

DOI: 10.19650/j.cnki.cjsi.J2108511

基于多维度选择性搜索的小样本缺陷识别方法*

卢森骧¹, 徐行¹, 张润江¹, 刘金海¹, 赵可天²

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

摘要: 超声内检测是油气管道缺陷的主要检测方式之一,目前超声内检测在工业小样本的情况下存在缺陷识别边界定位不准的问题。本文提出了一种基于多维度选择性搜索的小样本缺陷识别方法,该方法首先对超声回波进行特征提取,其中包含使用基于孤立森林的回波特征点提取,和基于自然断点法的特征点聚类;其次提出了风险相似性度量方法,并使用梯度提升树建立波形特征和风险程度的回归模型;然后将多维度缺陷相似性信息融合在选择性搜索算法中,实现小样本缺陷识别;最后使用异常分数等区域风险度量指标实现缺陷边界的精准定位。实验结果表明,本文设计的基于多维度选择性搜索的小样本缺陷识别方法的查全率和查准率分别高达 95.08% 和 85.46%,能有效解决超声信号缺陷识别边界定位不准的问题。

关键词: 缺陷识别;小样本;选择性搜索;孤立森林;自然断点法

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

Small sample defect recognition method based on multi-dimensional selective search

Lu Senxiang¹, Xu Hang¹, Zhang Runjiang¹, Liu Jinhai¹, Zhao Ketian²

(1. School of Information Science and Engineering, Northeastern University, Shenyang 110004, China;

2. CNOOC Energy Development Equipment Technology Co., Ltd., Tianjin 300452, China)

Abstract: Ultrasonic internal inspection is one of the main defect detection methods for the oil and gas pipeline. At present, the location of the defect boundary is inaccurate in the case of small industrial samples for ultrasonic internal inspection. This article proposes a small sample defect recognition method based on multi-dimensional selective search. Firstly, ultrasonic echo features are extracted by two steps, which are feature point extraction based on isolated forest and feature point clustering based on the natural breaks classification method. Secondly, the risk similarity measurement method is proposed. A regression model of waveform characteristics and risk degree is formulated by the boosting tree. Thirdly, multi-dimensional defect similarity is integrated information into a selective search algorithm to realize small sample defect identification. Finally, regional risk metrics such as anomaly scores are used to achieve precise positioning of defect boundaries. Experimental results show that the recall and precision of the small sample defect recognition method based on multi-dimensional selective search are 95.08% and 85.46%, which can effectively solve the problem of inaccurate positioning of the ultrasonic signal defect boundary detection.

Keywords: defect recognition; small samples; selective search; isolated forests; natural breaks classification method

0 引言

近年来,国家经济在快速发展,能源需求的国际竞争日益激烈,石油、天然气作为重要的能源,对国家建设和人民生活提升发挥着不可或缺的作用。管道运输是油气运输的主要方式,保证管道安全运行变得

至关重要^[1]。

然而管道运输的工作条件十分恶劣,在各种因素的影响下,易出现腐蚀和裂纹情况,甚至发生因管道破损造成油气泄露^[2]。油气的泄露不仅会造成巨大的经济损失,而且对生态环境会造成严重污染。所以对油气管道进行定期的无损检测十分重要,使油气运输的安全性得到保证^[3]。

收稿日期:2021-09-02 Received Date: 2021-09-02

* 基金项目:国家自然科学基金(61627809, 61973071, 61703087, 62003080)、辽宁省自然科学基金(2019-KF-03-04)项目资助

超声检测作为无损检测中重要的一种技术^[4],能够检测出管道内部细微缺陷,同时能够对缺陷尺寸进行测量,对缺陷进行定位分析,准确反映出管道的腐蚀和裂纹情况。超声检测被广泛应用于管道缺陷检测中,通过对超声内检测数据进行分析即可实现对管道缺陷的识别^[5]。目前,许多科研工作者围绕着缺陷识别做了大量的研究工作,主要包括基于图像学处理方法和深度学习方法。

图像学处理方法中,文献[6]提出了一种基于多形态成分分析的方法,该方法利用多纹理特征对遥感图像进行了有效的分类。文献[7]提出了一种纹理特征的提取方法,通过该纹理特征有效地识别了轴承保持架歪斜缺陷。文献[8]提出了一种基于信息融合的铁谱图像磨粒识别方法,该方法在滤波基础上提取了二值图像形态特征,识别准确率得到有效提升。基于图像学的处理方法多数可以解决小样本缺陷识别问题,但无法有效解决缺陷边界定位精度低的问题。然而在超声信号缺陷检测中,缺陷边界定位极为重要。

在深度学习方法中,文献[9]提出了一种端到端的钢铁表明缺陷识别方法。该方法将卷积神经网络提出的多尺度特征进行融合,融合后的特征包含了更多的缺陷位置信息,对缺陷位置检测有很强的指导作用。文献[10]提出了一种基于随机森林的缺陷识别方法,该方法使用抽象特征替代原始的显性特征进行训练,实验结果显示该方法具有很高的缺陷检测准确度。文献[11]提出了一种深度卷积神经网络级联结构,利用提出的网络结构将缺陷检测转换成两级目标检测问题,有效地解决了缺陷识别和定位问题。

上述用于缺陷检测的深度学习方法都需要大量的训练样本。然而在实际超声信号中,由于管道多数情况下是健康的,因此缺陷样本是很少的。在只有少量样本的情况下,深度网络往往无法被训练完备,以至于在缺陷识别或边界检测中误差很大。针对工业小样本问题,很多学者也做了深入研究。

文献[12]提出了一种基于自监督的小样本缺陷检测方法。该方法将视觉表示后的信号输入到使用SIMCLR预训练后的模型中,实现准确的缺陷检测。文献[13]提出了一种基于高斯概率分布回归模型的条件三元组缺陷识别方法,该方法在工业小样本缺陷识别中有很好的结果。文献[14]提出了一种基于注意力机制的生成对抗网络,用于小样本红外图像检测,实验结果相比于原始深度网络有很高的提升。

基于这些小样本缺陷识别研究方法,本文提出了一种基于多维度选择性搜索的小样本缺陷识别方法,解决了在单个尺寸管道少量缺陷样本情况下,缺陷边界定位精度低的问题。该方法在选择性搜索的基础上,提出了风险相似性度量方法,用异常分数等区域

风险度量指标,实现缺陷边界的精确定位。本文的主要创新如下:

1) 提出了一种基于多维度选择性搜索的缺陷识别方法。该方法通过对多种缺陷相似度信息的融合,实现对候选区域的合并,增强缺陷识别及定位的准确性。

2) 提出了一种缺陷风险相似性的度量方法。该方法通过对风险评估模型的训练,建立不同缺陷之间相似性评估,有效增强候选区域合并能力,提高缺陷定位的精度。

1 超声检测原理

超声波是超声探头压电晶片在电脉冲的激励下所产生的一种机械波,具备声波所具有的反射、折射和散射等属性,具有穿透力强的特点^[15]。

超声检测原理是利用材料内部缺陷声学性质发生变化的特点来探测材料的内部缺陷。由于超声波在不同的介质界面上传播时,部分能量将被反射和折射^[16],如果被测物体存在缺陷,超声波在缺陷表面会发生反射,其反射波到达传感器时间与正常管壁处不同,利用这一特点,通过对接收到的超声信号进行分析处理,即可完成对缺陷的定性定量分析。超声检测原理如图1所示。

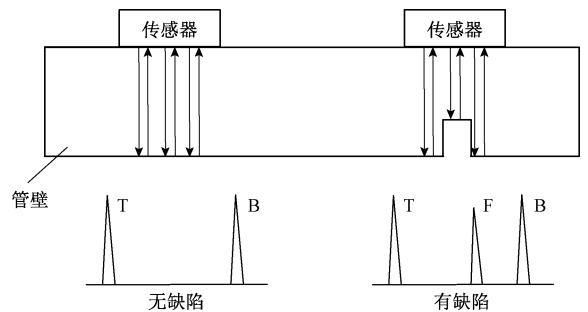


图1 超声检测原理图

Fig.1 Schematic diagram of ultrasonic detection

由图1可知,与无缺陷位置相比,缺陷处二次回波F和一次回波T的距离更短,即管壁厚度更小,因此可以对管壁进行多阵列检测,根据不同管壁厚度产生的波形信号特征来实现管道缺陷的定性定量分析。

2 整体方法介绍

本文提出了一种基于多维度选择性搜索的小样本缺陷识别方法,该方法主要包含两部分内容,分别是基于孤立森林的回波特征提取和基于多维度选择性搜索的缺陷识别方法,算法整体流程如图2所示。

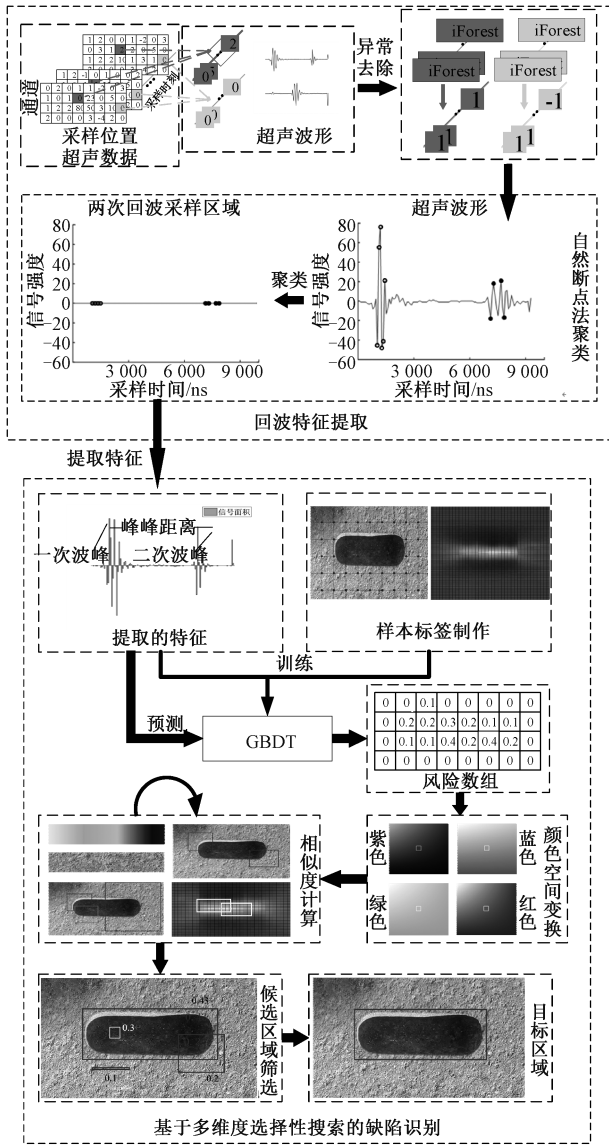


图2 算法整体流程图

Fig. 2 Algorithm overall flow chart

在基于孤立森林的回波特征提取中,首先使用孤立森林去除异常点并提取超声回波特征点,然后使用自然断点法对提取的峰值特征进行聚类,并通过聚类结果来计算壁厚。在基于多维度选择性搜索的缺陷识别中,首先使用提取的回波特征和样本标签训练一个特征提升树。并且根据信号临近区域危险程度无突变原则,提出了风险相似度量方法,为不同子区域合并增加一维重要信息。其次使用梯度提升树建立波形特征和风险程度的回归模型,然后将多维度缺陷相似性信息(包括颜色、纹理、尺寸、兼容和风险相似度)融合在选择性搜索算法中,为实现精准的小样本缺陷识别和定位提供了充足的信息;最后使用异常分数等区域风险度量指标实现缺陷边界的精准定位。

3 超声回波特征提取

超声回波特征提取主要包括基于孤立森林的回波特征点提取和基于自然断点法的特征点聚类两部分。孤立森林法主要通过多个随机数的训练完成对特征点的检测,若直接对特征点计算特征值,余波的特征点可能会影响特征值的确定,从而在缺陷识别中产生严重的误检问题,影响缺陷识别精度。自然断点法的应用很好的解决了这一问题,通过对特征点的聚类,从而准确计算出各类特征值。

3.1 基于孤立森林的回波特征点提取

对超声每个采样时刻所有通道的数据分别构造孤立森林,不同的森林扮演不同时刻超声回波峰值特征提取的专家。

使用孤立森林对超声检测数据 $X_{m \times n \times l}$ 进行特征提取,步骤如下:

1) 遍历每个采样时刻,循环执行以下步骤:

(1) 提取当前采样时刻所有通道的数据 $X_{m \times n}$;

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

(2) 指定孤立树数量 l 和下采样样本大小 Ψ , 训练孤立森林;

2) 循环结束,得到孤立森林集合 $\{iForest_k | 1 \leq k \leq l\}$;

3) 遍历每个超声波形 $X_l = [x_1, x_2, \dots, x_l]$, 对于每个采样时刻 k 的信号值 X_k , 使用对应的孤立森林模型 $iForest_k$ 进行预测, 得到波形对应的预测值 $Y_l = [y_1, y_2, \dots, y_k, \dots, y_l]$, $y_k \in \{-1, 1\}$, $1 \leq k \leq l$, 从而得到超声回波特征点。

3.2 基于自然断点法的特征点聚类

超声波在管道内外壁的回波信号强度会发生快速衰减,信号在不同回波之间的采样区域强度很弱,有明显的分界线,即同一次回波会集中在连续的一段采样时间内。利用这个特点,对超声信号强度较大的采样时刻进行聚类,聚类后同一回波范围内的采样时间会被划分为同一类,也就对不同次管道内外壁的超声回波划分了区域,从而寻找到回波峰值点位置。超声回波特征点聚类示意图如图3所示。

自然断点法主要包含以下步骤:

1) 确定聚类个数 K 。计算公式如式(2)所示。

$$num = \frac{T \times s \times v}{W_l} + 1 \quad (2)$$

其中, T 为超声检测器采样周期, s 为采样次数, v 为超声在钢管管壁中的传播速度, W_l 为壁厚。

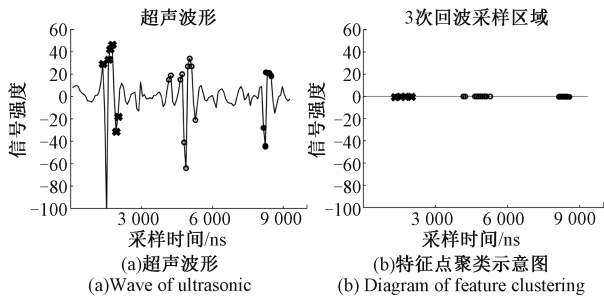


图3 超声回波特征点聚类示意图

Fig. 3 Diagram of ultrasonic echo feature point clustering

2) 确定聚类采样区域。根据超声在管壁回波处信号强烈,其他区域信号平稳,即超声波形的能量主要集中在回波位置的特点,选择强度最大的 m 个采样时刻 $\{t_1, t_2, \dots, t_m \mid m \geq K\}$ 并且满足式(3)所示条件。

$$\frac{\sum_{t=t_1}^{t_m} f^2(t)}{\sum_{t=1}^n f^2(t)} \geq T_1 \quad (3)$$

其中, T_1 为超声管壁回波信号能量在总能量中的比例阈值,在实际计算过程中,通常令 $T_1 = 90\%$ 。

3) 自然断点法聚类分析。根据管道设计壁厚,计算得到最大聚类个数如式(4)所示。

$$K_{\min} = \frac{T \times s \times v}{Wt_{de}} + 1 \quad (4)$$

其中, Wt_{de} 为当前管道设计壁厚。

由超声反射原理可知, K 次回波的中心区域应该是两两等间距的。利用这个特点,首先将步骤2) 筛选出的 m 个采样点使用自然断点法聚 $K = K_{\min}$ 类,得到聚类中心 $c_1, c_2, c_3, \dots, c_{K_{\min}}$, 判断:

$$T_2 \leq \frac{c_{k+1} - c_k}{c_k - c_{k-1}} \leq T_3 \quad (5)$$

其中, $k = 2, 3, \dots, K-1$; T_2, T_3 为比例阈值。

如果聚类中心满足式(5),则认为超声有 K 次回波。否则,将聚类次数加1,直到聚类中心满足式(5)或达到最大聚类次数为止,此时,前两次聚类中心 c_1, c_2 为超声信号的一、二次回波。

4) 计算特征。由步骤3)可知管道内壁和外壁回波时刻分别为 c_1 和 c_2 ,则可计算出各种等特征值。

4 基于多维度选择性搜索的小样本缺陷识别方法

4.1 风险评估模型建立

1) 训练样本建立

超声波形中包含了多种特征参数,可以真实的反应

出波形特点,选取波形均值、波形方差、信号能量、信号面积、一次波峰峰值、二次波峰峰值、一次波峰位置、二次波峰位置、波峰衰减系数、峰峰距离、回波数量共11个参数作为特征参数,以提升缺陷识别的准确率。

确定了模型的特征参数后,以每个采样点的危险程度作为标签:

$$y_{i,j} = \frac{Loss_{i,j}}{Wt}, y_{i,j} \in [0, 1] \quad (6)$$

其中, $Loss_{i,j}$ 为采样点 (i, j) 的金属损失, Wt 为管道正常壁厚。

2) 训练风险评估模型

提升树 (boosting tree) 是以决策树为基函数的线性组合对函数或数据进行拟合的集成模型,本文使用的提升树是由多个二叉回归树集成而成的。定义提升树的模型为:

$$f(x) = f_M(x) \quad (7)$$

损失函数为:

$$L = (y - f(x))^2 \quad (8)$$

本文使用随机梯度下降法来最小化损失函数,完成对模型参数的训练。

根据内检测的采样里程间隔和传感器轴向分布,按同等分布网格对人工标定的缺陷区域进行搜索,将采样波形与管壁实测金属损失对应,计算每个点的危险程度。

4.2 多维度特征相似度融合

除了颜色、纹理、尺寸和兼容4个相似度以外,特提出第5种相似度:风险相似度,会同前4种相似度并经加权后,得到总的相似度,以灰度图为例,具体如下:

1) 颜色相似度

采用 L_1 范数方法对数据进行归一化,建立全部超声数据的 $bins = n$ 直方图,进而各个子块的分布直方图具体为: $C_i = \{c_i^1, c_i^2, \dots, c_i^n\}$, 各子块之间颜色相似度为:

$$S_{color}(r_i, r_j) = \sum_{k=1}^n \min(c_i^k, c_j^k) \quad (9)$$

其中, r_i 和 r_j 代表着两个不同子块。如若两子块进行合并,得到合并后的子块 t , 对应的灰度分布直方图为:

$$C_t = \frac{size(r_i) \times C_i + size(r_j) \times C_j}{size(r_i) + size(r_j)} \quad (10)$$

2) 纹理相似度

采用尺度不变特征转换,采用 L_1 范数方法对数据进行归一化,在8个方向上做高斯微分,建立各子块纹理相似度数据的 $bins = n$ 直方图,进而各子块得到相应的纹理分布直方图为:

$$S_{texture}(r_i, r_j) = \sum_{k=1}^{2n} \min(t_i^k, t_j^k) \quad (11)$$

3) 尺寸相似度

若只采用颜色和纹理两种相似度进行子块的合并,

易出现小尺寸的子块被大尺寸子块吞并的现象,导致候选区域的均匀性得不到保证以及出现嵌套问题。针对这一问题,为小尺寸子块赋予较大的权重,各子块对应的尺寸相似度为:

$$S_{size}(r_i, r_j) = 1 - \frac{size(r_i) + size(r_j)}{size(image)} \quad (12)$$

其中, $size(image)$ 为输入图像尺寸。

4) 兼容相似度

对于两个要合并的子块来说,在颜色、纹理、尺寸3个维度上相似度高,易出现距离较远的两个子块发生合并问题。针对这一问题,特引出兼容相似度:

$$S_{fill}(r_i, r_j) = 1 - \frac{size(BB_{ij}) - size(r_i) - size(r_j)}{size(image)} \quad (13)$$

其中, BB_{ij} 是能框住子块 r_i 和 r_j 的最小矩形区域。

5) 风险相似度

考虑到管道缺陷的形成主要受自然环境持续腐蚀的影响,对于一个缺陷区域内两个位置的危险程度变化趋势是相近的,利用这一规律,将风险相似度引入到合并子块的规则中。因此,建立各子块危险程度数据的 $bins = n$ 直方图,各子块得到相应的风险分布直方图,具体为: $D_i = \{d_i^1, d_i^2, \dots, d_i^n\}$, 各子块间颜色相似度为:

$$S_{risk}(r_i, r_j) = \sum_{k=1}^n \min(d_i^k, d_j^k) \quad (14)$$

两个不同子块进行合并时,考虑上述5种相似度,并进行相似度融合,进而得到最终的相似度为:

$$S(r_i, r_j) = a_1 S_{color}(r_i, r_j) + a_2 S_{texture}(r_i, r_j) + a_3 S_{size}(r_i, r_j) + a_4 S_{fill}(r_i, r_j) + a_5 S_{risk}(r_i, r_j) \quad (15)$$

其中, $a_i \in \{0, 1\}$, 表示是否使用该相似度。

4.3 缺陷候选区域搜索

缺陷候选区域搜索首先利用基于图的贪心聚类算法将原始图像分割成多个小块,即初始化原始区域,然后使用贪心思想,在候选分割块里选择相似度最高的两个小块进行合并,直到只剩下一块区域为止,最后将每一步产生的合并小块保存下来,组成可能存在的目标候选区域。

在本文中首先通过风险评估模型得到风险数组,对风险数组进行颜色空间转换,将输入数据转换为常见的颜色空间,从不同的角度体现目标区域的特点,常用的颜色空间有 RGB、灰度空间 I、rgl(归一化 RGB 通道 rg 通道加上强度)、Lab、rgb(归一化 RGB)、C、HSV、HSV 的 H 通道。然后将原始数据分割成各个候选子块,综合考虑颜色、纹理、尺寸、兼容、风险这5个相似度,最后根据相似度最高的两个子块进行合并的原则,合并至只有一块区域,得到缺陷候选区域集合 $\mathbf{W} = \{W_1, W_2, \dots, W_N\}$ 。

4.4 区域筛选

由于改进选择性搜索算法得到的候选区域可能存在“回”字形嵌套、交叠和非缺陷区域,所以需要对候选区域进行筛选,除常规的缺陷长宽比、候选区域最小尺寸等规则外,主要筛选规则如下:

1) 异常分数

对于每个候选区域而言,其对应的风险数据反应了管道异常的真实情况,因此可以提取该区域的平均管道危险程度作为候选区域的异常分数,分数越高说明候选区域越危险,其定义为:

$$S_{W_k} = \frac{1}{m_k \times n_k} \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} Risk_{ij} \quad (16)$$

其中, m_k 和 n_k 分别为第 k 个候选区域的窗体宽度和高度。然后选取候选区域异常分数大于 λ_1 的区域作为候选区域,其中 $0 \leq \lambda_1 \leq 1$, 得到缺陷候选区域集合 $\mathbf{W} = \{W_{1'}, W_{2'}, \dots, W_{N'}\}$ 。

2) 边界精确

经过异常分数筛选之后,对于每个缺陷区域而言,可能存在多个候选区域相互交叠。因此,为了获取最合适的候选区域,使用空间交并比来衡量两个候选区域的重叠情况,其定义为:

$$IOU = \frac{s(W_{i'} \cap W_{j'})}{s(W_{i'} \cup W_{j'})} \quad (17)$$

其中, $s(\cdot)$ 表示缺陷候选区域的面积。如果两个候选区域的空间交并比大于 λ_2 , 则剔除异常分数低的候选区域,保留异常分数大的区域作为真实候选区域,最后得到缺陷候选区域集合 $\mathbf{W} = \{W_{1'}, W_{2'}, \dots, W_{N'}\}$ 。

3) 边界放缩

经过边界精确筛选之后,对于每个缺陷而言,最多只存在一个包含缺陷位置的候选区域,但是可能由于初始化分割子块时,缺陷位置的候选区域可能不完全包含缺陷或者还包括了除缺陷外的正常管壁位置。因此,为了进一步获取最合适的缺陷区域,再次利用波形风险数组对候选区域边界进行放缩。若候选区域 $W_{k'}$ 的边界在数据中的位置分别为 l_k, r_k, u_k 和 d_k , 首先对边界进行收缩,以左边界 l_k 为例,若:

$$\frac{\sum_{i=l_k}^{l_k+1} \sum_{j=d_k}^{u_k} Risk(i, j)}{size(l_k, l_k+1, d_k, u_k)} < \lambda_3 \times \frac{\sum_{i=l_k}^{r_k} \sum_{j=d_k}^{u_k} Risk(i, j)}{size(l_k, r_k, d_k, u_k)} \quad (18)$$

则将候选区域 $W_{k'}$ 的边界更新为 l_k+1, r_k, u_k 和 d_k , 其中, λ_3 为缩放倍数, $size(\cdot)$ 表示所在区域大小。同理,可对其它边界进行收缩。

由于候选区域可能不完全包含缺陷,因此需要对边界进行试探扩张,以左边界为例,若:

$$\frac{\sum_{i=l_k-1}^{l_k} \sum_{j=d_k}^{u_k} Risk(i,j)}{size(l_k-1, l_k, d_k, u_k)} > \lambda_4 \times \frac{\sum_{i=l_k}^{r_k} \sum_{j=d_k}^{u_k} Risk(i,j)}{size(l_k, r_k, d_k, u_k)} \quad (19)$$

则将候选区域 W_k 的边界扩张为 l_k-1, r_k, u_k 和 d_k , λ_4 为缩放倍数。同理,可对其它边界进行扩张。将所有候选区域按以上规则筛选后,得到最终的缺陷目标位置区域。

5 实验结果与分析

5.1 试验环境描述

仿真所采用的超声数据是超声内检测器经实际测量得到,管道实验场如图 4 所示。管道全长为 42.7 m,由 12 根 12 寸的管道组成,管道直径为 273 mm,壁厚为 12.7 mm,整个管道共包含 51 个缺陷。超声内检测器带有 128 个传感器,采样间隔为 6.70 mm。

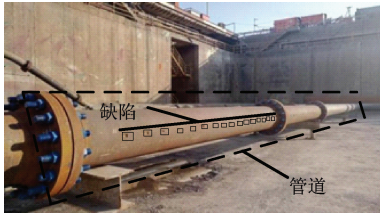


图 4 管道实验场
Fig. 4 Experimental platform of pipe

缺陷根据其长宽共分为 7 类,分别为一般金属损失,坑状金属损失,轴向凹沟,环向凹沟,孔,轴向凹槽和环向凹槽。

缺陷识别精度的评判采用查准率 (Precision)、查全率 (Recall) 和 F1 值 3 种分类评价指标,具体定义如下:

$$Precision = \frac{TP}{FP + TP} \quad (20)$$

$$Recall = \frac{TP}{FN + TP} \quad (21)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (22)$$

其中, TP 代表正例被正确分为正例的样本数,在本文中为包含人工标定缺陷的检出区域数量; FP 代表负例被错误分为正例的样本数,在本文中为不包含人工标定缺陷的检出区域数量; FN 代表正例被错误分成负例的样本数,在本文中为未检出的人工标定缺陷数量。

5.2 实验结果分析

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 5 个方面:

1) 特征提取方法评估

图 5 给出了管段有缺陷位置和无缺陷位置峰峰值特征计算结果。

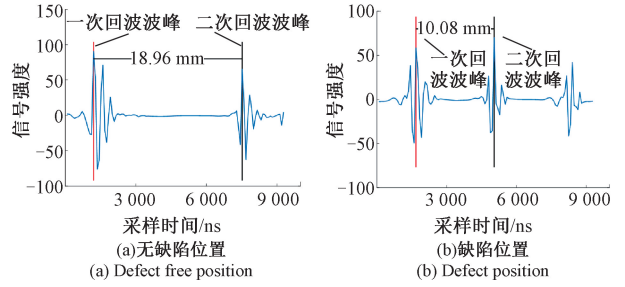


图 5 无缺陷位置和缺陷位置的峰峰值特征计算结果

Fig. 5 Results of feature of defect free position and defect position

基于孤立森林和自然断点法的特征提取方法的提出和应用能够准确确定各种特征值,避免余波特征点对特征提取的干扰。因此,该算法在实测超声数据中具有很强的应用价值。

2) 波形风险评估

使用制作的数据集,通过网格搜索和 10 折交叉验证得到梯度提升树模型最佳参数,为了衡量梯度提升树对管道采样点的危险程度预测效果,利用平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean square error, RMSE) 和决定系数 (R^2) 来评价该算法的拟合效果。

梯度提升树算法在训练集和测试集上的 MAE 和 RMSE 指标都很小,并且 R^2 指标都接近 1,表示模型拟合效果很好。整体来看,梯度提升树模型在未出现过拟合的情况下,能够满足后续选择性搜索算法的数据精度要求。

3) 候选区域搜索

使用选择性搜索和多维度选择性搜索算法对同一个含缺陷管段进行搜索,有效性对比结果如表 1 所示。

表 1 选择性搜索算法改进前后的有效性对比
Table 1 Comparison of effectiveness of selective search algorithm before and after improvement

方法	阈值	候选区域数量	包含缺陷数量	有效性/%
原始选择性搜索	$K = 30$	215	22	10.23
	$K = 60$	210	20	9.52
	$K = 90$	201	19	9.45
多维度选择性搜索	$K = 30$	64	23	35.94
	$K = 60$	64	22	34.38
	$K = 90$	62	22	35.48

由表1可知,相比于原始的选择性搜索算法,针对超声数据特点引入了风险相似度的多维度选择性搜索算法在减少生成候选区域数量的基础上保证了含缺陷候选区域的数量,提高了搜索的有效性。

4) 区域筛选

为经过缺陷长宽比为50、候选区域最小尺寸为5、异常分数阈值为0.1和边缘精确规则筛选后的有效性对比。

由表2可知,对候选区域根据超声数据特点设计的规则筛选后,在减少候选区域数量的基础上保证了含缺陷候选区域的数量,极大提高了缺陷的搜索准确率。

表2 缺陷区域筛选前后的有效性对比

Table 2 Effectiveness comparison of defect areas before and after screening

方法	阈值	候选区域数量	包含缺陷数量	有效性/%
筛选前	$K = 30$	64	23	35.94
	$K = 60$	64	22	34.38
	$K = 90$	62	22	35.48
筛选后	$K = 30$	24	21	87.50
	$K = 60$	24	21	87.50
	$K = 90$	23	21	91.30

5) 与其他超声缺陷识别方法的对比

为了能够验证多维度选择性搜索缺陷识别方法的有效性,特将其与基于形态学图像处理的缺陷识别方法进行对比,本节分别从计算时间、缺陷识别精度和边界定位精度来评价算法的缺陷识别效果,以下分别用方法1、方法2表示基于形态学图像处理的缺陷识别方法和基于多维度选择性搜索的缺陷识别方法。

(1) 计算时间

两种算法在超声数据集上随里程变化的时间消耗曲线如图6所示。

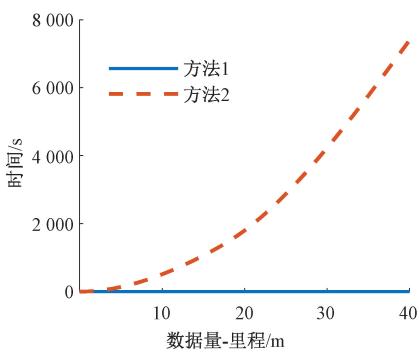


图6 算法时间对比

Fig. 6 Comparison of algorithm time

由图6可知,方法2的时间复杂度高于方法1,原因主要体现在风险评估时对波形特征的提取、对风险的预测等方面。

(2) 识别精度

就目前现有的3类超声数据集样本,对多维度选择性搜索缺陷识别方法和基于形态学图像处理的缺陷识别方法识别精度做对比,缺陷识别结果如表3所示。在本文中人工标定的缺陷数量表示为 $FN + TP$,检出区域缺陷数量用公式表示为 $FP + TP$,实际包含缺陷数量表示为 TP 。为了便于比较方法的有效性,两种方法的查全率、查准率和 $F1$ 值分别取各缺陷对应指标计算结果的平均值。

表3 两种缺陷识别方法结果对比

Table 3 Comparison of two defect recognition methods

方法	缺陷类型	人工标定缺陷数量 ($FN + TP$)	检出区域缺陷数量 ($FP + TP$)	实际包含缺陷数量 (TP)
形态学图像处理	所有损失	51	63	46
	一般金属损失	15	18	14
	坑状金属损失	28	35	25
多维度选择性搜索	轴向凹沟	8	10	7
	所有损失	51	57	48
	一般金属损失	15	16	14
多维度选择性搜索	坑状金属损失	28	32	26
	轴向凹沟	8	9	8

对应的评价指标计算结果如图7所示。由图7可知,方法1的查全率、查准率和 $F1$ 值分别为90.08%、73.06%和80.67%;方法2的查全率、查准率和 $F1$ 值分别为95.08%、85.46%和90.00%。从整体和缺陷类型分析,后者在查全率、查准率和 $F1$ 值方面都要强于前者。因此表明多维度选择性搜索方法检测精度更高,误检率更低。

(3) 边界定位精度

应用两种算法分别确定缺陷区域后,对缺陷区域进行反演即得到缺陷的长度和宽度,并将结果与人工标定数据做对比,应用MAE和RMSE评价指标得到各类缺陷长度和宽度误差如图8所示。

可知,无论从整体或缺陷类型分析,在边界长度定位精度方面,方法2优于方法1 53%,在宽度方面,方法2较方法1有30%的性能提升。

结果表明,多维度选择性搜索在边界定位精度上有着更突出的检测能力,在实际检测中样本量不足的情况下,缺陷的检测效果有着显著的提升。

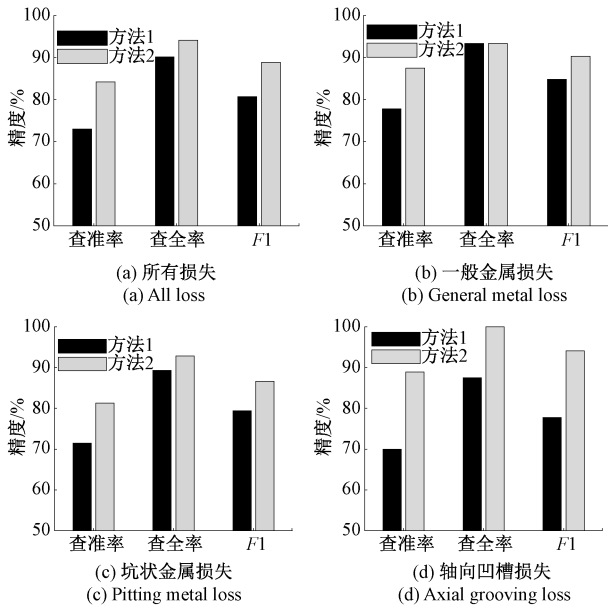


图 7 算法缺陷识别效果对比

Fig. 7 Comparison of defect recognition of different algorithms

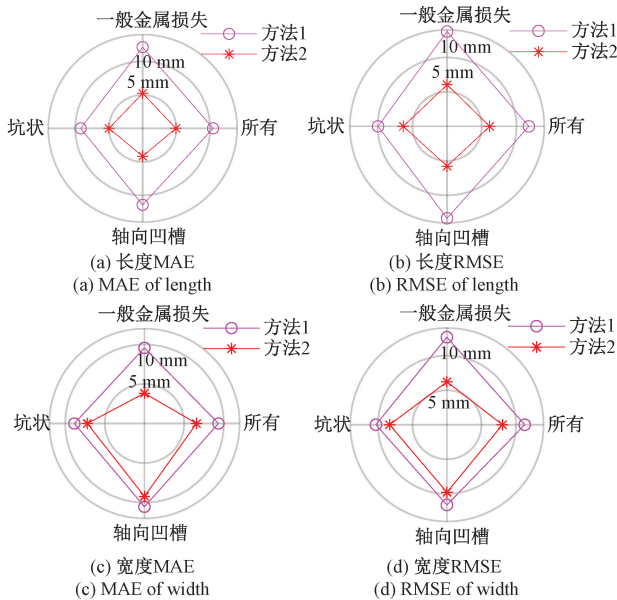


图 8 缺陷识别边界精度对比图

Fig. 8 Comparison of boundary accuracy of defect detection

6 结 论

针对超声数据内检测小样本缺陷识别定位不准确问题,本文提出了一种基于多维度选择性搜索的小样本缺陷识别方法。该方法首先通过孤立森林和自然断点法提取超声回波进行特征提取。然后将提出的缺陷风险相似性度量值等 5 种相似性信息融合到选择性搜索方法

表 4 两种方法缺陷识别结果各指标对比

Table 4 Comparison of each index of defect recognition results of two methods

方法	缺陷类型	长度 MAE	长度 RMSE	宽度 MAE	宽度 RMSE
形态学图 像处理	所有损失	10.42	11.76	9.32	11.26
	一般金属损失	12.13	13.76	9.57	12.66
	坑状金属损失	9.19	9.95	8.84	10.29
多维度选 择性搜索	轴向凹沟	11.42	13.32	10.53	11.61
	所有损失	4.93	6.14	6.56	8.03
	一般金属损失	5.23	6.04	3.83	6.20
	坑状金属损失	4.99	6.30	7.22	8.31
	轴向凹沟	4.16	8.77	9.21	9.77

中,通过一系列候选区域合并和筛选机制准确识别并定位超声缺陷。最后进行大量的对比实验。实验结果表明,在小样本超声信号缺陷识别中,提出的方法可准确识别并定位缺陷信号,效果优于其他对比方法。证明了所提方法在超声缺陷识别及定位中的有效性。

参考文献

[1] 杨理践, 郑文学, 李佳音, 等. 基于平衡电磁技术的管道裂纹全角度检测方法[J]. 仪器仪表学报, 2021, 42(6): 103-112.
 YANG L J, ZHENG W X, LI J Y, et al. Full-angle detection method of pipeline crack based on balanced electromagnetic technology [J]. Chinese Journal of Scientific Instrument, 2021, 42(6): 103-112.

[2] ZHANG H G, JIANG L, LIU J H, et al. Data recovery of magnetic flux leakage data gaps using multi-feature conditional risk [J]. IEEE Transactions on Automation Science and Engineering, 2021, 18(3): 1064-1073.

[3] 周琛, 何建, 胡暮平, 等. 基于 L(0,2) 超声导波的管段结构损伤定位研究 [J]. 仪器仪表学报, 2020, 41(12): 181-191.
 ZHOU CH, HE J, HU M P, et al. Research on damage location of pipe section based on L(0,2) ultrasonic guided wave [J]. Chinese Journal of Scientific Instrument, 2020, 41(12): 181-191.

[4] 尹昱, 陈振华, 肖峰, 等. 不锈钢疲劳微损伤的非线性超声检测方法 [J]. 电子测量与仪器学报, 2020, 34(1): 68-73.
 YIN Y, CHEN ZH H, XIAO F, et al. Nonlinear ultrasonic testing method for fatigue micro-damage of stainless steel [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(1): 68-73.

[5] YAN Y, LIU D, GAO B, et al. A deep learning-based

- ultrasonic pattern recognition method for inspecting girth weld cracking of gas pipeline[J]. IEEE Sensors Journal, 2020, 20(14): 7997-8006.
- [6] XU X, LI J, HUANG X, et al. Multiple morphological component analysis based decomposition for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(5): 3083-3102.
- [7] 郝勇, 耿佩, 温钦华, 等. 滚动轴承保持架缺陷的图像处理及模式识别方法研究[J]. 仪器仪表学报, 2019, 40(9): 162-169.
HAO Y, GENG P, WEN Q H, et al. Research on image processing and pattern recognition of skew defect of antifriction bearing cage[J]. Chinese Journal of Scientific Instrument, 2019, 40(9): 162-169.
- [8] 徐斌, 温广瑞, 苏宇, 等. 多层次信息融合在铁谱图像磨粒识别中的应用[J]. 光学精密工程, 2018, 26(6): 1551-1560.
XU B, WEN G R, SU Y, et al. Application of multi-level information fusion for wear particle recognition of ferrographic images [J]. Optics and Precision Engineering, 2018, 26(6): 1551-1560.
- [9] HE Y, SONG K, MENG Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1493-1504.
- [10] 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. MFL inner detection based defect recognition method[J]. Chinese Journal of Scientific Instrument, 2016, 37(11): 2572-2581.
- [11] TAO X, ZHANG D, WANG Z, et al. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2020, 50(4): 1486-1498.
- [12] 刘金海, 赵贺, 神祥凯, 等. 基于漏磁内检测的自监督缺陷检测方法[J]. 仪器仪表学报, 2020, 41(9): 180-187.
LIU J H, ZHAO H, SHEN X K, et al. Self-supervised defect detection method based on magnetic flux leakage internal detection [J]. Chinese Journal of Scientific Instrument, 2020, 41(9): 180-187.
- [13] 王伟, 余厚云. 基于改进深度度量学习算法的表面缺陷检测[J]. 计算机与现代化, 2021(6): 61-68.
WANG W, YU H Y. Surface defect detection based on improved deep metric learning algorithm[J]. Computer and Modernization, 2021(6): 61-68.
- [14] 吴晗, 张志龙, 李楚为, 等. 小样本红外图像的样本扩增与目标检测算法[J]. 控制理论与应用, 2021, 38(9): 1477-1485.
WU H, ZHANG ZH L, LI CH W, et al. Infrared image sample amplification and object detection method with small samples [J]. Control Theory & Applications, 2021, 38(9): 1477-1485.
- [15] 徐从裕, 胡宗久, 杨雅茹, 等. 基于FFT及IFFT的超声波相位差检测方法[J]. 电子测量与仪器学报, 2019, 33(5): 180-186.
XU C Y, HU Z J, YANG Y R, et al. Method of ultrasonic phase difference detection based on FFT and IFFT [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(5): 180-186.
- [16] WANG Y Q, LIAN M, LIU H B, et al. Incident angle identification based on first-echo energy attenuation in ultrasonic thickness measurement[J]. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2018, 65(11): 2141-2149.

作者简介



刘金海(通信作者), 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) received his Ph. D. degree from Northeastern University in 2009. He is currently a professor at Northeastern University. His main research interests include industrial artificial intelligence, data-driven fault diagnosis theory and technology.



卢森骧, 2018年于东北大学获得博士学位, 现为东北大学讲师, 硕士生导师。主要研究方向是工业人工智能, 故障诊断。

E-mail: lusenxiang@ise.neu.edu.cn

Lu Senxiang received his Ph. D. degree from Northeastern University in 2018. He is currently a lecturer and a master advisor at Northeastern University. His main research interests include industrial artificial intelligence and fault diagnosis.