hat{X}^{(i)}\|^2 + \text{KL}(q(z_V^{(i)} | Z^{(i)}) \| q(z_V^{(i)}) p(X)) + \alpha \text{KL}\left(q(z_V^{(i)}) \| \prod_j q(z_{V_j}^{(i)})\right) + \sum_j \text{KL}(q(z_{V_j}^{(i)}) \| p(z_{V_j})) \quad (9)$$

L_{cls} 和 L_{intra} 的组合推动模型同时提取类间判别特征 $z_I^{(i)}$ 并建模类内变异的分布。

对于输入图像 i ,提取其类别判别特征 $z_I^{(i)}$ 与类内变异均值 $\mu^{(i)}$ 和方差 $\sigma^{(i)}$,为该图像从分布 $\mathcal{N}(\mu^{(i)}, \sigma^{(i)})$ 中采样新的类内变异特征 $z_V^{(i)}$,并将其添加到 $z_I^{(i)}$ 中,得到增强特征 $\tilde{z}^{(i)} = z_I^{(i)} + z_V^{(i)}$ 。然后,使用以下交叉熵训练模型:

$$L_{\text{aug}}(X^{(i)}) = L_{\text{cross-entropy}}(W(\tilde{z}^{(i)}), y^{(i)}) \quad (10)$$

最终得到的总体损失函数是上述项的加权组合:

$$L = L_{\text{cls}} + L_{\text{intra}} + \beta \cdot L_{\text{aug}} \quad (11)$$

式中: β 是 L_{aug} 的系数。

2.2 锐度感知最小化

为满足工业小样本场景下分类模型对泛化性的需求,本文引入锐度感知最小化(SAM)^[8]。SAM 借助损失函数景观的几何特性及其与泛化性的联系,同时最小化损失值与损失锐度,以此提升模型的泛化能力。

图 4^[8]展示了 SGD 优化器和 SAM 优化器的损失超平面。在优化同一模型时,SGD 会使模型收敛至较为尖锐的最小值处,而 SAM 则收敛于更为平坦的最小值区域。

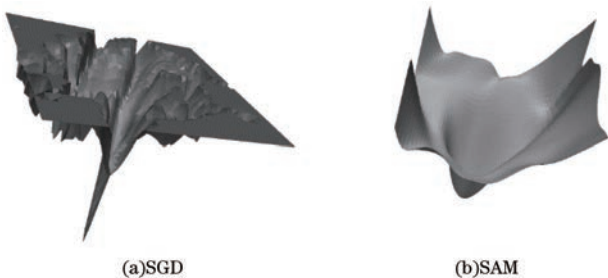


图 4 SGD 和 SAM 的损失超平面示意图

Fig. 4 Schematic diagram of loss hyperplane for SGD and SAM

SAM 的主要目标是解决以下优化问题:

$$\min_w \left\{ \max_{\|\epsilon\|_2 \leq \rho} L_S(w + \epsilon) + \lambda \|w\|_2 \right\} \quad (12)$$

式中: $L_S(w)$ 是训练集上的损失函数; ϵ 是扰动项; ρ 是控制邻域大小的超参数; λ 是正则化项的系数,用

于控制模型复杂度。

在每次迭代中,SAM 算法首先计算当前参数的梯度,随后依据梯度的方向与大小生成“对抗性扰动”,此扰动会使参数沿着梯度朝损失增加最多的方向移动。接着,SAM 算法计算添加对抗性扰动后的参数所对应的损失值,将其作为 SAM 损失。最后,SAM 算法运用梯度下降优化器,最小化 SAM 损失,以更新参数。

SAM 算法的关键步骤之一是计算对抗性扰动^[20],这可通过一阶泰勒展开近似。具体来说,对于给定的参数 w ,SAM 算法的损失计算如下:

$$\epsilon^*(w) \approx \underset{\|\epsilon\|_2 \leq \rho}{\operatorname{argmax}} \epsilon^T \nabla_w L_S(w) \quad (13)$$

由上可知,SAM 并非通过直接计算 Hessian 矩阵获取信息,而是在参数空间施加微小扰动,观察损失函数的变化,以此间接考量损失函数的局部几何性质。SAM 巧妙设计扰动与更新机制,绕开 Hessian 矩阵的显式计算,高效计算梯度并完成优化,进而在一定程度上提升模型的泛化能力与性能。

3 实验验证

3.1 缺陷数据集

为了检验本文方法在工业场景下的有效性,使用公开的轧钢表面缺陷数据集 NEU-CLS^[22]、来自天津大学的金属表面缺陷数据集 GC10-DET^[23] 以及自制螺栓紧固件缺陷数据集进行实验。

3.1.1 轧钢缺陷数据集

NEU-CLS 轧钢缺陷数据集涵盖 6 种缺陷类型,分别为划痕(Sc)、斑块(Pa)、凹坑表面(Ps)、夹杂物(In)、龟裂(Cr)和卷积鳞片(Rs)。每种缺陷类型包含 300 张图片,共计 1 800 张。图 5 展示了 NEU-CLS 的 6 种缺陷示意图。

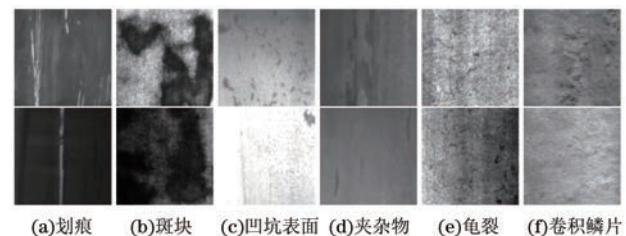


图 5 轧钢缺陷样本

Fig. 5 Steel rolling defect samples

从图中可明显看出,NEU-CLS 数据集中类内缺陷外观差异较大,如划痕(最后一列)包含水平、垂直、倾斜等多种形态。同时,类间缺陷存在相似之处,像滚入式水垢、裂纹和点蚀表面。此外,受光照及材料变化影响,类内缺陷图像灰度值存在波动。

总体而言,NEU-CLS 数据集存在两大难题:一是类内缺陷外观差异大;二是类间缺陷具有相似性,且缺陷图像受光照和材料变化干扰较大。

3.1.2 金属缺陷数据集

GC10-DET 数据集总共包含 10 种缺陷,包括冲孔(Pu)、焊缝(Wl)、新月形缝隙(Cg)、水斑(Ws)、油斑(Os)、丝斑(Ss)、夹杂物(In)、轧坑(Rp)、折痕(Cr)、腰部折痕(Wf)。图 6 为 GC10-DET 的缺陷示意图。

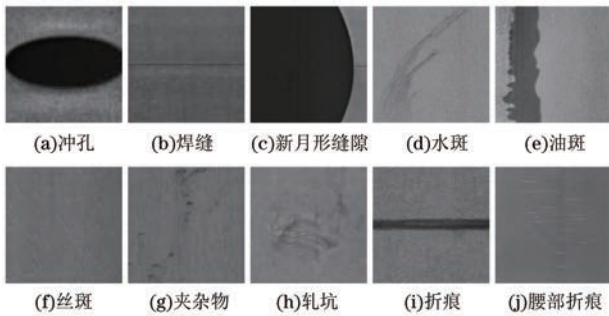


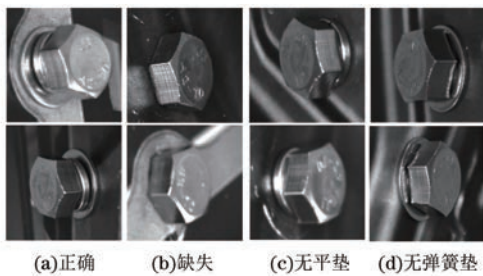
图 6 金属缺陷样本

Fig.6 Metal defect samples

GC10-DET 原本为检测数据集,经过数据预处理后,将此数据集用于分类,共 2 837 张图像,这是一种不平衡数据集,如 Ss 类和 Rp 类数量甚至相差超 10 倍。

3.1.3 自制紧固件缺陷数据集

紧固件缺陷数据集由实地拍摄的 4 类螺栓紧固件照片组成,涵盖 1 类正确样本与 3 类缺陷样本。因螺栓需同时装配平垫片(最下层)和弹簧垫片(平垫片上层),才能确保安装安全有效,故未装配平垫片和弹簧垫片的样本归为“缺失”,仅未装配平垫片的样本归为“无平垫”,仅未装配弹簧垫片的样本归为“无弹簧垫”。该数据集共有 1 673 张图像,4 类样本数量较为均衡。图 7 为螺栓缺陷数据集示意图。



(a)正确 (b)缺失 (c)无平垫 (d)无弹簧垫

图 7 自制紧固件缺陷样本

Fig.7 Self-made fastener defect samples

3.2 实验环境及参数

本实验采用的硬件配置和软件配置如表 1 所示。

表 1 实验平台参数

Table 1 Experimental platform parameters

配置项	配置规格
操作系统	Windows 11
处理器	铭瑄 H610M
显卡	NVIDIA GeForce RTX 2080Ti
显存/GB	11
深度学习框架	PyTorch 2.2.1+cu118 CUDA 11.8
编程语言	Python 3.8
集成开发环境	PyCharm

在本文实验中,选用 ResNet18 作为主干网络,输入图像尺寸为 224×224 像素,并通过随机裁剪等方式进行数据增强。对比模型均采用 Adam 优化器更新模型参数,实验均运用指数衰减学习率策略优化模型训练过程。本算法与所有对比方法均采用初始学习率 $r_{lr}=0.001$,进行 100 个 Epoch 的训练。学习率衰减公式如下:

$$r_{new_lr} = r_{initial_lr} \times (f_{decay_factor}^{E_{Epoch}}) \quad (14)$$

式中: r_{new_lr} 表示更新后的学习率; $r_{initial_lr}$ 是初始学习率; f_{decay_factor} 是学习率衰减因子; E_{Epoch} 表示当前的训练轮次。在实验设置中,衰减因子设定为 0.96,意味着每完成一个 Epoch,学习率将自动降低到前一个 Epoch 的 96%。

3.3 实验结果分析

3.3.1 性能对比

为了证明所提算法在钢铁表面缺陷分类方面的优越性,采用 8:2 的比例划分数据集,并在 3 个实验数据集上进行性能对比实验。

在金属缺陷数据集 GC10-DET 上,将本文所提分类模型与 VGG16、ResNet18 等经典模型进行对比。图 8 呈现了不同模型在 GC10-DET 数据集上训练过程中的测试正确率。从图 8 可知,本文模型收敛速度较快,且能实现最高准确率,可达 93.52%,相较于经典模型 ResNet18 和 VGG16,分别提升 4.55 和 4.74 百分点。同时,VFD-SA 在训练早期便达到较高准确率,且随着训练的推进,准确率稳步上升,最终达到最优性能,这表明本文模型能加速收敛至较为平坦的损失区域,提升准确性。

图 9 为 GC10-DET 测试集的混淆矩阵可视化,可直观展示各类缺陷的测试状况。从中可以看出,即便 GC10-DET 数据集极不平衡,本文算法对每个类别仍具有较高的分类准确率。

在轧钢缺陷数据集 NEU-CLS 上,将本文方法与最新的表面缺陷分类方法进行性能对比,结果如表 2 所示。文献[24]运用小波变换和灰度共生矩

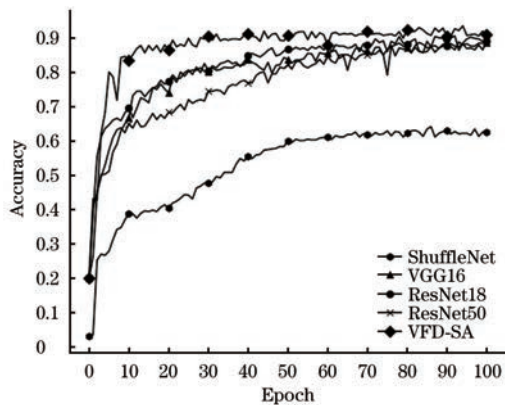


图 8 各模型在测试集 GC10-DET 上的分类准确率

Fig.8 Classification accuracy of each model on test set GC10-DET

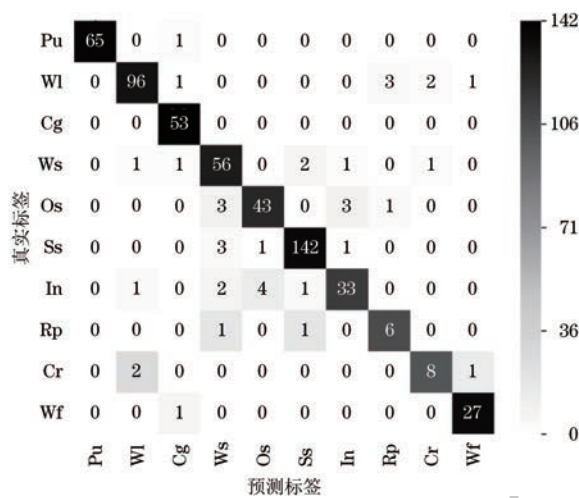


图 9 GC10-DET 测试集的混淆矩阵可视化

Fig.9 Visualization of confusion matrix for GC10-DET test set 阵,通过提取低频子带信息,计算 4 个方向的灰度共生矩阵特征值,达成了超 96% 的分类精度。文献[25]引入软注意力机制,改进深度残差网络 ResNet 模型,准确率达 98.61%。文献[13]引入注意力机制 CBAM 和 Inception 网络结构,对 VGG16 网络进行改进,准确率为 98.84%。ConCNN^[5] 作为新型并行卷积神经网络,针对实时缺陷分类应用进行优化,实现了 99.63% 的分类准确率。而本文所提方法凭借 VFD-SA,准确率达到 100%,超越了其他方法,充分证明了 VFD-SA 的有效性。

表 2 NEU-CLS 数据集上的性能对比

Table 2 Performance comparison on the NEU-CLS dataset %

Method	Accuracy
文献[24]方法	96.17
文献[25]方法	98.61
文献[13]方法	98.84
ConCNN ^[5]	99.63
VFD-SA 方法	100.00

在紧固件数据集上开展实验,结果如表 3 所示。所有模型在“缺失”类别上均达到 100% 的准确率,这一显著成果源于“缺失”类别特征显著,即同时缺少平垫片和弹簧垫片,模型易于学习此类特征。“无垫片”类别缺少关键且视觉范围较大的平垫片,存在一定视觉差异,模型也大多能达到较高准确率。总体而言,VFD-SA 在多个类别上展现出最优性能,这表明 VFD-SA 不仅能有效区分表面缺陷,在结构缺陷分类方面同样具备优势,凸显了本文方法在实际工业应用中的巨大潜力。

表 3 自制紧固件缺陷数据集上的准确率对比

Table 3 Accuracy comparison on the self-made fastener defect dataset %

Model	Accuracy				
	正确	缺失	无平垫	无弹簧垫	平均
ShuffleNet	97.67	100.00	97.73	98.31	97.92
VGG16	100.00	100.00	100.00	96.61	98.96
RepVGG	97.67	100.00	100.00	96.61	98.44
ResNet18	97.67	100.00	100.00	96.61	98.44
ResNet50	95.35	100.00	100.00	100.00	98.96
VFD-SA	100.00	100.00	100.00	98.31	99.48

图 10 展示了基于 NEU-CLS 数据集的 T-SNE 可视化,其中图 10(a)为基准模型,图 10(b)为 VFD-SA 算法。经 T-SNE 可视化可知,VFD-SA 算法在特征空间中实现了更清晰的类别分离。相较于基准模型,VFD-SA 算法的点群分布更为分散,这表明其能学习到更具判别性的特征,进而在特征空间有效区分不同缺陷类别。

图 10 中这种分散的点群分布,彰显了 VFD-SA 算法处理类间区分度小、类内方差大问题优势。VFD-SA 借助变分特征解纠缠技术,可捕捉更细微的类别差异,同时降低类内不必要的方差,这在模型分类性能上得以直观体现。

3.3.2 不同比例的训练样本

为验证 VFD-SA 算法在工业缺陷分类任务中的泛化能力,本文充分模拟真实工业场景下表面缺陷分类可能遭遇的少样本难题。在 NEU-CLS 数据集与自制紧固件数据集上,分别开展了 3 个不同比例训练条件下的实验(以 20%、50% 和 80% 的数据作为训练样本,其余样本用作测试数据集),以此检验本文所提算法在不同规模训练数据集上的性能,并与多个主流基准模型进行对比。

如表 4 所示,NEU-CLS 数据集上各类别具有独特视觉特征,当训练样本占比 80% 时,多数模型可实现较高准确率。VFD-SA 算法在训练样本占比

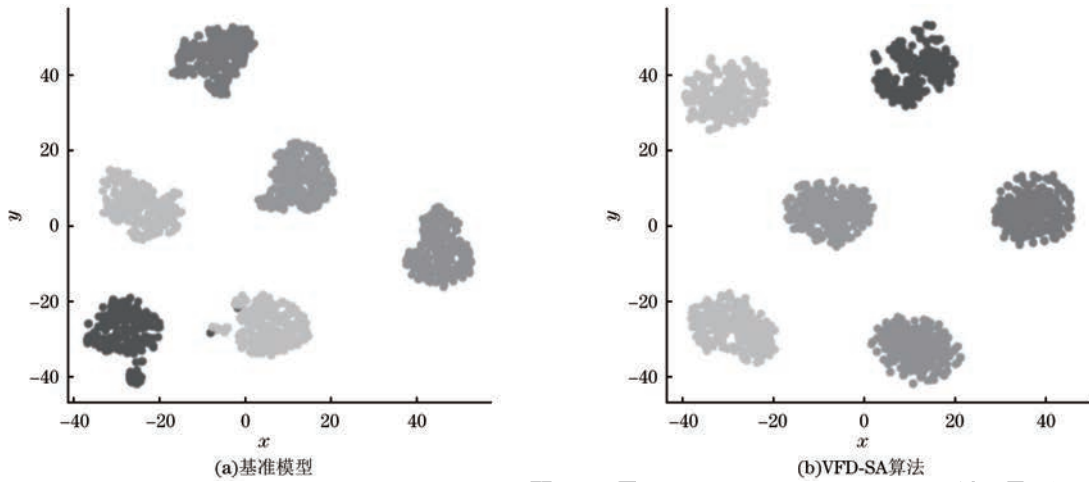


图 10 基于 NEU-CLS 数据集的 T-SNE 可视化
Fig.10 T-SNE visualization based on the NEU-CLS dataset

20%、50%和 80%的情况下,均获得最高准确率,即便训练样本仅占 20%,准确率仍高达 98.12%。其中,模型规模最小(1.46×10^6)的 ShuffleNet 准确率最低,仅为 89.17%,VGG16 模型大小达 268×10^6 ,性能也仅能达到 95.76%。

表 5 展示了 NEU-CLS 数据集在 20%训练样本下各模型分类准确率。VFD-SA 模型在各类别上均保持较高的准确率,平均准确率达到 98.12%,展现出在少样本训练时对不同缺陷特征强大的提取与分类能力。相较而言,ShuffleNet 在 Ps 类别上的准确率仅为 86.25%。

表 4 不同训练样本比例下的模型性能对比
(NEU-CLS 数据集)

Model	Accuracy/%			Size/ 10^6
	20%样本	50%样本	80%样本	
ShuffleNet	89.17	97.00	100.00	1.46
VGG16	95.76	98.22	99.72	268.00
RepVGG	95.63	98.33	100.00	30.10
ResNet18	92.15	97.22	99.44	11.69
ResNet50	95.83	99.67	100.00	23.50
VFD-SA	98.12	100.00	100.00	11.85

表 5 NEU-CLS 数据集上的具体分类准确率对比(20%训练样本)

Table 5 Comparison of specific classification accuracy on the NEU-CLS dataset (with 20% training samples)								%
Model	Accuracy							
	Sc	Pa	Ps	In	Cr	Rs	Average	
ShuffleNet	90.00	90.83	86.25	87.92	91.67	88.33	89.17	
VGG16	96.67	96.67	94.17	95.00	96.25	95.83	95.76	
RepVGG	95.42	95.83	95.00	95.42	97.08	95.00	95.63	
ResNet18	88.75	93.75	89.58	92.50	95.41	92.92	92.15	
ResNet50	97.58	96.25	95.42	96.25	96.67	95.83	95.83	
VFD-SA	97.92	98.75	96.25	97.50	99.58	98.75	98.12	

表 6 数据进一步验证了 VFD-SA 模型的泛化性能。在该数据集上,当训练样本比例为 20%、50%、80% 时,VFD-SA 模型准确率分别达到 88.86%、95.10%、99.48%,泛化能力出色,尤其在样本受限场景下优势更为突出。

本文的缺陷分类模型仅 11.85×10^6 ,约为 VGG16 的 1/20。依托变分特征解纠缠和锐度感知技术,该模型能够高效提取缺陷特征,快速收敛至最优损失区域,同步提升准确性与泛化性,其轻量化特性为广泛应用奠定了坚实基础。

表 6 不同训练样本比例下的模型性能对比
(自制紧固件数据集)

Table 6 Comparison of model performance under different training sample ratios (self-made fastener dataset)				
Model	Accuracy/%			Size/ 10^6
	20%样本	50%样本	80%样本	
ShuffleNet	76.05	92.34	97.92	1.46
VGG16	79.71	93.18	98.96	268.00
RepVGG	79.97	93.90	98.44	30.10
ResNet18	72.77	91.99	98.44	11.69
ResNet50	83.90	93.54	98.96	23.50
VFD-SA	88.86	95.10	99.48	11.85

3.3.3 消融实验

为验证本文算法中 VFD 和 SA 模块对分类性

能的作用,对比多个类别(Sc、Pa、Ps、In、Cr、Rs)下的准确率及平均准确率,结果如表 7 所示。

表 7 NEU-CLS 数据集上的消融实验结果(20%训练样本)

Table 7 Ablation experiment results on the NEU-CLS dataset (with 20% training samples)

Method	Accuracy						%
	Sc	Pa	Ps	In	Cr	Rs	
Baseline	91.67	95.00	81.25	90.42	97.08	97.50	92.15
+VFD	97.91	98.75	94.17	96.25	98.75	97.08	97.08
+SA	97.08	97.92	90.42	94.58	96.25	94.58	95.14
VFD-SA	97.92	98.75	96.25	97.50	99.58	98.75	98.12

整体而言,相较于基准模型,仅采用+VFD 便实现了显著提升,平均准确率提高 4.93 百分点;+SA 带来的性能提升幅度虽小于+VFD,但也较基准模型提升了 2.99 百分点;本文方法融合 VFD 和 SA 两者的优势,表现最佳,平均准确率大幅提升 5.97 百分点,达到 98.12%。实验结果证实了这些改进模块的有效性,尤其在处理困难类别(如 Ps 类)时。本文方法针对 Ps 类的准确率从 81.25%提升至 96.25%,充分表明其对难分类样本的处理效果显著。VFD 性能的显著提升,说明它能够分离出不同的特征表示,助力模型更好地理解各类缺陷的独特特征。SA 则使模型在参数空间中探寻到更平坦、泛化能力更强的解,帮助模型在仅利用 20%训练样本的情况下,也能获得较好的泛化能力。本文将两者的优势相融合,产生协同效应,VFD 助力模型学习到更具区分度的特征表示,SA 则确保这些特征在不同样本上稳定有效。

4 结束语

本文提出一种基于变分特征解纠缠的锐度感知表面缺陷分类算法,用于工业缺陷分类。该算法运用 VAE 对缺陷图像进行特征解纠缠,通过重采样和重构,有效增强缺陷图像特征。同时,引入锐度感知最小化,以寻找更平坦的最小损失区域,显著提升模型在小样本场景下的泛化性能。实验结果表明,在 GC10-DET、NEU-CLS 以及自制紧固件数据集上,该算法显著提升了模型分类效果,充分验证了算法的有效性与可行性,在钢铁表面缺陷检测领域展现出一定的应用潜力。

然而,当前算法在特征解纠缠策略上存在局限,仅简单地对特征进行分离与重构,未深入挖掘分离特征间的内在联系。后续研究拟设计更高效的策略,改进特征解纠缠方法,充分发挥特征间的关联作用。此外,鉴于现实场景中缺陷图像普遍存在数据不平衡问题,后续将重点提升模型在极端不平衡数

据场景下的性能,进一步完善算法设计。

参考文献

[1] 宋文琦, 吴龙, 黎尧. 小样本条件下的带钢表面缺陷检测[J]. 计算机系统应用, 2024, 33(5): 85-93.
SONG W Q, WU L, LI Y. Surface defect detection of strip steel with few shots [J]. Computer Systems and Applications, 2024, 33(5): 85-93. (in Chinese)

[2] KHANAM R, HUSSAIN M, HILL R, et al. A comprehensive review of convolutional neural networks for defect detection in industrial applications[J]. IEEE Access, 2024, 12: 94250-94295.

[3] LI F, LI F, XI Q G. DefectNet: toward fast and effective defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 2507109.

[4] 杨春龙, 吕东瀚, 张勇, 等. 融合自适应下采样的带钢表面缺陷检测算法[J]. 钢铁研究学报, 2024, 36(6): 806-816.
YANG C L, LÜ D H, ZHANG Y, et al. Fusion of adaptive down-sampling for strip steel surface defect [J]. Journal of Iron and Steel Research, 2024, 36 (6): 806-816. (in Chinese)

[5] LIU Y, YUAN Y C, BALTA C, et al. A light-weight deep-learning model with multi-scale features for steel surface defect classification[J]. Materials, 2020, 13(20): 4629.

[6] LI Q N, YANG Z P, SUN H X. Fine-grained classification of rail fastener images based on B-CNN[C]//Proceedings of the 5th IEEE Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Washington D.C., USA: IEEE Press, 2021: 1018-1024.

[7] XU J Y, LE H, HUANG M Z, et al. Variational feature disentangling for fine-grained few-shot classification [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2022: 8792-8801.

[8] FORET P, KLEINER A, MOBAHI H, et al. Sharpness-aware minimization for efficiently improving generalization [EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2010.01412>.

[9] 余蓉, 熊邦书, 欧巧凤. 基于改进 LeNet-5 优化算法的轴承故障诊断研究[J]. 南昌航空大学学报(自然科学版), 2023, 37(4): 82-87, 114.
YU R, XIONG B S, OU Q F. Bearing fault diagnosis based on the improved LeNet-5 optimization algorithm[J]. Journal of Nanchang Hangkong University (Natural Sciences), 2023, 37(4): 82-87, 114. (in Chinese)

[10] ZENG N Y, WU P S, WANG Z D, et al. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3507014.

[11] 谢敬峰, 王玲, 尹湘云, 等. 基于卷积神经网络的钣金件表

- 面缺陷分类识别方法[J]. 计算机测量与控制, 2020, 28(6): 187-190, 196.
- XIE Z F, WANG L, YIN X Y, et al. Classification and recognition method of sheet metal parts surface defects based on convolution neural network[J]. Computer Measurement & Control, 2020, 28(6): 187-190, 196. (in Chinese)
- [12] LIU T, YE W. A semi-supervised learning method for surface defect classification of magnetic tiles[J]. Machine Vision and Applications, 2022, 33(2): 35.
- [13] 胡坤, 吴国庆, 胡祖辉, 等. 基于改进的 VGG16 网络金属表面缺陷图像分类研究[J]. 计算机应用与软件, 2024, 41(6): 175-180.
- HU K, WU G Q, HU Z H, et al. Metal surface defect image classification based on improved VGG16 network[J]. Computer Applications and Software, 2024, 41(6): 175-180. (in Chinese)
- [14] 罗晶, 周威, 张昱中, 等. 基于对抗性弱化的多阶段钢材表面缺陷分类算法[J]. 组合机床与自动化加工技术, 2024(7): 170-176, 181.
- LUO J, ZHOU W, ZHANG Y Z, et al. Adversarial weakening based multi-stage classification for steel surface defects[J]. Modular Machine Tool & Automatic Manufacturing Technique, 2024(7): 170-176, 181. (in Chinese)
- [15] 王亚, 甘青松, 沈琦, 等. 基于动态联合加权的带钢表面缺陷分类方法[J]. 计算机工程, 2025, 51(6): 286-296.
- WANG Y, GAN Q S, SHEN Q, et al. Classification method for surface defects of strip steel based on dynamic joint weighting[J]. Computer Engineering, 2025, 51(6): 286-296. (in Chinese)
- [16] KINGMA D P, WELING M. Auto-encoding variational Bayes[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/1312.6114>.
- [17] GONZALEZ-GARCIA A, VAN DE WEIJER J, BENGIO Y. Image-to-image translation for cross-domain disentanglement[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/1805.09730>.
- [18] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/1803.07728>.
- [19] FENG Z Y, XU C, TAO D C. Self-supervised representation learning by rotation feature decoupling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 10356-10366.
- [20] LIU Y H, SANGINETO E, CHEN Y J, et al. Smoothing the disentangled latent style space for unsupervised image-to-image translation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 10780-10789.
- [21] CHEN R T Q, LI X C, GROSSE R, et al. Isolating sources of disentanglement in variational autoencoders[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/1802.04942>.
- [22] SONG K C, YAN Y H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects[J]. Applied Surface Science, 2013, 285: 858-864.
- [23] LÜ X M, DUAN F J, JIANG J J, et al. Deep metallic surface defect detection: the new benchmark and detection network[J]. Sensors, 2020, 20(6): 1562.
- [24] 单东日, 童灿, 乃学尚, 等. 基于小波和灰度共生矩阵的带钢表面缺陷识别[J]. 制造技术与机床, 2020(2): 120-123.
- SHAN D R, TONG C, NAI X S, et al. Recognition of surface defects on strip based on wavelet and gray level co-occurrence matrix[J]. Manufacturing Technology & Machine Tool, 2020(2): 120-123. (in Chinese)
- [25] 陆雅诺, 陈炳才, 陈德刚, 等. 一种基于注意力模型的带钢表面缺陷识别算法[J]. 激光与光电子学进展, 2021, 58(14): 1410014.
- LU Y N, CHEN B C, CHEN D G, et al. Recognition algorithm of strip steel surface defects based on attention model[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410014. (in Chinese)

文字编辑 吴云芳
栏目编辑 宋 圆