

基于格拉姆角场与并行滑动卷积注意力网络的轴承故障诊断方法

谢胜灵¹ 荆晓远^{2*} 杜杰宾¹ 姚永芳² 易 鹏² 王平雁²
XIE Shengling JING Xiaoyuan DU Jiebin YAO Yongfang YI Peng WANG Pingyan

摘 要

针对当前轴承故障诊断领域中存在故障特征提取困难的问题,提出了一种基于格拉姆角场(gramian angular field, GAF)与并行滑动卷积注意力网络的轴承故障诊断方法。首先,利用 GAF 技术将一维时间序列数据转换为二维图像表示,生成格拉姆和场与差场。这些图像标识不仅保留了时间序列中的时间依赖性和变化信息,而且便于后续的图像处理。然后,采用并行滑动卷积方法对转化的格拉姆和场与差场进行处理,通过卷积核的滑动和运算提取其所包含的关键故障特征信息。同时,为了进一步提升模型诊断性能,在模型中引入了卷积块注意力模块(convolutional block attention module, CBAM),其能够自适应地调整输入特征图的通道和空间权重,使模型能够聚焦于更为重要的故障特征中的关键信息,忽略无关的背景噪声,进而提升模型的故障诊断准确率。实验结果表明,与其他方法相比,所提方法在轴承故障诊断任务中具有较高的分类识别准确率和鲁棒性。

关键词

格拉姆角场; 并行滑动卷积; 注意力机制; 故障诊断; 卷积神经网络

doi: 10.3969/j.issn.1672-9528.2025.05.043

0 引言

轴承作为现代机械设备中的关键传动元件,其运行状态直接影响设备的性能与寿命。随着工业技术向高精度、高效率、高自动化的方向发展,对轴承的性能和可靠性要求日益严格^[1-2]。然而,轴承的工作环境复杂多变,加之多种潜在因素,如高温、重载、高速运转、材料疲劳、润滑不良及安装不当等,导致轴承故障频发。这些故障严重时可能引发设备停机,给企业带来巨大的经济损失和安全隐患。因此,对轴承故障进行及时、准确地诊断至关重要。

随着人工智能等高新技术的不断发展,数据驱动的方法逐渐成为故障诊断的主流方法^[3-4]。王贡献等人^[5]提出了一种基于多分支卷积神经网络(MBCNN)的轴承变工况故障诊断方法,通过多尺度均值化重构振动信号并利用 MBCNN 的多个并行分支提取故障特征,结合残差学习和全局平均池化层改善模型训练性能,并采用基于微调的迁移学习策略以适应变工况。郝洪涛等人^[6]提出了一种多尺度特征自适应融合网络用于气动控制阀的故障诊断,通过自注意力机制和权重自适应特征融合网络提高了故障特征

的表征能力。郭宝良等人^[7]针对双轴惯性直线振动机械中滚动轴承的单一局部损伤故障诊断问题,建立了基于赫兹接触理论的内、外环局部损伤故障的冲击振动模型,并通过仿真和试验验证了模型的有效性。Li 等人^[8]结合超像素分割方法与多尺度卷积注意力网络实现了轴承的故障诊断有效地提升了故障诊断的精度。Hu 等人^[9]采用自注意力机制的故障诊断模型,结合信号处理方法,同时实现了故障类型和故障程度的识别。

上述方法虽在轴承故障诊断中已取得一定成果,但模型在特征提取过程中仍存在一些问題,如传统的 CNN 结构进行池化操作导致的特征丢失和特征模糊。为了解决这些问題,本文提出了一种基于格拉姆角场(gramian angular field, GAF)与并行卷积的 CBAM(convolutional block attention module)注意力机制的故障诊断方法。首先利用格拉姆角场将一维振动信号转换为二维图像,通过计算时间序列的累积与差异生成格拉姆和场与差场,有效保留时序依赖性与变化特征。设计并行卷积结构同步处理两种场信息,结合滑动卷积替代传统池化层,通过卷积核局部采样实现特征细节保留。进一步引入 CBAM 注意力机制,在通道与空间维度自适应调整特征权重,强化关键故障信息。

1 相关理论

1.1 格拉姆角场

格拉姆角场的概念来源于线性代数中格拉姆矩阵的概

1. 吉林化工学院信息与控制工程学院 吉林吉林 132000
2. 广东石油化工学院计算机学院 广东茂名 525000
[基金项目] 基于类不平衡深度特征学习的石化动设备故障信号分类研究(62176069); 软件缺陷预测关键技术和新方法研究(2023A1515012653)

念,是一种将时间序列数据转换为图像的新方法^[10-11]。主要通过将时间序列映射到极坐标系中,并利用三角函数来生成 GASF 与 GADF 两类图像。通过式(1)将时间序列数据归一化到 $[-1, 1]$ 的区间,后通过式(2)将归一化后的数据映射到笛卡尔坐标系中得到每个时间下的极角 ϕ_i 和极径 r_i 。

$$\tilde{x} = \frac{2 \cdot x_i - \max(x) - \min(x)}{\max(x) - \min(x)} \quad (1)$$

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i) \\ r_i = \frac{i}{n} \end{cases} \quad (2)$$

式中: i 为时间戳; n 为正则化常数。

格拉姆角场定义了两种特有的带有惩罚项的内积形式,用以消除高斯噪声的影响,其定义为:

$$\langle x_i, x_j \rangle = \cos(\phi_i + \phi_j) \quad (3)$$

$$\langle x_i, x_j \rangle = \sin(\phi_i - \phi_j) \quad (4)$$

1.2 卷积神经网络

卷积神经网络(CNN)是一种强大的深度学习模型,特别适用于处理具有网格结构的数据,尤其是图像数据。其核心架构由卷积层、池化层和全连接层等关键组件构成,这些组件共同协作以实现对输入数据的特征提取、空间尺寸缩减和全局特征综合。卷积层通过滑动窗口操作和可学习的卷积核,捕捉图像中的局部空间特征,而池化层则通过减小特征图的空间尺寸来降低计算量和内存使用,同时增强模型对输入数据微小形变和位置变化的鲁棒性。全连接层则负责将提取的高层次特征进行全局综合和分类,完成最终的识别或预测任务。由于其强大的特征学习和端到端训练的能力,CNN在图像分类、目标检测、语义分割、图像生成、语音识别、

自然语言处理等多个领域都取得了显著的成功,成为现代机器学习和人工智能领域不可或缺的重要工具^[12-13]。CNN的基本结构如图1所示。

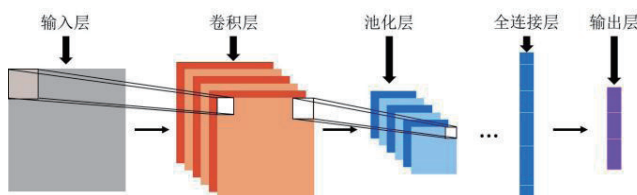


图1 卷积神经网络基本结构

2 故障诊断技术框架

在故障诊断过程中,首先利用 GAF 技术将一维的时间序列数据转化为二维的图像表示,生成格拉姆场与差场。这两种图像表示不仅保留了时间序列中的时间依赖性和变化信息,而且为后续的图像处理提供了便利。接下来,采用并行卷积的方式对格拉姆场与差场进行并行处理。通过多个卷积核在图像上的滑动和运算,能够同时提取出场与差场中的关键故障特征,实现特征的全面捕获。为了克服传统池化操作可能导致的特征丢失和模糊问题,采用滑动卷积的方式替代了池化操作。这种方法在保留图像局部特征的同时,还能通过逐步滑动卷积核进行特征聚合,从而避免了信息的损失和模糊。最后,在模型的后端加入 CBAM 注意力机制。CBAM 通过自适应地调整特征图的通道和空间权重,使模型能够聚焦于更重要的故障特征,而忽略无关的背景噪声。通过这种方式,能够进一步提高故障诊断的准确性和鲁棒性,实现更可靠的故障检测和识别。表1为所提出诊断模型的具体结构参数。图2为所提出诊断方法的整体框架图。

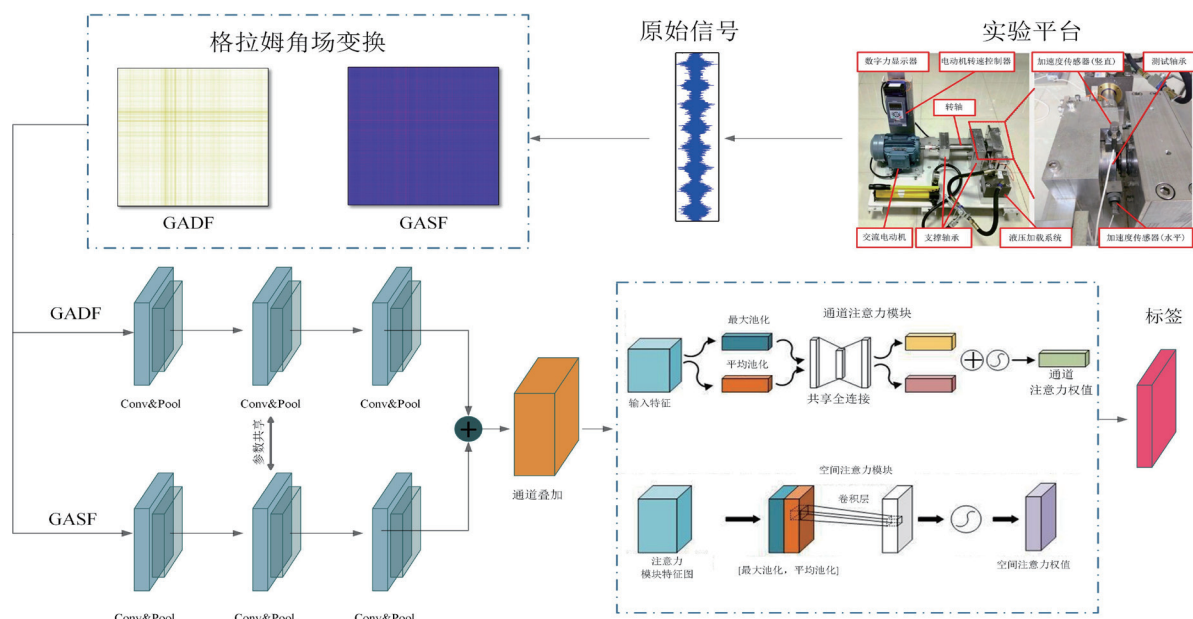


图2 诊断方法整体结构图

表 1 诊断模型的具体结构参数

层次	结构名称	结构参数	通道数	输出数据大小
	输入	(150,150)	1	150×150
采样 模块	卷积层	(3,3,1)	16	150×150
	正则化	—	16	150×150
	滑动卷积层	(3,3,2)	16	75×75
	ReLu	—	16	75×75
	卷积层	(3,3,1)	32	75×75
	正则化	—	32	75×75
	滑动卷积层	(3,3,2)	32	38×38
	ReLu	—	32	38×38
	卷积层	(3,3,1)	64	38×38
	正则化	—	64	38×38
空间 注意力 机制	卷积层	(7,7,1)	64	38×38
	Sigmoid 激活	—	64	38×38
通道注 意力机 制	自适应平均池化	—	64	1×1
	自适应最大池化	—	64	1×1
	卷积层	(1,1,1)	64	1×1
	Relu 激活	—	64	1×1
	卷积层	(1,1,1)	64	1×1

3 实验验证

3.1 实验数据集介绍

本次试验使用的数据集来源于西安交大轴承加速寿命试验（XJTU-SY）数据集^[14]。该实验采用 LDK UER204 滚动轴承作为测试轴承，提供的数据集包括 3 种工况，每种工况有 5 个故障轴承。XJTU-SY 测试平台由交流电机、电机调速控制器、转轴、支撑轴承、液压加载系统和测试轴承等部件组成。实验平台的具体结构如图 3 所示。

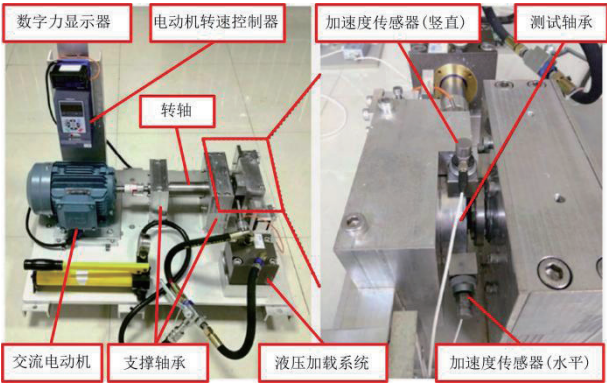


图 3 轴承加速寿命试验台

3.2 数据处理

本次试验数据选取工况为径向力 12kN 采样频率 35 kHz 下的实验数据。对数据进行以 2 500 个样本点为一组的数据切割，并将分割数据进行格拉姆角场变换生成对应的格拉姆图，同时，进行数据打乱以 8:2 的比例进行训练集与测试集的划分。具体的试验数据信息如表 2 所示，各故障模式所对应的 GASF 与 GADF 图如图 4 所示。

表 2 试验数据详细信息

故障类型	故障标签	训练样本数	测试样本数
外圈故障 1	A	160	40
外圈故障 2	B	160	40
外圈故障 3	C	160	40
保持架故障	D	160	40
外圈、内圈混合故障	E	160	40

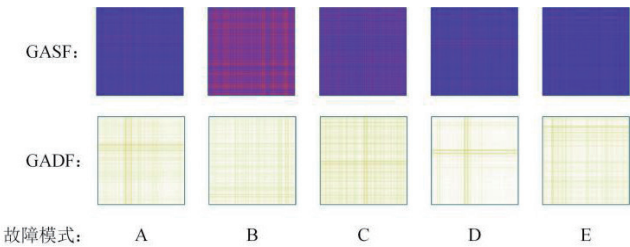


图 4 各故障模式下的格拉姆图

3.3 诊断效果对比分析

为验证所提出的网络模型相较于其他模型的诊断优势，本文选取 ResNet、MnasNet 及 Vgg16 作为对比分析。所有模型均采用 Python 语言进行编写，采用 PyTorch 框架进行搭建。采用交叉熵损失函数进行损失衡量，并采用 Adam 算法进行参数优化，学习率设置为 0.01，计算机配备了 NVIDIA 4060Ti GPU 和 32 GB RAM 硬件配置。训练过程中将数据批大小设置为 64，训练轮数设置为 100，学习率定义为 0.001。图 5 为各模型的准确率迭代曲线，图 6 为各模型测试 10 次得到的平均准确率的柱状图。

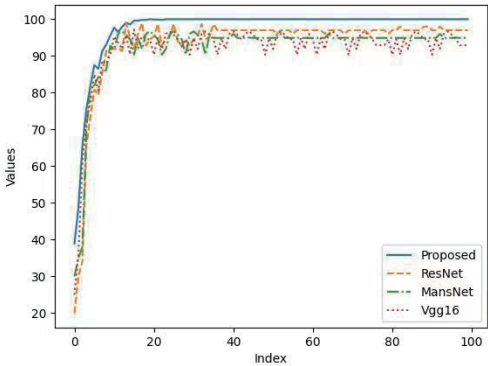


图 5 各模型准确率迭代曲线

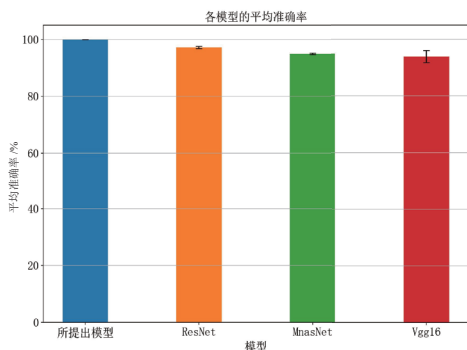


图6 各模型平均准确率

由图5、6可以得出，所提模型具有最快的收敛速度基本在迭代20次后即可达到收敛，并且可达到100%的诊断精度并且稳定性良好。而其他三种模型的收敛速度较慢，且在诊断精度及稳定性上均逊色于所提出模型，其中Vgg16的诊断精度只有93.5%，且稳定性最差。为更好地分析所提出模型的诊断性能，图7、8分别给出了模型诊断的三维混淆矩阵图以及经过T-SNE降维后的特征可视化图。由三维混淆矩阵图可以看出，所提出的故障诊断模型在各分故障模式下，均具有较高的分类识别准确率。通过三维特征可视化图的分类结果可知，各故障特征聚类紧凑，边界清晰，不同故障类型下特征距离较远，特征提取效果良好，进一步证明了本文所提方法的诊断优势。

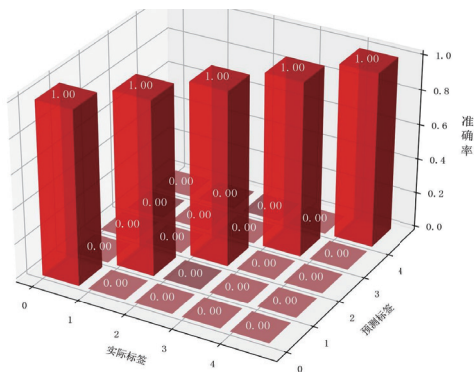


图7 混淆矩阵图

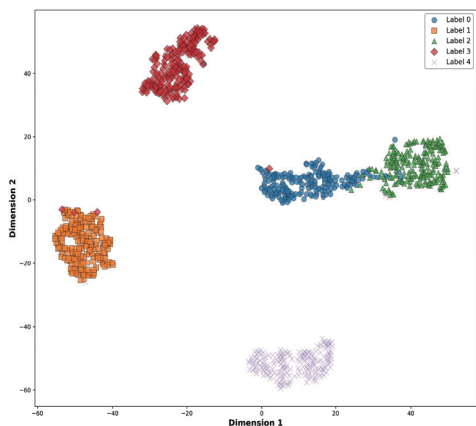


图8 特征可视化图

3.4 消融实验

为验证改进模型相较于原始模型的有效性，采用消融实验进行试验验证。分别将并行卷积改为单通道卷积，将CBAM注意力机制去掉进行故障诊断。具体诊断结果如表3所示。实验结果表明，若采用单通道进行诊断准确率会明显下降，并且稳定性也会变差，此外去掉CBAM模块模型的性能会进一步下降。因此，该结果进一步证明了所提出诊断方法的优越性。

表3 消融实验结果

模型	准确率/%	标准差
所提出模型	100	0.00
差场-单通道	95.3	1.42
和场-单通道	96.1	1.31
差场-单通道-去掉CBAM	94.8	1.05
和场-单通道-去掉CBAM	94.5	1.87
双通道-去掉CBAM	97.6	0.76

4 结论

本文将信号转换为格拉姆图，并构建并行卷积注意力网络进行故障诊断。在诊断模型中采用并行卷积的策略分别对GADF及GASF进行采样，以滑动卷积的方式代替池化操作并加入CBAM注意力机制，经过实验验证后得到以下结论：

（1）通过与ResNet、Vgg16以及最新的诊断模型进行对比分析后得出，所提出诊断模型收敛速度快，稳定性良好，诊断精度高，可达到100%的诊断准确率。

（2）采用三维混淆矩阵进行综合性诊断性能分析后可以得出，所提出诊断模型在矩阵对角线上的值均为1.00，对各故障模式分类效果良好。此外，采用三维T-SNE方法进行特征可视化后发现，模型对于各故障特征提取效果良好，聚类明显，边界清晰。

（3）采用消融实验对模型改进策略进行进一步验证后发现，在分别采用单通道采样及去掉CBAM注意力机制后，诊断性能均得到了不同程度的下降。可知，并行采样的方法以及CBAM注意力可以提升模型的诊断精度并增强稳定性。

参考文献：

- [1] 王玉静, 夏林, 康守强, 等. 基于多源域异构模型迁移的滚动轴承故障诊断方法[J]. 振动与冲击, 2023, 42(24): 257-266.
- [2] 杜浩飞, 张超, 李建军. 基于SENet-ResNext-LSTM的风机轴承故障诊断[J]. 机械强度, 2023, 45(6): 1271-1279.
- [3] 董会锦, 劳胜领, 修素朴, 等. 基于CFA参数优化DAE方法的滚动轴承故障诊断[J]. 机械设计与研究, 2023, 39(6): 109-112.
- [4] 郭绍强. 基于GOA-VMD和ISVM的滚动轴承故障诊断[J]. 化工设备与管道, 2023, 60(6): 72-79.

[5] 王贡献,付泽,胡志辉,等.基于多分支卷积神经网络的轴承变工况故障诊断[J].噪声与振动控制,2023,43(6):135-141.

[6] 郝洪涛,王凯,张炳建,等.多尺度特征自适应融合的气动控制阀故障诊断[J].仪器仪表学报,2023,44(10):167-178.

[7] 郭宝良,赵玉秀,史丽晨,等.双轴惯性直线振动机械滚动轴承单一局部损伤故障诊断[J].振动与冲击,2023,42(24):24-31.

[8] LI Y H, MEN Z H, BAI X N, et al. A bearing fault diagnosis method based on M-SSCNN and M-LR attention mechanism[J/OL]. Structural health monitoring, 2024[2024-11-10].https://www.researchgate.net/publication/380202307_A_bearing_fault_diagnosis_method_based_on_M-SSCNN_and_M-LR_attention_mechanism. DOI:10.1177/14759217241244477.

[9] HU C Q, LI Y H, CHEN Z, et al. Research on fault diagnosis of rolling bearing based on multi-sensor bi-layer information fusion under small samples[J]. Review of scientific instruments, 2023, 94(11): 115106.

[10] LI G, AO J X, HU J Y, et al. Dual-source gramian angular field method and its application on fault diagnosis of drilling pump fluid end[J]. Expert systems with applications, 2024, 237: 121521.

[11] MEMARIAN A, DAMARLA S K, MEMARIAN A, et al. Detection of poor controller tuning with gramian angular field (GAF) and stack auto encoder (SAE)[J]. Computers and chemical engineering, 2024, 185: 108652.

[12] GUO Y J, LEI X J, PAN Y. An encoding-decoding framework based on CNN for circRNA-RBP binding sites prediction[J]. Chinese journal of electronics, 2024, 33(1): 256-263.

[13] JORGES C, BERKENBRINK C, GOTTSCHALK H. Spatial ocean wave height prediction with CNN mixed-data deep neural networks using random field simulated bathymetry[J]. Ocean engineering, 2023, 271: 113699.

[14] 雷亚国,韩天宇,王彪,等. XJTU-SY 滚动轴承加速寿命试验数据集解读[J]. 机械工程学报, 2019, 55(16): 1-6.

【作者简介】

谢胜灵(1999—)男,湖南长沙人,硕士,研究方向:故障诊断、类不平衡, email: 3347313819@qq.com。

荆晓远(1971—),通信作者(email: jingxy_2000@126.com),男,江苏南京人,博士,教授、博士生导师,研究方向:模式识别、计算机视觉、故障诊断。

杜杰宾(2000—),男,广东深圳人,硕士研究生,研究方向:故障诊断、类不平衡。

姚永芳(1975—),女,湖南常德人,硕士,助理研究员,研究方向:模式识别、机器学习、人工智能与应用。

易鹏(1986—)男,湖北武汉人,博士,研究方向:计算机科学、人工智能研究。

王平雁(1990—)男,广东湛江人,博士,研究方向:软件安全、云数据安全、人工智能。

(收稿日期:2025-01-09 修回日期:2025-05-16)

(上接第 186 页)

[2] 张傲,李涌睿,张至涵,等.星载 SAR 实时成像系统矩阵转置方法综述[C]//中国高科技产业化研究会智能信息处理产业化分会.第十七届全国信号和智能信息处理与应用学术会议论文集.北京:北京理工大学信息与电子学院雷达技术研究所,2023:111-235.

[3] 谢应科,张涛,韩承德.实时 SAR 成像系统中矩阵转置的设计和实现[J].计算机研究与发展,2003(1):6-11.

[4] 卢世祥,韩松,王岩飞.合成孔径雷达实时成像转置存储器的两页式结构与实现[J].电子与信息学报,2005(8):1226-1228.

[5] 林桐,谢宜壮,刘伟.实时 SAR 成像系统矩阵原位转置的实现[J].计算机工程,2013,39(6):319.

[6] 边明明,毕福昆,汪精华.实时 SAR 成像系统矩阵转置方法研究与实现[J].计算机工程与应用,2011,47(22):117-119.

[7] 江帆,刘光平,周志敏.多计算机上分布式矩阵转置[J].微处理机,2002(2):34-37.

[8] 远远,陈亮.基于 SMP 的矩阵转置算法研究与实现[J].计算机工程与设计,2016,37(10):2690-2694.

[9] 任志国,侯永艳.基于伪地址存储结构的稀疏矩阵快速转置算法[J].仪器仪表与自动化装置,2019(5):125-128.

[10] 肖汉,李彩林,李琦,等.CPU+GPU 异构并行的矩阵转置算法研究[J].东北师大学报(自然科学版),2019,51(4):70-77.

[11] 裴向东,王庆林,廖林玉,等.多核数字信号处理器并行矩阵转置算法优化[J].国防科技大学学报,2023,45(1):57-66.

[12] 刘玉坤,单冬红.矩阵转置算法的研究[J].计算机时代,2011(9):30-33.

【作者简介】

王莉丽(1981—),女,安徽临泉人,硕士研究生,副教授,研究方向:智能信息系统。

(收稿日期:2024-06-26 修回日期:2025-04-30)