

DOI: 10.11835/j.issn.2096-6717.2023.080



开放科学(资源服务)标识码 OSID:



桥梁拉索动力响应监测数据质量评价方法

邓扬¹, 张强¹, 钟国强², 柳尚², 徐润², 鞠翰文¹

(1. 北京建筑大学土木与交通工程学院, 北京 100044; 2. 山东省交通规划设计院集团有限公司, 济南 250101)

摘 要:桥梁拉索动力响应监测数据中存在大量低质量数据, 现有的监测数据检测研究集中于时域波形异常的明显异常数据, 然而监测数据中还存在时域波形正常但频域特征混乱的数据, 这类数据无法准确地反映桥梁拉索动力特性。针对该问题, 将现有的异常数据检测拓展为数据质量评价, 同时对明显异常数据和频域混乱数据进行检测。采用卷积神经网络(CNN)和数据频域特征建立桥梁拉索动力响应监测数据质量评价方法, 实施流程包括: 采用快速傅里叶变换(FFT)将时域数据序列转化为功率谱密度函数(PSDF), 利用格拉姆角场(GAF)方法对 PSDF 序列进行可视化, 进而搭建 CNN 模型, 对监测数据质量进行自动化评价。以某斜拉桥的拉索加速度监测数据为例开展应用研究, 结果表明, 与时域序列检测方法相比, PSDF 序列检测方法能够更好地区分正常与频域混乱数据, 评价准确率更高; 利用两个传感器监测数据建立的 CNN 模型对所有 26 个传感器监测数据质量评价准确率均在 94% 以上; 此外, 将该评价模型应用于另一座类似桥梁的监测数据质量评价中, 准确率也达到 95%。

关键词:桥梁健康监测; 数据质量评价; 卷积神经网络; 机器视觉

中图分类号: TU317 **文献标志码:** A **文章编号:** 2096-6717(2025)05-0208-10

Data quality evaluation method for dynamic response monitoring of bridge cables

DENG Yang¹, ZHANG Qiang¹, ZHONG Guoqiang², LIU Shang², XU Run²,
JU Hanwen¹

(1. School of Civil and Transportation Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, P. R. China; 2. Shandong Provincial Communications Planning and Design Institute Group Co., Ltd., Jinan 250101, P. R. China)

Abstract: There is a large amount of low-quality data in the monitoring data of bridge cable dynamic response.

收稿日期: 2023-04-19

基金项目: 国家自然科学基金(51878027); 北京市教委青年拔尖人才培养计划(CIT & TCD201904060); 北京建筑大学基本科研业务费项目(X20174、X21073); 山东省交通运输厅科技计划(2021B66)

作者简介: 邓扬(1984-), 男, 教授, 博士生导师, 主要从事结构健康监测研究, E-mail: dengyang@bucea.edu.cn。
鞠翰文(通信作者), 男, E-mail: jhwrrr@sina.com。

Received: 2023-04-19

Foundation items: National Natural Science Foundation of China (No. 51878027); Beijing Municipal Education Commission (No. CIT & TCD201904060); Fundamental Research Funds for Beijing University of Civil Engineering and Architecture (Nos. X20174, X21073); Science and Technology Plan Project of Shandong Provincial Department of Transportation (No. 2021B66)

Author brief: DENG Yang (1984-), professor, doctoral supervisor, main research interest: structural health monitoring, E-mail: dengyang@bucea.edu.cn.

JU Hanwen (corresponding author), E-mail: jhwrrr@sina.com.

The existing data detection research focuses on obviously abnormal data with abnormal time-domain waveforms. However, there is chaotic data in frequency-domain characteristics with normal time-domain waveforms in the monitoring data, which can't accurately reflect the dynamic characteristics of bridge cables. Aiming at this problem, the existing abnormal data detection is extended to data quality evaluation, and obvious abnormal data and frequency-domain chaotic data are detected at the same time. The data quality evaluation method of bridge cable dynamic response monitoring is established by using a convolutional neural network (CNN) and data frequency-domain features. The implementation process includes: the time-domain data sequence is transformed into a power spectral density function (PSDF) by fast Fourier transform (FFT); the Gramian angular field (GAF) method is used to visualize the PSDF sequence, and a CNN model is built to evaluate the data quality automatically. Taking the cable acceleration monitoring data of a cable-stayed bridge as an example, the proposed method is validated. The results show that compared with the time-domain sequence detection method, the PSDF sequence detection method can better distinguish normal and pseudo-normal data, and has a higher evaluation accuracy; the accuracy of the CNN model, established by using the monitoring data of two sensors, to evaluate the quality of all 26 sensor monitoring data is above 94%. In addition, the evaluation model, established by this method, is applied to the monitoring data quality evaluation of another similar bridge with an accuracy of 95%.

Keywords: bridge health monitoring; data quality evaluation; convolutional neural network; machine vision

拉索作为缆索承重体系桥梁中的关键构件,在外界环境因素激励下易产生振动。拉索的大幅振动会引起其保护层脱落,拉索梁锚固区出现疲劳裂纹等病害^[1-2],大大降低了其使用寿命。因此,需要对拉索的振动及状态进行监测,获取其响应特征,为拉索健康状态的准确评估和安全预警提供科学依据。

由于电源故障、电磁干扰等因素的影响,传感器不可避免地产生大量低质量或异常监测数据^[3-5],严重影响结构健康监测系统的准确性。因此,有必要对监测系统产生的海量数据进行质量评估,从而将低质量数据从监测系统数据库中剔除,避免在结构健康监测过程中对结构损伤预警的错报和状态评估的误判。以桥梁拉索动力响应为例,低质量监测数据主要包括以下两类:1)数据的时域波形明显不同于正常数据,通常表现为平波、漂移、方波等^[6-7],可以视为明显异常数据,根据时域波形就可以识别,学者们对此做了大量研究^[6-10];2)从数据的时域波形看不出明显异常,但其频域特征混乱,称为频域混乱数据。受环境因素影响,这类数据无法清晰地识别桥梁拉索的动力特性,如模态频率和阻尼比等,所以是无意义和无用的。因此,在现有异常数据检测识别的基础上,进一步开展桥梁拉索监测数据的质量评价。

高斯模型是异常监测数据检测的常用方法。在模型训练阶段,通常使用最大似然估计方法来估计高斯分布的均值和方差;在测试阶段,采用一些统计测试方法,如均方根误差等。Yang等^[11]提出了

一种基于全局最优高斯混合模型的无监督离群点检测方法,将每个数据点的离群因子定义为混合比例的加权,权重表示与其他数据点的相似性,从而检测数据的离群点。为了提高异常数据检测的准确性,Tang等^[12]提出了一种基于子空间学习和高斯混合模型(GMM)的异常数据检测方法,利用子空间学习的局部保持投影(LPP)来有效保持邻域结构,该方法提高了检测结果的真阳性率。然而,在处理结构健康监测系统产生的大数据样本时,这些基于统计模型的方法适用性可能较差^[13],在数据量较大的情况下很难获取数据统计特征和精确的模型参数。

随着人工智能技术的发展,深度学习和机器视觉已被应用到结构异常监测数据检测中。Bao等^[6]提出了一种基于深度学习的异常监测数据检测方法,通过对数据序列图像进行标记并训练深度神经网络(DNN)模型来识别异常数据。作为一种具有卷积计算和深层结构的前馈神经网络^[14],卷积神经网络(CNN)也被用于结构异常监测数据的检测中。高照明^[15]将时域数据序列可视化后输入CNN中,进行异常数据的分类和检测。Tang等^[7]进一步堆叠桥梁加速度数据的时域图像和频域图像,并通过多次信息输入提高了CNN对异常数据图像的识别精度。为了解决不平衡训练集的问题,Mao等^[16]提出了一种基于无监督学习的深层神经网络方法,该方法结合了生成对抗网络和自动编码器,虽然避免了不平衡训练集的问题,并且不需要为数据标记标签,但模型采用二分类,不能关注不同类型异常数

据的特征。

传统异常监测数据检测研究大多数关注了时域波形异常的明显异常数据^[6-7,15-16],但对于低质量的频域混乱数据研究较少,并未特别重视。大跨桥梁拉索动力响应监测数据异常特征较复杂,建立高精度的预测模型较困难,同时,为了提高模型的泛化性,大量训练样本的分类和标记需要大量人力和时间成本。为了克服上述缺点,进一步提高低质量监测数据的识别精度,提出一种基于 CNN 的数据质量评价方法,提取桥梁加速度监测数据的频域特征,训练 CNN 模型,对低质量监测数据进行自动化检测和分类。在此基础上,以国内某大型斜拉桥的拉索加速度数据为例,验证该方法对于数据质量的评价准确性,并进行跨桥梁对象应用,使用已建立的 CNN 模型对另一座斜拉桥的加速度数据进行质量评价。与传统的异常监测数据检测方法相比,本文提出的数据质量评价方法通过增强数据图像特征,提高 CNN 模型对数据质量的评价精度,解决传统的时域序列检测方法不能评价频域混乱数据的问题,在跨桥梁对象的应用中具有良好的泛化能力和鲁棒性。

1 监测数据质量评价方法

1.1 方法流程概述

现有的异常监测数据检测常采用高斯模型,而这种基于统计模型的方法在处理结构健康监测系统产生的大数据样本时可能适用性较差,在数据量较大的情况下,很难获取数据统计特征和精确的模型参数。本文针对桥梁拉索动力响应开展监测数据质量评价方法研究,利用桥梁拉索加速度监测数据的频域特征,构造监测数据的二维表达式,增强各类型数据的质量特征,利用机器视觉技术对各类型低质量监测数据进行自动化检测与分类,具有较高的精确度和较好的泛化能力。监测数据质量评价方法如图 1 所示。具体步骤为:1)通过短时滑动窗口将桥梁拉索响应的长期监测数据划分为较短的时域数据序列。2)通过快速傅里叶变换(FFT)将每个划分的时域数据序列转化为功率谱密度函数(PSDF),基于 PSDF 序列中包含的频域特征来判断每个划分的时域数据序列的数据质量类型。3)每个划分的 PSDF 序列通过格拉姆角场(GAF)方法转换为 GAF 图像^[17],根据上述判断的数据质量类型对 GAF 图像进行分类和标注。4)建立包括输入层、卷积层、池化层、全连接层和输出层的卷积神经网络(CNN)模型,用于对划分的时域数据序列进行数据质量评价。CNN 模型在训练过程

中提取 GAF 图像特征并学习分类。5)将训练后的 CNN 模型用于自动评价桥梁拉索动力响应监测数据质量。

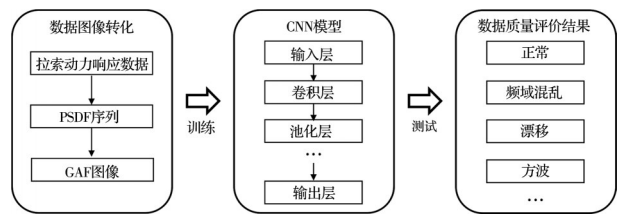


图 1 桥梁拉索动力响应监测数据质量评价方法

Fig. 1 Quality evaluation method of monitoring data for dynamic response of bridge cables

1.2 GAF 图像转化

为了充分利用 CNN 在机器视觉和图像分类方面的优势,将分割后的 PSDF 数据序列转换为格拉姆角场(GAF)图像,然后将图像输入 CNN 模型进行训练和分类。GAF 在极坐标系中对数据序列进行编码,在 Gramian 矩阵中的每个值都是角度的三角函数值。给定数据序列 $S=(s_1, s_2 \cdots s_d)$, 对数据序列进行归一化,使所有值都落在区间 $[-1, 1]$ 之间,并获得归一化序列 $\tilde{S}=(\tilde{s}_1, \tilde{s}_2 \cdots \tilde{s}_d)$ 。根据式(1)、式(2),将归一化的数值编码为角度的余弦值,时间戳编码为半径,以极坐标表示时间序列 \tilde{S} 。

$$\phi = \arccos(\tilde{s}_h), -1 \leq \tilde{s}_h \leq 1, \tilde{s}_h \in \tilde{S} \quad (1)$$

$$r = \frac{t_h}{D}, t_h \in D \quad (2)$$

式中: \tilde{s}_h 为 \tilde{S} 中的数值; t_h 为对应于 \tilde{s}_h 的时间戳; D 为时间戳的总长度; ϕ 为极角; r 为极径。一维数据序列的 Gramian 矩阵 G 定义为

$$G = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_d) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_d) \\ \vdots & \ddots & \vdots \\ \cos(\phi_d + \phi_1) & \cdots & \cos(\phi_d + \phi_d) \end{bmatrix} \quad (3)$$

在矩阵 G 中,对角线由缩放的原始序列值组成,时间随着对角线从左上角向右下角移动,因此,时间维度也被编码到矩阵 G 中。GAF 方法可以保留数值之间的时间依赖性,得到的矩阵 G 为双射映射,转换方法如图 2 所示。利用上述特点将 PSDF 数据序列转换成 GAF 图像,基于 GAF 进行数据转换后,各种质量数据具有明显的图像特征,有利于 CNN 模型的学习和识别。选择一个短时间窗口来划分数据序列,包括 1 200 个数据样本点,每个 GAF 图像具有 24 位灰度。为了降低模型训练成本,对图像像素进行压缩,将图像分辨率降低到 100×100 。

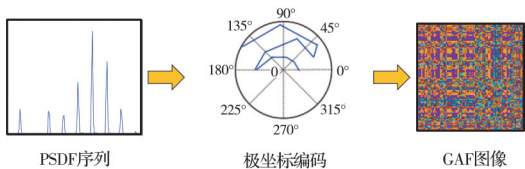


图 2 GAF 映射说明

Fig. 2 Illustration of the encoding map of GAF

1.3 卷积神经网络模型

作为前馈神经网络的主要代表,卷积神经网络(CNN)在图像分析中具有局部接受域、权值共享和空间域子采样的特点^[18]。所选择 CNN 模型^[19]的结构和相关参数如表 1 所示,第 1 层为尺寸为 100×100 的输入层图像;第 2 层为第 1 卷积层,图像尺寸为 $94 \times 94 \times 6$,包括 6 个用于卷积运算的卷积核,尺寸为 7×7 ;第 3 层采用最大池化层,子采样尺寸为 2×2 ;第 4 层为第 2 卷积层,图像尺寸降低到 $42 \times 42 \times 9$,包括 9 个用于卷积运算的卷积核,尺寸为 6×6 ;第 5 层为最大池化层,子采样尺寸为 3×3 ;第 6 层是全连接层;第 7 层是一个尺寸为 6×1 的分类器,用于对各种类型的数据质量进行分类。

表 1 CNN 模型参数

Table 1 CNN model parameters

层数	网络层	尺寸(像素)	操作	尺寸(像素)
1	输入层	100×100		
2	卷积层 1	$94 \times 94 \times 6$	卷积核 1	7×7
3	最大池化层 1	$47 \times 47 \times 6$	子采样 1	2×2
4	卷积层 2	$42 \times 42 \times 9$	卷积核 2	6×6
5	最大池化层 2	$14 \times 14 \times 9$	子采样 2	3×3
6	全连接层	$1\ 764 \times 1$		
7	输出层	6×1		

2 实例验证:斜拉桥拉索加速度监测数据

为了验证提出的数据质量评价方法的准确性,提取了某斜拉桥的拉索加速度监测数据。如图 3 所示,该桥是一座空间四索面独塔斜拉桥,拉索加速度监测系统包含 26 个加速度传感器通道,加速度传感器的采样频率为 50 Hz,各通道的安装位置和编号如图 3 所示。

滑动窗口的序列长度设置为 1 200,利用时间滑动窗口将加速度监测数据划分为长度为 1 200 的短时段时域数据序列,利用 FFT 得到分割后时域数据序列的 PSDF 序列。根据上述方法,将 PSDF 数据序列转换为 GAF 图像,经过压缩和调整后,GAF 图像分辨率为 100×100 。

2.1 CNN 模型训练样本

使用传感器通道 7 和 21 的监测数据构建 CNN

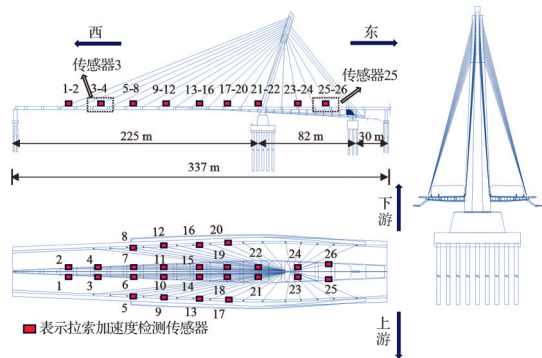


图 3 斜拉桥拉索加速度监测系统

Fig. 3 Cable acceleration monitoring system of a cable-stayed bridge

模型的数据样本,表 2 给出了桥梁拉索监测数据中 6 种典型数据类型(正常、频域混乱、漂移、倾斜、方波和平波)、各数据类型时域数据序列、相应数据类型的 PSDF 数据序列和转换后的 GAF 图像。如表 2 所示,频域混乱数据的时域波形正常,根据时域波形的特征可被视为正常数据,但由于频域特征混乱,这类数据无法识别桥梁拉索的模态频率。频域混乱数据和明显异常数据共同构成了桥梁拉索的低质量数据,其中,明显异常数据包括漂移、倾斜、方波和平波 4 种类型。

2.2 CNN 模型训练和测试

在分类任务的统计分析中,精确率、召回率和 F_1 分数是衡量分类结果准确性的常用指标^[20],其计算公式为

$$P = \frac{t_p}{t_p + f_p} \quad (4)$$

$$R = \frac{t_p}{t_p + f_n} \quad (5)$$

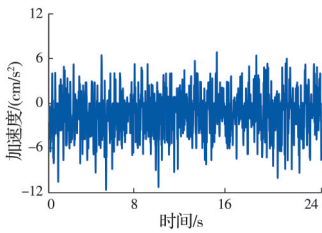
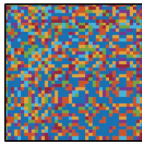
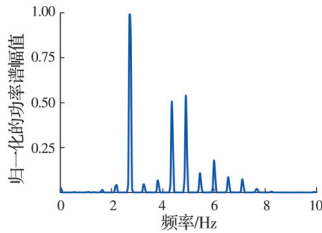
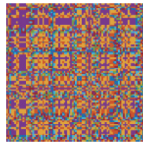
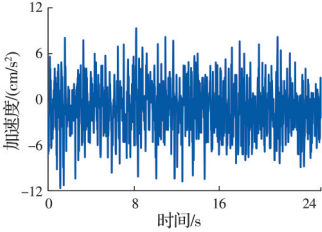
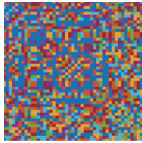
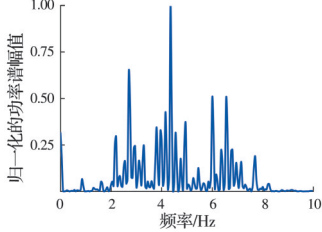
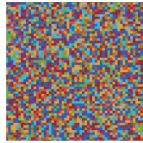
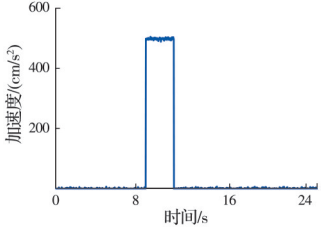
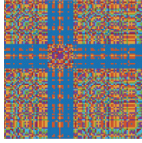
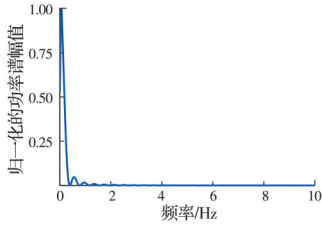
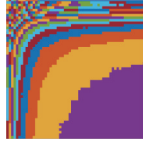
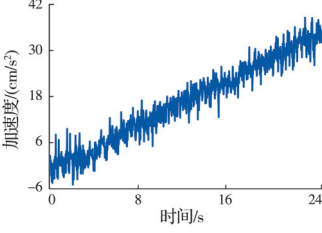
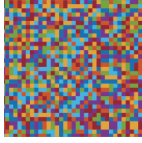
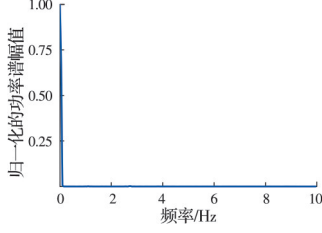

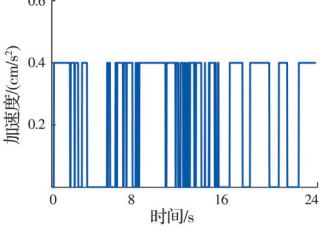

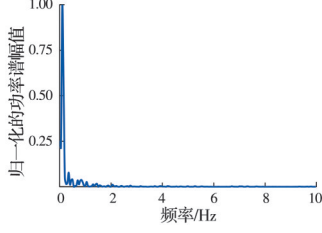
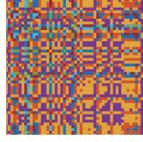
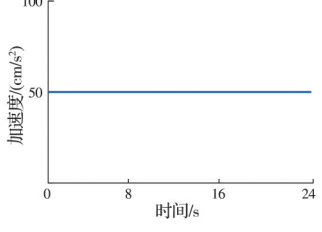

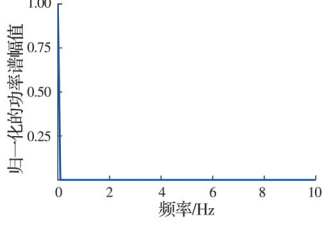

$$F_1 = \frac{2}{1/P + 1/R} \times 100\% \quad (6)$$

式中: P 为精确率; R 为召回率; t_p 为真阳性,表示每种类型数据正确分类的样本数量; f_p 为假阳性,表示模型将其他类型图像分类为本类型图像的样本数量; f_n 为假阴性,表示从本类图像分类到其他类别图像的样本数量; F_1 分数为精确率与召回率的调和均值。精确率针对预测结果,表示预测为某一类别的图像样本中有多少真正属于这一类别的图像样本。召回率针对真实样本,表示有多少真实图像样本被正确预测。在对分类结果进行评价时,利用 F_1 分数对分类效果进行综合评价。

选择 2020 年 5 月 1 日至 2 日传感器通道 7 和 21 的加速度监测数据进行 CNN 模型训练,将数据转化为 GAF 图像并对图像进行标记。研究中通过不断地增加训练样本数量使 CNN 模型达到较高且稳定的分类准确率。在所有类型的数据中,共提取了

表 2 时域数据序列和 PSDF 数据序列转换为 GAF 图像

Table 2 Converting time-domain data sequences and PSDF data sequences into GAF images

类型	时域数据			PSDF 序列		
	时域数据序列	描述	GAF 图像	PSDF 数据序列	描述	GAF 图像
正常		时域波形正常, 数值在正常范围内			拉索的模态频率可以在频域中清晰地识别	
频域混乱		时域波形正常, 数值在正常范围内, 频率特性无法确定			无法在频域中识别模态频率	
漂移		数据具有随机漂移的非平稳性			PSDF 序列波动较大, 然后变成平波	
倾斜		数据在时域中具有明显的趋势			PSDF 序列从较高的初始值迅速下降, 随后出现一些小波动	
方波		数据波形呈方形			PSDF 序列起初波动较大, 然后波动减小	
平波		数据为常值或空值			PSDF 序列从较高的初始值迅速减小, 然后变成平波	

7 200 个样本图像, 其中 4 800 个用于模型训练, 2 400 个用于模型测试。

对于 CNN 模型的分类任务, 平衡的训练样本集对于 CNN 模型学习不同数据类型的特征非常重要^[7]。为了避免训练样本不平衡的问题, 将每种质量数据类型的样本数量限制在相同比例。为了验

证本方法的有效性, 将该方法与时域数据序列直接转换为 GAF 图像的方法进行比较。

图 4 给出了 CNN 模型对于两种图像转化方法的测试样本分类结果混淆矩阵。从图 4(a) 可以看出, CNN 模型对于 PSDF 数据序列转化的 GAF 图像具有较高的识别精度, 平波、倾斜和漂移的召回

率达到100%,方波的精确率达到100%,这表明明显异常的低质量数据转换到频域后仍保持了明显的异常特征;对于正常和频域混乱数据,虽然分类精确率低于其他类型,但精确率和召回率均达到96%以上,表明CNN对频域混乱数据也具有较好的识别和分类效果。图4(b)给出了将时域数据序列直接转换为GAF图像的分类结果,从图中可以看出,CNN模型在正常数据和频域混乱数据之间存在严重混淆,无法准确评价这两种数据类型。同时,因为明显异常数据在时域序列中具有明确的数据类型特征,CNN模型可以较好地对其进行评价,明显异常数据分类结果的精确率和召回率均达到92%以上。

真实类	正常	388 10.22%	12 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	97.0% 3.0%
	频域混乱	11 0.5%	389 10.22%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	97.3% 2.7%
	缺失	0 0.0%	0 0.0%	400 10.77%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	方波	0 0.0%	1 0.0%	3 0.1%	385 10.10%	2 0.1%	9 0.4%	96.3% 3.7%
	漂移	0 0.0%	0 0.0%	0 0.0%	0 0.0%	400 10.77%	0 0.0%	100% 0.0%
	倾斜	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	400 10.77%	100% 0.0%
		97.2% 2.8%	96.8% 3.2%	99.3% 0.7%	100% 0.0%	99.5% 0.5%	97.8% 2.2%	98.4% 1.6%
		正常	频域混乱	缺失	方波	漂移	倾斜	
		预测类						

(a) PSDF序列转化的GAF图像

真实类	正常	273 14.3%	119 5.0%	0 0.0%	5 0.2%	0 0.0%	3 0.1%	68.3% 31.7%
	频域混乱	102 4.3%	280 14.7%	0 0.0%	18 0.8%	0 0.0%	0 0.0%	70.0% 30.0%
	缺失	0 0.0%	0 0.0%	400 10.77%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	方波	0 0.0%	14 0.6%	3 0.1%	370 15.4%	5 0.2%	8 0.3%	92.5% 7.5%
	漂移	0 0.0%	0 0.0%	0 0.0%	0 0.0%	400 10.77%	0 0.0%	100% 0.0%
	倾斜	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	399 10.6%	99.8% 0.2%
		72.8% 27.2%	56.8% 43.2%	99.3% 0.7%	93.9% 6.1%	98.7% 1.3%	97.3% 2.7%	88.4% 11.6%
		正常	频域混乱	缺失	方波	漂移	倾斜	
		预测类						

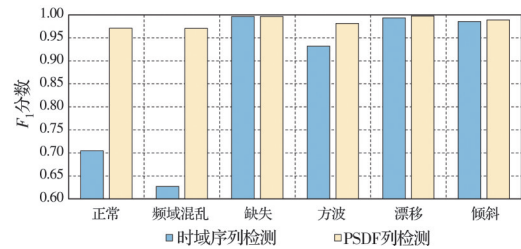
(b) 时域序列转化的GAF图像

图4 分类结果混淆矩阵

Fig. 4 Confusion matrices of classification results

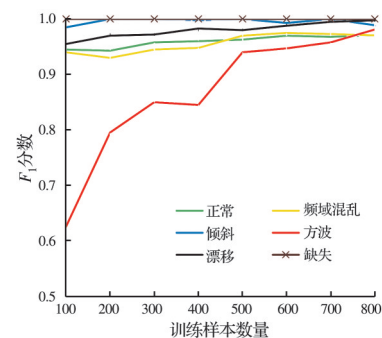
图5给出了两种方法对不同类型数据分类结果的 F_1 分数对比。在频域混乱数据的评价中,时域数据序列检测方法对正常和频域混乱数据分类结果的 F_1 分数分别为70%和62%,PSDF序列检测方法对这两类数据分类结果的 F_1 分数达到97%。在对明显异常数据的评价中,PSDF数据序列检测方法也具有相同或更高的 F_1 分数。以上结果表明,CNN模型无法通过时域数据序列准确评价频域混乱数据,而PSDF序列检测方法对各种数据的分类精度

都优于时域数据序列检测方法。此外,时域数据序列检测方法对于方波数据的分类精确率和召回率明显低于其他明显异常数据,而PSDF序列检测方法在很大程度上弥补了这一点。

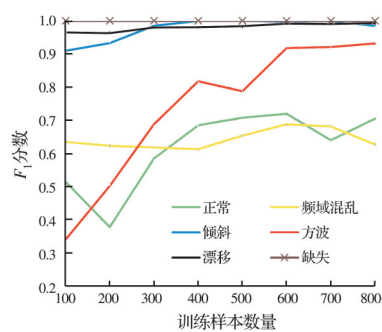
图5 两种分类方法 F_1 分数对比Fig. 5 Comparison of F_1 scores of the two classification methods

2.3 训练样本数量的影响

为了进一步验证该方法在少量训练样本下的分类精度,分别利用时域序列和PSDF序列的GAF图像对两个CNN模型进行分类精度测试,研究训练样本数量对分类精度的影响。随机选取每种数据类型的图像400张,共2400张图像来构建测试样本集,并在每种数据类型训练样本100~800张的情况下测试两个CNN模型对2400幅图像的分类精度,分类结果的 F_1 分数如图6所示。从图6中可以看出,在相同的训练样本数下,CNN模型对于时域序列转化的GAF图像分类精度明显低于PSDF序



(a) PSDF序列转化的GAF图像



(b) 时域序列转化的GAF图像

图6 不同训练样本数量下的 F_1 分数Fig. 6 The F_1 scores under different numbers of training samples

列,当每种数据训练样本图像为 600 张时,4 种明显异常数据类型的 F_1 分数都达到 91% 以上;时域数据序列检测方法在任何训练样本数下都无法有效地对正常数据和频域混乱数据进行分类,而 PSDF 数据序列检测方法解决了这一问题,可以对频域混乱数据进行有效分类;在明显异常数据评价中,PSDF 数据序列检测方法在样本较少的情况下具有更高分类精度。

2.4 大规模测试

构建包含所有传感器通道数据的训练样本可以使 CNN 模型具有良好的泛化能力和鲁棒性,但需耗费大量人力和时间成本,并且会面临训练样本不平衡问题。如果网络模型可以用少量传感器通道的数据样本进行训练,并且在所有通道中都能获得较好的分类精度,那么上述问题就可以得到解决。因此,仍然使用 2.2 节中训练的 CNN 模型来测试所有传感器通道的加速度数据。为了滤除之前模型训练和测试中的数据样本,选择 2020 年 7 月 21 日拉索加速度传感器 12 h 的监测数据,以测试所有 26 个加速度传感器通道产生的监测数据,包括 5 616 000 个数据样本点和 46 800 个转换后的 GAF 样本图像。

利用 7 号和 21 号两个传感器通道的数据构建训练样本集,对所有传感器通道的加速度监测数据进行评价。26 个通道大致分为 5 组:通道 1~4、5~6、7~18、19~21 和 22~26,图 7 给出了每个加速度传感器通道各种数据类型的数量和比例。从图 7 中可以看出,相邻通道之间的数据质量分布相互影响,集群模式出现时间较长,可能是因为传感器受到外部环境的干扰,相邻传感器会产生类似的异常反应;单个传感器的低质量数据有时会出现,这可能是由传感器自身故障或小的外部干扰造成的;1、2、4 组有大量明显异常数据,特别是第 2 组,出现了 10 h 以上的明显异常数据,可能主要受到方波数据的影响。第 1 组和第 4 组的明显异常数据也达到了 1.5 h 左右,可能主要受到平波和倾斜数据的影响。表明 3 组传感器长期受到干扰,产生了较多的明显

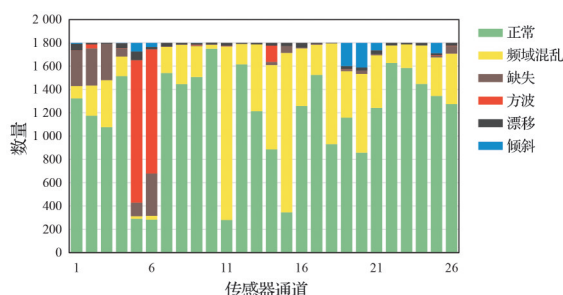


图 7 各通道数据质量统计

Fig. 7 Data quality statistics of each channel

异常数据;漂移数据的产生通常是偶然的,分散在每个通道中;频域混乱数据存在于所有通道中,尤其是通道 11 和 15 中,出现时间超过 9 h,这些频域混乱数据不能用来获取拉索的动态特性,缺乏使用价值。

为了验证测试结果中各种数据类型的分类准确性,图 8 给出了大规模测试的分类结果混淆矩阵。从图 8 中可以看出,所有 6 种数据类型都得到了很好的识别,精确率为 84.8%~98.6%,召回率为 92.4%~97.3%,整个分类结果的准确率为 96.6%。表明当使用少量通道数据作为训练样本时,该方法可以评价所有通道的数据质量;在明显异常数据中,平波、方波和倾斜的分类精度最高,精确率和召回率均达到 95% 以上;在大规模测试中,漂移类数据的精确率和召回率显著降低,这可能是因为不同通道中的漂移数据特征有较大差异,并且模型对不同通道中漂移数据的泛化能力相对较低;CNN 模型对频域混乱数据的精确率和召回率均在 92% 以上,表明该方法能有效评价频域混乱数据。结果表明,CNN 模型对大规模测试的分类保持了较高准确性。Tang 等^[7]发现,不平衡的训练样本会严重影响 CNN 模型分类的准确性,尤其是漂移数据,大规模测试下漂移数据的分类精度降低到 57%,大量倾斜数据被错误地归类为漂移,而本文提出的通过少量通道数据构造平衡训练样本的方法可以避免上述问题,提高 CNN 模型分类精度。

真实类	正常	29583 93.2%	876 1.9%	0 0.0%	3 0.0%	16 0.0%	0 0.0%	97.1% 2.9%
	频域混乱	403 0.9%	10313 22.0%	0 0.0%	20 0.0%	40 0.1%	0 0.0%	95.7% 4.3%
	缺失	1 0.0%	2 0.0%	1679 3.6%	25 0.1%	19 0.0%	0 0.0%	97.3% 2.7%
	方波	0 0.0%	9 0.0%	57 0.1%	2373 5.1%	23 0.0%	28 0.1%	95.3% 4.7%
	漂移	5 0.0%	10 0.0%	14 0.0%	14 0.0%	596 1.3%	6 0.0%	92.4% 7.6%
	倾斜	0 0.0%	0 0.0%	13 0.0%	7 0.0%	9 0.0%	657 1.4%	95.8% 4.2%
		98.6% 1.4%	92.0% 8.0%	95.2% 4.8%	97.2% 2.8%	84.8% 15.2%	95.1% 4.9%	96.6% 3.4%
		预测类						

图 8 大规模测试的分类结果混淆矩阵

Fig. 8 Confusion matrix of large-scale test classification results

为了测试 CNN 模型对不同通道的数据质量评价准确性,验证模型的泛化能力,提取了所有通道分类结果的准确率。图 9 给出了每个通道 CNN 模型分类准确率的分布,从图中 9 可以看出,通道 7 和通道 21 的模型分类准确率最高,分别达到 98.2% 和 97.3%。这是因为使用这两个通道的传感器数据进行了 CNN 模型训练,该模型对这两个通道的

数据特征具有较高的分类准确率;与通道 7 和通道 21 相邻的传感器分类准确率较高,但准确率随空间距离的增加而降低,因此,相邻传感器通道通常具有相似的数据质量特征,训练后的 CNN 模型对相邻传感器具有较高的分类精度;利用通道 7 和通道 21 的数据作为训练集的 CNN 模型,对所有通道的分类准确率均达到了 94% 以上。结果表明,训练后的 CNN 模型具有较强的泛化能力和鲁棒性,进一步验证了该方法的数据质量评价准确性。

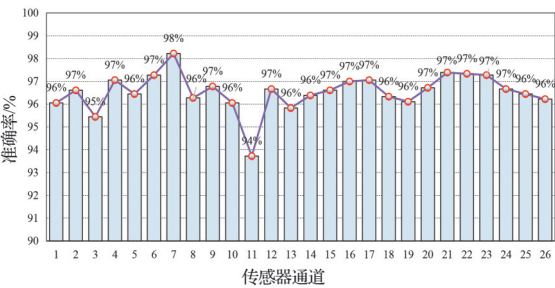


图 9 各通道分类准确率

Fig. 9 Classification accuracy of each channel

3 跨桥梁对象应用

由于各桥梁拉索振动监测数据的质量特征存在差异,同时,对不同桥梁对象分别建立数据质量评价模型需要大量人力和时间成本,当前对于桥梁拉索异常数据检测的研究较少涉及将已建立桥梁拉索模型应用到其他类似桥梁拉索数据检测中。

利用某斜拉桥两根拉索的加速度监测数据对本文所提出方法进行跨桥梁对象的应用,测试跨对象应用的效果。如图 10 所示,该斜拉桥是一座双塔钢箱梁斜拉桥,总跨度 1 150 m,桥塔高度 163 m,拉索的加速度监测频率为 50 Hz。提取该斜拉桥从 2020 年 7 月 1 日至 7 日两根拉索传感器 AS-1 和 AS-2 的监测数据,根据本文所提出方法,利用 2.2 节中建立的 CNN 模型对该桥加速度监测数据质量进行评价。

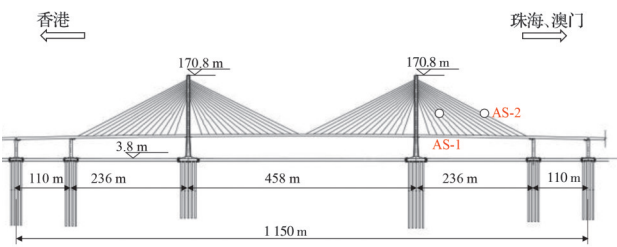
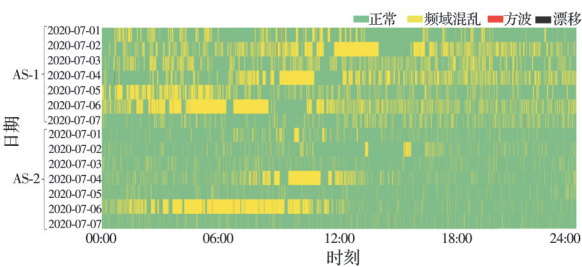


图 10 斜拉桥

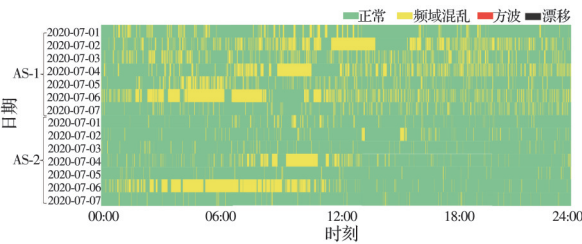
Fig. 10 The cable-stayed bridge

图 11(a) 给出了该斜拉桥加速度监测数据质量评价结果,图 11(b) 给出了该斜拉桥的实际数据质量分布,可以看出, CNN 模型的分类结果与实际数

据质量分布一致性较好。



(a) CNN 模型评价数据质量分布



(b) 实际数据质量分布

图 11 数据质量时空分布

Fig. 11 Temporal and spatial distribution of data quality

为了进一步定量分析 CNN 模型的跨桥梁对象应用效果,图 12 给出了 CNN 模型的分类结果混淆矩阵,从图 12 中可以看出,该桥梁两个拉索加速度监测数据中不存在明显异常的监测数据;频域混乱数据中分别有 21 张图片被错误分类为方波,3 张被错误分类为漂移;正常与频域混乱两种类型数据的精确率和召回率均达到 84% 以上;相比原桥本身的准确率, CNN 模型分类结果虽然出现了下降,但依然达到 95% 以上。因此,该方法可被应用于跨桥梁对象,具有良好的泛化能力和鲁棒性。

真实类	正常	39658 78.7%	1549 3.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	96.2% 3.8%
	频域混乱	678 1.3%	8491 16.8%	0 0.0%	21 0.0%	3 0.0%	0 0.0%	92.4% 7.6%
	缺失	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0.0% 0.0%
	方波	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0.0% 0.0%
	漂移	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0.0% 0.0%
	倾斜	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0.0% 0.0%
		98.3% 1.7%	84.6% 15.4%	0.0% 0.0%	0.0% 0.0%	0.0% 0.0%	0.0% 0.0%	95.5% 4.5%
		正常	频域混乱	缺失	方波	漂移	倾斜	预测类

图 12 跨桥梁对象的分类结果混淆矩阵

Fig. 12 Confusion matrix for the classification results of the cross-bridge object

4 结论

未经数据质量评价就直接进行结构状态评估时,在结构健康监测系统中的海量监测数据中,大

量低质量数据会对评估结果的准确性产生较大影响,甚至导致工程结构事故的漏报和错报。因此,通过对监测数据评价和分类后,识别和剔除低质量数据,提取准确有效的数据对结构状态进行评估。提出了一种基于CNN和频域特征的数据质量评价方法,利用CNN在机器视觉和分类任务中的优势,训练了一个CNN模型,该模型可用于自动化评价桥梁拉索动力响应监测数据质量。所提出方法具有以下优点:

1)提出的桥梁拉索健康监测系统中的低质量数据包括明显异常和频域混乱数据,但现有时域数据序列检测方法不能有效评价频域混乱数据。该方法通过FFT将时域数据序列转换为PSDF序列,并通过格拉姆角场法将PSDF序列转换为GAF图像,增强了各种数据的质量特征,解决了时域序列检测方法无法评价频域混乱数据的问题。

2)比较了不同训练样本数量时CNN模型对PSDF序列图像和时域序列图像分类准确性的影响。在相同数量的训练样本下,时域数据序列检测方法的分类准确性明显低于PSDF序列检测方法,且在训练样本数量较少的情况下,PSDF序列检测方法依然具有较好的分类准确性。

3)使用少量通道的监测数据训练CNN模型,并对所有通道数据质量进行评价,每个通道的低质量数据分布不均衡,呈现集群模式,相邻传感器通道的数据特征相似。大规模数据质量评价表明,分类精度随距离的增加而降低,但该方法仍保持了较高的分类准确率。

4)利用某斜拉桥的拉索加速度监测数据对所提出方法进行跨桥梁对象应用,准确率达到95%以上。该方法不仅适用于大跨桥梁拉索动力响应监测数据质量的自动化评价,也适用于结构健康监测的其他振动信号的质量评估,可根据所需评估的桥梁振动信号监测数据来训练数据,评价CNN模型,并对其振动信号的数据类型进行检测和分类。该方法可以准确区分正常数据、明显异常数据和频域混乱数据,从而将低质量数据从监测系统数据库中剔除,同时,在跨对象应用方面,具有较好的泛化能力,减少了大量训练样本的分类和标记需要的人力和时间成本。

参考文献

[1] WU Q, TAKAHASHI K, NAKAMURA S. Non-linear response of cables subjected to periodic support excitation considering cable loosening [J]. *Journal of Sound and Vibration*, 2004, 271(1/2): 453-463.

[2] LILIE J L, COSTA A P D A. Vibration amplitudes caused by parametric excitation of cable stayed structures [J]. *Journal of Sound and Vibration*, 1994, 174(1): 69-90.

[3] NIE Z H, LIN J, LI J, et al. Bridge condition monitoring under moving loads using two sensor measurements [J]. *Structural Health Monitoring*, 2020, 19(3): 917-937.

[4] 岳加利, 郝静, 卢海林, 等. 桥梁结构监测数据异常值处理方法[J]. *武汉工程大学学报*, 2022, 44(1): 107-111.

YUE J L, HAO J, LU H L, et al. Outliers processing method of bridge structure monitoring data [J]. *Journal of Wuhan Institute of Technology*, 2022, 44(1): 107-111. (in Chinese)

[5] TAN X Y, SUN X X, CHEN W Z, et al. Investigation on the data augmentation using machine learning algorithms in structural health monitoring information [J]. *Structural Health Monitoring*, 2021, 20(4): 2054-2068.

[6] BAO Y Q, TANG Z Y, LI H, et al. Computer vision and deep learning-Based data anomaly detection method for structural health monitoring [J]. *Structural Health Monitoring*, 2019, 18(2): 401-421.

[7] TANG Z Y, CHEN Z C, BAO Y Q, et al. Convolutional neural network-Based data anomaly detection method using multiple information for structural health monitoring [J]. *Structural Control and Health Monitoring*, 2019, 26(1): e2296.

[8] 范时泉, 张金辉, 张其林. 结构健康监测系统的异常数据识别[J]. *计算机辅助工程*, 2016, 25(5): 60-65.

FAN S X, ZHANG J H, ZHANG Q L. Abnormal data recognition in structural health monitoring system [J]. *Computer Aided Engineering*, 2016, 25(5): 60-65. (in Chinese)

[9] 吴迪. 基于深度学习的结构健康监测异常数据识别方法研究[J]. *智能城市*, 2020, 6(20): 10-13.

WU D. Research on abnormal data identification method of structural health monitoring system based on deep learning [J]. *Intelligent City*, 2020, 6(20): 10-13. (in Chinese)

[10] ZHANG Y X, LEI Y. Data anomaly detection of bridge structures using convolutional neural network based on structural vibration signals [J]. *Symmetry*, 2021, 13(7): 1186.

[11] YANG X W, LATECKI L J, POKRAJAC D. Outlier detection with globally optimal exemplar-based GMM [C]//*Proceedings of the 2009 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2009: 145-154.

[12] TANG X M, YUAN R X, CHEN J. Outlier detection in energy disaggregation using subspace learning and Gaussian mixture model [J]. *International Journal of Control and Automation*, 2015, 8(8): 161-170.

[13] CHANDOLA V, BANERJEE A, KUMAR V. Anom-

- aly detection: A survey [J]. ACM Computing Surveys, 2009, 41(3): 15: 1-15: 58.
- [14] 沙爱民, 童峥, 高杰. 基于卷积神经网络的路表病害识别与测量[J]. 中国公路学报, 2018, 31(1): 1-10.
SHA A M, TONG Z, GAO J. Recognition and measurement of pavement disasters based on convolutional neural networks [J]. China Journal of Highway and Transport, 2018, 31(1): 1-10. (in Chinese)
- [15] 高照明. 结构健康监测数据异常诊断与修复方法研究[D]. 辽宁 大连: 大连理工大学, 2020.
GAO Z M. Research on abnormal diagnosis and repair method of structural health monitoring data [D]. Dalian, Liaoning: Dalian University of Technology, 2020. (in Chinese)
- [16] MAO J X, WANG H, SPENCER B F Jr. Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders [J]. Structural Health Monitoring, 2021, 20(4): 1609-1626.
- [17] 骆勇鹏, 王林堃, 郭旭, 等. 利用单传感器数据基于GAF-CNN的结构损伤识别[J]. 振动测试与诊断, 2022, 42(1): 169-176, 202.
LUO Y P, WANG L K, GUO X, et al. Structural damage identification using single sensor data based on GAF-CNN [J]. Journal of Vibration, Measurement & Diagnosis, 2022, 42(1): 169-176, 202. (in Chinese)
- [18] 全钰, 庞新宇, 魏子涵. 基于GADF-CNN的滚动轴承故障诊断方法[J]. 振动与冲击, 2021, 40(5): 247-253, 260.
TONG Y, PANG X Y, WEI Z H. Fault diagnosis method of rolling bearing based on GADF-CNN [J]. Journal of Vibration and Shock, 2021, 40(5): 247-253, 260. (in Chinese)
- [19] OH B K, GLISIC B, KIM Y, et al. Convolutional neural network-based wind-induced response estimation model for tall buildings [J]. Computer-Aided Civil and Infrastructure Engineering, 2019, 34(10): 843-858.
- [20] 余加勇, 李锋, 薛现凯, 等. 基于无人机及Mask R-CNN的桥梁结构裂缝智能识别[J]. 中国公路学报, 2021, 34(12): 80-90.
YU J Y, LI F, XUE X K, et al. Intelligent identification of bridge structural cracks based on unmanned aerial vehicle and mask R-CNN [J]. China Journal of Highway and Transport, 2021, 34(12): 80-90. (in Chinese)

(编辑 王秀玲)