



仪器仪表学报
Chinese Journal of Scientific Instrument
ISSN 0254-3087, CN 11-2179/TH

《仪器仪表学报》网络首发论文

题目：基于漏磁内检测的自监督缺陷检测方法
作者：刘金海，赵贺，神祥凯，鲁丹宇，唐建华
收稿日期：2020-07-20
网络首发日期：2020-09-27
引用格式：刘金海，赵贺，神祥凯，鲁丹宇，唐建华. 基于漏磁内检测的自监督缺陷检测方法. 仪器仪表学报.
<https://kns.cnki.net/kcms/detail/11.2179.TH.20200925.1713.010.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于漏磁内检测的自监督缺陷检测方法

刘金海¹, 赵贺¹, 神祥凯¹, 鲁丹宇², 唐建华³

(1.东北大学信息科学与工程学院 沈阳 110004;2.沈阳智谷科技有限公司 沈阳 110004;3.中海油能源发展装备技术有限公司, 天津 300452)

摘要：漏磁内检测是长输管道主要检测方式。目前漏磁内检测数据分析中的缺陷检测方法环境适应性差，需要大量样本为不同环境分别建立检测模型，难以满足实用需求。本文提出一种自监督的缺陷检测方法，可以在少量样本下建立精确缺陷检测模型，克服目前缺陷检测方法需要大量样本才能训练精确模型的问题，并且在不同环境下检测效果都有所提升，因此具有良好的适用性和迁移性。首先对漏磁内检测缺陷数据进行预处理，接着将缺陷数据自适应的可视化，然后利用视觉表示对比学习的简单框架(Simple framework for contrastive learning of visual representations, SIMCLR)对可视化后的缺陷进行训练获得预训练权重，最后采用深度学习完成对缺陷的识别与定位。试验研究表明，本文设计的自监督检测方法能够有效解决可标记数据少的问题，具有检测精度高，迁移性好，泛化能力强的特点。

关键词：缺陷检测；深度学习；自监督；视觉表示对比学习的简单框架

中图分类号：TE88 TP277 TH878+3 **文献标识码：**A **国家标准学科分类代码：**440.55

Self-supervised defect detection method based on magnetic flux leakage internal detection

Liu Jinhai¹, Zhao He¹, Shen Xiangkai¹, Lu Danyu², Tang Jianhua³

(1. School of Information Science and Engineering, Northeastern University, Shenyang 110004, China; 2. Shenyang Zhigu Technology Company, Shenyang 110004, China; 3. Energy Development Equipment Technology Co. Ltd, CNOOC, Tianjin 300452, China)

Abstract: Magnetic flux leakage internal detection is the main detection method for long-distance pipelines. Current defect detection method in magnetic flux leakage internal detection data analysis has poor environmental adaptability, require a large number of samples to establish detection models for different environments, which is difficult to meet practical requirement. This paper proposes a self-supervised defect detection method, which can establish an accurate defect detection model with a small number of samples, overcomes the problem that current defect detection method requires a large number of samples to train an accurate model, and the detection effects are all improved in different environments. So the proposed method has good applicability and migration. Firstly, the defect data in the magnetic flux leakage internal detection are pre-processed, the defect data are adaptively visualized, then simple framework for contrastive learning of visual representations (SIMCLR) is used to train the visualized defects to obtain pre-training weights, and finally deep learning is used to complete the identification and localization of the defects. Experiment research shows that the self-supervised detection method designed in this paper can effectively solve the problem of less tagged data, and has the characteristics of high detection accuracy, good migration and strong generalization ability.

Keywords: defect detection; deep learning; self-supervision; simple framework for contrastive learning of visual representations

0 引言

近年来，由于各国经济的快速发展，对能源的需求也越来越迫切。其中石油和天然气是主要能源，因此为这两种能源运输提供可靠保证就很重要。只有管道

的安全运行，才能够保障油气等能源的有效运输及充分利用，同时保障人民的生活^[1]。

然而随着我国管道保有量的快速增加和使用时间的逐步延长^[2]，风险也在逐渐增加。一旦未能及时发现

收稿日期：2020-07-20 Received Date: 2020-07-20

基金项目：国家重点研发计划（2017YFF0108800），国家自然科学基金（61973071, 61627809, 61703087）和辽宁省自然科学基金（2019-KF-03-04）资助

并采取必要措施,不仅会造成严重的能源浪费、经济损失,而且会严重污染环境^[3]。因此对管道定期检查,确保管道的安全状态至关重要。在管道的检查中,最常用的是无损检测,作为预防管道泄漏的关键步骤之一,无损检测应用十分广泛^[4-5]。

漏磁检测属于无损检测的一种,漏磁检测对环境要求不高,可靠性好^[6],可以同时检测内外缺陷并且不受管道输送介质的影响,所以目前管道的内检测技术主要采用漏磁检测,根据测得的漏磁数据就可判别管道的内部情况。

对于漏磁数据的后续数据分析主要分为以下步骤:数据的预处理、目标的识别、目标的反演以及安全评估。其中目标识别是核心步骤之一,目前国内外已经有较为成熟的目标识别技术^[7-8],但是所需要的样本量一般较大。

在漏磁数据的目标识别中,缺陷识别是必不可少的重要组成部分,准确的识别缺陷能够有效保证管道的安全运行。缺陷识别分为阈值法和机器学习方法,机器学习进一步分为传统机器学习和深度学习。

在传统的机器学习算法中,文献^[9]在采用随机森林等方法进行缺陷的识别,在理想的牵拉试验条件下,效果比较理想。文献^[10]采用贝叶斯等方法对缺陷信号进行检测,取得了不错的效果。文献^[11]采用基于数据融合的小波变换,用来检测异常边缘区域。文献^[12]采用极限学习机对缺陷信号进行识别与分类。上述文献都采用机器学习方法进行检测,所需要的样本量少,但是精度一般也较差。

在深度学习算法中,文献^[13]通过采用时空正则化等方法实现了在复杂环境中的异常检测。文献^[14]提出利用深度学习提取特征网络,有效的完成了管道漏磁缺陷信号识别。由深度学习相关文献可知,采用深度学习算法来检测缺陷时,取得的精度一般较高,但样本需求也较大。

在内检测中,不同的检测环境(例如管材不同、内检测器不同)缺陷特征差别巨大,导致检测方法在未知的环境中检测能力显著下降。实际内检测的通常做法是针对新的环境参数,事先做模拟新环境的试验,获得少量的缺陷样本(称之为少量有标记样本)。然后根据少量有标记样本有限更新检测模型,再在实际管道检测数据上应用。这种方法由于样本量有限,且难以完全模拟实际环境,检测模型的精度提升有限。

近年来自监督得到快速发展,其基本思想是利用未标记数据自我训练模型,从而实现少量样本下模型

精度的提高。文献^[15]将自监督与深度学习相结合,使自监督能够学习到良好的特征。文献^[16]将原图旋转一定角度,通过预测旋转的角度来学习特征,学习到的表示方法能够在下游任务中产生良好的作用。文献^[17]提出的 SIMCLR,其性能已经逼近于有监督学习。

针对不同环境下缺陷样本获取困难导致的检测模型精度低问题,受到自监督思想的启发,本文提出一种基于漏磁内检测缺陷的自监督检测方法,将少量有标记样本和未经分析的大量实际管道检测的数据(称之为大量无标记数据)联合用于更新检测模型,能够显著提高检测模型精度。本文的主要创新如下:

(1)提出了缺陷数据自适应的转换方法。该方法通过图像增强对数据进行处理,能够在不同场景下都得到较为清晰的训练样本,具有良好的通用性。

(2)提出了自监督的缺陷检测方法。通过这种方法,在使用少量有标记样本的条件下就能训练出精确的检测模型,并且在不同检测环境下都具有良好的性能,具有检测精度高,迁移性好,泛化能力强的特点。

本文的论文结构为:第一节给出了平均中值,三次样条插值等数据预处理的方法。第二节提出了缺陷数据自适应的可视化方法,可以将小缺陷放大,大缺陷缩小,为之后的训练提供清晰的样本。第三节设计了生成缺陷检测模型方法。

1 缺陷数据预处理

在实际检测中,传感器采集到的漏磁数据首先要进行数据预处理,缺陷数据在进行预处理操作后才能减少干扰的影响。数据预处理分为:基线校正,异常值判别,数据插补和滤波这四个部分。缺陷数据首先要进行基线校正,在进行异常值判别后如果有异常数据要进行数据修正,数据缺失检测后如果数据存在缺失要进行数据插补,最后将处理好的数据进行整体滤波。

基线校正主要解决基线漂移或传感器异常,采用平均中值的算法,首先,计算每一个通道的均值:

$$S_j = \frac{1}{K} \sum_{i=1}^K a_{ij} \quad (1)$$

式中: S_j 是第 j 个通道的数据均值, $j=1, 2, \dots, N$, N 是通道数量, K 是信号轴向采样点的数量, a_{ij} 是第 j 个通道在第 i 个采样点的原始幅值。

然后利用通道均值数据完成一次校准:

$$a'_{ij} = a_{ij} - S_j + S \quad (2)$$

式中: S 为所有通道的中值。

异常值判别能够有效解决数据剧烈震荡的情况,当最终数值过高时为异常值。具体公式如式(3)所示。

$$Z = (a'_{ij} - \mu) / \sigma \quad (3)$$

式中 Z 代表着标准差范围内的原始值与总体均值之间的距离, $a'_{ij} - \mu$ 为离均差, σ 表示标准差。

数据插补中采用三次样条插值对数据进行插补,能够有效解决数据缺失的问题。

滤波采用二维高斯滤波对校正后的数据进行高斯滤波处理。具体公式如式(4)所示。

$$\Phi(x) = G(x) \cdot L \quad (4)$$

式中: $\Phi(x)$ 是滤波后的数据, G 是高斯滤波函数, L 是校正后的数据, x 表示每一个数据点。

2 缺陷数据可视化

2.1 缺陷数据灰度化

本文设计了一种缺陷数据自适应的转换方法,能够完整地将缺陷自适应的显示出来。在将缺陷自适应显示之前,需要将数据灰度化,灰度化之后才能突出缺陷边缘特征,达到自适应的显示效果。

将缺陷数据灰度化的方法采用的是基于中值的灰度化。由于轴向信号在缺陷处波动较大,单纯采用中值灰度化会导致整体的灰度偏移,因此基于灰度偏移的情况设计了分段映射的方法,以每段数据的中值为基准进行分段灰度映射,能够完整的将经过预处理的数据映射到 $[0, 255]$ 的灰度区间。其具体公式如公式(5)所示。

$$G_{i,j} = \begin{cases} (G_m - G_{cv}) \cdot \frac{V_{i,j} - V_{m_i}}{V_{max} - V_{m_i}} + G_{cv} & V_{i,j} > V_{m_i} \\ G_{cv} & V_{i,j} = V_{m_i} \\ G_{cv} \cdot \frac{V_{i,j} - V_{min}}{V_{m_i} - V_{min}} & V_{i,j} < V_{m_i} \end{cases} \quad (5)$$

式中, $G_{i,j}$ 为转换后灰度值; G_m 为最大灰度值, 8 位像素下取值为 255; G_{cv} 为中值点的邻接灰度值, 常取值为 50; $V_{i,j}$ 为位置 (i, j) 信号幅值, V_{m_i} 为通道 i 幅值中值, V_{max} 为信号幅值最大值, V_{min} 为信号幅值最小值。

2.2 缺陷数据自适应转换法

由于单纯的将数据灰度化并不能很好的体现缺陷信号,训练样本并不清晰,效果也不理想。因此考虑将灰度化后的缺陷数据伪彩色化。针对缺陷数据的特殊性,设计了缺陷数据自适应的转换法,能够将缺陷特征

自适应的显示出来,为之后的训练提供清晰的样本。

文献^[18]表明,采用数据增强能够有效的将缺陷信号更明显的表现出来,因此采用图片增强对数据进行处理。图片增强主要分为空间域增强与频率域增强两个方面,考虑到实际信号的特征,利用图像平滑与图像锐化的方式实现缺陷数据的图片增强。

图像平滑可以针对图像中的噪声进行不同程度的抑制,图像平滑采用的方式有均值滤波、中值滤波以及高斯滤波的滤波信号,针对缺陷信号的特殊性,其缺陷分布主要接近为高斯噪声,因此利用高斯滤波将高斯模板中的每一个偏重与实际的管道数据做运算进而去除管道数据中的高斯噪声。

图像锐化可以使灰度值跳跃的区域,通过退化手段实现其增强,因此可以将小缺陷更好的凸显出来。

对于一个离散的二维离散函数 $f(i, j)$, 可以用一个有限差分作为梯度值的近似值,在实际应用中采用 Robert 交叉梯度,来实现基于一阶导数的图像增强,从而达到图片锐化的效果,其表达式如公式(6)所示。

$$|\nabla f(i, j)| = [f(i+1, j) - f(i, j)] + [f(i, j+1) - f(i, j)] \quad (6)$$

式中, $f(i, j)$ 为二维离散函数

这里采用了一阶导数微分算子与二阶导数微分算子相互结合,从而达到更好的图片增强的效果。对于离散的二维图像 $f(i, j)$, 用于图像锐化的拉普拉斯算子如式(7)所示。

$$\nabla^2 f = (f(i, j+1) + f(i, j-1) + f(i+1, j) + f(i-1, j)) - 4f(i, j) \quad (7)$$

式中, $f(i, j)$ 为二维离散函数

缺陷数据在图片增强之后,缺陷特征能够体现的更加明显,此时需要进行对数据进行伪彩色化,通过不同的映射关系,将灰度值转换为 R、G、B 三个通道的值,分别对应的映射关系为 f_R, f_G, f_B , 所代表的公式分别为如式(8), (9), (10)所示:

$$C_R(i, j) = \begin{cases} 0 & 0 \leq G_{i,j} < 96 \\ 255(G_{i,j} - 96) / 64 & 96 \leq G_{i,j} < 160 \\ 255 & 160 \leq G_{i,j} < 256 \end{cases} \quad (8)$$

$$C_G(i, j) = \begin{cases} 0 & 0 \leq G_{i,j} < 32 \\ 255(G_{i,j} - 32) / 64 & 32 \leq G_{i,j} < 96 \\ 255 & 96 \leq G_{i,j} < 160 \\ 255(192 - G_{i,j}) / 32 & 160 \leq G_{i,j} < 192 \\ 255(G_{i,j} - 192) / 64 & 192 \leq G_{i,j} < 256 \end{cases} \quad (9)$$

$$C_B(i, j) = \begin{cases} 255 & 0 \leq G_{i,j} < 96 \\ 255(128 - G_{i,j}) / 64 & 96 \leq G_{i,j} < 128 \\ 0 & 128 \leq G_{i,j} < 192 \\ 255(G_{i,j} - 192) / 64 & 192 \leq G_{i,j} < 256 \end{cases} \quad (10)$$

式中, $G_{i,j}$ 为转换后灰度值;

通过图像平滑与图片锐化等方法, 并经过伪彩色化后, 可以使缺陷达到自适应显示, 使小缺陷得到凸显, 大缺陷进行压缩, 为之后的训练提供了清晰的样本。

3 自监督建模方法

3.1 基于 SIMCLR 的检测模型预训练

由于缺陷数据在同一场景中可标记量较少, 并且有大量的不同背景下的无标记数据, 本文因此结合自监督理念提出了一种 SIMCLR 与深度学习结合的一种检测算法。SIMCLR 为视觉表象对比学习框架, 其核心思想为对比学习。本文利用 SIMCLR 生成预训练模型的主要步骤为: 将伪彩色化后的漏磁图片裁剪为 64×64 大小, 选用缺陷数据作为正样本, 其他背景区域作为负样本制作数据集, 图片数量为 10 万张。随机送入未标记的 64×64 的实测数据集数据, 对里面的每张图片做两种数据增强, 通过神经网络编码器 $f(\cdot)$ 提取出数据增强后的图片特征。将同一张图中相似的进行拉近, 不同的进行拉远。最后通过一个非线性的神经网络映射架构, 能够将编码后的数据映射到损失空间, 从而使最终网络能够进行损失函数的计算。利用交叉熵对每个互异的目标计算损失函数, 损失函数的计算方法如公式(11)所示。

$$l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j) / \tau)}{\sum_{k=1}^{2N} L_{[k \neq i]} \exp(\text{sim}(z_i, z_j) / \tau)} \quad (11)$$

式中, $L_{[k \neq i]}$ 是一个指示函数, 其默认值为 1, τ 代表着温度参数, 其数值影响着模型的好坏, $\text{sim}(z_i, z_j)$ 表示两个向量之间的余弦相似性。

通过公式计算使整体的损失函数达到最小, 从而扩大同类图像的相似性, 扩大不同类图像的差异性, 从而增加检测模型的迁移能力。

通过充分训练 SIMCLR, 能够得到良好的预训练模型, 并得到最终的神经编码器 $f(\cdot)$, 能够作为下游任务用于深度学习的训练中。

3.2 生成检测模型

用 SIMCLR 得到预训练权重之后, 利用深度学习对缺陷进行识别检测, 利用的是高速区域卷积神经网络

(Faster Region Convolutional Neural Networks, Faster R-CNN) 算法, 其主要步骤为:

(1) 标注缺陷数据用作数据集, 并设置训练集与测试集的比例。

(2) 将 SIMCLR 训练得到的预训练模型替换到 Faster R-CNN 特征提取网络中, 作为其特征提取网络参数的初始值, 以此基准用 Faster R-CNN 训练漏磁缺陷数据, 得到缺陷数据的最终检测模型 $f(\cdot)$ 。

用 Faster R-CNN 训练好的模型, 用来测试测试集, 并得到相应的里程与类型, 根据漏磁信号还原管道缺陷轮廓, 便能够在实际场地中准确的找到缺陷的位置, 进行修补与排查工作。

3.3 整体算法流程

Algorithm 1: 整体算法流程

Input: 传感器第 j 个通道第 i 个采样点采集的值 N_{ij} , 异常值阈值 R , batch size N , constant τ , structure of f, g, T

```

1: for  $j=1 \rightarrow N_{ij}$  do
2:   for  $i=1 \rightarrow N_{ij}$  do
3:      $S_j = \frac{1}{K} \sum_{i=1}^K N_{ij}$ 
4:      $a'_{ij} = a_{ij} - S_j + S$ 
5:     if  $|(a'_{ij} - \mu) / \sigma| > R$  then
6:        $a'_{ij} = R$ 
7:     if 数据缺失 then
8:        $g_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$ 
9:        $\Phi(x) = G(x) \cdot L$ 
10:    until 整个通道采样点处理结束
11:  end for
12: end for
13:  $V_{ij} \rightarrow G_{ij}$ 
14:  $G_{ij} \rightarrow C_R(i, j), C_G(i, j), C_B(i, j)$ 

```

15: **for all** $i \in \{1, \dots, 2N\}$ and $j \in \{1, \dots, 2N\}$ **do**

16: $s_{i,j} = z_i^T z_j / (\|z_i\| \|z_j\|)$

17: **Define** $l(i,j)$ as $l(i,j) =$

$-\log \frac{\exp(s_{i,j} / \tau)}{\sum_{k=1}^N [l(2k-1, 2k) + l(2k, 2k-1)]}$

18: **until** 1 最小

19: **end for**

20: **return** 神经编码器 $f(\cdot)$

21: $f(\cdot) \rightarrow f'(\cdot)$

4 实验结果与分析

4.1 实验环境

通过用 SIMCLR 训练得到预训练模型, 用 Faster R-CNN 得到检测模型, 两者的实验环境都是如下表格所示。

表 1 实验环境

Table 1 Experiment environment

名称	参数
系统	LINUX64 Ubuntu18.04
框架	pytorch
语言	python
神经网络	Resnet50
CPU	Intel Xeon Silver 4110(2.1 GHz)
GPU	2 颗 Tesla P4
RAM	40G

4.2 试验环境描述

一个约 800 米的管道环路试验平台如图 1 所示。该管道回路由近 67 条 8 英寸管道组成, 内径 200mm, 壁厚 10mm。管道材料为 X65 碳钢, 介质流速为 0.5m-3m/s, 操作压力为 3MPa。本文中所有缺陷数据来源于由磁偶极子模型产生的人工缺陷和试验场中的真实缺陷, 数量总共为 3500 个。其缺陷长度为 9.7mm-60mm, 缺陷宽度为 9.5mm-39.6mm, 缺陷深度为 0.335mm-9.3mm。



图 1 管道环路实验平台图片

Fig. 1 Picture of the pipeline loop experiment platform

本文中训练集缺陷样本伪彩色化后的图片大小为 200×144 , 测试集中缺陷样本伪彩色化后的图片大小为 500×144 与 200×144 , 通过测试集检测训练模型能够有效证明模型的检测精度高与迁移效果好的特点。

4.3 实验结果分析

4.3.1 伪彩色图结果分析

通过数据增强等方法能够明显的将小缺陷显示出来, 能够为之后的训练提供足够清晰的样本。实验证明, 采用缺陷数据自适应的转换方法, 能够将管道内的各个缺陷的特征表现得更加明显。缺陷数据对比图如图 2, 图 3 所示。

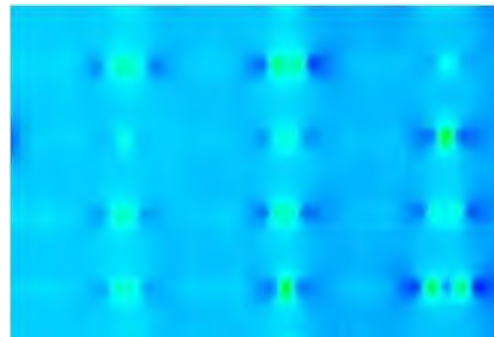


图 2 原始缺陷数据转伪彩色图

Fig. 2 Converting original defect data into pseudo color image

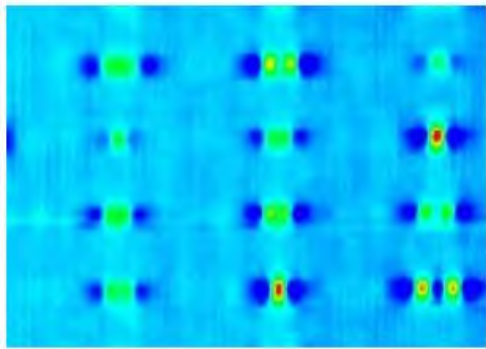


图 3 改进后缺陷数据转伪彩色图

Fig. 3 Converting the improved defect data into pseudo color image

4.3.2 检测模型结果分析

为了充分验证本文方法的可行性, 因此采用了多次实验并进行测试, 具体实验验证分为以下三个方面:

(1) P-R 曲线与 AP 值对比

为了能体现出本文设计的自监督缺陷检测方法的优越性, 检验模型的性能, 采用文献^[19]引入的标准来评估缺陷检测精度, 如果检测出的缺陷是真实的缺陷, 叫做正检缺陷(true positive, TP); 真实缺陷但未被检出, 则叫做漏检缺陷(false negative, FN); 非缺陷但被检出为缺陷, 叫做误检缺陷(false positive, FP)。精度和召回率计算公式如式(12), (13)所示。

$$\text{精度} = \frac{TP}{TP+FP} \cdot 100\% \quad (12)$$

$$\text{召回率} = \frac{TP}{TP+FN} \cdot 100\% \quad (13)$$

精度表示缺陷检测的准确性, 召回率表示缺陷检测的全面性。高精度和召回率表示良好的检测性能。因此通过精确率-召回率(Precision-Recall, P-R)曲线往往能得到检测模型的好坏。选取 1000 个、2000 个缺陷样本作为训练集, 分别用基于 SIMCLR 生成的预训练权重与在公共数据集下进行训练, 并绘制 P-R 曲线进行对比。

选用 1000 个缺陷样本用作训练集, 在没有考虑目标置信度的前提下, 其两种不同模式下的 P-R 曲线如图 4 所示。

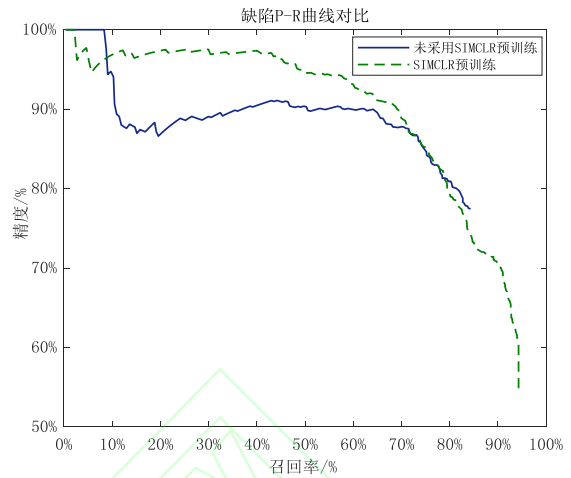


图 4 1000 个缺陷样本 P-R 曲线对比图

Fig. 4 P-R curve comparison chart of 1000 defect samples

由图 4 可知, 采用 SIMCLR 预训练与未采用 SIMCLR 预训练相比, SIMCLR 召回率最终可达到 94% 左右, 未采用 SIMCLR 最终召回率仅在 83% 左右。SIMCLR 下模型平均精确度(Average Precision, AP)为 0.86, 未采用 SIMCLR 预训练生成的检测模型 AP 为 0.76, 且在相同召回率的前提下, 采用基于 SIMCLR 训练模型的 Faster R-CNN 平均精度要高于未采用 SIMCLR 预训练下的 Faster R-CNN 检测模型, 因此表明 SIMCLR 与 Faster R-CNN 结合的方式对于少量缺陷样本参与训练时的情况, 模型最终检测精度得到提升。

选用 2000 个缺陷样本用作训练集, 在没有考虑目标置信度的前提下, 其两种不同模式下的 P-R 曲线如图 5 所示。

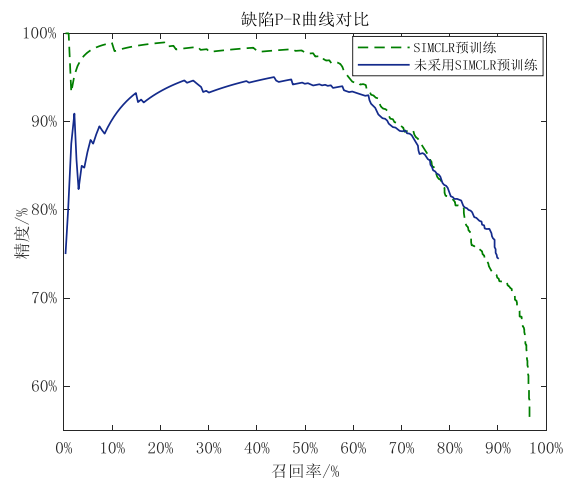


图 5 2000 个缺陷样本 P-R 曲线对比图

Fig. 5 P-R curve comparison chart of 2000 defect samples

由图 5 的 P-R 曲线图可知, 采用基于 SIMCLR 的 Faster R-CNN 与未采用 SIMCLR 的 Faster R-CNN 相

比, SIMCLR 的最终召回率在 96%左右, 未采用 SIMCLR 的召回率接近 91%。SIMCLR 下模型的 AP 为 0.89, 未采用 SIMCLR 的预训练模型 AP 为 0.82, 在相同召回率前提下采用 SIMCLR 预训练的精度要更高。通过比较前者的检测结果整体来看检测精度较高, 其查全率相对而言更高, 相对而言前者模型的漏检概率更低。

(2) 检测精度的对比

为了更加充分证明基于 SIMCLR 的训练模型在实际检测中能够取得更好的检测效果, 因此分别选用、500 个、1000 个、1500 个、2000 个缺陷样本用作训练集, 训练集的图片大小为 200×144 。用 1106 个未参与训练的缺陷样本用作测试集, 测试集的图片大小为 500×144 与 200×144 。缺陷目标置信度 80%以上设置为正确检测, 记录下正确检测缺陷的数量。具体对比效果如图 6 所示。

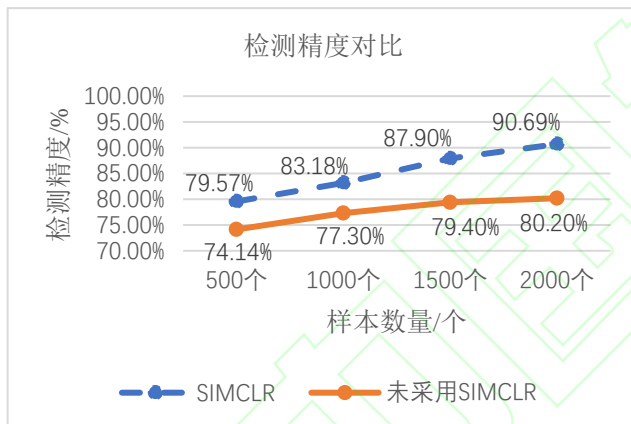


图 6 检测精度对比图

Fig. 6 Comparison chart of detection accuracy

由图 6 两者可以发现, 用自监督与深度学习结合的方法, 能够有效提高目标的检测精度, 在实际检测中能够取得更好的检测效果, 能够有效解决缺陷数据由于样本少而造成的检测精度不高的问题, 并且迁移效果得到明显改善, 提高了模型的迁移能力和整体检测精度。

(3) 与其他漏磁缺陷识别方法的对比

为了能够验证自监督检测方法的有效性, 采用基于 SIMCLR 的 Faster R-CNN 与文献^[20]采用的区卷积神经网络(convolutional neural network, CNN)检测方法进行对比。该文献采用 CNN 进行特征提取用于分类, 用 22800 个训练样本达到了 95.4%的精度。本文对该文献提出的方法进行改进, 使其能够应用于目标检测中, 并分别选用 500 个、1000 个、1500 个、2000 个缺陷用作

CNN 的训练集, 测试样本为 1106 个, 缺陷目标置信度 80%以上设置为正确检测, 记录下正确检测缺陷的数量。具体对比效果如图 7 所示。

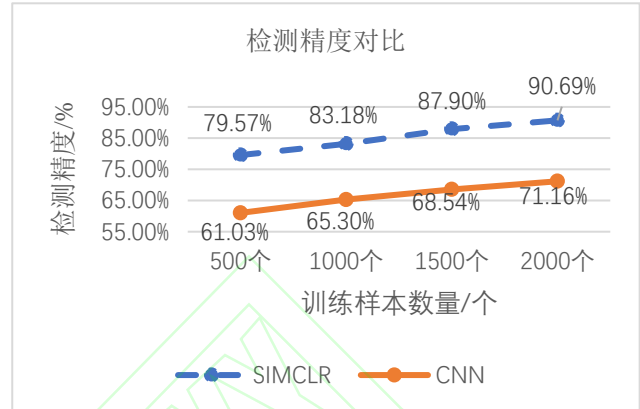


图 7 检测精度对比图

Fig. 7 Comparison chart of detection accuracy

由图 7 可见, 当采用 CNN 提取图像特征时, 训练样本太少会明显减少检测的精度, 采用基于 SIMCLR 的 Faster R-CNN, 能够明显提升缺陷的检测效果, 在实际检测中能够取得更理想的结果。

5 结论

针对目前缺陷检测需要大量样本才能得到精确模型, 环境适用性差的问题, 本文设计了一种自监督的缺陷检测方法。首先, 对漏磁内检测缺陷数据进行预处理, 接着提出一种缺陷数据自适应的可视化方法, 可以得到清晰的缺陷样本图片。最后, 设计出了自监督与深度学习相结合的检测算法, 并与常规深度学习算法和其他漏磁缺陷检测方法进行了性能对比分析, 完成了方法的验证。实验结果表明, 在少量缺陷样本条件下采用自监督的缺陷检测方法能明显优于只采用深度学习算法和其他漏磁缺陷检测方法, 检测精度明显提高, 具有良好的迁移能力。证明所提出的自监督缺陷检测方法在缺陷样本量较少情况下能够具有更好的检测效果。

参考文献

- [1] 杨理践, 耿浩, 高松巍. 长输油气管道漏磁内检测技术[J]. 仪器仪表学报, 2016, 37(08): 1736-1746.
- [2] 李忠虎, 张琳, 闫俊红. 管道腐蚀视觉测量图像边缘检测算法研究[J]. 电子测量与仪器学报, 2017, 31(11): 1788-1795.
- LI ZH H, ZHANG L, YAN J H. Research on image edge detection algorithm for pipeline corrosion visual measurement[J].

- Journal of Electronic Measurement and Instrumentation, 2017,31(11):1788-1795.
- [3] 杨理践,曹辉.基于深度学习的管道焊缝法兰组件识别方法[J].仪器仪表学报,2018,39(2):193-202.
- YANG L J, CAO H. Recognition method of pipeline weld flange components based on deep learning[J]. Chinese Journal of Scientific Instrument, 2018,39(2):193-202.
- [4] ALI F, MOHAMMAD T G, KRISTEN M. Characterization of corroded reinforced steel bars by active microwave thermography[J]. IEEE Transactions on Instrumentation and Measurement, 2015,64(9):2583-2585.
- [5] ALI F, CODY A E, MOHAMMAD T G, et al. Active microwave thermography for defect detection of CFRP-strengthened cement-based materials[J]. IEEE Transactions on Instrumentation and Measurement, 2016,65(11):2612-2620.
- [6] 杨理践,耿浩,高松巍,等.高速漏磁检测饱和场建立过程及影响因素研究[J].仪器仪表学报,2019,40(10):1-9.
- YANG L J, GENG H, GAO S W, et al. Research on establishment process and influencing factors of saturation field in high-speed magnetic flux leakage detection[J]. Chinese Journal of Scientific Instrument, 2019,40(10):1-9.
- [7] 周显恩,王耀南,朱青,等.基于机器视觉的瓶口缺陷检测方法研究[J].电子测量与仪器学报, 2016,30(5):702-713.
- ZHOU X E, WANG Y N, ZHU Q, et al. Research on bottle mouth defect detection method based on machine vision[J]. Journal of Electronic Measurement and Instrumentation, 2016,30(5):702-713.
- [8] 陈彦彤,徐伟,朴永杰,等.基于快速视网膜局部特征的遥感图像目标识别[J].仪器仪表学报,2016,37(4):852-859.
- CHEN Y T, XU W, PU Y J, et al. Remote sensing image target recognition based on fast local features of retina[J]. Chinese Journal of Scientific Instrument, 2016,37(4):852-859.
- [9] 刘金海,付明芮,唐建华.基于漏磁内检测的缺陷识别方法[J].仪器仪表学报,2016,37(11):2572-2581.
- LIU J H, FU M R, TANG J H. Defect recognition method based on magnetic flux leakage internal inspection[J]. Chinese Journal of Scientific Instrument, 2016,37(11):2572-2581.
- [10] LIU J H, FU M R, LIU F L, et al. Window feature-based two-stage defect identification using magnetic flux leakage measurements[J]. IEEE Transactions on Instrumentation and Measurement, 2018,67(1):12-23.
- [11] 曹辉,杨理践,等.基于数据融合的小波变换漏磁异常边缘检测[J].仪器仪表学报,2019,40(12):71-79.
- CAO H, YANG L J, et al. Wavelet transform magnetic leakage abnormal edge detection based on data fusion[J]. Chinese Journal of Scientific Instrument, 2019,40(12):71-79.
- [12] LIU J H, FU M R, WU Z N, et al. An ELM-based classifier about MFL inspection of pipeline[C]. Chinese Control and Decision Conference(CCDC), 2016:1952-1955.
- [13] FU M R, LIU J H, ZANG D, et al. Anomaly detection of complex MFL measurements using low-rank recovery in pipeline transportation inspection[J]. IEEE Transactions on Instrumentation and Measurement, 2020,69(9): 6776-6786.
- [14] 王宏安,陈国明.基于深度学习的漏磁检测缺陷识别方法[J].石油机械,2020,48(05):127-132.
- WANG H A, CHEN G M. Defect recognition method of magnetic flux leakage detection based on deep learning[J]. Petroleum Machinery, 2020,48(05):127-132.
- [15] Caron M, Bojanowski P, Joulin A, et al. Deep Clustering for Unsupervised Learning of Visual Features[C]. ECCV, 2018, arXiv:1807.05520
- [16] Gidaris S, Singh P, Komodakis N. Unsupervised Representation Learning by Predicting Image Rotations[C]. ICLR, 2018, arXiv:1803.07728
- [17] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]. ICML, 2020, arXiv:2002.05709
- [18] FU M R, LIU J H, ZHANG H G, et al. Multi-sensor fusion for magnetic flux leakage defect characterization under information incompleteness[C]. IEEE Transactions on Industrial Electronics, 2020, doi: 10.1109/TIE.2020.2984444.
- [19] T. Fawcett, An introduction to ROC analysis[J]. Pattern Recognit. Lett., 2006, 27(8): 861-874.
- [20] FENG J, LI F M, LU S X, et al. Injurious or noninjurious defect identification from MFL images in pipeline inspection using convolutional neural network[J]. IEEE Transactions on Instrumentation and Measurement, 2017,66(7): 1883-1892.

作者简介



刘金海(通讯作者), 2009 年于东北大学获得博士学位, 现为东北大学教授, 博士生导师。主要研究方向是工业人工智能, 数据驱动故障诊断理论与技术。

E-mail: jh_lau@126.com

Liu Jinhai (corresponding author) obtained Ph.D. degree in 2009 from Northeastern University. Now, he is a professor and PhD supervisor in Northeastern University. His main research direction is industrial artificial intelligence, data-driven fault diagnosis theory and technology.



赵贺, 2019 年于山东科技大学获得学士学位, 现为东北大学信息科学与工程学院硕士研究生, 主要研究方向为数据驱动故障诊断理论与技术。

E-mail: 781458112@qq.com

Zhao He received bachelor degree from Shandong University of Science and Technology in 2019. He is currently a master student in School of Information Science and Engineering, Northeastern University. His main research direction is data-driven fault diagnosis theory and technology.