

基于一维密集连接卷积网络的故障诊断研究

赵志宏^{1,2}, 赵敬娇¹, 李晴², 李乐豪²

1. 石家庄铁道大学 信息与科学技术学院, 石家庄 050043;

2. 石家庄铁道大学 省部共建交通工程结构力学行为与系统安全国家重点实验室, 石家庄 050043

摘要: 近年来, 由于深度学习的不断发展, 基于数据驱动的轴承故障诊断被广泛研究, 卷积神经网络(CNN)等深度神经网络模型逐渐被应用到故障诊断中. 针对传统方法人工提取故障特征困难, 深层 CNN 网络模型训练效率低, 过拟合严重的问题, 提出了一种基于一维密集连接卷积网络的轴承故障诊断模型. 轴承的一维振动数据作为输入, 利用卷积网络自动提取故障特征信息, 采用密集连接机制实现高层次特征与低层次特征相结合, 从而更有效地提取故障特征信息. 通过批归一化等方法避免过拟合现象, 最后利用 SoftMax 层对故障进行分类, 从而实现智能故障诊断. 对凯斯西储大学轴承数据集及高铁轮对轴承数据集上的不同故障类型的实测数据进行实验, 实验结果表明, 该方法在数据集上的准确率均能达到 98% 以上, 能有效识别故障类型, 具有一定的工程应用价值.

关键词: 深度学习; 卷积网络; 密集连接; 故障诊断

中图分类号: TH133.33; TH165.3

文献标志码: A

文章编号: 1673-9868(2020)12-0025-09

在机械领域, 航空发动机、大型电机等重大装备正在朝着高速、高精、高效的方向发展. 为保障人身安全, 防止事故造成重大经济损失, 建立可靠的故障诊断系统十分必要^[1]. 轴承是旋转机械设备必不可少的也是极易出现损伤的部件, 对轴承的故障诊断具有重要的经济和社会意义.

随着机器学习的发展, 以及工业监测数据的积累, “大数据”时代下传统的故障诊断方法越来越不能满足当今的需求, 数据驱动的故障诊断方法越来越受国内外专家、学者们的重视^[2]. 深度学习理论作为模式识别和机器学习领域最新的研究成果, 在机器视觉、语音识别、情感分析、语言翻译、个性化推荐等领域取得了优异的成果. 卷积神经网络(convolutional neural network, CNN)是深度学习方法中的典型网络, 由于其强大的特征提取能力而被广泛应用于图像处理、视频分析、目标检测等场景.

近年来, CNN 及其发展的神经网络也越来越多地被应用于故障诊断. 文献[3]研究了一种基于 CNN 和振动信号峭度的故障诊断方法, 有效识别了滚动轴承故障. 文献[4]提出了一种自适应的一维卷积网络(ACNN-FD)算法, 实现了端到端的故障诊断. 文献[5]将原始振动信号分别输入 1-DCNN 和 LSTM 两个通道, 将空间和时间维度上特征信息融合, 从而对轴承进行了故障分类. 文献[6]提出了一种能处理一维频谱序列的改进的一维 DenseNet 网络结构. 以上方法虽然取得了较好的诊断效果, 但其中有些并没有很好地发挥 CNN 网络强大的特征提取能力, 有些转化为二维数据后会损失一些时序信息, 而有些网络结构较浅, 网络泛化能力以及对小数据集的诊断效果可能并不理想. 因此, 本文提出一种基于一维密集连接的卷积网络的故障诊断方法, 直接输入一维振动数据, 通过多层密集连接的卷积网络对故障进行特征提取, 实

现智能故障诊断.

1 基础理论

1.1 CNN

CNN 主要包括输入层、卷积层、池化层、输出层,有以下三大优点:局部感受野、权值共享以及池化^[7]. 一维 CNN 对于序列数据的特征提取非常重要,比较适用于端到端的轴承故障诊断.

卷积层的主要功能是对输入数据进行特征提取,是 CNN 的重要组成部分,其内部包含多个卷积核. 在前馈运算时,卷积运算将卷积核与输入数据相应位置的对应元素相乘,并按一定的步长扫描整个输入数据^[8]. 不同卷积核可以学习出不同特征,对应得到多个特征图. 一维卷积公式为

$$Z^l = [Z^{l-1} \otimes w^l] + b^l \quad (1)$$

池化层:卷积层的主要作用是特征提取,卷积得到的特征图会被传至池化层进行特征选择和过滤,使用 pooling 技术将卷积后得到的小邻域内的特征点整合得到新的特征,改变卷积层的特征图并降低特征图的维度^[9]. 池化处理一方面可以防止无用参数增加时间复杂度,另一方面也增加了特征的整合度.

1.2 DenseNet

CNN 的发展一般通过层数的加深或者加宽进行. 如 ResNet^[10] 网络,通过残差块堆叠加深网络,利用跳跃连接机制减弱网络退化;如 Inception^[11] 在加宽网络的同时一定程度上解决了梯度消失现象和过拟合现象,而 DenseNet^[12] 通过对特征的复用提出了一种新的结构,不但减缓了梯度消失的现象,参数量也更少. 3 个密集块的 DenseNet 网络结构如图 1 所示,它是采用跨通道的形式来连接的,其公式为

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (2)$$

其中 x_l 为第 l 层的输出,即通过特征在通道上的连接来实现特征复用,提升效率. 由于特征复用,最后的分类器使用了低级特征,较标准卷积网络只学习深层特征的诊断效果更好.

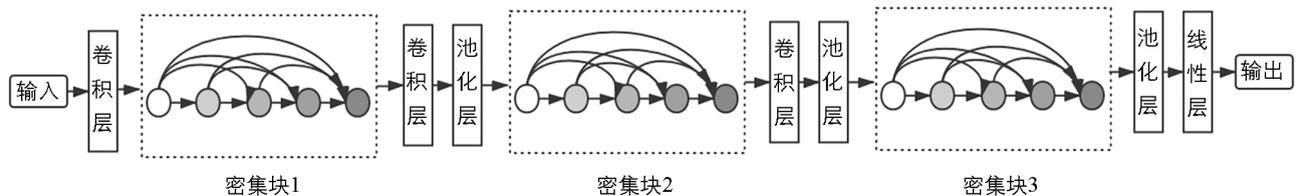


图 1 3 个密集块的 DenseNet 网络结构图

1.3 Dropout

深层网络模型的参数太多,而训练样本小的情况下训练出来的模型很容易产生过拟合的现象. 文献^[13]提出了 Dropout 方法,对解决此问题有较好的效果. 其原理为:为了防止过拟合,可以通过在训练过程中随机删掉网络中一半的隐藏神经元,阻止特征检测器的共同作用来提高神经网络的性能.

2 故障诊断模型

DenseNet 网络一般由多个 BN+ReLU+Conv 的三层结构组成的密集块和由 Conv+AvgPooling 组成的 Transition 层两部分组成. 但网络过深容易产生过拟合问题,不利于轴承的故障诊断,因此,本文利用不同卷积核大小的卷积层,以密集连接的方式构造密集块. 基于一维密集连接卷积网络的故障诊断模型如图 2 所示,轴承加速度数据经过预处理输入到模型中,经过深层网络不断地提取特征信息,最终经过分类层输出概率分布,根据概率分布输出故障类型.

本文网络结构各参数如表 1 所示. 一维数据进入 1 个卷积层提取浅层信息;再经过由 BN+Tanh+Conv 组成的密集块,其中一维卷积层是卷积核大小分别为 12×1 , 6×1 和 3×1 的卷积层,以提取不同尺度的特征,密集连接能够使每一层与前面的层建立连接,从而利用不同层次的特征,充分提取故障信息;后

经过一个 BN 层、一个一维卷积层和一个最大池化层作为过渡层; 扁平层能够将所提取的特征信息压平以输入到 SoftMax 分类层. 为减小网络过拟合的影响以及提高泛化能力, 本文网络加入了 Dropout 层. 已有研究表明 Dropout 的取值为 0.2 或 0.5 时能得到较好的结果^[14], 取 0.5 的时候 Dropout 随机生成的网络结构最多, 因此将抛弃率取值设为 0.5.

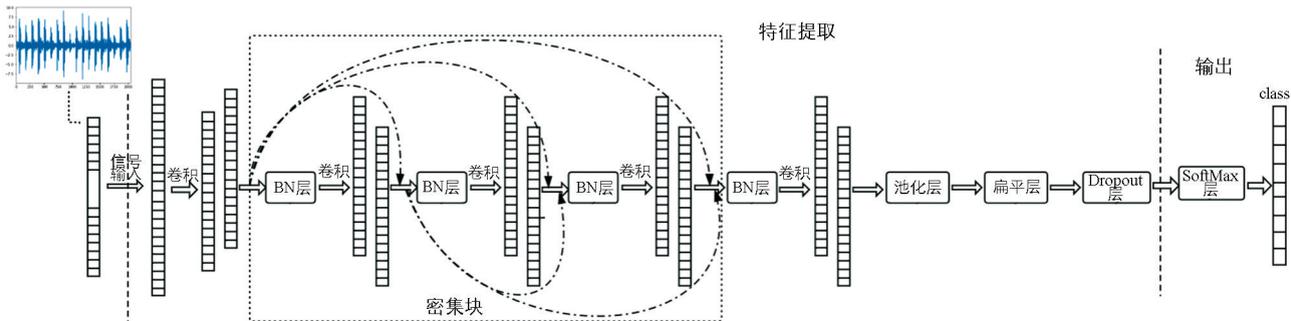


图 2 故障诊断模型

表 1 网络结构参数

模型参数	输入层	卷积层 C0	BN 层	卷积层 C1	BN 层	卷积层 C2	BN 层
特征图数目	1	64	64	64	64	64	64
特征图大小	2 048×1	2×3	2×3	12×1	12×1	6×1	6×1
模型参数	卷积层 C3	BN 层	卷积层 C4	池化层	扁平层	Dropout 层	输出层
特征图数目	64	64	32	32	1	1	1
特征图大小	3×1	3×1	2×1	4×4	5 440	5 440	10

3 实验结果与分析

3.1 实验数据

实验数据来自美国凯斯西储大学轴承数据^[15], 轴承型号为 SKF6205, 1 hp 负载, 12 kHz 采样频率下的驱动端加速度数据. 其中故障类型共 9 种, 包括故障程度分别为 0.177 8 mm、0.355 6 mm、0.533 4 mm 的滚动体故障、内圈故障和外圈故障, 以及 1 类无故障数据. 实验数据集划分如表 2 所示, 滚动轴承振动数据共 10 种类型, 训练集每类 1 750 个样本, 测试样集每类 500 个样本, 验证集每类 1 000 个样本.

表 2 实验数据集

信号类型	故障程度/mm	训练样本个数	测试样本个数	标签
无故障	无	1 750	500	Normal
滚动体故障	0.177 8	1 750	500	0.18_B
	0.355 6	1 750	500	0.36_B
	0.533 4	1 750	500	0.53_B
内圈故障	0.177 8	1 750	500	0.18_I
	0.355 6	1 750	500	0.36_I
	0.533 4	1 750	500	0.53_I
外圈故障	0.177 8	1 750	500	0.18_O
	0.355 6	1 750	500	0.36_O
	0.533 4	1 750	500	0.53_O

3.2 实验结果

本文训练和测试环境为 12 核, 24G 内存, Intel Xeon (Cascade Lake) Platinum 8269 处理器. 编程环境为 keras2.2, tensorflow1.10, python3.6. 输入样本长度取为 2 048, batchsize 为 128, 训练迭代次数为 150. 学习率初始值设为 0.000 5, 采用动态调整策略, 每 50 次衰减为原来的 1/5, 优化器为 Adam.

基于一维密集连接卷积网络故障诊断模型在数据集上的故障分类结果如图 3 所示, 图 3(a)为该模型在数据集的准确率曲线图, 可以看到模型在经过 2 到 3 次迭代训练后在验证集上的准确率就达到了 97% 左右, 在迭代 20 次左右便可稳定在 99% 以上, 而经过 150 次的迭代后本文模型在验证集上的准确率可达 100%. 图 3(b)为模型在测试集上的混淆矩阵, 实验结果表明各类均分类正确.

模型的学习能力和其复杂度有关, 模型复杂度越高, 其学习能力越强. 对神经网络来说, 提高复杂度一般有两条途径: 加深模型和加宽模型. 在一定程度上, 网络越深特征提取的效果越好. CNN 具有强大的特征提取能力, 本文的方法选用了多层一维卷积提取深层特征, 利用密集连接机制来提高模型的复杂度, 并将深层特征与浅层特征进行融合, 充分提取故障特征, 因此能取得较好的故障诊断结果.

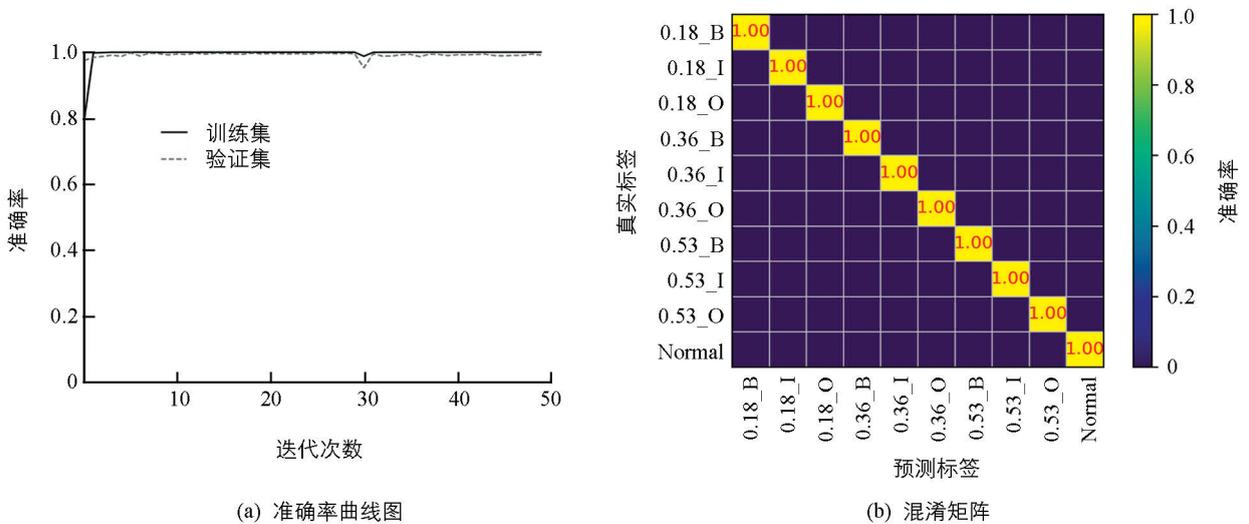


图 3 轴承故障分类结果

3.3 对比实验

机器学习中的评测指标有准确率(Accuracy)、召回率(Recall)、 F 值(F-Measure)等^[16]. 其中准确率为最常见和直观的评价指标; 召回率能反映模型预测值与真实值的差距, 体现算法的鲁棒性; F 值综合准确率和召回率, F 值较高时说明该方法较为有效. 为验证本文方法的有效性, 以准确率、召回率、 F 值为评价指标, 在凯斯西储大学轴承数据集上将本文方法与基于以下深度学习的故障诊断方法进行对比:

CNN: 该算法结构采用表 1 的输入层、卷积层 C0、BN 层、卷积层 C4、池化层、扁平层、Dropout 层、输出层. 其中卷积层 C0 由 64 个卷积核为 2×1 , 步长为 3 的神经元组成, 激活函数为 relu; 卷积层 C4 由 32 个卷积核为 2×1 , 步长为 3 的神经元组成, 激活函数为 relu; 池化层采用最大池化, 池化尺寸为 4; Dropout 层的参数为 0.5; 最后输出层采用 SoftMax 对故障进行分类.

LSTM: 该算法结构采用表 1 的输入层、LSTM 层、扁平层、Dropout 层、输出层. 其中输入层参数与表 1 相同, LSTM 层单元数设为 128, 激活函数为 tanh; Dropout 层的参数为 0.5; 最后输出层采用参数为 SoftMax 的全连接层对故障进行分类.

ResNet: 该算法结构采用图 2 的结构, 密集块中一维卷积层和 BN 层参数不变, 将密集连接结构换为 skip connection.

不同方法的诊断结果如表 3 所示, 由于 CNN 网络较浅, 训练时间以及诊断时间最短, 特征提取不理

想, 准确率仅为 82.76%。LSTM 诊断准确率在 98% 以上, 但其准确率很大程度上与 LSTM 的单元数取值有关, 单元数大, 虽然诊断结果更好但训练代价过大, 诊断时间也较长。ResNet 和本文模型的准确率最为接近, 达到了 99.7%, 从训练时间上看两者差异不大, 但由于本文模型利用密集连接机制将各层不同层次特征信息融合, 准确率较高一些。

为避免实验偶然性, 本文将以上 4 种方法进行多次训练, 5 次训练准确率如图 4 所示, 5 次训练准确率差别不大, LSTM 和 CNN 的准确率明显劣于 ResNet 和本文方法, ResNet 和本文方法的准确率差别并不明显。从 5 次实验数据实验准确率均值看, ResNet 的平均准确率为 99.8%, 而本文方法的平均准确率为 99.9%, 只有少量错误。两种方法在准确率上的差异较小, 但本文方法只构造了一组密集块和过渡层, 从整体上来看本文方法的诊断效果稍好一些。

表 3 不同方法的诊断结果

网络结构	准确率	召回率	F 值	训练时间/min
CNN	82.76%	0.692 2	80.4%	31
LSTM	98.44%	0.118 2	98.35%	872
ResNet	99.7%	0.015 8	99.7%	177
本文方法	100%	0.000 5	100%	166

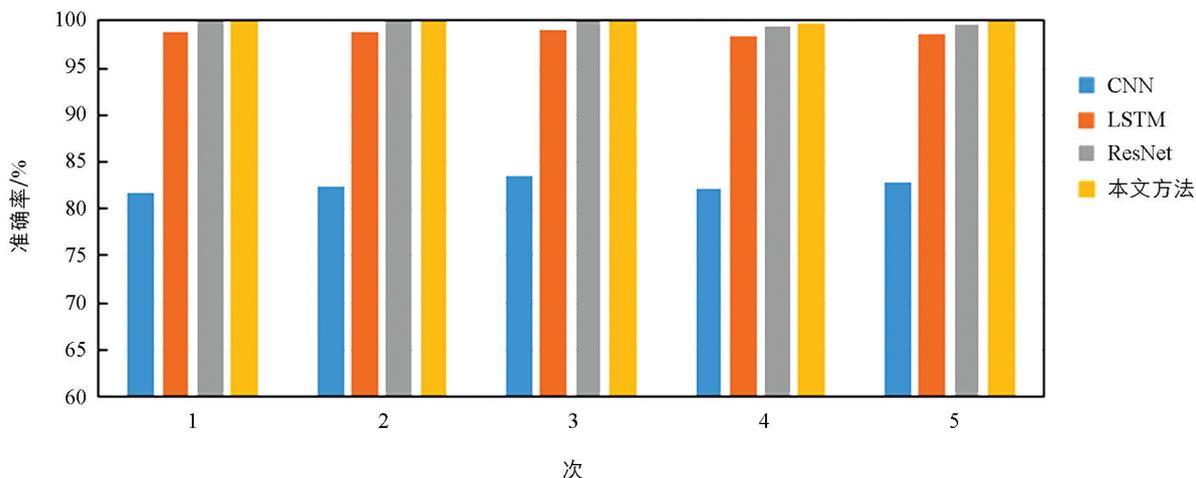


图 4 5 次训练准确率

为了直观地看出模型分类结果的优劣, 对模型特征提取情况进行可视化。利用 t-SNE 算法将多维向量降到二维, 将最后一层降维可视化, 不同算法最后一层 t-SNE 可视化结果如图 5 所示, 10 种颜色代表 10 种类别的数据, 图 5(a) 中有 3 个类别混杂在一起, 而图 5(b) 中蓝色和绿色混杂, 图 5(c) 中有个别几个数据分类错误, 而图 5(d) 中分类完全正确。

3.4 高铁轮对轴承实验

高铁轮对轴承故障实验数据, 轴承型号为 FAG F-80781109 TAROL 130/240-B-TVP, 采样频率为 51.2kHz, 采样时间为 60 s, 高铁轮对轴承实验台如图 6 所示。对轴承加工尺寸为长 5 mm, 宽 1 mm, 深 0.7 mm 的故障, 故障类型包括外圈故障、内圈故障和无故障轴承数据。测试工况包括定转速 1 200 r/min, 1 500 r/min, 1 800 r/min, 2 100 r/min 4 种。每类工况下的 3 种状态共 12 类, 并分别标记为 0~11。对数据集进行划分, 训练集每类 3 500 个样本, 测试样集每类 600 个样本, 验证集每类 1 200 个, 高铁轮对轴承数据集如表 4 所示。数据集中 1 200 r/min 和 2 100 r/min 下的无故障、内圈故障和外圈故障的部分数据如图 7 所示。

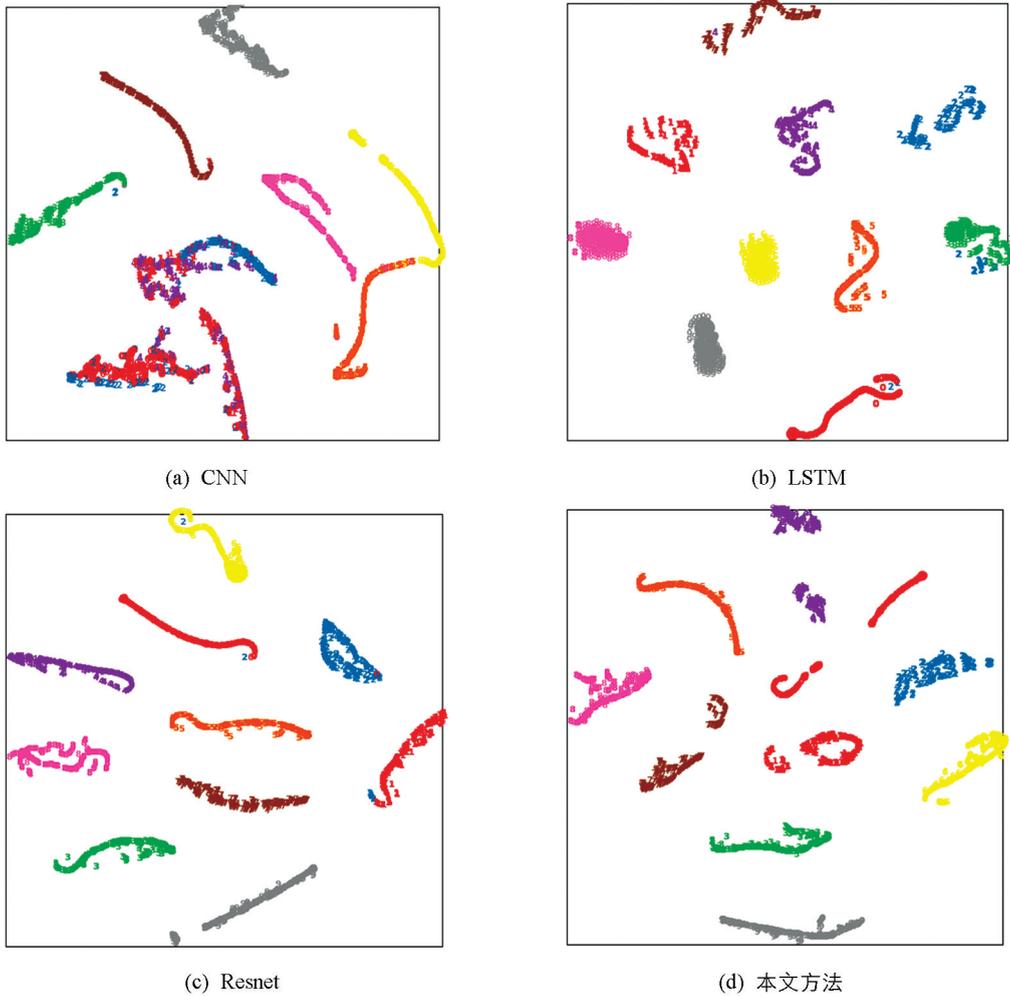


图 5 不同算法最后一层 t-SNE 可视化结果

表 4 高铁轮对轴承实验数据集

轴承类型	转速/ ($r \cdot \min^{-1}$)	训练样本 个数	测试样本 个数	标签
无故障	1 200	3 500	600	12_H
	1 500	3 500	600	15_H
	1 800	3 500	600	18_H
	2 100	3 500	600	21_H
内圈故障	1 200	3 500	600	12_O
	1 500	3 500	600	15_O
	1 800	3 500	600	18_O
	2 100	3 500	600	21_O
外圈故障	1 200	3 500	600	12_I
	1 500	3 500	600	15_I
	1 800	3 500	600	18_I
	2 100	3 500	600	21_I



图 6 高铁轮对轴承试验台

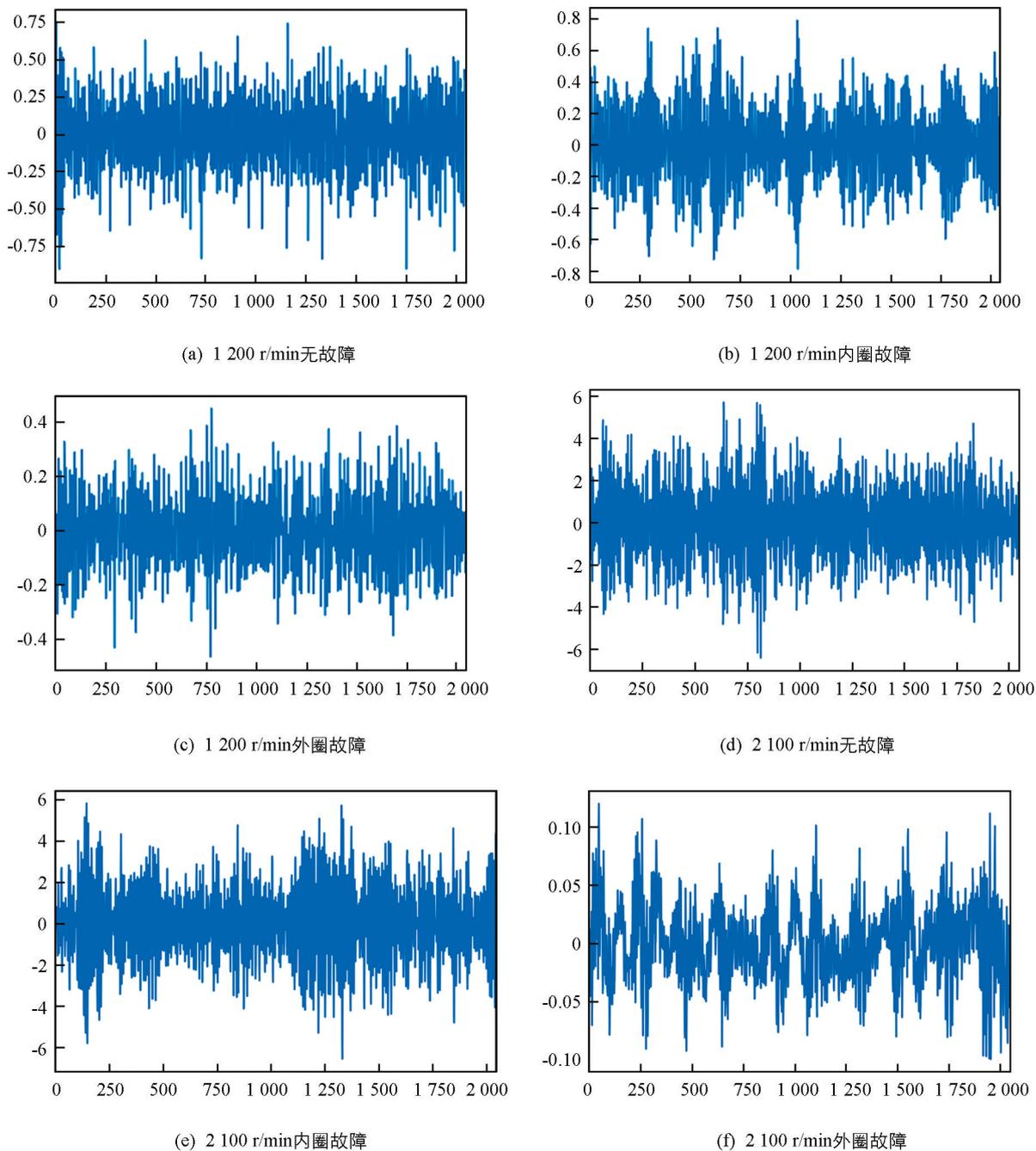
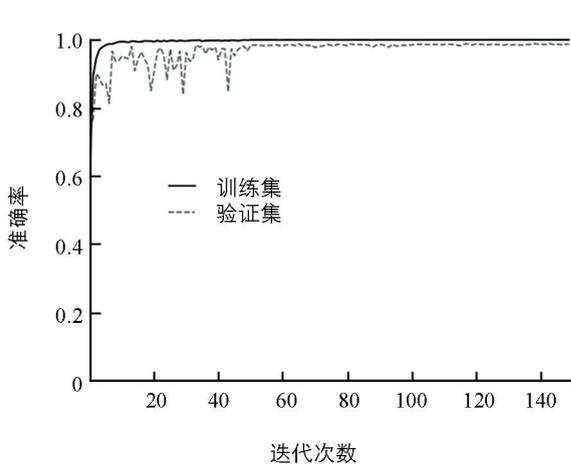


图 7 部分样本数据

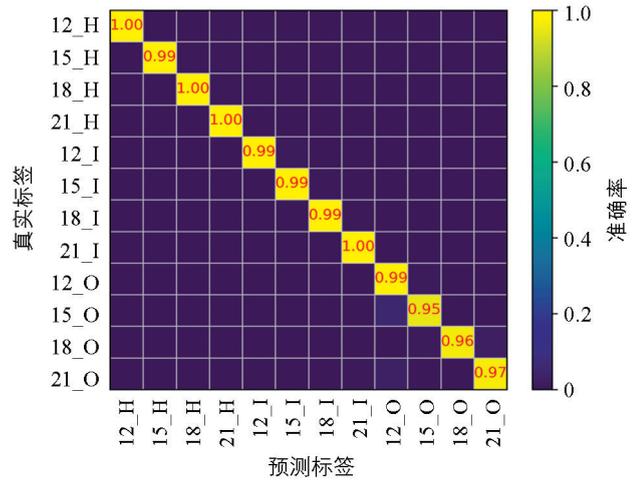
不同方法在高铁轮对轴承数据集上的故障诊断效果如表 5 所示, 从表 5 中可以看到, 比起其他故障诊断方法, 本文所提方法准确率更高, 效果更好. 高铁轮对轴承实验结果如图 8 所示, 其中图 8(a)为模型在高铁轮对轴承数据集上的训练准确率曲线图. 前 20 次迭代准确率较低, 保持在 90% 以下. 虽然训练次数不断增加但准确率波动较大. 在迭代 50 次时学习率缩减为 0.000 1, 模型能够稳定学习特征信息, 准确率缓慢增加. 迭代 150 次后, 最终在验证集上的准确率为 98.6%, 损失值为 7.1%. 图 8(b)为模型在高铁轮对轴承数据集中的诊断混淆矩阵, 可以看到模型对于健康状态的诊断较为准确, 内圈故障和外圈故障虽然有一定的错误, 但准确率都达到 93% 以上, 诊断总准确率在 98% 以上. 实验结果表明该方法能有效识别故障.

表 5 不同方法在高铁轮对轴承数据集的故障诊断结果

方法	CNN	LSTM	CNN+LSTM	本文方法
准确率	88.1%	89.3%	89.7%	98.6%



(a) 准确率曲线图



(b) 混淆矩阵

图 8 高铁轮对轴承实验结果

4 总 结

本文提出了一种基于密集连接卷积网络的故障诊断方法。利用 CNN 提取局部特征，密集连接机制使网络模型复杂化的同时避免产生过多的参数，提高了模型的学习能力，解决了传统故障诊断方法人工提取特征的不足，实现了端到端的故障诊断。实验结果表明，本文方法得到的故障诊断模型在美国凯斯西储大学的数据集上故障诊断准确率可高达 99.9%，对于轴承不同负载下的故障类型均能诊断良好。

在高铁轮对轴承数据集实验中，本文提出的方法准确率也达到了 98% 以上，实验结果表明该方法在机械故障诊断中具有一定的应用价值。

参考文献:

- [1] 常 勇, 包广清, 程思凯, 等. 基于 VMD 和 KFCM 的轴承故障诊断方法优化与研究 [J]. 西南大学学报(自然科学版), 2020, 42(10): 146-155.
- [2] 郑怀亮, 王日新, 杨远涛, 等. 数据驱动故障诊断方法泛化性能的经验性分析 [J]. 机械工程学报, 2020, 56(9): 102-117.
- [3] 李 俊, 刘永葆, 余又红. 卷积神经网络和峭度在轴承故障诊断中的应用 [J]. 航空动力学报, 2019, 34(11): 2423-2431.
- [4] 曲建岭, 余 路, 袁 涛, 等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法 [J]. 仪器仪表学报, 2018, 39(7): 134-143.
- [5] GU X, TANG X H, LU J G, et al. Adaptive Fault Diagnosis Method for Rolling Bearings Based on 1-DCNN-LSTM [J]. Machine Tool & Hydraulics, 2020, 48(6): 107-113.
- [6] 王裕峰. 融合多传感器的故障诊断方法研究 [D]. 北京: 北京交通大学, 2019.
- [7] 李 恒, 张 氢, 秦仙蓉, 等. 基于短时傅里叶变换和卷积神经网络的轴承故障诊断方法 [J]. 振动与冲击, 2018, 37(19): 124-131.
- [8] 安 晶, 艾 萍, 徐 森, 等. 一种基于一维卷积神经网络的旋转机械智能故障诊断方法 [J]. 南京大学学报(自然科学版), 2019, 55(1): 133-142.
- [9] 李鸿雁, 苏庭波. 基于贝叶斯网络和卷积神经网络的手绘草图识别方法 [J]. 西南师范大学学报(自然科学版), 2019,

44(9): 96-102.

- [10] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, US: IEEE, 2016: 770-778.
- [11] SZEGEDY C, LIU W, JIA Y Q, et al. Goingdeeper with Convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, US: IEEE, 2015: 1-9.
- [12] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely Connected Convolutional Networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, US: IEEE, 2017: 4700-4708.
- [13] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors [EB/OL]. [2012-10-20] [2020-09-20]. <https://arxiv.org/abs/1207.0580>.
- [14] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks From Overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [15] The Case Western Reserve University Bearing Data Center Website. Bearing Data Center Seeded Fault Test Data [EB/OL]. [2007-11-27] [2020-09-20]. <http://www/eecs/cwru/edu/laboratory/bearing/>.
- [16] 荆云建. 基于改进型卷积神经网络的动车组滚动轴承故障诊断方法研究 [D]. 北京: 北京交通大学, 2019.

Fault Diagnosis Based on One-Dimensional Densely Connected CNN

ZHAO Zhi-hong^{1,2}, ZHAO Jing-jiao¹, LI Qing², LI Le-hao²

1. School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;

2. State Key Laboratory of Mechanical Behavior and System Safety of Traffic Engineering Structures,

Shijiazhuang Tiedao University, Shijiazhuang 050043, China

Abstract: In recent years, data-driven bearing fault diagnosis has been widely studied due to the continuous development of deep learning. Models of deep neural networks such as convolutional neural network (CNN) gradually find applications in fault diagnosis. In view of the difficulties in manual extraction of fault features, and low training efficiency and serious over fitting of deep CNN network, a bearing fault diagnosis model based on one-dimensional densely connected convolutional network is proposed and implemented in this paper. The one-dimensional vibration data of bearing is used as input, and the fault feature information is extracted automatically by the convolutional network. High-level feature and low-level feature are combined by a dense-connection mechanism so that the fault feature information is more effectively extracted. Then, the over fitting phenomenon is avoided through the batch normalization and dropout methods. Finally, the SoftMax layer is used to classify the faults, so as to achieve intelligent fault diagnosis. The results of an experiment with the measured data of various types of bearing faults in CWRU (case-western-reserve-university) bearing data set and in the bearing data set of high-speed trains in China show that this method can effectively identify fault types, with an accuracy of 98% or more.

Key words: deep learning; CNN(convolutional neural network); dense connection; fault diagnosis