



江苏师范大学

硕士学位论文

(2021 届)

基于深度迁移学习的风电机组
齿轮箱故障诊断方法研究

Research on wind turbine gearbox fault diagnosis with
deep transfer learning method

作 者 王欣

导 师 刘文艺 副教授

江苏师范大学机电工程学院


二〇二一年五月


江苏师范大学

学位论文版权使用授权书

本学位论文作者完全了解 江苏师范大学 有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权 江苏师范大学 可以将本学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等手段保存、汇编、出版本学位论文。

保密的学位论文在解密后适用本授权书。


作者签名： 
2024年5月25日

导师签名： 
2024年5月25日

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是我个人在导师指导下进行的研究工作及取得的研究成果。论文中除了特别加以标注和致谢的地方外，不包含任何其他个人或集体已经公开发表或撰写过的研究成果。其他同志对本研究的启发和所做的贡献均已在论文中作了明确的声明并表示了谢意。

本人学位论文与资料若有不实，愿意承担一切相关的法律责任。

学位论文作者签名： 
2024年5月25日

中图分类号: TM-315
UDC: 621

单位代码: 10320
密 级: 公开

江苏师范大学

硕士学位论文

基于深度迁移学习的风电机组 齿轮箱故障诊断方法研究

Research on wind turbine gearbox fault diagnosis
with deep transfer learning method

作 者 王欣 导 师 刘文艺 副教授
申 请 学 位 工学硕士 培 养 单 位 机电工程学院
学 科 专 业 机械工程 研 究 方 向 机械故障诊断
答辩委员会主席 李威 教授 评 阅 人 _____

二〇二一年六月

摘要

风电机组齿轮箱作为传动系统中运行工况最为复杂的部件，其运行状态对于风电机组设备的健康状态监测与运行控制有着非常巨大的影响。风电机组的异常导致停机的情况不仅仅会造成电力能源的损失，同时伴随着维修费用的大幅增加。以风电机组齿轮箱作为主要的研究对象，进行了以下相关的研究：

针对齿轮箱齿轮系统中微观局部工况，提出了利用单齿故障进行有限元仿真分析获得模态数据的方法。并且利用深度自编码结构和传统 BP 结构结合进行二次训练的策略，综合考虑性能与效率的关系，提出了线性和非线性的性能评估方法。

深度迁移结构中的超参数配置往往具有随意性，所以在处理齿轮箱行星轮系故障分类问题时提出了一种层级迁移网络结构超参数寻优方法。利用经典的 LeNet-5 进行重构迁移应用在行星轮系模态数据集上进行算法验证。最后，进行了该算法性能的稳定性验证与结果分析。

提出了基于压缩感知的稀疏信号分解方法，并且重新设计了迁移网络的结构实现了从滚动轴承故障到齿轮故障的深度迁移学习。利用即插即用的注意力模块设计了新的网络结构。针对故障数据设计和制作了预训练模型，提高了故障诊断模型分类的准确率和识别速度。最后验证了迁移学习中源域和目标域样本数量相同，样本特征分布不同情形下对迁移学习方法的性能影响和超参数对于网络结构最终性能的影响。

关键词：深度迁移学习；风电机组；机械故障诊断；超参数搜索

Abstract

As the most complex component in the transmission system, the operating state of the wind turbine gearbox has a tremendous impact on the monitoring of the health status and operation control of the wind turbine equipment. Abnormalities in wind turbines that lead to downtime not only result in a loss of electrical energy, but also a significant increase in maintenance costs. Therefore, with the wind turbine gearbox as the main object of study, the following studies were carried out:

For microscopic local conditions in gearbox gear systems, a method for obtaining modal data using finite element simulation analysis of single tooth faults is proposed. Using a combination of deep auto-encoder structures and BP structures for secondary training strategies, a linear and non-linear performance evaluation method is proposed, which takes into account the relationship between performance and efficiency.

Hyper-parameter configuration in deep transfer structures is often arbitrary, so a hierarchical transfer network structure hyper-parameter searching method is proposed to address the gearbox planetary system fault classification problem. The algorithm is validated using the classical LeNet-5 reconfiguration transfer application on a modal dataset of the planetary system. Finally, a stability validation and results analysis of the algorithm performance is carried out.

A compressed sensing-based sparse signal decomposition method is proposed, and the structure of the transfer network is redesigned to achieve deep migration learning from rolling bearing faults to gear faults. A new network architecture was designed using a plug-and-play attention module. Pre-training models were designed and produced for fault data to improve the accuracy and recognition speed of fault diagnosis model classification. Finally, the effects of the same number of samples in the source and target domains and different distributions of sample features on the performance of the transfer learning method and the effects of hyperparameters on the final performance of the network structure are verified.

Keywords: deep transfer learning; wind turbine; mechanical fault diagnosis; hyper-parameter search

目录

摘要.....	I
Abstract.....	III
目录.....	V
图清单.....	IX
表清单.....	XIII
1 引言.....	1
1.1 课题来源.....	1
1.2 研究意义.....	1
1.3 风电机组齿轮箱中的关键部件.....	3
1.4 风电机组机械故障诊断研究现状.....	6
1.4.1 信号处理方法.....	8
1.4.2 传统机器学习方法.....	9
1.4.3 深度学习方法.....	9
1.5 本文主要研究内容.....	12
1.5.1 本文的整体架构.....	12
1.5.2 主要研究内容和研究方法.....	12
1.5.3 技术路线.....	13
1.6 小结.....	14
2 深度迁移学习理论.....	15
2.1 深度学习中神经网络的基本结构.....	15
2.1.1 自编码网络.....	15
2.1.2 卷积神经网络.....	17
2.1.3 循环神经网络.....	18
2.1.4 基本结构在分类任务与回归任务应用上的适用性.....	21
2.2 深度学习理论中的迁移方法.....	21

2.2.1 基于样本集的迁移特性	23
2.2.2 基于特征图的迁移特性	23
2.2.3 以网络结构为主的迁移理论	24
2.3 小结	25
3 齿轮箱单齿故障分类模式识别	27
3.1 单齿故障的 3D 模型的设计与构建	27
3.2 利用有限元分析及仿真产生数据集	29
3.2.1 模态分析	30
3.2.2 谐波响应分析	32
3.3 深度神经网络进行分类和优化性能	33
3.3.1 单一深度神经网络分类器应用于单齿故障诊断	33
3.3.2 多神经网络组合的性能优化策略	37
3.4 小结	40
4 齿轮箱行星轮系齿轮故障识别	41
4.1 行星轮系模态分析与特征提取方法	41
4.1.1 行星轮系模型的构建与模态分析位移数据的获得	41
4.1.2 断齿与点蚀关键特征图的生成	43
4.2 网络结构的方法研究	46
4.2.1 经典卷积网络的模型复用与分类器层重构	47
4.2.2 动态网格随机化超参数寻优方法	49
4.3 数值实验结果分析与讨论	51
4.4 小结	54
5 滚动轴承稀疏特征提取与迁移故障诊断方法	55
5.1 滚动轴承信号的稀疏特征提取	54
5.1.1 压缩感知的基本理论与方法	55
5.1.2 信号的 L1 与 L2 正则化	57
5.1.3 振动信号的获取与数据描述	58
5.1.4 振动信号的稀疏特征提取与试验数据结果分析	60
5.2 深度迁移方法的设计和实验验证	63
5.2.1 深度迁移结构的参数设计与结构解释	63
5.2.2 神经网络的迁移策略	68

5.3 小结	70
6 结论与展望	71
6.1 结论	71
6.1.1 主要结论	71
6.1.2 主要创新点	72
6.2 展望	72
参考文献	73
致谢	81
作者简介	83

图清单

图序号	图名称	页码
图 1-1	全球风电机组装机容量的变化	1
Figure 1-1	Changes in the global wind turbine installed capacity	1
图 1-2	陆地风电机组和离岸风电机组的装机容量占比	2
Figure 1-2	Proportion of installed capacity of onshore wind turbine and offshore wind turbine	2
图 1-3	风电机组基本结构	4
Figure 1-3	Wind turbine basic structure	4
图 1-4	增速齿轮系与行星轮系组成的结构	5
Figure 1-4	Speed increasing gear and planetary gear train defined structure	5
图 1-5	滚动轴承的基本结构	6
Figure 1-5	Rolling bear basic structure	6
图 1-6	风电机组机械故障诊断研究进展	7
Figure 1-6	Wind turbine mechanical fault diagnosis research progress	7
图 1-7	章节内容的结构与层次关系	12
Figure 1-7	Structure and hierarchy of each chapter	12
图 1-8	本文技术路线	14
Figure 1-8	Roadmap in this article	14
图 2-1	自编码器的基本结构	16
Figure 2-1	Auto-encoder basic construct	16
图 2-2	卷积算子的计算方法	17
Figure 2-2	Calculation method on convolution operator	17
图 2-3	不同大小的池化算子计算结果	18
Figure 2-3	Calculation results of pooling operators with different sizes	18
图 2-4	循环神经网络的单神经元模型	19
Figure 2-4	Recurrent neural network single neuron model	19
图 2-5	长短期记忆网络的基本结构	20
Figure 2-5	LSTM basic structure	20
图 2-6	门极可控单元网络的基本结构	20
Figure 2-6	GRU basic structure	20
图 2-7	迁移学习基本方法	22
Figure 2-7	Transfer-learning basic method	22
图 2-8	基于样本集的迁移方法示意图	23
Figure 2-8	Instances-based deep transfer learning	23

图序号	图名称	页码
图 2-9	基于特征图的迁移方法示意图	24
Figure 2-9	Mapping-based deep transfer learning	24
图 2-10	以网络结构为主的迁移方法示意图	25
Figure 2-10	Network based deep transfer learning	25
图 3-1	研究结构的流程图	27
Figure 3-1	The flow chart of research structure	27
图 3-2	单齿建模过程	28
Figure 3-2	Single gear tooth modeling process	28
图 3-3	四种类型的齿轮故障	29
Figure 3-3	Four types of gear fault	29
图 3-4	模态分析的流程	30
Figure 3-4	Modal analysis flow chart	30
图 3-5	网格划分与边界约束	31
Figure 3-5	Meshing and boundary constraints	31
图 3-6	模态分析获得的前十阶频率	31
Figure 3-6	First tenth order frequency obtained by modal analysis	31
图 3-7	谐波响应外部均匀载荷的位置	33
Figure 3-7	Position of external uniform load applied by harmonic response	33
图 3-8	数据处理过程与深度神经网络作为分类器的设计	34
Figure 3-8	Date processing and DNN model designed as classifier	34
图 3-9	设计的结构	35
Figure 3-9	Architectures designed	35
图 3-10	使用 FNN 和 NN based on Deep AE 进行综合性能优化	37
Figure 3-10	Performance optimization on two datasets using FNN and NN based on Deep AE	37
图 4-1	行星轮系统建模及其模态分析	42
Figure 4-1	Modeling and modal analysis of planetary gear system	42
图 4-2	行星轮系破损与点蚀模态分析	43
Figure 4-2	Modal analysis of damage and pitting in planetary gear train gear and ring gear	43
图 4-3	高斯白噪声信号的仿真结果	45
Figure 4-3	Simulation results of Gaussian white noise signal	45
图 4-4	单通道特征图与三通道特征图	45
Figure 4-4	Single-channel feature map and three-channel feature map	45
图 4-5	行星轮系样本集可视化及其密度分布	46
Figure 4-5	Visualization of planetary gear aystem sample set and its density distribution	46
图 4-6	LeNet5 模型	47
Figure 4-6	LeNet5 model	47
图 4-7	AlexNet 模型	47
Figure 4-7	AlexNet model	47

图序号	图名称	页码
图 4-8	堆栈冻结方法	48
Figure 4-8	Stack freezing method	48
图 4-9	超参数网格搜索与随机搜索方法	50
Figure 4-9	Grid search and random search for hyperparameter	50
图 4-10	层次网格超参数搜索算法流程图	51
Figure 4-10	Hierarchical grid scaling hyperparameter random search flowchart	51
图 4-11	三个阶段搜索空间的网格划分	52
Figure 4-11	Three-stage meshing of search space	52
图 4-12	三个阶段超参搜索的结果	52
Figure 4-12	Three-stage hyperparameter search results	52
图 4-13	七种不同的优化器的训练过程和测试过程的可视化	53
Figure 4-13	The training process and testing process visualization of seven optimizers	53
图 4-14	七种优化算法的损失与准确率示意图	53
Figure 4-14	The loss and accuracy of seven optimizers	53
图 5-1	L0~L2 范数的二维平面解析示意图	57
Figure 5-1	L0~L2 norm 2D plane diagram	57
图 5-2	风力发电机组实验台	58
Figure 5-2	Test-bed of wind turbine	58
图5-3	振动信号的采集	59
Figure 5-3	Acquisition of vibration signals	59
图 5-4	五类信号的时域与其解析信号的时域表达	60
Figure 5-4	Time domain of five types of signals and the time domain expression of their analytical signals	60
图 5-5	L0~L2 滚动轴承信号的信号处理设计流程	60
Figure 5-5	L0~L2 Signal processing design process of rolling bearing signal	60
图 5-6	L0~L2 滚动轴承信号的上包络和下包络曲线	61
Figure 5-6	L0~L2 Upper envelope and lower envelope curve of rolling bearing signal	61
图 5-7	不同故障状态下的密度分布曲线	62
Figure 5-7	Density distribution under different fault states	62
图 5-8	V0 卷积神经网络结构示意图	64
Figure 5-8	V0 convolutional neural network structure	64
图 5-9	V1 卷积神经网络结构示意图	67
Figure 5-9	V1 convolutional neural network structure	67
图 5-10	滚动轴承数据集作为源域，齿轮数据集作为目标域	68
Figure 5-10	Rolling bearing dataset is used as the source domain and gear dataset is used as the target domain	68
图 5-11	三种卷积神经网络在随机超参数分布下的两种迁移策略效果	69
Figure 5-11	Results of two transfer strategies of three kinds of convolutional neural networks under random hyper-parameter distributions	69

图序号	图名称	页码
图 5-12	三种网络结构在不同的超参数下的迁移结果对比的平行线图	70
Figure 5-12	Parallel line graphs comparing the transfer results of three network structures under different hyper-parameters	70

表清单

表序号	表名称	页码
表 3-1	二维单齿平面几何尺寸	28
Table 3-1	Two-dimensional single tooth plane geometry size	28
表 3-2	普通碳钢相关材料参数	29
Table 3-2	Common carbon steel related parameters	29
表 3-3	总应变数据	32
Table 3-3	Total strain data(Datasets A)	32
表 3-4	谐波响应总应变数据	33
Table 3-4	Data generated by harmonic response (Datasets B)	33
表 3-5	两种数据集和网络结构下的结果	36
Table 3-5	Three kinds of outputs including total two different datasets and NN architectures	36
表 3-6	多次神经网络的组合线性优化性能	38
Table 3-6	Combination linear optimization performance of multiple neural networks	38
表 3-7	多次神经网络的组合非线性优化性能	39
Table 3-7	Combination nonlinear optimization performance of multiple neural networks	39
表 4-1	太阳-行星轮系传动系统中参数	42
Table 4-1	The planetary train system parameters	42
表 4-2	40CrMnMo 材料参数	43
Table 4-2	40CrMnMo material parameters	43
表 4-3	层次网格超参数搜索	54
Table 4-3	Hierarchical grid scaling hyperparameter random search	54
表 5-1	故障信号不同类型的香农熵计算结果	63
Table 5-1	Shannon entropy calculation results of different types of fault signals	63
表 5-2	V0 卷积神经网络结构具体参数	65
Table 5-2	V0 design parameters of convolutional neural network structure	65
表 5-3	V1 卷积神经网络结构具体参数	66
Table 5-3	V1 design parameters of convolutional neural network structure	66

1 引言

1.1 课题来源

本课题来源：

国家自然科学基金：基于多维核域谱的风机传动链系统微征兆健康状态识别方法。

江苏省基础研究计划（自然科学基金）项目：基于深度迁移学习的风电机组故障诊断方法研究。

1.2 研究意义

从 1996 年开始世界各国的风电机组装机容量开始了迅速的增长，依据全球风能理事会 GWEC 的报告来看，到 2018 年为止的装机容量达到了 591000MW 以上的容量。从图 1-1 来看，装机容量的增速尽管在降低，但是总量已经非常巨大。尽管全球风电机组的装机容量已经拥有较大的规模，但是仍然存在比较大的市场和可拓展空间。同时，陆上风电机组和离岸风电机组的数量存在着明显的数量差异。



图 1-1 全球风电机组装机容量的变化
Figure 1-1 Changes in the global wind turbine installed capacity

图 1-2 展示了我国的风电机组的两种类型数量的显著区别，陆上风电机组因为开发较早，技术也比较成熟，数量上占主导。离岸风电机组的数量占比较少，尽管发展增速在加快，但是受限于成本和技术积累其占比在较短时间内不可能出现较大的提升。无论是离岸风电机组还是陆上风电机组其日常维护和维修的资金消耗都是一个不可忽略的支出。未来的几年时间风电机组的装机容量毫无疑问会快速增长。

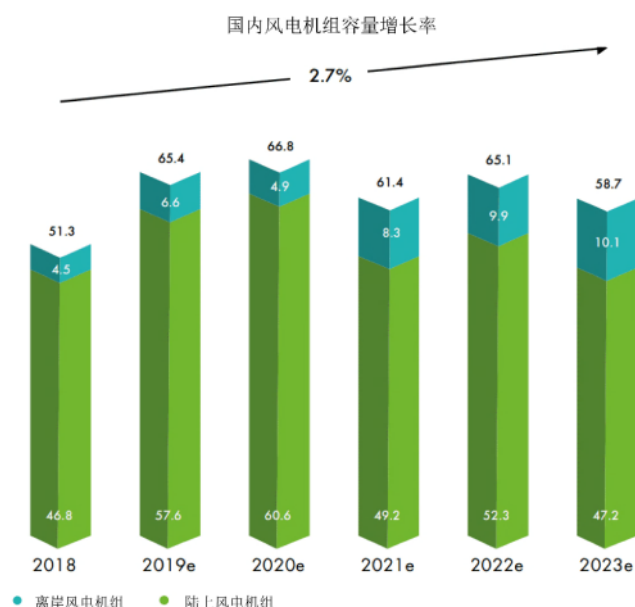


图 1-2 陆地风电机组和离岸风电机组的装机容量占比
Figure 1-2 Proportion of installed capacity of onshore wind turbine and offshore wind turbine

风电机组的各种故障类型主要涉及两个主要的方面电气故障和机械故障^[1]。当然，通常情况下这两方面的故障会互相影响，且有定的因果性和耦合性，并不存在明显的界限。风电机组产生的电能通常在接入电网之前，会经过一系列的存储和转换，对其中产生的故障类型相关研究比较丰富^[2-3]。而风电机组中的机械故障主要是指构成风电机组的传动系统中的部件不正常振动产生的这部分故障。风电机组运行过程中各种部件之间由于安装设备中可能出现的装配误差，导致部件与部件之间存在磨损、松动和失效^[4-5]的状况。针对其运行状态的监测和控制中重点关注的研究对象主要是行星齿轮箱与滚动轴承^[6]。相关的研究报告表明这两个关键部件发生的故障占到风电机组停机的故障原因的绝大部分比例。因此，若是能针对这两个部件进行非常准确和精密的监测和故障诊断，那么将会极大的提高风电机组的正常运行时长，进而提高经济效益。

实际上不仅仅是陆上风电机组有运维和维护的困难和巨大的经济利益的损失，离岸风电机组的维修成本更加高昂。考虑对风电机组在其设计寿命中进行故障诊断和运行状态监测是十分必要的，既有学术研究方面的价值同时兼具社会经济价值。概括如下：

(1) 风电机组通常会通过一定的电力设备接入电力网络，因此其瞬间的接入和断开势对电网造成瞬态的冲击。为了电力系统整体运行状态的稳定，在风电机组端进行监测和控制是成本低廉且易于操作的手段。

(2) 风电机组在其设备的正常运行周期内的停机时间实际上意味着能源收益的减少，从这个角度考虑，风电机组的运行状态所带来的经济方面的考量也是一个需要对其进行故障诊断和运行状态监测的重要原因之一。

(3) 风电机组运行过程中产生的数据对于设备的生产企业来说是一笔重要的“资产”。数字经济时代，数据本身就是一种资源。利用产生的数据持续改进风电机组大型设备的设计生产和制造过程逐渐成为一种新的潮流。

(4) 风电机组的运行状态监测和故障诊断问题，从广义的角度来考虑，其实是几乎所有大型设备都具有的问题。探究现象背后的机理，是学术追求的最主要目标之一。一方面增强对风电机组机械故障诊断的理论认识，另一方面开发依据理论成果的相关应用。

1.3 风电机组齿轮箱中的关键部件

风电机组的分类方法多种多样，比较典型的分类是按照发电机的类型或者传动方式的差异来进行的。图 1-3 表达了风电机组的基本构成，一般包括风速风向仪、发电机、机舱控制柜、制动系统、齿轮箱等部分。风速风向仪进行风向和风力的监测，将获得的信息传递给机舱控制柜，机舱控制柜发送指令给变桨系统。变桨系统依据信号调整至最佳角度。传动系统方面，叶片转动带动轴系的旋转，通过齿轮箱中的增速齿轮系统和行星轮系实现较高的转速，满足发电机所需要的转速。转速过高时，制动系统会起到相应的限制作用。

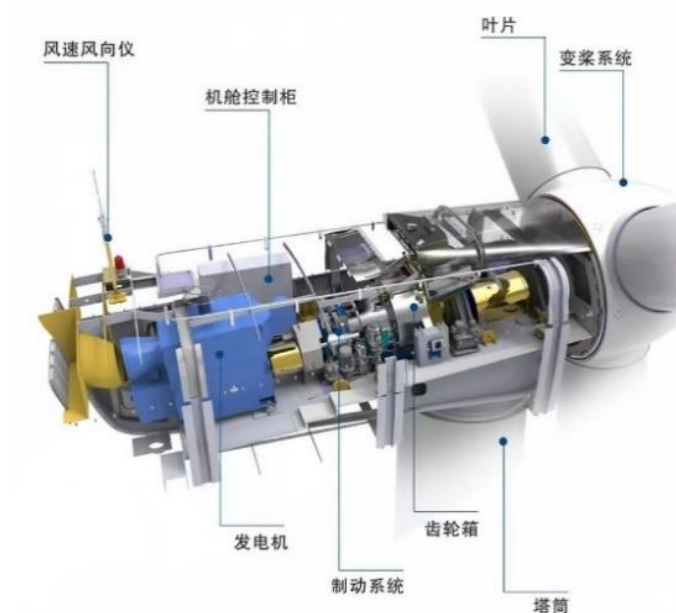


图 1-3 风电机组基本结构

Figure 1-3 Wind turbine basic structure

风电机组的外部机械结构和其一般所处的野外环境，在叶片的运行过程中产生随机振动的现象^[7]。此外，位于传动链系统的末端的双馈异步发电机，也常常受到传动系统的传递效应影响产生很多异常^[8]。异步发电机，作为电场和磁场进行能量交换的器件，伴随着噪声和发热的问题。当然，内部高速运转的轴承是振动产生的主要原因。针对异步发电机的振动测量，通常在距离其一定距离的支撑底座布置振动传感器。叶片和异步发电机运行中不正常对整个风电机组正常运行造成的影响较小，发生故障最多的部件是传动系统中的齿轮箱。此外，齿轮箱中的故障形式和模式也是最多样的。

风电机组中齿轮箱的结构简化如图 1-4 所示。典型的传动系统通常由增速齿轮部分和行星轮系部分组合而成。增速齿轮部分实现低转速到高转速的变化，获得一个较高的转速。该转速被传递进一个由太阳轮-行星轮-齿圈构成的行星轮系。经过行星轮系之后，最后作为双馈异步电机的输入转速用来持续供给电能。低速轴直接与叶片相连接，而叶片的转动具有随机性大且载荷变化无规律的特点，在这种工况下运行，容易产生各种故障。所以，齿轮箱是故障诊断和运行监测系统中需要重点研究的对象。

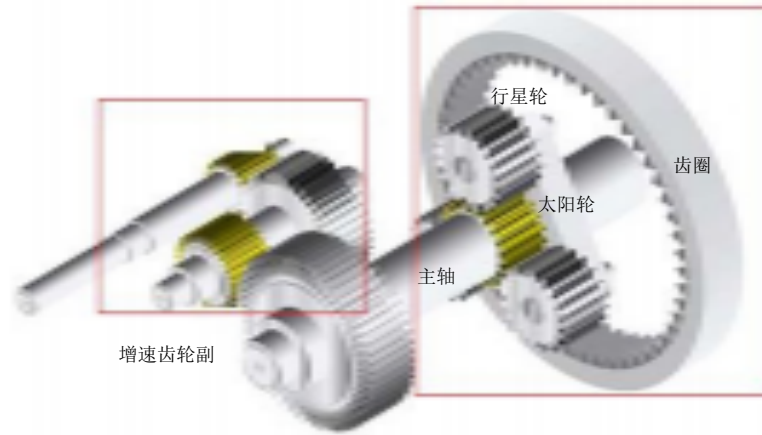


图 1-4 增速齿轮系与行星轮系组成的结构

Figure 1-4 Speed increasing gear and planetary gear train defined structure

齿轮箱内部的结构复杂度较高且振动与噪声的来源不容易定位到具体的部位。振动信号的特点与高斯白噪声非常相似伴随非线性强耦合非平稳杂波信号的特点。高速旋转的主轴，与之相连传递动能的各部件经常会出现各种故障。各个部件可能出现的故障类型有很多，比如油液温度过高、轴系的不对中不平衡断裂和齿轮的相关故障。

齿轮和轴承是齿轮箱中各种故障类型中最复杂和最丰富的部件。大部分研究论文的研究对象均集中在这两个部件。从零部件的生命周期来说，零部件从设计生产到装配使用，直至最后零部件的维护保养更换环节均有可能出现问题。在设计生产环节，最容易对后来使用环节产生影响的是轮齿本身的制造缺陷和装配误差。这些环节主要由生产齿轮的企业其制造技术和工艺决定了齿轮的使用寿命和精度。除了齿轮出厂时的产品质量之外，在装配过程中产生的轴线不对中^[9-10]等等问题也可能造成其出现失效现象。另外可能造成齿轮出现故障的原因更多集中于风电机组的运行工况过程中。风电机组的正常运行时，产生的振动来源主要是变工况变载荷的正常发电过程和齿轮故障产生引起的传递函数和刚度变化。变工况变载荷对于风电机组来说是比较常规的现象，因为风力本身具有的高度随机性，这是无法避免的现象。齿轮失效的原因具体包括点蚀、断齿、磨损等等。而这些失效现象最终可以通过在齿轮箱合理的位置设置传感器进行测量，获得其对应产生的振动信号利用信号处理的手段进行分析，可以得到粗略的判断。

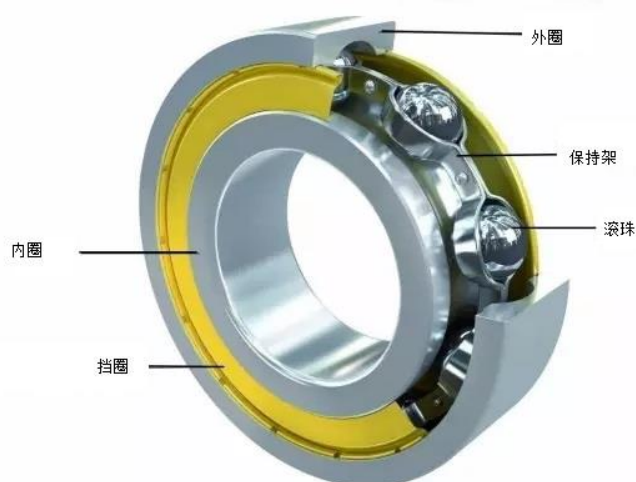


图 1-5 滚动轴承的基本结构
Figure 1-5 Rolling bearing basic structure

风电机组齿轮箱中另一个需要关注的研究对象是滚动轴承。滚动轴承在齿轮箱中的数量很多，其运行工况更加复杂。因为周期性的载荷变化，其运行工况产生的振动波形往往和正弦波有很大的关联性。在滚动轴承出现故障的初期即使出现了异常的振动信号，往往比较能量较微弱会被淹没在齿轮箱其他的振动信号中，很难被监测和控制。因此，经常会错过最佳的故障发现时间，可能会有后续维修和诊断不及时的问题。当轴承故障引起的振动能被测量的时候，一般出现的问题都比较严重了。这时候进行维修的成本也相对较高。滚动轴承的基本结构如图 1-5 所示。其主要包括内圈、外圈、保持架和滚珠四个部分，通常也存在挡圈。挡圈的存在主要是防止滚珠的高速转动中防止外部的灰尘等杂质进入，影响其正常的转动。滚动轴承的故障现象主要是内圈故障、外圈故障、保持架断裂和滚珠破损等几种。另外，滚珠和保持架之间还有可能会出现磨损导致的断裂和失效破坏现象。滚珠与内外滚道之间的摩擦在高速滑动情形下也是一个容易发生故障的情形，通常伴随着温度的急速升高，产生胶合现象。

1.4 风电机组机械故障诊断研究现状

风电机组机械故障诊断的方法分类有很多种。从测量方法和被检测的物理量这个角度来考虑，比较常用的有基于加速度和位移的振动测量方法，基于声学的噪声监测和控制方法^[11]，基于温度的红外线温度测量方法，基于物质物理或者化学性质的油液分析^[12]方法和通过光学利用 X-ray 进行扫描的无损检测方法等等。从故障发

生的来源可以分为单一来源故障、多源故障和系统故障（单一来源故障和多源故障都含有，或者多个故障源存在互相影响）^[13]。根据机械故障造成的影响和对设备造成的损坏程度可分为微小故障和严重故障。另外，根据故障在运行机械的生命周期^[14-15]内的不同阶段可以分为初期故障、中期故障和末期故障。

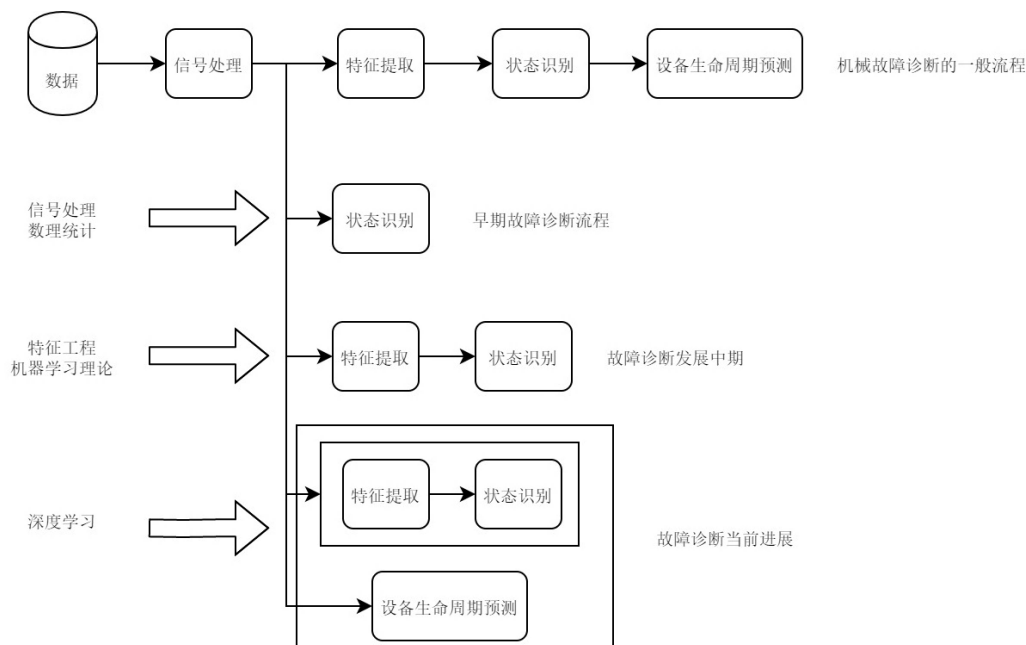


图 1-6 风电机组机械故障诊断研究进展

Figure 1-6 Wind turbine mechanical fault diagnosis research progress

风电机组机械故障诊断的一般处理过程由四个部分组成，分别是信号处理、特征提取、状态识别和设备生命周期预测。在故障诊断技术与方法的发展初期，数据获取和信号处理方法作为主要的环节，通过时域、频域或者时域-频域等现代信号处理方法^[16]，结合数理统计中的部分理论，直接进行设备的运行状态判断。中期阶段，随着信息系统中的模式识别理论的引入，将获取到的故障信号数据进行特征工程的预处理，再结合机器学习与模式识别中的经典分类方法，进而实现机械设备的运行状态识别。与早期的故障诊断方法相比，这一时期的主要提升之处在于特征工程的应用^[17]，手工设计的特征使样本之前的差异性增大，便于学习算法的计算效率和准确率的提升。针对少量样本和特征工程优异（模式识别术语中“数据清洗彻底”）确实有着不错的效果。然而这个时期的方法局限性也是存在的，在数据量较大或者样本的特征过高的时候，效果并不理想。现代深度学习理论，伴随着计算机的算力不断提升，在图像分类与识别领域^[18]、自然语言处理和语音识别^[19-20]等方面的应用呈现出巨大的前景。深度学习领域中主要模型目前仍然以神经网络为主，但不再是曾经的浅层 BP 神经网络。取而代之，更多的是卷

积神经网络和循环网络为代表的深层结构。网络结构层数的增长使神经网络的表达能力得到了提高。很明显的一个想法是利用网络本身的深层结构直接进行状态识别,特征工程的部分与最后的状态判断综合实现,实现端到端的功能^[21]。此外,在进行风电机组设备剩余寿命的预测这个任务的时候,也可以直接利用信号本身进行。这一切都得益于网络结构的复杂性和计算的高效性。设备剩余生命周期预测与机械设备的状态识别在这种角度思考下,可以视作两个单独处理的问题。

在下面三个小节中,将分别从信号处理方法、传统机器学习方法和深度学习方法(图 1-7),三个不同的方法论角度进行简单综述,梳理用于风电机组机械故障诊断的相关研究。

1.4.1 信号处理方法

信号处理的方法针对的主要是齿轮和轴承在运行过程中的振动信号。除了最简单的时域直接观察之外,通过信号的相关频谱分析进行幅值增强和啮合频率及其边频带来进行故障诊断也是一种方法^[22]。行星齿轮箱的中轮齿的啮合频率及其谐波成分的幅频谱非常复杂,行星轮的传递效应,甚至是时变的振动信号传递路径改变均有可能是其成因。因此,采集到的振动信号往往需要进行消噪或者滤波。其中,信号的分解方法可以将信号中的故障特征突出表现。许多学者针对信号的分解方法采取了非常多的理论研究与应用。Antoniadou 使用了经验模态分解(EMD, Empirical Mode Decomposition)分析了变载荷情况下的风电机组齿轮箱的振动信号。并且将特征信号进行希尔伯特变换,通过 Teager 能量算子结合能量分离算法实现波形分离^[23]。冯等提出了使用集成经验模态分解(EEMD, Ensemble empirical mode decomposition)将扰动信号分解成多个本征模态函数,并将其中的噪声信号去除^[24]。而局部均值分解(LMD, Local mean decomposition)算法将行星齿轮箱中获得的振动信号分解为一系列乘积函数^[25,26]。与 EMD 算法比较, LMD 方法保留了信号更多的幅值-频率特征。王^[27]等将单通道的振动信号使用 EEMD 分解为本征模态函数的组合,然后利用独立成分分析方法将滚动轴承故障信号从齿轮啮合信号中分离出来。并且利用实际的数据进行了相关的算法验证工作。

机械运行过程中产生的振动信号在传感器测量的初期一般需要利用信号处理的基本方法,进行预处理,包括滤波等操作。近年来,有关运用在机械故障诊断方面的信号处理相关研究成果并没有太多进展。两个原因造成了这种现状,其中一个原因是以深度神经网络代表的深度学习方法热度过高,另一个原因是信号处理方法相

关理论发展比较成熟，理论上的创新并不突出。

1.4.2 传统机器学习方法

专家系统是传统机器学习近代发展的开端之一。专家系统的主要特点是将专家具有的故障诊断知识集成在机械系统中以解决机器故障诊断的问题，解放了大量的人力资源。其通常由以专家知识为基础结合动态数据库和解释系统，前端应用配合用户交互接口。在这一阶段以模糊逻辑方法^[28,29]为其中的典型代表。Gelgele 将 If-Then 规则应用在发动机引擎的专家系统中^[30]。Lee 等人依据模糊逻辑设计了一个针对电力能源系统的，这是模糊逻辑推理在机械故障诊断领域的首次应用^[31]。Berredjem 将模糊专家系统用在滚动轴承的故障诊断上并且取得了较高的准确率^[32]。Jayaswal 等人将神经网络和模糊规则结合使用构建了专家使用并进行了滚动轴承的故障诊断^[33]。

传统的 BP 神经网络在也在故障诊断领域中有着非常多广泛的应用。以传统的神经网络作为起点，后续相关学者相继开始了 RBF 神经网络、小波神经网络和 ELM 网络等网络结构的应用和理论研究。Merainani 使用了自组织网络进行了齿轮箱的健康状态的辨识^[34]。陈等人将概率神经网络用来诊断液压变压器的故障^[35]。钟等人提出了一种分层的神经网络用于电动机系统的故障诊断，在识别复合故障问题上取得了不错的效果^[36]。

除了模糊逻辑和传统神经网络在故障诊断领域广泛应用之外，其他出现的被广泛应用的方法包括支持向量机(SVM)、k 最近邻算法(kNN)、概率图模型和选择树算法等等。支持向量机的优点是完整的数学理论提供了工程保障，但是其应用的数据量一般较小，许多学者和专家针对这个问题提出了改进方法^[37-39]。kNN 是一种无监督中心聚类方法，且需要指定聚类中心的个数。从故障诊断的任务来说，需要提前知道故障类型的个数^[40,41]。

传统的机器学习方法在过去很长一段时期内非常有效，主要受限于当时的计算能力较小且获取的振动数据量也小。大部分算法都有完备的数学理论作为支撑，其误差和准确率利用解析方法也可以比较容易进行分析和比较，这是这些方法流行的原因。

1.4.3 深度学习方法

多层感知器已经在机械故障诊断领域应用已经很多年了^[42-45]。深度学习技术在

最近几年才大规模的应用在机械故障诊断领域。基于深度自编码器(AE)与受限玻尔兹曼机(RBM)的层级深度神经网络(DNN)被证明可以提高深度神经网络的训练速度和效率,并且可以在一定程度上将设备运行中产生数据的特征强化,增强各个特征之间的差异性。卷积神经网络和循环神经网络则提供了从数据中获得复杂和抽象特征的代表方法。在上述基于深度学习的机械设备运行状态监测与故障诊断的系统中,网络的最终层通常是给定的标签。并且在最后一层完全是离散的给定类别的时候,其上层通常是被用来表征概率的 softmax 层。另一种情形是最终的输出是连续的数值,它所对应的上层是 Linear regression 层。此外,这种端到端的结构使得基于深度学习的机械设备运行状态监测与故障诊断的系统花费更少的人力和需要更少的专家知识。因此,深度学习所构建的模型不会因为设备的不同,领域的差异受限。重点对下列三类深度神经网络做一个简短的梳理。

(1) 自编码器及其变种

自编码器模型中当前比较流行的堆栈自编码器,可以自动的从机械运行数据中提取特征。孙等人提出一种单层自编码网络用来进行感应电机的故障辨识^[46]。受限于训练样本的数量,他们将研究重点放在避免过拟合等方面。陆等提出了一种关注微小数据特征的多层堆栈降噪自编码结构。这种网络具有三层隐藏层,用于旋转机械设备的部件故障诊断^[47]。陶等设计了基于深度神经网络的变结构两层稀疏自编码(SAE)架构。其主要的特点是隐藏层基本计算单元的大小和层数是可变的,并且其满足一定的分布规律,这提高了用于故障诊断的性能和效率^[48]。贾等人提出了将时间序列的时频谱作为自编码器的输入来进行旋转机械故障诊断^[49]。孙等人利用压缩感知技术从时间序列中提取低维特征来训练堆栈自编码器^[50]。周等人设计了三个变尺度的自编码器分别用于模态分类、故障源分类和故障破坏程度的分类^[51]。谭等人利用数字滤波器框架和非线性软阈值方法对振动信号进行预处理,而后基于处理后的信号进行了滚动轴承的故障诊断^[52]。将故障数据的时域特征、频域特征和时频域特征结合起来作为综合的特征在文献^[53]中也有提及。在文献^[54]中,陈利用时间序列的时域特征来训练堆栈自编码,并且将支持向量机作为最终的分类器。为了避免过拟合现象的出现,其团队在训练中掺入了高斯噪声^[55]。从以上的研究文章来看,自编码器在其中的主要作用是用于微弱特征的提取。

(2) 卷积神经网络及其变种

机械运行产生的数据通常是一维的时域信号,同时也可以利用一些变换将其转化为二维的时频谱等形式。CNN 最早应用在计算机视觉领域,天然的具有处理二维信号的能力。卷积神经网络与传统神经网络的不同之处在于,卷积层可以从数据中

学习到抽象的特征并且利用更少的参数来达到相同甚至更优的准确率。刘等人提出了一种交错时间序列的卷积神经网络，并把该网络结构用于电气设备的故障诊断^[56]。在另一种卷积神经网络中，研究者将时间信号进行小波分解，把得到的各级信号进行时频域变换，最终将频谱图像作为输入进行机械故障诊断的研究^[57]。陆等人将含有噪声的一维信号进行滑动插值，把信号的尺寸变成二维方阵^[58]。丁等人提出将信号经过小波包处理得到的能谱图作为卷积网络的输入用于滚动轴承的故障诊断^[59]。为了能尽可能的提取数据中的特征，通常在卷积层之后加入多尺度层。该层同时具有连接各个池化层的作用。郭等人^[60]提出了一种层级自适应深度卷积网络。其主要特点是进行了特征图的大小自适应调整。孙利用双树复变小波变换将时间序列转换成 2D 特征图，将序列的时移特性作为特征进行卷积神经网络的构建^[61]。在文献^[62]中，陈将时间序列的 256 个统计特征由一维序列制作成尺寸时 16*16 的二维方阵，将其作为每一个样本，取得了误差很小的效果。Weimer 等人做了一些列的数值实验进行了不同尺寸的卷积网络来进行相同的缺陷检测任务^[63]，试图从中寻找最优的卷积网络参数。文献^[64]中提到的卷积神经网络被用在风电机组早期微弱故障检测的目标上，这种网络的输入是水平和竖直的两组相互正交放置的传感器所获取的信号。You 等人在常规的卷积神经网络最后一层加入了支持向量机作为最终的分类判别器^[65]。在 Lee 等人^[66]的研究中，他们第一次将卷积神经网络用在半导体生产过程中的故障检测。卷积神经网络在机械故障诊断领域的应用特点主要在信号的特征预处理和网络的结构增强。总体来说，深度卷积网络的特征提取能力和作为分类器的判别能力都比较强，大部分研究将其一体化处理，很少进行严格的区分。

(3) 循环神经网络及其变种

目前大多是的机械运行数据都是从传感器获得的，本质是是一维时间序列。循环神经网络模型主要包括 LSTM 和 GRU，这是当前最流行的用来处理时间序列的结构。其主要的特点是可以实现对时间的隐藏信息进行编码。近年来，许多高级 RNN 模型相继被提出也开始逐渐应用在机械故障诊断方面。Yuan 等人将常规 RNN，LSTM 和 GRU 三种结构结合使用用在飞行器引擎的故障诊断与运行监测方面^[67]。其中一个比较有趣的发现是，上述三种模型的集成并不能显著的增强 LSTM 的性能。赵等人设计了一种更为复杂的深度神经网络，将 CNN 和 LSTM 结合利用它们的强特征提取和记忆存储能力实现网络的更强性能^[68]。文献^[69]中，另外一些研究人员将一些手工设计的特征和网络自动提取的特征组合实现对机械运行工况的监测。Malhotra 等人提出了一种有趣的结构用于设备的剩余生命周期的预测^[70]。他们设计的 LSTM 吸收了自编码器中的编码——解码部分组件，并且在预测任务中取得了超

越的单一结构的效果。

1.5 本文主要研究内容

1.5.1 本文的整体架构

本文的整体结构如图 1-7 所示。第一章绪论部分介绍了风电机组相关的研究背景及意义；第二章部分对深度迁移学习相关的理论进行了介绍和分析；第三章部分进行了齿轮箱单齿故障分类模式识别与性能评估；第四章主要的研究内容是基于层次网络超参数寻优方法的齿轮箱行星轮系齿轮故障识别；第五章提出了齿轮箱滚动轴承稀疏特征提取与迁移故障诊断方法。其中三四两章主要研究对象是齿轮，而第五章的主要研究对象是滚动轴承。最后，第六章是对所有工作的总结和未来工作展望。

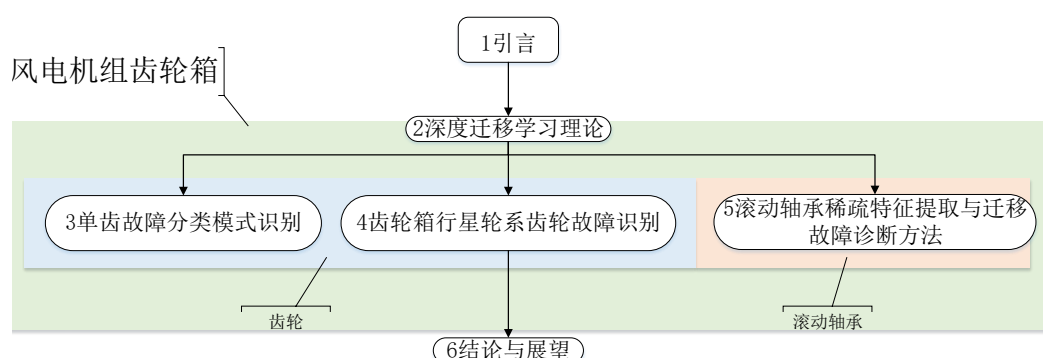


图 1-7 章节内容的结构与层次关系
Figure 1-7 Structure and hierarchy of each chapter

1.5.2 主要研究内容和研究方法

本文研究内容对象为风电机组齿轮箱，使用的主要方法为深度神经网络及其迁移方法，试验分析中使用的数据来源分别是虚拟仿真软件产生的按照节点分布的数据、风电场风机的长期监测和搭建的故障模拟试验台数据。包括以下几个方面：

(1) 数据采集和相关处理。虚拟仿真实体产生的数据主要来源是通过三维建模软件，给与一定的材料特性，通过有限元数值分析算法，获得理想状态下的振动信号。是对实验室环境中无法实现的故障类型的相关数据进行补充。已经获得的相关齿轮箱数据，例如断齿和点蚀，通过模型分析得到的数据和实验测量得到的数据共

同利用，实现更加真实和合理的数据分析过程。同时注意到一点，无论是来源于仿真模拟或者实验直接测量，数据源的数据都是多维的。因此，考虑利用比较高维度的数据也是一个能提高深度神经网络分类能力的一种思路。

(2) 数据特征的选择与提取。总体上关于特征的相关研究主要三个特点，第一类专家利用专家知识来设计特征，第二类综合利用多维特征并且由专家选择比较显著的特征，第三类直接利用神经网络来进行特征提取和选择。将所获得的时间序列进行分片处理，利用具有强特征提取能力的网络自编码器(AE)进行特征提取，然后再利用任一种弱分类器模型进行最后的类别判定。

(3) 深度神经网络的构建、调整与效率评估。当前的深度神经网络的复杂度都比较高，进行神经网络训练和测试所消耗的资源比较大。尽管 GPU 的利用可以加速网络的训练过程，但是仍然存在比较大的浪费。所以构建结构较浅的网络或者将现存的网络结构合理的改造，可以改善这个问题。同时，用于故障诊断的神经网络在训练过程中拟合情况的判断也是一个问题。依据经验判断是许多研究采用的办法，但这不是一种定量的比较容易利用的方法，所以希望找出一种能判断神经网络的状态的指标。

(4) 基于压缩感知的信号稀疏特征提取与不同零部件之间基于深度迁移学习方法的故障诊断。设计了一种针对滚动轴承故障和齿轮故障信号的卷积网络，通过加入注意力结构，提高用于故障诊断的网络结构的准确率和识别速度。将超参数进行随机搜索，在随机参数的情形下比较深度迁移学习两种策略的性能优劣。

1.5.3 技术路线

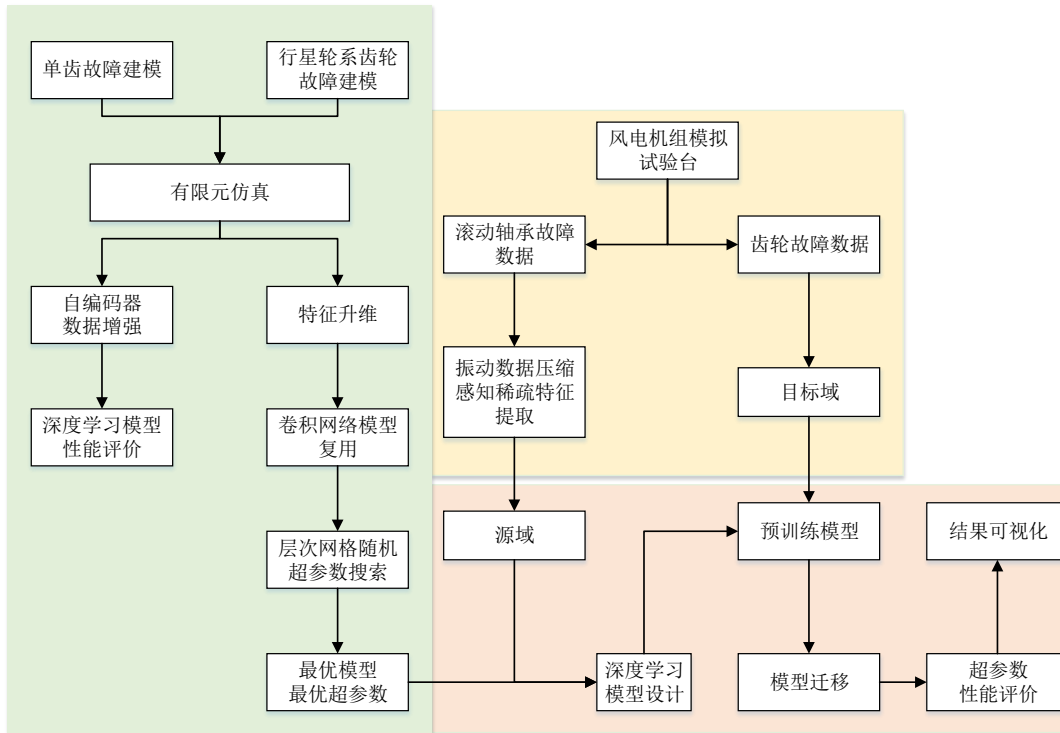


图 1-8 本文技术路线

Figure 1-8 Roadmap in this article

图 1-8 是本文的技术路线。从研究的背景与方法角度，结合具体的实验设备和数据。从仿真到具体的实验数据，从齿轮箱中的滚动轴承和齿轮系的分析，进行风电机组机械故障诊断的方法，提出本论文的研究路线。希望其能为新的故障诊断方法提供新的角度和思考路径。

1.6 小结

本章从多种研究方法的角度针对风电机组故障诊断的背景和意义进行了阐述。对风电机组故障诊断中的研究方法发展与不足进行了分析，在已有的相关工作基础上确定了本论文的研究主要内容。此外，还确定了本文的研究架构和研究技术路线，对后续文章的进行了连接。

2 深度迁移学习理论

深度学习理论的核心结构是深度神经网络。与传统的 BP 网络的隐藏层的个数和节点数目较少,对于数据的表示能力较弱相比,深度神经网络的层次与节点更多,对于数据的特征提取能力和表示能力更强。除此之外,在激活函数、连接结构、优化算法与超参数选取等几个方面,深度神经网络的研究进展也在持续推进。

2.1 深度学习中神经网络的基本结构

深度神经网络中研究理论基础比较完备的是自编码器(AutoEncoder,AE)、卷积神经网络(Convolutional Neural Network,CNN)和循环神经网络(Recurrent Neural Network,RNN)。尽管网络的结构和计算方法始终处在不断的改进和创新过程中,当时经典的结构始终在发挥作用。从 2.1.1 到 2.1.3 的三个小节,简要介绍三种主要结构。

2.1.1 自编码网络

以神经元之间的全连接为基础单元的自编码网络,同样具有一般编码器最常见的特点。如图 2-1 所示,由对称的编码器和解码器组成,且编码器和解码器的网络结构具有对称性。编码器和解码器部件之间的连接部分是将原始样本进行多层特征提取之后的与样本集的高斯分布进行比较,进而将潜在空间的分布再次解码。反复重复上述的编码解码过程,最后无监督学习的结果是重构的样本与原始样本基本达到相同的分布,同时潜在的样本空间也接近于高斯分布。

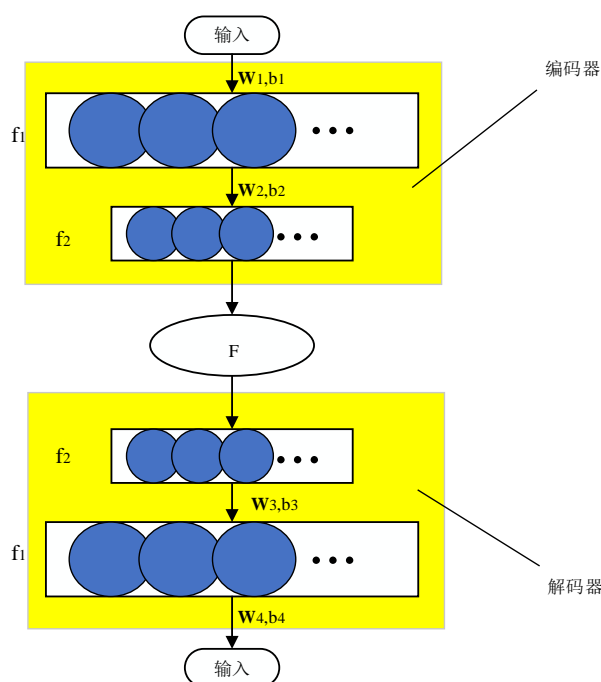


图 2-1 自编码器的基本结构
Figure 2-1 Autoencoder basic construct

事实上这样的网络结构其所能表达的样本特征非常有限，但是因为编码器与解码器之间存在的样本潜在空间满足特定的分布 F （通常情况下是高斯分布），样本的多样性并没有被改变，只是对样本的分布进行了调整，使其更加接近正态分布。自然中的大量数据均满足的一个天然分布就是高斯分布，这并不是一个偶然。尽管缺乏理论来说明样本的自然分布为什么和正态分布如此接近，然而无数的事实证明其存在着某些相似性或者共同特征。图 2-1 所展示的网络结构自顶而下的计算表达式可以用式(2-1)来表达。神经网络的权值更新过程包括前向传播的权重调节和误差方向传播机制。式(2-1)这样一个迭代的公式表达了自编码网络具有的基本结构，与一般的神经网络的主要不同点是样本空间到样本特征的潜在空间的映射和特征潜在空间到最后输出的重构样本空间的映射互为逆映射。

$$y = f_1 \left(\underbrace{W_4^T f_2 \left(W_3^T f_2 \left(\overbrace{W_2^T f_1 \left(W_1^T x + b_1 \right) + b_2}_{f} \right) + b_3 \right)}_{f^{-1}} + b_4 \right) \quad (2-1)$$

从式(2-1)同时可以看到神经网络的另一个特点，多重迭代的高维非线性结构。网络的表达能力的强弱表现与该特点在某种程度上存在关联。自编码网络由原始样本空间、编码器层、潜在样本空间、解码器层与重构样本空间构成。

$$\begin{aligned}
 f &: X \rightarrow F \\
 f^{-1} &: F \rightarrow \bar{X} \\
 f, f^{-1} &= \arg \min_{f, f^{-1}} \|\bar{X} - f^{-1}(f(X))\|^2
 \end{aligned} \tag{2-2}$$

如果将原始样本空间到潜在样本空间的映射定义为编码过程，而由潜在样本空间到重构样本空间的映射定义为解码过程，那么自编码网络的目标就是在寻找最合适的编码和解码映射，使原始样本和重构样本之间的分布差异最小。以数学语言表述如式(2-2)。从自编码网络的最终表达式，可以观察到神经网络的设计和训练过程本质上是构造一个非线性迭代函数和其针对其最优化的求解过程。深度学习中利用的神经网络中的各种变形和改进通常也是围绕着这两个方面来进行。

2.1.2 卷积神经网络

卷积神经网络与传统的 BP 网络的一个根本区别是引入了针对特征图的卷积算子和池化算子。卷积算子在信号处理领域是非常常见的算子，然而神经网络结构中的卷积计算尽管名字与其相同，但实际上含义并不完全相同。

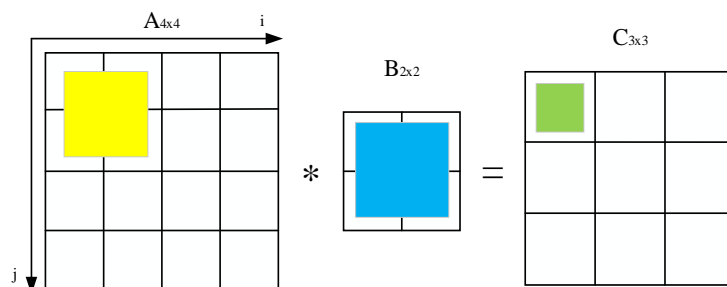


图 2-2 卷积算子的计算方法
Figure 2-2 Calculation method on convolution operator

考虑一个特征图 $A(4 \times 4)$ 的大小，将一个卷积算子 $B(2 \times 2)$ 与其进行卷积计算，在步长为 1 且恰好特征图与卷积算子均为方的（长和宽相等）情形下，卷积结果是 $C(3 \times 3)$ 。卷积结果的大小计算， $4 - 2 * 1 + 1 = 3$ 。所以实际上卷积神经网络中的卷积运算与信号处理中的相关运算从计算的角度来说是等价的。此外，经历过卷积运算的特征图的大小是降低的。换句话说，提取高维特征的代价是将低维特征进行了融合并且适度抛弃一些特征。经历过卷积的特征图不仅仅是维度上出现了缩减，与同等数量的全连接网络相比，更是表现出了一种等间隔采样的特性。池化算子同样是一种减小特征图尺寸的算子，比较特别的是池化算子没有任何参数，因此也就没有任何计算量可言。图 2-3 中的(a)和(b)分别是在池化算子的尺寸是 2×2 和 3×3 的情形下运

算结果的表示。可见，不同的池化算子在相同的特征图上应用有可能产生相同的新的特征图。与卷积算子相同，池化算子也是一种维度缩小的计算方法。池化算子的概念和计算方法与信号处理中的滤波方法非常相似，尤其是均值池化与中值滤波，完全是相同的概念和计算方法。因此，在一定程度上来说，卷积神经网络中的两种关键算子——卷积算子和池化算子，均来源于信号处理领域。从这个角度，也可以将卷积神经网络视作一系列滤波器的结合，其目标是逐步寻找样本的高维特征表示。

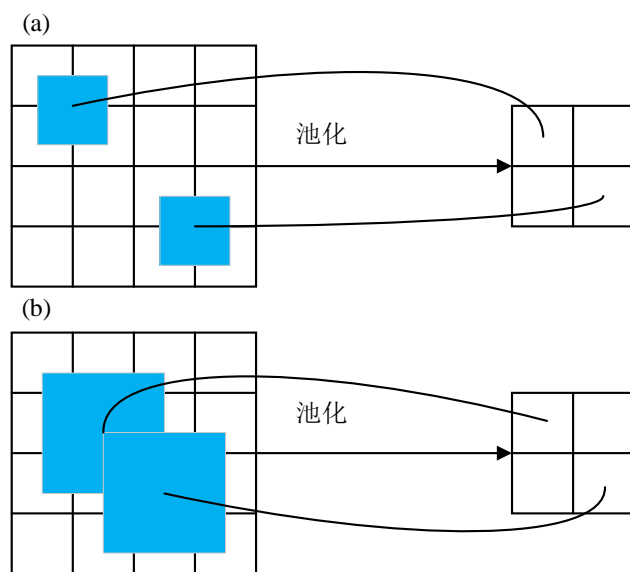


图 2-3 不同大小的池化算子计算结果

Figure 2-3 Calculation results of pooling operators with different sizes

卷积网络的一般由卷积层、池化层、全连接层全部或者其中的几层组成。通常的卷积神经网络的参数还是大量集中在全连接部分。卷积层和池化层的相关操作已经有很多替代方法，例如 attention 机制及其变种和 transformer 组合等等结构。关于卷积和池化的计算观点，学术界没有一个统一的共识，其机制并没有详细的理解。尽管新的机制如 attention 可以在计算上取代卷积和池化，但对其实际作用机制的解释仍旧不统一。对于特征图的缩放和高维特征提取能力的辨析也还在改造和升级中。卷积神经网络的输入以二维数据为主，少数用来进行 1D 卷积网络的研究和应用。尤其在计算机视觉领域，二维卷积神经的应用已经逐渐成为主流。

2.1.3 循环神经网络

循环神经网络最早的设计来源于时序的 BP 神经网络，均是用来处理输入样本的特征是序列的数据。循环神经网络的主要特点是当前的状态和输出，依赖于上一个时间点的状态和输入。从贝叶斯概率的角度思考，默认先验信息对未知信息

存在影响和有一定的联系。简而言之，在回归任务时承认经验的积累的作用。单一神经元的时序网络的模型如图 2-4 所示，其计算公式和更新规则如式(2-3)。

$$a_t = \tanh(U * x_t + b_{ih} + w * a_{t-1} + b_{hh}) \quad (2-3)$$

需要注意的是，公式(7)中的 b_{ih} 和 b_{hh} 分别代表了输入偏置项和隐藏层偏置项。另外，图 2-4 所展示的单神经元模型是仅有一层隐藏层的模型，所以其计算规则仅有一层的迭代模式，相对来说计算模式比较简单。在多数循环网络模型的应用中，单隐藏层的结构并不常见。多个隐藏层多个输出的结构扩展了网络的表达能力，对计算能力的要求也更高。

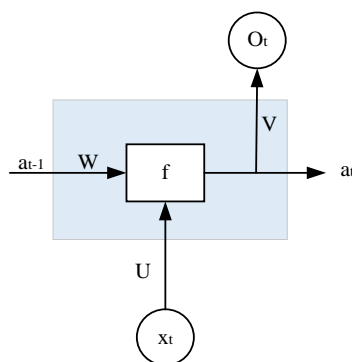


图 2-4 循环神经网络的单神经元模型

Figure 2-4 Recurrent neural network single neuron model

多个单隐藏层神经元结构的叠加组合成为复合的时序神经网络。随之而来的问题就是网络复杂性的提高带来的训练一个时序网络所需的计算力的提高和训练过程中出现梯度消失的可能性增大，继而带来的网络的不收敛现象。不管是卷积神经网络还是循环神经网络，梯度爆炸的现象都有可能发生。只是在循环神经网络中，由于时间步长的选择通常间隔不会很大，就导致了不收敛问题发生的可能性变得很大。选取更合适的激活函数或者将同一个批次的数据进行正则化处理都可以减少不收敛现象的发生。然而想要从根本上解决这个问题，必须从网络结构本身出发。

长短期记忆网络(LSTM)模型的提出就是为了从结构上解决梯度消失的问题。其结构特点是非常巧妙的设计了循环体架构。与典型的 RNN 网络相比,LSTM 模型摆脱了对长期状态的依赖，对于梯度问题进行了改进。

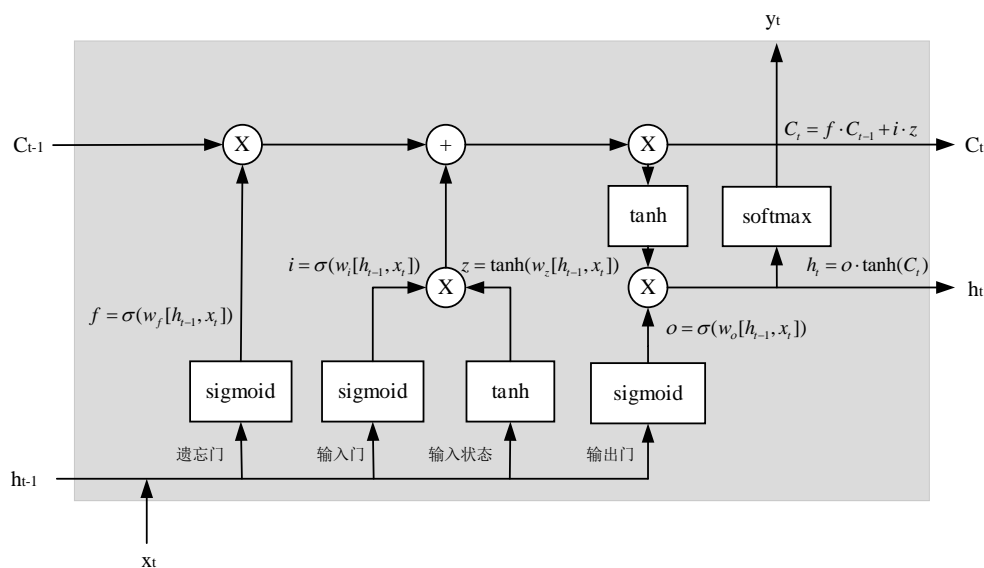


图 2-5 长短期记忆网络的基本结构
Figure 2-5 LSTM basic structure

尽管 LSTM 模型相对于 RNN 模型来说有相当多的改进，但仍然存在可以进一步改进的地方。基于 LSTM 的基本结构，研究人员提出了 GRU 结构，如图 2-6 所示。其结构相对于 LSTM 结构来说进一步得到了简化，去掉了一个门极可控单元，提高了对计算资源的利用效率。并且在输出端口上也做了简化，将时序步单元状态和输出合并为单一输出状态。

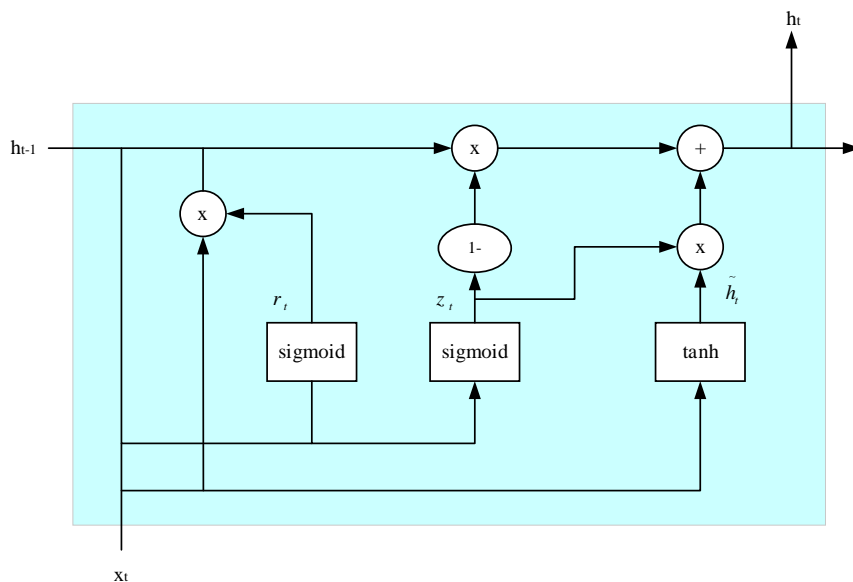


图 2-6 门极可控单元网络的基本结构
Figure 2-6 GRU basic structure

GRU 门极可控单元网络的基本结构其对应的计算方法和更新如式(2-4)，从计算结构上来看确实减少了一部分的计算量：其中 z_t 和 r_t 分别代表了更新门控结构和重

置门控结构，最基本的运算仍旧是矩阵乘法和向量的点乘。

$$\begin{aligned}
 z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\
 r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\
 \tilde{h}_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]) \\
 h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t
 \end{aligned} \tag{2-4}$$

从 RNN 到 LSTM 和 GRU 的改进使得针对序列的神经网络计算越来越流行。并且在应用上的扩展也越来越广泛。从时间序列到空间序列，从早期的手动特征设计到现在有直接使用原始数据的相关研究出现。目前在语音识别领域的应用已经比较成熟，并且已经产生了相关的产品和应用软件。最常见的输入数据和输出数据均由序列组成，这也是序列网络最常见的应用场景之一。而输入数据为序列和输出数据为类别的应用以典型的情绪识别为代表。

2.1.4 基本结构在分类任务与回归任务应用上的适用性

回归任务和分类任务在使用神经网络进行处理与相关数据集的构造既有不同点也存在一些共性。深度神经网络的能力体现主要是无限的拟合能力。回归任务和分类任务区别在于，回归任务的拟合目标是特定的数值而分类的主要目标是数值标定的分布。以非学术化的语言来说，回归要求固定的目标而分类只对目标的分布存在要求。从统计理论的角度来说，分类任务的要求更容易达到，因为是针对整个样本集而言达到某几个样本分布的要求远比满足特例的要求低。若从优化计算的计算量上进行比较，显然追求目标集的分布要比目标集的特例要容易的多。卷积神经网络和循环神经网络的输入样本集的存储结构和中间运算使用了不同的算子，因而他们从不同的低维样本中提取高维特征的能力也有着显著的差异。

2.2 深度学习理论中的迁移方法

迁移学习或者迁移方法的提出主要是用来解决训练数据不足的问题。其主要思想是在训练数据和测试数据满足分布独立的假设下，将在训练数据下也就是源域中学习到领域知识迁移至目标域^[71]。这样的想法拥有实际的工程应用背景，在很多领域中获得足够的训练数据是一个非常苛刻的条件^[72]。或者数据采集的困难，或者采集到数据收到“污染”的可能性很大，甚至数据在运行环境中根本就无法进行合理

的采集。

在进行完整的迁移方法介绍之前，需要对一些使用到的概念进行定义和解释。其中，最主要的概念是域和任务。域的表达用公式 $D:=\{X,P(X)\}$ 来定义。其中 X 表示样本的特征空间, $P(X)$ 表示样本集的边缘概率分布。样本集 $X:=\{x_1,x_2,\dots,x_n\}$ 的概念是域中样本的无序集合。一个需要进行训练或者测试的任务由公式 $T:=\{Y,f(x)\}$ 。同样，它是由样本对应的标签空间 Y 和目标预测函数 $f(x)$ 构成。从概率统计的角度理解， $f(x)$ 也可以视作是条件概率分布函数 $P(y|x)$ 的另一种表达。有了上述的几个定义的铺垫和准备，终于可以给出迁移理论一个相对准确的定义。

给定一个学习任务基于目标域 D_t 的待完成任务 T_t , 需要我们从源域 D_s 上训练完成的任务 T_s 上获得性能改进。在源域与目标域不同，或者源任务和目标任务 T_t 不同的情况下，迁移学习的目标是从源域和源任务获得知识将其迁移到目标任务，并且改善在目标任务上预测函数的性能。此外，一般情况下源域的数据量远远大于目标域的数据量。图 2-7 所示流程对一般的迁移学习进行了表达。

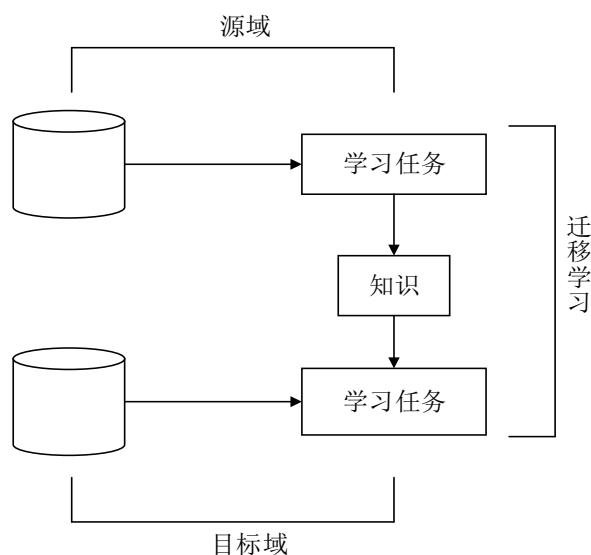


图 2-7 迁移学习基本方法

Figure 2-7 Transfer-learning basic method

迁移学习方法从大类上可以分为两类。一类是关于数据集分类即源域与目标域，或者新的数据集之间的关系运算为主，另一类的主要特点是用于迁移的神经网络结构的变动，例如特征提取网络的直接迁移和深度网络的部分结构移植^[73]。当前主要的研究热点集中在域适配和多源域迁移等方向^[74]。深度迁移的定义可以由迁移学习理论进一步延伸。深度迁移学习的理论可以视作由 $\langle D_s, T_s, D_t, T_t, f_t(X) \rangle$ 组成的五元组，其中 $f_t(X)$ 表示由一个深度神经网络表达的非线性映射^[75]。从 2.2.1 至 2.2.3 个小节，分别对基于样本集、特征图和网络结构的迁移方法进行理论解释。

2.2.1 基于样本集的迁移特性

基于样本集的迁移方法利用特定的权重适应方法，从源域中选择部分样本作为源域训练集的补充，通过赋予给这部分样本合适的权重。尽管源域和目标域的样本的分布不同，但是部分源域中的样本可以通过进行一定的权重调整在目标域中使用^[76]。基于样本集的迁移学习方法如图 2-8 所示。淡黄色表示的源域样本和浅蓝色表示的目标域样本存在一定的交集，同时源域样本和目标域样本的分布明显不同。

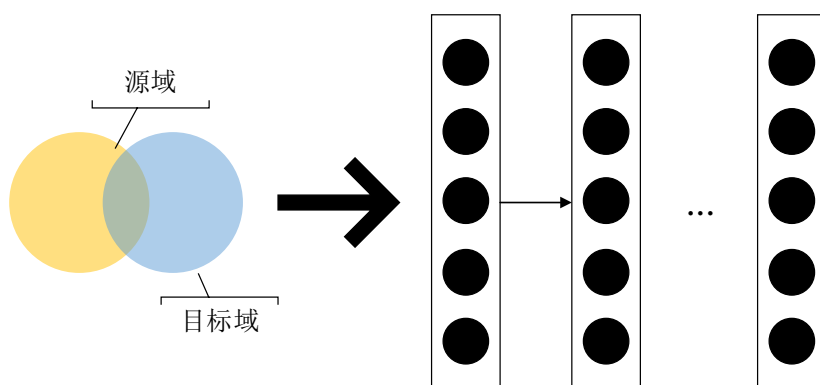


图 2-8 基于样本集的迁移方法示意图

Figure 2-8 Instances-based deep transfer learning

2.2.2 基于特征图的迁移特性

基于特征图的深度迁移学习方法指的是将源域和目标域中的样本数据组成新的数据空间。在新的数据空间中，两个域中的样本是相似的并且易于联合成新的深度神经网络，如图 2-9 所示。基于两个原始域中的数据不同，然而在新的数据域中数据分布的相似度更高这样一个假设，特征图迁移方法产生了正向迁移的影响。将源域和目标域中数据共同组成新的数据空间，作为神经网络的训练集^[77]。

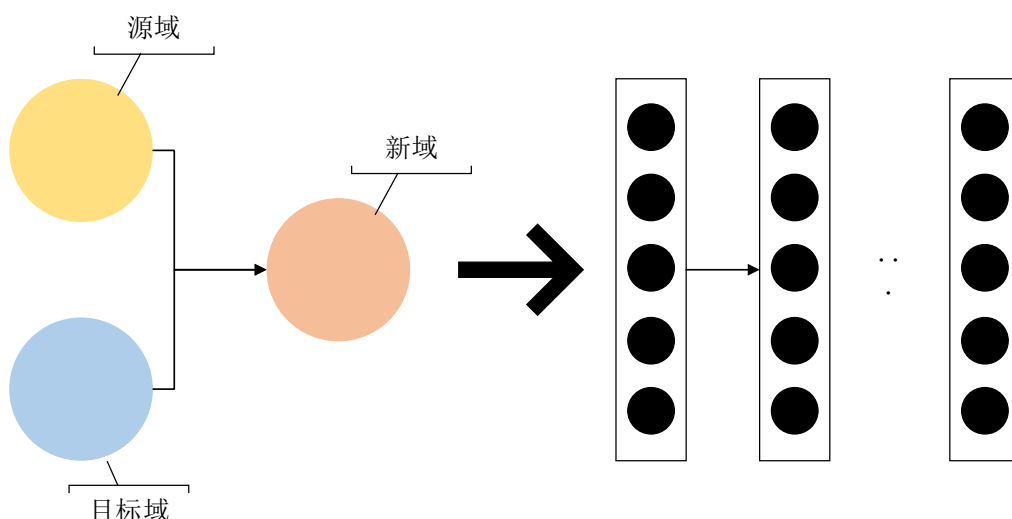


图 2-9 基于特征图的迁移方法示意图
Figure 2-9 Mapping-based deep transfer learnin

迁移成分分析(TCA)和许多基于 TCA 方法的应用在传统的迁移领域已经非常流行。一个比较自然的想法是将 TCA 方法应用在深度神经网络上。此外还有一种方法是将 MMD 方法扩展用于比较不同的样本特征分布。通过引入一个自适应层和联合域损失函数来进行表示学习探索域的不相关性。MMD 距离的定义如下:

$$D_{MMD}(X_s, X_t) = \left\| \frac{1}{|X_s|} \sum_{x_s \in X_s} \phi(x_s) - \frac{1}{|X_t|} \sum_{x_t \in X_t} \phi(x_t) \right\| \quad (2-5)$$

其对应的损失函数由式(2-6)表达:

$$L = L_c(X_L, y) + \lambda D_{MMD}^2(X_s, X_D) \quad (2-6)$$

其中, 隐藏层相关的学习任务在卷积神经网络中组合用来进行核希尔伯特空间表达。不同域之间的空间距离通过多核优化方法进行寻优。这种空间距离在不存在唯一的设定, 函数空间中常见的几种表示均可以用来进行域之间距离的估计。

2.2.3 以网络结构为主的迁移理论

基于网络结构的迁移学习指的是复用利用源域中的样本集进行训练得到其预训练模型, 并将预训练模型的架构或者结构参数迁移到针对目标域设计的深度神经网络模型。神经网络的学习与运算机制与人类的大脑相似, 存在更新与进化等连续迭代等抽象过程。从源域中迁移至新的网络结构的部分被视作是一个特征提取的网络, 该网络的作用是对样本进行了高级的特征表示。图 2-10 对以网络结构为主的迁移方法进行了图形化展示。

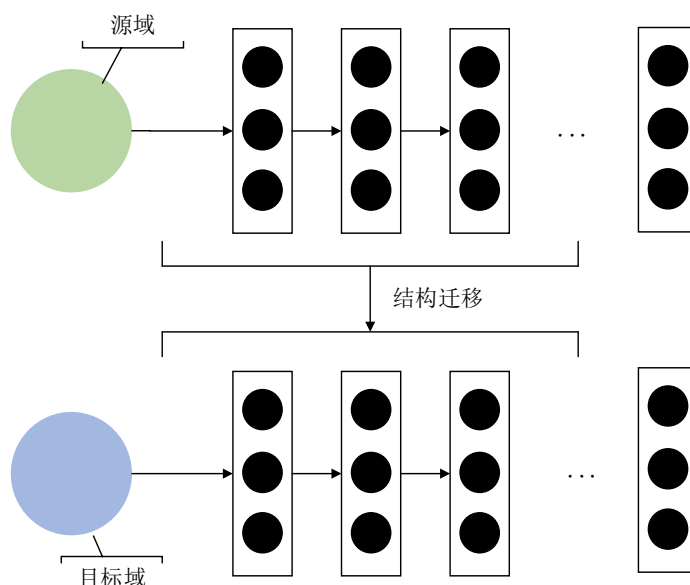


图 2-10 以网络结构为主的迁移方法示意图
Figure 2-10 Network based deep transfer learning

针对目标域的新结构包括两部分，一部分是从源域进行训练得到的特征提取网络，另一部分是对提取的特征进行分类的分类器部分。对于卷积神经网络来说，将预训练模型在已知的数据集上直接进行特征提取或者在标准数据集上进行再训练，用在其他的目标检测或者图像视频分类上均有比较高效和优异的性能表现。此外，一些更有价值的观点和创新性的方法也在相继研究中。网络结构与网络的可迁移性质也对相关的研究有促进作用。其中有些 CNN 中的经典模块的迁移对性能的影响具有不确定性，同时其迁移能力也受到影响。从不同的预训练模型中迁移过程获得的结构具有不同的特征提取能力，样本的高层表示也截然不同。

2.3 小结

本章的主要内容包括两部分，其一是深度学习理论中基本网络结构与方法的总结和概述，应用在回归和分类任务的不同特点和区别。其二是基于深度学习的主流迁移方法概括和分析。迁移理论的研究方法集中在数据集、特征维度与网络结构三个角度。本章是后续章节的理论基础。

3 单齿故障分类模式识别

利用神经网络进行故障诊断在机械故障诊断领域已经有很广泛的应用。然而现场的故障数据的采集一直是一个不能被忽视的问题，利用有限元分析进行机械故障诊断中振动数据的模拟是一个思路。同时，考虑到利用组合优化的方法来评估神经网络用于机械故障诊断的性能是自然的。尝试对采取的分类器进行能力的度量，以方便对网络结构的设计和优化。第 3 章的研究示意图如图 3-1 所示。

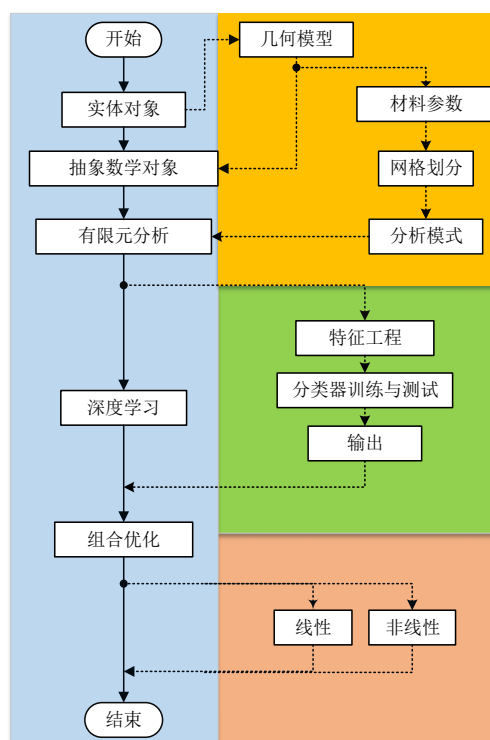


图 3-1 研究结构的流程图

Figure 3-1 The flowchart of research structure

3.1 单齿故障的 3D 模型的设计与构建

齿轮故障发生故障的情况在实际的风电机组增速齿轮箱中是非常常见的。根据实际故障发生时的记录进行模拟，或者说通过数字化的手段建立相对真实的模型来

模拟现实生活中真实发生的各种故障。

表 3-1 二维单齿平面几何尺寸
Table 3-1 Two-dimensional single tooth plane geometry size

线段	长度
AB	2.2m
BC	1.5m
CD	0.9m
EF	0.4m

考虑图 3-2(a)中展示的标准直齿轮二维平面图形，一个完整的 3D 立体单齿模型可以通过将该图形进行纵向方向的拉伸得到。图 3-2 中的模型的尺寸参数按照表 3-1 的比例进行配置。其中， m 是齿轮的模数。所以实际上表格中的尺寸是相对的长度。在论文中该参数取值为 2.5mm。

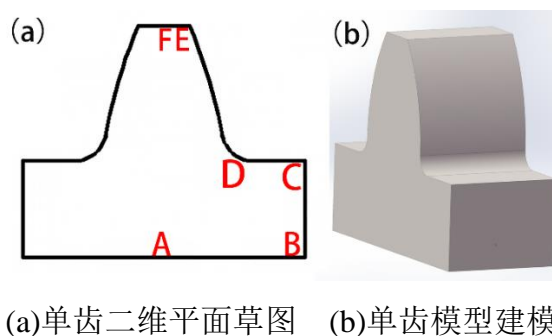
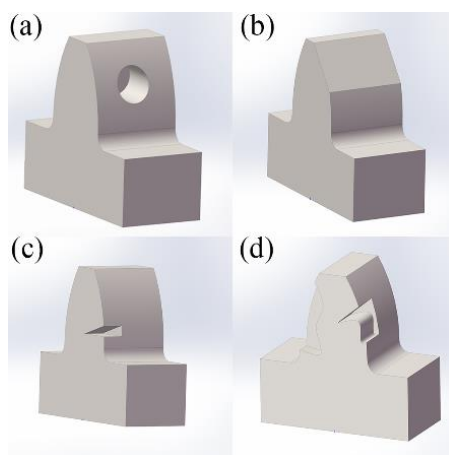


图 3-2 单齿建模过程
Figure 3-2 Single gear tooth modeling process

将该平面以深度为 5mm，方向为垂直该平面方向进行拉伸，就得到了一个三维局部轮齿的几何模型。至此，单轮齿的几何模型就创建完成了。利用建模软件进行设计，依据上述尺寸进行 3D 几何建模，做出的单齿模型如图 3-2(b)所示。利用 3D 建模软件在模型上进行一系列的打孔与切割等操作，在一定程度上可以模拟实际工况中的轮齿可能出现的损伤情况。图 3-3 中的四个子图(a)、点蚀(b)、尖端断裂、(c)裂纹与(d)磨损分别模拟了四种不同类型的故障。其中，磨损这种故障的几何模型的构建比较复杂，理论上应该利用随机性来产生相应的曲线进行拉伸，然而发现这些软件对于产生随机的磨损这种情况并没有相应的解决方法。最终，通过两个三次样条函数，在草图上随机取点并且任意拉伸一定的深度来近似，勉强达到可以替代的程度。



(a)点蚀 (b)尖端断裂 (c)裂纹 (d)磨损

图 3-3 四种类型的齿轮故障

Figure 3-3 Four types of gear fault

在几何模型确立之后，下面进一步确立轮齿的材料相关参数。文中选择了比较常见的一种材料普通碳钢，在表 3-2 中重点考虑下列三个属性：弹性模量、泊松比和质量密度。至此，建立了故障轮齿的局部几何模型，并且定义了材料参数。在下一个小节，将利用有限元这一数学工具对轮齿进行非常详细的分析。

表 3-2 普通碳钢相关材料参数

Table 3-2 Common carbon steel related parameters

材料特征	数值	单位
弹性模量	2.2e+11	N/m ²
泊松比	0.28	/
密度	7800	kg/m ³

3.2 利用有限元分析及仿真产生数据集

针对上述第 3.1 节的建立的轮齿局部故障的微小模型，在不考虑外部温度等力学之外的参数情况下，分别进行结构动力学中的模态分析和谐响应分析。其中，结构的模态分析(固有频率与振型)将是所有振动分析的基础。描述结构动力学特征的基本力学变量和方程与静力问题类似，但增加了惯性力项和阻尼力项，且所有的变量都将随时间而变化。系统的整体有限元方程，即

$$M\ddot{q}_t + C\dot{q}_t + Kq_t = P_t \quad (3-1)$$

其中，

$$\left. \begin{aligned} q_t &= \sum_{e=1}^n q_t^e, P_t = \sum_{e=1}^n P_t^e \\ M &= \sum_{e=1}^n M^e, C = \sum_{e=1}^n C^e, K = \sum_{e=1}^n K^e \end{aligned} \right\} \quad (3-2)$$

n 是单元的数量, 上标字母含有 e 的参数是系统划分单元的单节点元素。考虑其中的一种无阻尼自由振动情形, 即 $C = 0$ 与 $P_t = 0$ 的情况。将方程(3-1)转化为(3-3),

$$M\ddot{q}_t + Kq_t = 0 \quad (3-3)$$

这种振动形式叫做自由振动, 解的形式如下(14),

$$q_t = \hat{q} \cdot e^{i\omega t} \quad (3-4)$$

将式(3-4)代入(3-3)中, 得到(3-5)。

$$(K - \omega^2 M)\hat{q} = 0 \quad (3-5)$$

式(3-5)有非零解的条件是

$$|K - \omega^2 M| = 0 \quad (3-6)$$

式(3-6)就是无阻尼自由振动方程的特征方程, ω 是自然频率(rad/sec), 即圆频率, 对应的频率是 $f = \omega/2\pi$ (Hz)。求解得到圆频率之后, 将其带入到式(3-5)中, 得到对应的特征向量 \hat{q} , 这就是振动频率 ω 对应的振型。以上的分析基于以下的假定: 参数 K 与 M 都是常量; 假设所选择的材料是线弹性材料特性; 仅使用小挠度特性, 即不考虑非线性特性; $C=0$, 即不考虑阻尼; $P_t=0$, 即作结构不存在外部激励的假设。另外, 利用软件进行有限元分析计算时, 根据其物理方程, 结构仅在完全不受约束或者部分受约束的情况下得到的解是合理的解。

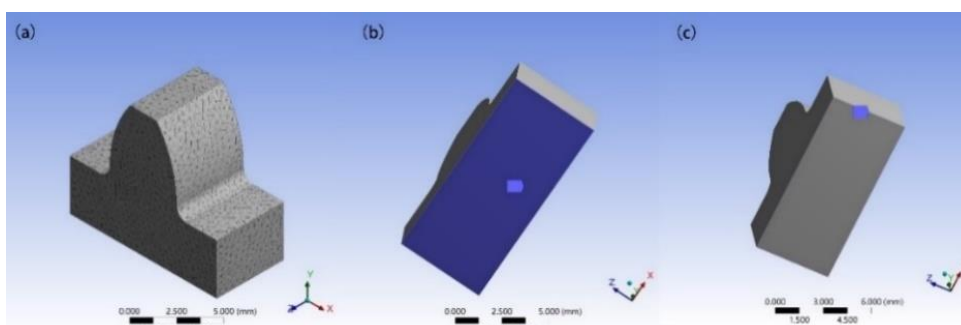
3.2.1 模态分析

通常来说, 典型的有限元分析过程包括以下的环节: 几何建模、材料定义、网格划分、载荷与边界条件的设置以及对最终结果的后处理, 如图 3-4 所示。几何建模和材料定义在 3.1 节中已经进行了详细的说明, 所以在这个小节中不在重复说明。简要的对网格划分和边界约束进行解释, 并且对模态分析得到的结果加以说明。



图 3-4 模态分析的流程
Figure 3-4 Modal analysis flow chart

由于前面建立的模型相对来说，几何结构并不复杂且不存在明显的不连续的情形，所以并没有进行特定的网格划分过程。利用软件的网格自动划分功能，进行了自适应的划分方法选取，同时将网格密度设置到精密的状态。载荷设置与边界条件定义的部分，将模型的底部的四个边线和一个底面固定。这种设置是为了模拟轮齿在实际的工况过程中，内部的响应几乎不受到影响或者完全受到微弱的载荷作用的情形。所以，该部分分析是整体系统处在部分受约束的情形下进行的。以完整的轮齿模型作为例子，网格划分的结果和边界约束的范围标记在图 3-5 中。前述的四种故障模型也分别进行相同的操作，后续将进行相应的数据预处理过程。



(a)网格划分 (b)与(c) 面约束与四个边线约束

图 3-5 网格划分与边界约束

Figure 3-5 Meshing and boundary constraints

在软件中设置前十阶模态作为求解器求解的范围，通过 GPU 与 CPU 协同计算得到固有频率的数值解。在图 3-6 中可以发现，完整的轮齿与故障的模型之间的频率差异并不是很大。模型的尺寸是按照实际使用中的大小 1:1 制作的，属于比较微小的模型，所以这种现象是正常的。因此，从固有频率的角度很难进行轮齿故障模式的判别与分析。从总应变数据的角度出发进行分析，可能是一种分辨轮齿故障模式的切入点。

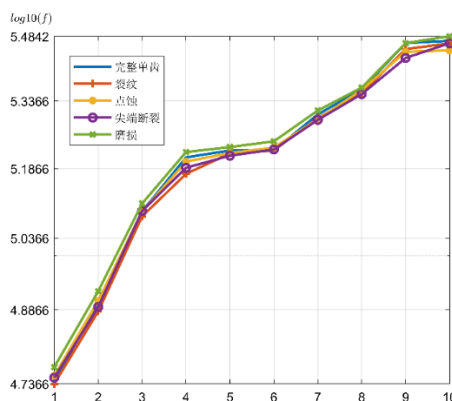


图 3-6 模态分析获得的前十阶频率

Figure 3-6 First tenth order frequency obtained by modal analysis

从完整轮齿与四种故障模式的模态分解中，获得模态分析合成应变数据点数与模态如表 3-3 所示。去除所有约束的数据点数一致，很方便进行后续的进一步利用神经网络分类。

表 3-3 总应变数据
Table 3-3 Total strain data(Datasets A)

	原始数据	预处理数据
完整单齿	26744×10	25000×10
裂纹	27811×10	25000×10
点蚀	28039×10	25000×10
尖端断裂	29444×10	25000×10
磨损	26744×10	25000×10

3.2.2 谐波响应分析

对于运行工况中的轮齿来说，通常来说，会承受周期性的冲击力这种负载。在现实的物理世界中很难满足这种条件，因此考虑一种理想的实验环境。并且假定轮齿的正常运行过程中，随机产生的载荷和线性给定恒力与谐波响应过程可以剥离。式(3-1)作为动力学求解的基本方程，假设外部载荷 P_t 和位移 q_t 具有相同的频率 Ω ：

$$\{P_t\} = \{P_t e^{i\psi}\} e^{j\Omega t} = \{P_{t\max} (\cos \psi + i \sin \psi)\} e^{j\Omega t} = \{P_{t1} + iP_{t2}\} e^{j\Omega t} \quad (3-7)$$

$$\{q_t\} = \{q_t e^{i\psi}\} e^{j\Omega t} = \{q_{t\max} (\cos \psi + i \sin \psi)\} e^{j\Omega t} = \{q_{t1} + iq_{t2}\} e^{j\Omega t} \quad (3-8)$$

那么，位移 q_t 的一阶导数速度 V 与二阶导数加速度 A 分别是(3-9)和(3-10)。

$$V = \{\dot{q}_t\} = i\Omega \{q_{t1} + iq_{t2}\} e^{j\Omega t} \quad (3-9)$$

$$A = \{\ddot{q}_t\} = -\Omega^2 \{q_{t1} + iq_{t2}\} e^{j\Omega t} \quad (3-10)$$

本小节的谐波响应分析是在 3.2.1 小节的模态分析的基础上进行的，在轮齿的侧面选定一个侧曲面，施加 50000Mpa 的均匀载荷(如图 3-7 所示)，外部激励的频率设置为 20Hz 至 200Hz，间隔设置为 20Hz，总计 10 个频率响应分别求解其总位移、速度和加速度。文中选定的轮齿模型属于尺寸比较小的，在有限元方法的求解过程中的位移改变并不明显，不同的 10 个频率作用下的激励产生的总位移近乎于不存在。但是速度与加速度响应过程产生的数值解，每一个频率下产生的情况

可以观察到明显的差异性。

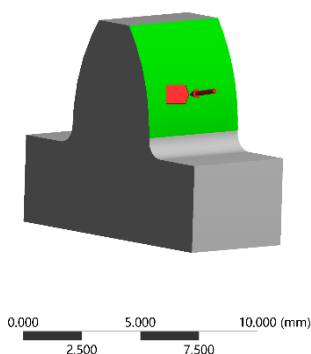


图 3-7 谐波响应外部均匀载荷的位置

Figure 3-7 Position of external uniform load applied by harmonic response

将过程中产生的所有数据整理如表 3-4，原始数据因为每个模型的几何构造不同，再加上都采用自适应网格划分的策略，其网格会根据几何结构自适应的进行或稠密或稀疏的调整。所以维度都是不同的。其中，含有大量的连续零值，这些值的出现是因为最初分析过程中边界条件的设定，但是在接下来的模式分析中并没有利用到这部分数据的特征。反而，多余的零值点会对判别的效果产生不利的影 响。因此，在初步的数据整理过程中，将其删除。此外，响应产生的位移数据之间的差异过小，这样的数据特征在判别算法利用数据的过程中，实际上也是无用的，有时甚至降低了判别算法的性能。出于上述的理由，最终经过一系列数据清洗和数据预处理之后的数据，具有相同的维度。

表 3-4 谐波响应总应变数据

Table 3-4 Data generated by harmonic response (Datasets B)

	原始数据	预处理数据
完整单齿	26660× 30	25000× 20
裂纹	27811× 30	25000× 20
点蚀	28039× 30	25000× 20
尖端断裂	29444× 30	25000× 20
磨损	26744× 30	25000× 20

3.3 深度神经网络进行分类和优化性能

3.3.1 单一深度神经网络分类器应用于单齿故障诊断

将第 3 小节的数据整理成数据集(Datasets A and Datasets B)，在本小节将利用这个数据集进行分类器的训练和测试。本文中轮齿的故障诊断问题，实际上是一个五分类问题。深度神经网络在这种问题上具有天然的优势。然而，针对上面小节的数据集，全连接的神经网络效果并不能达到令人满意的效果。尤其是，在数据集中的数据存在几个数量级的差异，会造成深度神经网络的效果极差，这种情况也是极其常见的。对于这个问题的处理方法是，减少针对数据的预处理过程，只剔除部分很明显对于分类来说是无效的数据，例如连续零值的部分。进行较少的数据预处理的好处是明显的，避免了数据处理造成的数据之间差异的特征消失，最大程度上保留了数据的原始特性。然而，完全不进行任何的数据预处理过程也是具有很大的风险的，常常导致深度神经网络的训练周期过长或完全发散，导致完全达不到理想的结果。深度自编码器的提出，对于解决这个问题有很大的启发意义。因此，尝试直接将其和全连接深度神经网络结合起来，形成一种新的模式。并且采取二次训练的方法对其进行利用，两个层次的网络，每一层次单独训练，第一层的训练结果作为第二的输入。这种方法使得相对复杂的神经网络在计算能力很普通的 PC 上也能进行训练和调试过程。对于模拟实验和数值仿真来说，大大降低了进行研究所必需的基本硬件条件。

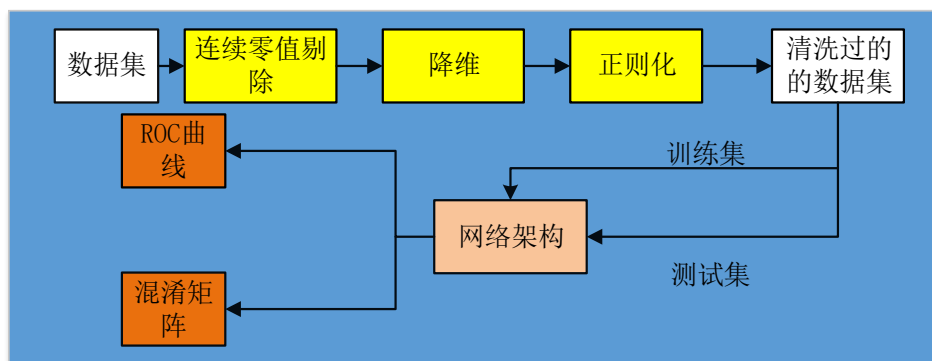
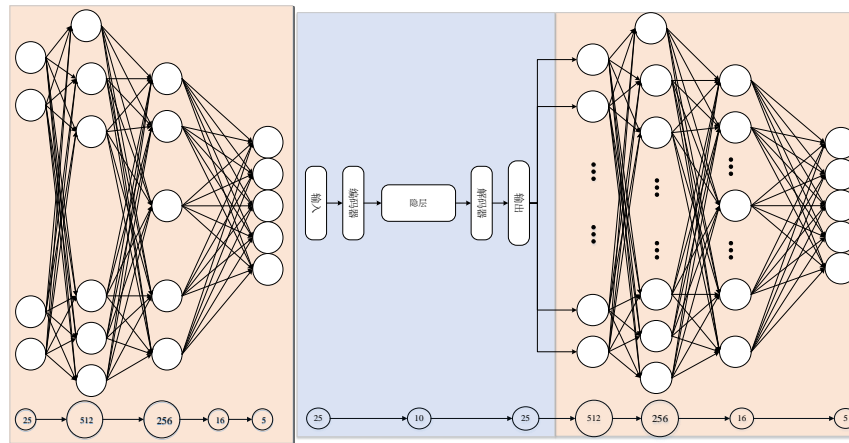


图 3-8 数据处理过程与深度神经网络作为分类器的设计
Figure 3-8 Date processing and DNN model designed as classifier

图 3-8 展示了一个深度神经网络用于故障数据的预测和分类的典型过程。利用有限元分析获取不同故障类型的数据，组织和形成数据集。进行一系列的数据预处理动作，主要的数据处理包括清除连续的零值、维度的降低和数据的正则化。经过上述的数据清洗过程之后，获得了相对比较“干净”的数据特征。用于神经网络的数据集通常依据一定的比例划分为训练集、验证集与测试集。在研究过程中，将这一比例选择为 75%，15%和 15%。原始数据经历降维处理之后的样本数量较大，所以数据集的比例的选择对于最终的神经网络的效果并不会有什么特别

的影响。

在我们的研究过程中，设计了两种不同的网络架构。图 3-9 展示了这两种架构的区别。子图(a)设计了一个全连接的深度神经网络，输入层和输出层的节点个数分别是 25 和 5，并且输出层采用的是 one-hot 编码(形如[0 0 1 0 0])，即一个输出向量中仅有一个非零值且标定为 1)。隐层激活函数选择最常见的逻辑函数(logistic function)，总用 3 层隐层，节点个数是 512-256-16。该设计并不考虑数据中的干扰部分，或者说不存在一个明显的滤波过程。子图(b)设计了一个深度自编码器作为数据滤波过程的结构。自编码器的结构是输入和输出拥有相同的节点个数 25，隐层节点个数设定是 10。将自编码器看作是滤波器，进行数据的第一遍清洗（也称为预训练），值得注意的是该过程要保证不能过拟合，容易丧失部分关键的数据特征。接下来的 FNN 网络基于自编码网络的训练结构进行二次训练，以此进行数据的预测和分类。这种结构的优越之处也是明显的，主要表现在两个地方。其一，增加了自编码过程用于数据的滤波过程，显著的降低了对数据的特征要求，避免了复杂和具有专业背景知识的特征工程。其二，训练的过程是分层级的，从第一层次到第二层次，逐级进行训练避免了大规模神经网络的训练对计算速度的要求，并且次级层次的 FNN 收敛速度和模型准确率有显著的提升。



(a)全连接网络 (b)基于自编码器的二次训练网络

图 3-9 设计的结构

Figure 3-9 Architectures designed

利用 Datasets A 和 Datasets B 分别对上述两种网络架构进行训练和测试得到的结果如表 3-5。非常容易的从测试的结果观察到一些细节。用于故障特征 FNN 网络的训练时间明显比经过预训练的网络要长，Datasets A 和 Datasets B 数据集的表现力在这种网络下的准确率均不超过 79.80%，达不到可用的程度。而经过预训练的深度自编码与全连接深度神经网络在准确率上有明显的优势，平均准确率在两

个数据集上分别达到了 99.69% 和 99.49%。另外，仔细观察实验的结果，发现 FNN 用在 Datasets B 上的训练时间和更迭次数都远超其他的三个模型，这说明 FNN 网络结构对于数据不是很完美的情况，适应能力不是很强。需要较长的代价来进行多次的训练，并且在实验过程中还发生了不收敛的情况。经过多次测试，发现还是不能避免，存在一定的几率会发生这种情况。为了避免这种情况，需要反复的加大测试的数量进行重复实验。增加了预训练过程的 FNN 正确率极高，训练速度也很快，但是也很容易出现过拟合的情形。为了尽量避免出现过拟合这种问题，可以在进行预训练的过程中采取提前结束的策略，即不必等到深度自动编码器的误差完全为零，指定一定的训练周期，在实验中该值取为 300。除此之外还可以考虑的一种做法是，深度自编码预训练网络中的用于压缩感知的中间节点个数可以进行适当的调整，避免失去过多的数据有效特征。通过保持较低的压缩强度，也是在最大程度上保留较多的有用信息。

表 3-5 两种数据集和网络结构下的结果

Table 3-5 Three kinds of outputs including total two different datasets and NN architectures

	Datasets A						Datasets B					
	FNN		FNN with deep encoder				FNN		FNN with deep encoder			
	准确率	迭代次数	计算时间	准确率	迭代次数	计算时间	准确率	迭代次数	计算时间	准确率	迭代次数	计算时间
1	0.792	151	1055.052	0.990	395	579.018	0.660	383	2022.371	0.998	219	1053.225
2	0.792	163	886.821	0.999	393	545.244	0.661	419	1909.342	0.996	162	1249.516
3	0.794	163	893.238	0.997	311	338.093	0.736	882	4926.478	0.995	103	760.655
4	0.798	200	1103.801	0.998	260	412.248	0.629	412	2205.636	0.994	105	753.338
5	0.794	186	998.917	0.997	276	363.730	0.671	460	2042.183	0.997	116	832.044
6	0.795	144	782.685	0.999	603	676.354	0.662	394	1861.258	0.997	132	935.026
7	0.796	161	876.938	0.998	287	383.248	0.782	444	2354.272	0.992	94	687.211
8	0.794	143	783.963	0.998	380	458.975	0.739	481	2595.919	0.993	94	682.598
9	0.795	190	1053.824	0.995	355	435.164	0.564	474	3638.746	0.993	106	761.219
10	0.791	136	1065.482	0.998	381	464.765	0.543	200	1368.790	0.994	113	836.406
均值	0.794	164	950.072	0.997	364	465.684	0.665	455	2492.499	0.995	124	855.124
最大值	0.798	200	1103.801	0.999	603	676.354	0.782	882	4926.478	0.998	219	1249.516
最小值	0.791	136	782.685	0.990	260	338.093	0.543	200	1368.790	0.992	94	682.598

综上，FNN 的特点是泛化性高，对于数据的扩展性很好。而进行深度自编码器过程的 FNN 训练速度过快，准确率也高，但是其发生过拟合的风险极大，并且是否过拟合判定的标准也没有一个固定的标准。因此，在应用过程中随意性比较大，理论上还不存在严格的推理过程能进行详细和合理描述过拟合的程度。因此，在理论进行拟合程度估计这个方法进行有一定难度的时候，考虑使用一定的优化策略来尽可能的降低该风险。同时，还希望设计出一种稳定性强，准确率和速度

都比较均衡的结构。既然单一的 NN 很难同时兼顾以上的种种问题，所以考虑了一种优化策略，通过优化来避免上述的问题。

3.3.2 多神经网络组合的性能优化策略

众所周知的是，神经网络的复杂性和多样性在数学模型的表现上主要是任意神经元权值的随机性和激活函数的非线性共同组成。所以，在这次优化过程中采用类似的神经元结构，稍有不同的是这里的每一个节点实际上是一个由每一个独立的神经网络组成，这种节点的输入和输出分别是由不同的数据集和独立的神经网络训练完成之后的准确率、训练时间和训练速度(Accuracy, Epoch Time, Training Speed)。该向量的三个分量需要三个神经元的权值进行匹配，即表示三个特征和神经元组进行内积运算。定义一个度量指标，在论文中该指标用来度量整个组合神经网络的稳定性和能效的浮动范围。比较简单的一种方式就是将每个神经网络的输出进行线性比例相加，并且在输出时通过单个神经网络结构所用的相同激活函数。因此，实际上该组合策略的形成相当于将神经网络在两个层级进行了深度扩展。式(3-11)是对每个神经网络结构输出结果进行的一个简化的表达，注意该表达式仅仅使用矩阵来存储对应的信息，而不是进行数学意义上的矩阵计算。

$$\begin{pmatrix} Acc_1 & T_1 & V_1 \\ Acc_2 & T_2 & V_2 \\ Acc_3 & T_3 & V_3 \\ Acc_4 & T_4 & V_4 \end{pmatrix}_{4 \times 4} * \begin{pmatrix} W_{11} & W_{12} & W_{13} & W_{14} \\ W_{21} & W_{22} & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33} & W_{34} \\ W_{41} & W_{42} & W_{43} & W_{44} \end{pmatrix}_{4 \times 4} = \begin{pmatrix} Obj_1 \\ Obj_2 \\ Obj_3 \\ Obj_4 \end{pmatrix}_{4 \times 1} * \begin{pmatrix} W_1 & W_2 & W_3 & W_4 \end{pmatrix}_{1 \times 4} = \sum_{i=1}^4 W_i \cdot Obj_i \quad (3-11)$$

图 3-10 刻画了上一个小节假设的组合优化神经网络的表情式，值得关注的是，在该优化的最后一个步骤，采用了神经网络中最为常见的激活函数(logistic function)，使其更具有一般意义上的普遍性和通用性，避免特例对优化效果的影响。

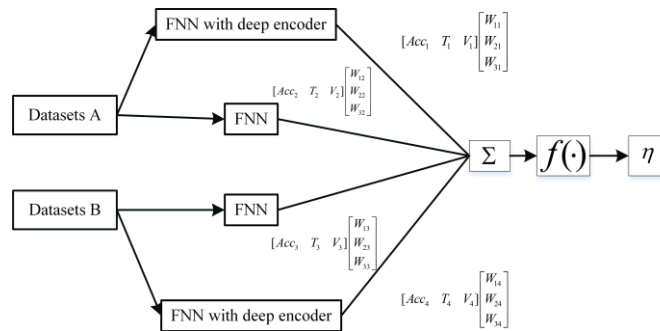


图 3-10 使用 FNN 和 NN based on Deep AE 进行综合性能优化
Figure 3-10 Performance optimization on two datasets using FNN and NN based on Deep AE

下面将上述优化过程中描述语言利用数学进行一个规范的表达，即式(3-12)。式(3-12)其实作了一个假定:准确率、训练时间和训练速度是按照一定的比例进行线性组合，并且决定了神经网络的性能。因为每一个特征都是最终决定神经网络性能参数的一部分，所以也就自然而然的是正数，这是毫无疑问的。

$$\begin{aligned} \min_{W_{ij}} \eta = \min f(W_{i1}, W_{i2}, W_{i3}, W_{i4}) = \min \frac{1}{1 + e^{-\sum_{i=1}^4 W_i \cdot Obj_i}} \\ s.t. \sum_{i=1}^3 W_{in} = 1, n = 1, 2, 3, 4 \\ W_{ij} \geq 0 \end{aligned} \quad (3-12)$$

将外部的激活函数剥离，实际上这是一个线性规划问题。利用求解线性规划通用的单纯形方法，从图 3-10 中的四组数据中，进行独立随机抽取 4 个神经网络的性能参数，分别进行从 100 到 10000 次的求解过程，得到最终的性能参数的优化解表 3-6。最小值接近 0.5 也就是性能的下界，这是很容易理解的。最终得到的性能度量范围在 0.7386~0.9502，说明大多数神经网络的稳定性和性能是足够用的，而且分类失误的可能性很低。

表 3-6 多次神经网络的组合线性优化性能
Table 3-6 Combination linear optimization performance of multiple neural networks

N	最小值	最大值	中位数
100	0.5709	0.9303	0.7532
200	0.5551	0.9483	0.7443
500	0.5551	0.9205	0.7353
1000	0.5377	0.9443	0.7450
2000	0.5377	0.9393	0.7395
5000	0.5377	0.9502	0.7388
10000	0.5377	0.9443	0.7386

式(3-12)作为一个比较容易想到的线性组合的性能度量方法是比较容易想到的。然而，实际上神经网络的训练速度和总的训练时长并不是严格线性相关的。由于使用的计算机不是多线程处理的运算方式，因此这两个参数的非线性关联才是比较一般的情形。因此考虑引入一组参数 α, β 。其中， α 设定为准确率的度量系数， β 设定为训练时间的度量系数，同时考虑到训练时间短的网络结构，速度也相对来说更快。所以形成式(3-13)这种矩阵的元素，用来存储神经网络的训练结果。

$$\left(\alpha \begin{pmatrix} Acc_1 \\ Acc_2 \\ Acc_3 \\ Acc_4 \end{pmatrix} \beta \begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{pmatrix} \frac{1}{\beta} \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{pmatrix} \right) = \begin{pmatrix} \alpha Acc_1 & \beta T_1 & \frac{1}{\beta} V_1 \\ \alpha Acc_2 & \beta T_2 & \frac{1}{\beta} V_2 \\ \alpha Acc_3 & \beta T_3 & \frac{1}{\beta} V_3 \\ \alpha Acc_4 & \beta T_4 & \frac{1}{\beta} V_4 \end{pmatrix} \quad (3-13)$$

将上述矩阵(3-13)中的数据进行与式(3-12)的优化模型进行组合,得到式(3-14)。与式(3-12)的优化模型有着明显不同的部分是该模型的内部也是一个非线性的表达式。并且增加了两个优化参数 α, β , 其组合系数也有非线性的特征。这在一定程度上保证了训练时间和速度的关联性,但不是简单的线性关联。

$$\begin{aligned} \min_{W_{ij}, \alpha, \beta} &= \min \frac{1}{1 + e^{-\sum_{i=1}^4 (\alpha W_{1i} Acc_i + \beta W_{2i} T_i + \frac{1}{\beta} W_{3i} V_i)}} \\ s.t. & \alpha W_{1n} + \beta W_{2n} + \frac{1}{\beta} W_{3n} = 1, n = 1, 2, 3, 4 \\ & \alpha + \beta = 1 \\ & \alpha, \beta, W_{ij} \geq 0 \end{aligned} \quad (3-14)$$

表 3-7 多次神经网络的组合非线性优化性能
Table 3-7 Combination nonlinear optimization performance of multiple neural networks

N	最小值	最大值	中位数
100	0.5398	0.9127	0.7628
200	0.5514	0.9219	0.7544
500	0.5377	0.9339	0.7509
1000	0.5621	0.9483	0.7552
2000	0.5394	0.9339	0.7544
5000	0.5377	0.9483	0.7530
10000	0.5377	0.9483	0.7540

利用内点法进行该优化模型的求解,初始点选择为[0,0,0,0,0,0,0,0,0,0,0,0],进行独立随机抽取 4 个神经网络的性能参数,分别进行从 100 到 10000 次的求解过程,得到最终的性能参数的优化解如表 3-7。下界值非常接近 0.5,说明这是一个非常合理的估计。性能度量的变化范围从 0.7530~0.9483,非线性的模型虽然比线性模型的性能度量上限稍低一些,但是也非常接近 0.95,稳定性和性能都是属于足够好的范围。该模型相比线性模型实际上是更加真实可靠。

3.4 小结

包括神经网络在内的许多数据驱动的分类模型，对数据的依赖性不言而喻。因此，数据的种类、数量以及数据的质量都对最终的算法分类效果有很大的影响。对于论文中的这种模拟方法来说，尽可能多的做出种类和数量更丰富的故障模型，必然会对子分类器性能的提升有巨大的作用使用了有限元进行了动力学中的模态分析与谐波响应分析，并且进行了深度神经网络的架构探索和滤波过程的增强。同时，将组合优化的思想引入了对神经网络的性能估计过程，

4 齿轮箱行星轮系齿轮故障识别

智能故障诊断中的深度学习方法的应用越来越广泛。然而，实际应用中的故障诊断中的深度模型常常会出现过拟合或者欠拟合的现象。此外，对这些模型的训练过程需要配置大量的超参数，而这些超参数的选取过分依赖于经验。因此，提出一种新的针对网络模型的超参数搜索方法，并将其应用在针对风电机组行星轮系的故障诊断方面。首先，通过不同的故障模型得到不同的模态数据，并将其构建对应的特征图。其次，对 LeNet5 模型进行复用，利用已经存在的预训练模型加速近似最优超参数和模型的寻找。最后，在行星轮系模态数据上验证提出的算法，并进行实验结果的讨论。提出的方法对于获得更优的超参数和模型具有重要的意义。

第 4 章主要包括如下几个部分。4.1 小节进行了行星轮系的建模和模态特征的分析，也进行了样本特征图的生成。4.2 小节的内容主要包括了 CNN 网络的复用与冻结，提出了一种新的超参数搜索方法。4.3 小节基于样本特征图进行了层次超参数寻优算法的验证，并且进行了相关结果的讨论。

4.1 行星轮系模态分析与特征提取方法

4.1.1 行星轮系模型的构建与模态分析位移数据的获得

行星齿轮箱是风电机组的传动系统中比较关键的部件之一。作为风电机组的故障发生频率最高的部位，对于其的研究一直是热点问题。实际发电系统中，对于行星齿轮箱内部进行非常详细的监测和故障诊断是困难的。因此，试图通过数字仿真手段来构建比较接近真实的行星齿轮模型来模拟几种故障发生的齿轮系统。把这种方式可以获得的数据和实际传感器测量的数据作补充，提高分类算法对于数据的认识挖掘更深层次的信息。

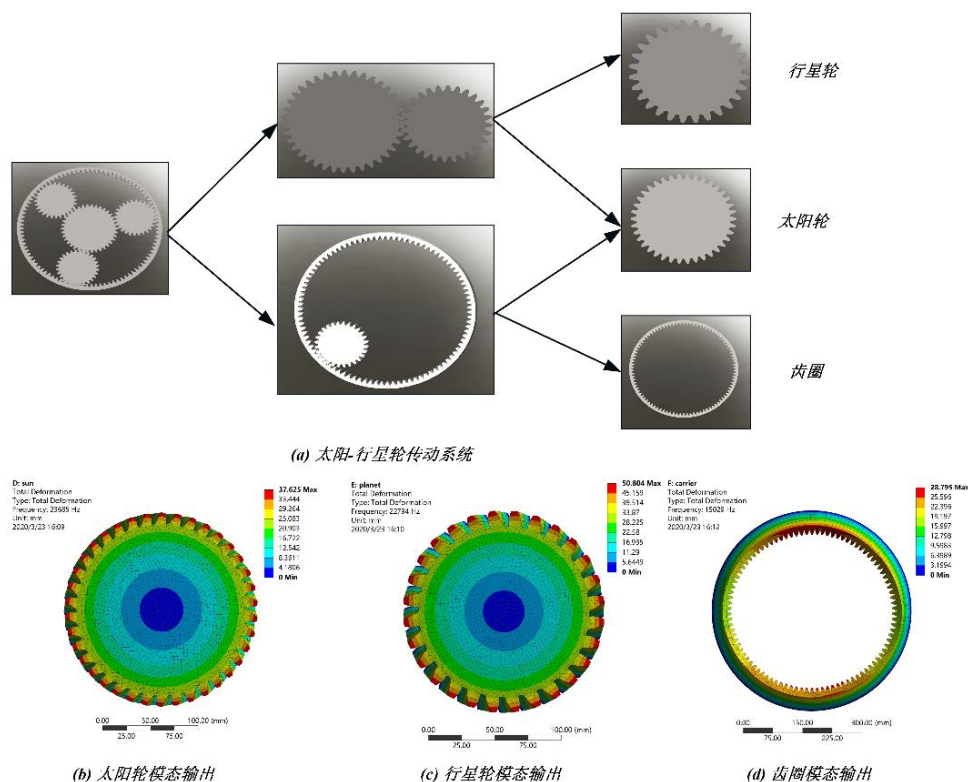


图 4-1 行星轮系统建模及其模态分析

Figure 4-1 Modeling and modal analysis of planetary gear system

考虑图 4-1(a)中的行星齿轮系统，主要由太阳轮、行星轮和齿圈组合而成。为了更具有有一般性，考虑将行星轮的个数设置为 3 个。可以将一个单级行星轮系拆解为太阳轮-行星轮和行星轮-齿圈的耦合系统。当然，也可以进一步分解为三个独立齿轮的耦合系统。在本小节中，仅考虑最简单的情形，将三个齿轮视作单独的部分。图 4-1 中(b)、(c)和(d)分别是完整的没有施加外部载荷下的太阳轮、行星轮与齿圈的位移云图。利用三维设计软件进行建模，依据表 4-1 的参数制作出几何模型。相关齿轮的齿数、模数压力角和齿宽在表 4-1 中详细的列出。

表 4-1 太阳-行星轮系传动系统中参数
Table 4-1 The planetary train system parameters

	太阳轮	行星轮	齿圈
齿数	38	28	94
模数		5mm	
压力角		20°	
齿宽		30mm	

在建立了完整的几何模型之后，另一个步骤是要确定齿轮的相关材料参数。对于标准的模态分析来说，通常不存在外部载荷，仅考虑齿轮的材料参数对其相应的形变影响。在研究中选择了 40CrMnMo7 这种材料，其主要的三个属性弹性模量、泊松比和质量密度的数值在表 4-2 中详细的展示出来。仿真模拟的目的主要是

为了验证方法的可行性，因此材料的种类或者说差异在进行模态分析时的影响对于每一个待分析的模型来说是具体的，但是对于深度学习分类算法来说实际上是影响并不明显的。模态分析对于一个固体结构来说具有识别其类型的功能，不同的结构具有不同的固有频率和振型。因此通过有限元分析的方法来将发生故障的部件与完整的部件区分成为可能。

表 4-2 40CrMnMo 材料参数
Table 4-2 40CrMnMo material parameters

特征	数值	单位
弹性模量	2.05e+11	N/m ²
泊松比	0.28	/
密度	7800	kg/m ³

图 4-2 中展示了分别进行的太阳轮、行星轮和齿圈破损和点蚀状态下的模态分析下产生出的位移云图结果。通过在齿面接近齿顶的位置任意选择一点，将其与齿顶中点连线，以这条线作草图，沿垂直于齿面方向进行切割几何体的切割，用剩下的几何体近似模拟齿轮破损的情形。点蚀这种故障在我们的模型主要是在齿轮侧面绘制圆形孔，选择一定的深度进行操作通过这种操作来模拟。通过对磨损和点蚀情形下的故障模拟，利用有限元方法进行模态分析，得到相应的固有频率下的位移数据。产生的数据对应的是模型网格划分之后，每个节点的相对于原始位置的偏移量。

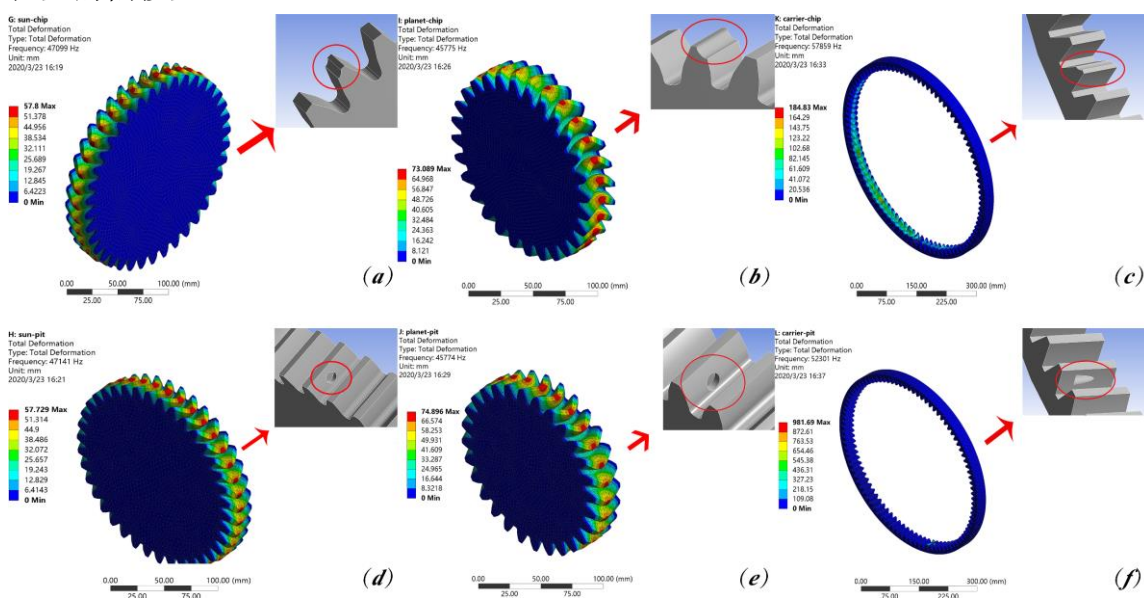


图 4-2 行星轮系破损与点蚀模态分析

Figure 4-2 Modal analysis of damage and pitting in planetary gear train gear and ring gear

4.1.2 断齿与点蚀关键特征图的生成

卷积神经网络的输入是二维与三维张量，而振动数据，无论是来源于数字仿真还是实际的传感器测量到的振动位移，都是一维的序列。因此，将一维数据转化为二维数据是实现卷积神经网络充分利用所必需的过程。从另一个角度来理解，即将一个序列编码为一张或多张图片。整个过程的数学表达如下式(4-1):

$$x \in R^1 \xrightarrow{T_x} X \in R^2 \quad (4-1)$$

其中一种类型的变换是格雷角域(GAF, Gramian Angular Field)。对于一个一维序列（时间序列或者空间序列） $x=\{x_1, x_2, \dots, x_n\}$ 来说，主要的计算过程分为三个过程：

(a)考虑将一维序列 x 的 n 点观测值进行放缩，取值范围缩放到 $[-1,1]$ 范围内：

$$\bar{x}_i = \frac{(x_i - \max(x)) + x_i - \min(x)}{\max(x) - \min(x)} \quad (4-2)$$

(b)将归一化处理之后的序列依据采样点的位置进行反余弦函数编码，如果是时间序列这种具有次序的序列，按照均匀分布的准则进行处理。空间序列按照一般的序列处理即可。具体的计算方法如(4-3):

$$\begin{cases} \phi = \arccos(\bar{x}_i), -1 \leq \bar{x}_i \leq 1, \bar{x}_i \in \bar{x} \\ r = \frac{i}{N}, i \in N \end{cases} \quad (4-3)$$

(c)得到序列的广义余弦角度之后，将角度序列按照式(4-4)和(4-5)从一维信号到二维信号的升维计算。最终得到 $N \times N$ 点的二维矩阵 XS 和 XD 。

$$XS = \left[\cos(\phi_i + \phi_j) \right] = \bar{x}_i^T \cdot \bar{x}_j - \sqrt{1 - \bar{x}_i^2} \cdot \sqrt{1 - \bar{x}_j^2} \quad (4-4)$$

其中， $XS_{ij} = \bar{x}_i \cdot \bar{x}_j - \sqrt{(1 - \bar{x}_i^2)(1 - \bar{x}_j^2)}$ 。

$$XD = \left[\cos(\phi_i + \phi_j) \right] = \sqrt{1 - \bar{x}_i^2} \cdot \bar{x}_j - \sqrt{1 - \bar{x}_j^2} \cdot \bar{x}_i^T \quad (4-5)$$

其中， $XD_{ij} = \sqrt{(1 - \bar{x}_i^2)} \cdot \bar{x}_j - \sqrt{(1 - \bar{x}_j^2)} \cdot \bar{x}_i$ 。

考虑一个 512 点的白噪声序列，其波形如图 4-3(a)所示。将其依次进行上述的计算过程，得到最终的 XS 和 XD 并且绘制出对应的矩阵图。从图 4-3(b)和(c)中明显可以观察到通过变换得到的二维信号具有稀疏性，实际上对于神经网络来说，数据集中的数据具有一定形式的稀疏表达是非常有助于提高网络的拟合和泛化能力。图 4-3(b)中矩阵的数值比较接近-1，整体的数值分布比较集中，而图 4-3(c)中矩阵元素的分布则比较均匀。

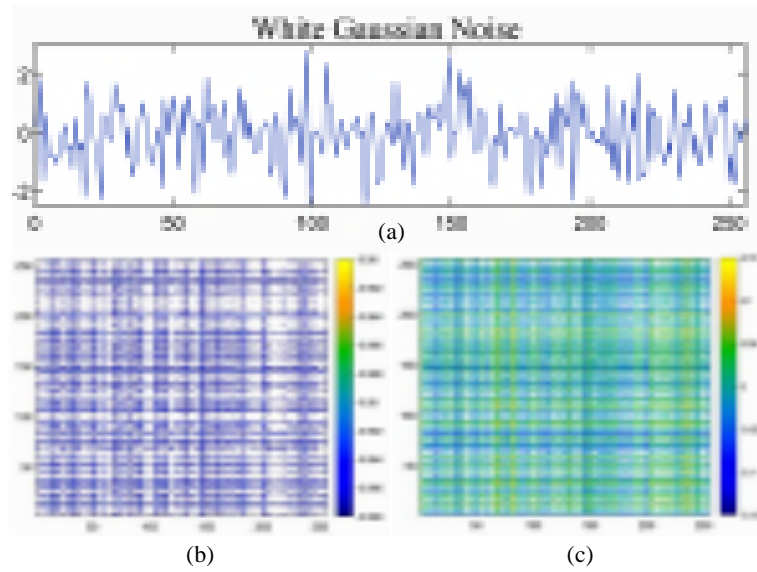


图 4-3 高斯白噪声信号的仿真结果

Figure 4-3 Simulation results of Gaussian white noise signal

将上述 XS 与 XD 矩阵进行组合得到最终需要的样本特征, 考虑到后续卷积网络有单通道和三通道两种作为输入的模型, 将其分别表示如图 4-4。图 4-4 中(a)(b)和(c)分别对应了对 XS 和 XD 特征图进行了加法乘法和取模操作, 即

$$\begin{aligned} channel_a &= XS + XD \\ channel_b &= XS \times XD \\ channel_c &= \sqrt{XS^2 + XD^2} \end{aligned} \quad (4-6)$$

对应的单通道数据获得方法是计算(4-6)中 a b c 三个通道之和, 单通道的样本特征计算定义如下(4-7):

$$channel = (XS + XD) + (XS \times XD) + (\sqrt{XS^2 + XD^2}) \quad (4-7)$$

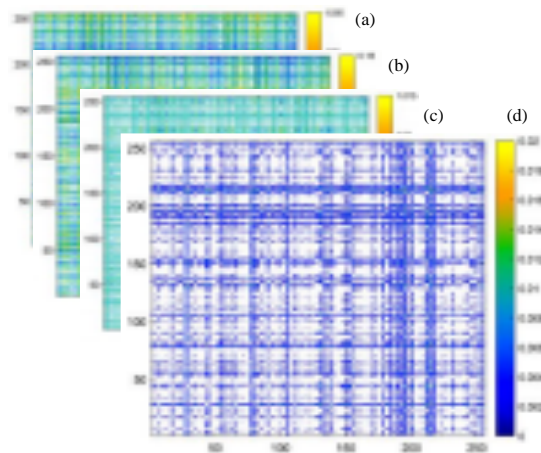


图 4-4 单通道特征图与三通道特征图

Figure 4-4 Single-channel feature map and three-channel feature map

将模态分析获得的数据进行循环均匀采样，利用上述生成特征图的方法，得到用于行星轮系故障诊断的数据集。使用 T-SNE 进行低维度流形嵌入的操作，将其数据投影到二维平面上。并且利用核密度方法进行无参数密度估计，得到其而二维的数据分布，并将其投影到 X-Y 平面到。利用这种可视化方法，可以比较清楚的观察到数据的分布信息和密集程度。图 4-5 中的六张子图分别是其结果。

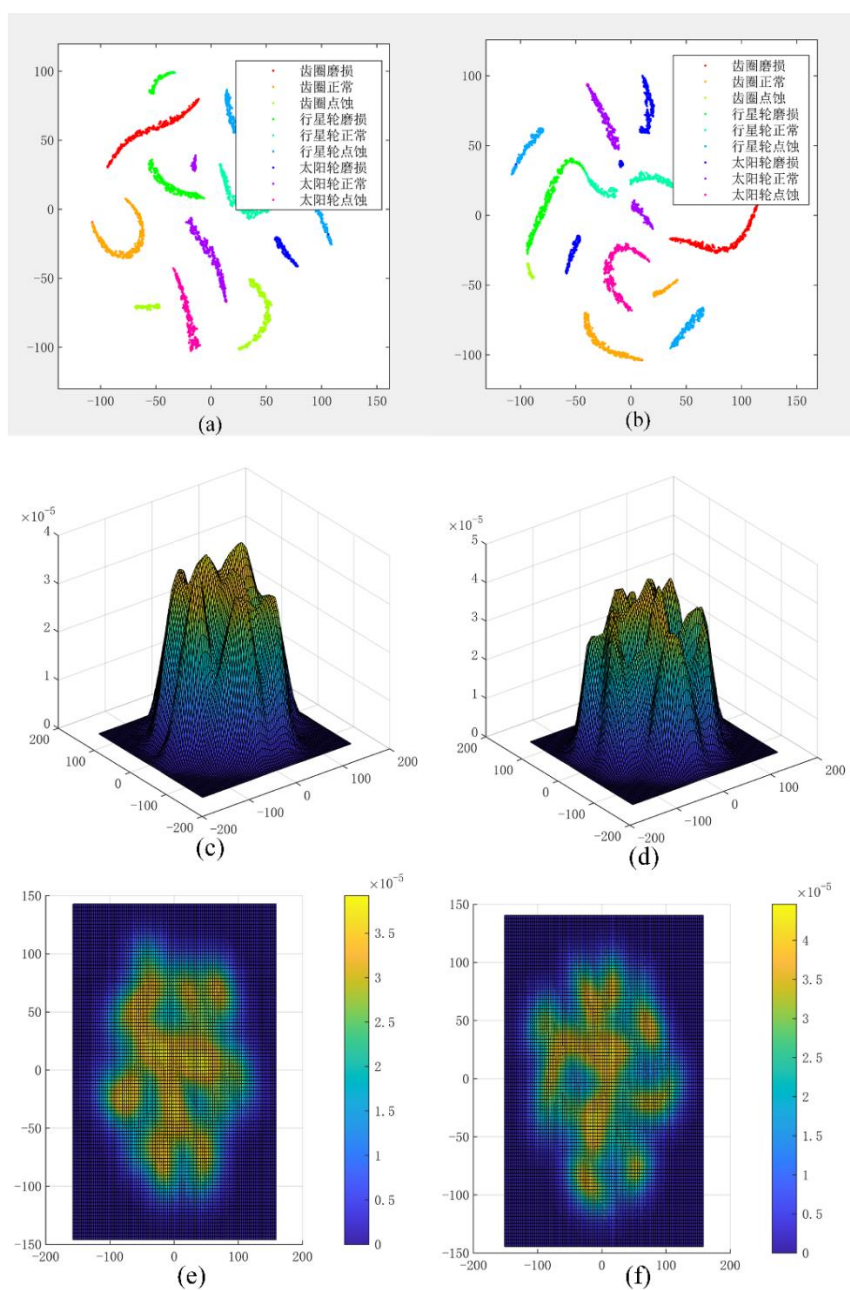


图 4-5 行星轮系样本集可视化及其密度分布

Figure 4-5 Visualization of planetary gear system sample set and its density distribution

4.2 网络结构的方法研究

4.2.1 经典卷积网络的模型复用与分类器层重构

这一小节，简要介绍两个典型卷积神经网络。第一个是用于卷积网络最优超参数组寻找的 LeNet5，第二个是用于验证我们的超参数组有效性的 AlexNet。与经典的 LeNet5(图 4-6)和 AlexNet(图 4-7)稍有不同的一点是，在原始版本的 LeNet5 和 AlexNet 网络中最后一层分别是 10 和 1000。这是因为它们最初被设计用来做手写数字的分类和超大规模的物品分类，而在本研究中正常的运行工况、磨损和点蚀工况总共分为 9 类，因此将这两个网络的最后一层全连接层节点个数设置为 9 个。另外在原始的 AlexNet 结构中是将卷积与池化结构分开，在两个单独的显卡上分别计算，而本研究中将其合并，这种合并并不改变原始网络的计算性能和效率，仅仅是为了研究和应用的方便。池化层本质上是对特征图的均匀随机采样，而且不存在可变参数，所以将其与卷积层视作共同的可变参数层是合理的。

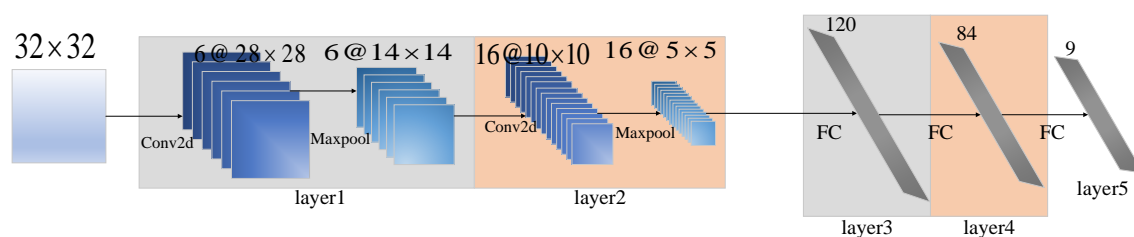


图 4-6 LeNet5 模型
Figure 4-6 LeNet5 model

选择 LeNet5 与 AlexNet 作为研究使用的基本网络结构并不是一个随意的行为。LeNet5 是最简单的卷积神经网络，具有所有卷积网络的共同特点，也是现代复杂网络结构的起点。当然，它的参数数量是比较小的，从参数量小和结构比较简单这个特点来看，很适合在资源有限的情况下做大规模重复实验。而超参数寻优这个研究方向需要数量比较大的重复实验，从计算量和效率的角度考虑，LeNet5 是小型网络中相对来说比较适合数值实验的。此外，由于 LeNet5 网络是最早出现的卷积神经网络，因此对其进行超参数寻优的过程对后续的几乎所有的网络结构都具有一定的参考意义。

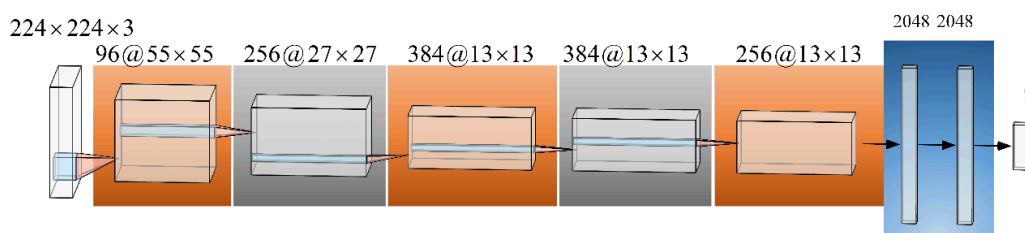


图 4-7 AlexNet 模型
Figure 4-7 AlexNet model

参数量较大且结构复杂的网络结构，无论是从零开始进行训练和测试，还是利用已经现存任务完成预训练过程的模型，在后期的二次训练和调优过程中，依旧要经历反复微调过程。如果模型参数量特别大并且同时需要多次调优的话，这个过程效率非常低并且也耗时非常久。将整个大型网络分别特征提取和分类器两个层次，迁移策略主要是针对特征提取层，而分类器层往往是从零开始训练。众所周知，卷积网络的较大参数量主要集中在分类器层，也就是传统的全连接结构。事实上，这种迁移方法是比较粗略的，它起作用的部分是共用一个特征提取层，暗含了源数据集和目标数据集的分布存在一定的关联。但是迁移的目的显然是为了适应更丰富的数据分布，而不是局限于某一个分布。根据迁移学习的一般做法和它对数据分布的限制，提出了如图 4-8 展示了这种一次迁移全部参数，逐层冻结，依据有限层级反复训练的方法。

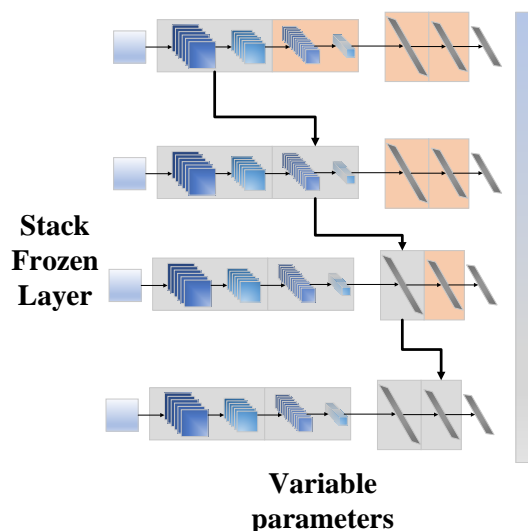


图 4-8 堆栈冻结方法
Figure 4-8 Stack freezing method

这种方法的具体操作主要是两个部分。首先，将原始模型训练或者从现存的预训练模型中完整复制一个结构和参数完全相同的模型。然后，如果是直连前馈型的网络结构，将卷积层和池化层视作同一个层级，全连接层每一层视作一个层级。至此，整个迁移模型的构建和分层就结束了。接下来，开始逐层冻结和训练过程。以图 4-8 中的 LeNet5 为例，除了最后一层外，共有四个层级，所以要进行至少四次的训练。每次模型训练过程之后，冻结的层数逐次增加，网络的可变参数减少。当然，网络的训练所需要的时间在这个过程中也是逐渐减少的。逐层冻结逐层训练的方法用于大型网络结构的训练具有显而易见的优势，与数学理论中高维函数的不动点迭代方法有相似之处。寻找全局最优点的其中一种比较直观方

法是尽可能比较所有的局部最优点。

4.2.2 动态网格随机化超参数寻优方法

卷积神经网络的参数主要由结构参数和超参数两部分组成。结构参数包括激活函数、网络结构的层数、节点之间的连接权值和偏置值等部分。这一部分的设定很大一部分取决于该网络所要完成的任务，由任务所要达成的目标决定。而我们本小节讨论的超参数是指在网络本体的结构和参数确定之后，在神经网络训练过程中进行配置的那些参数。这些超参数可以主要由三个部分构成。第一部分是迭代次数。第二部分是大批次训练下的每个测试批次和训练批次的样本个数，最后一个是优化器的参数包括优化器的种类、学习率、动量与正则化的类型等等。理想情况下，假设拥有足够多的计算资源，那么仅仅要做的就是遍历所有配置的组合，手工的从其中选择最佳的配置即可。但是，实际上遍历所有的可能组合的所需要的时间代价是不可忍受的，所以寻找最优超参数的想法只能降低一下要求。实践中超参数搜索目标实际上进行了一定程度的弱化，即在一定时间和计算资源内尽可能寻找到最优的超参数。这个问题与最优化理论的目标是相同的，所以很多研究者在一定程度上参考了优化问题求解的思路。然而，神经网络的训练问题与一般的优化问题（凸优化问题与连续优化）有明显的差异性。神经网络并不存在类似梯度（导函数）这种表达方式。因而，现存的基于梯度的优化方法并没有太多的应用空间，而已知的很多研究透彻的优化求解算法和软件包恰恰利用了这些方法。

超参数搜索所使用的基本方法包括网格搜索算法和随机搜索算法。在图 4-9 中以两个参数为例子，简要说明一下上述两种方法的具体实施步骤。图 4-9(a)中将两个待搜索变量的搜索范围按照 d_1 与 d_2 进行等间隔划分，形成了 d_1*d_2 个网格点，然后逐一穷尽所有格点，根据提前设定好的优化目标函数比较最小值。网格中最小值所在的格点坐标就是线性搜索所能找到的最优参数。理论上网格的密度越大，搜索到最优参数组的可能性就越大。计算资源足够的情况下，蛮力搜索是一个可以保证一定能寻找到最优参数的基本方法。为了避免计算资源的浪费和提高线性搜索的效率，采用随机搜索图 4-9(b)是顺其自然的想法。通过在划分网格内多次独立的随机取点，既保证了条件下的搜索空间不变，同时提高了搜索效率，使得找到最优参数所用的计算资源大幅度的减少。

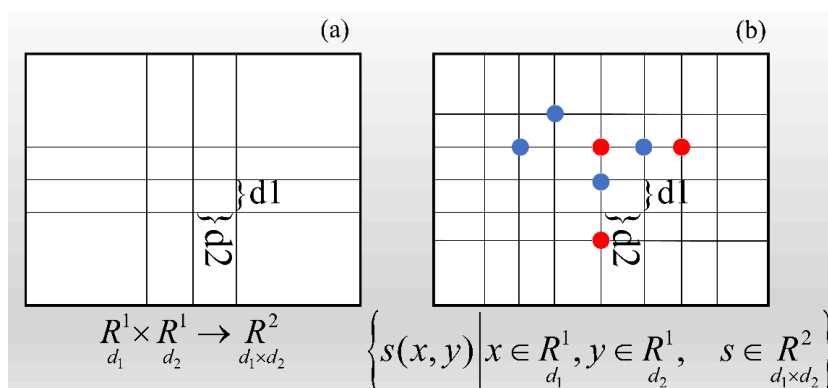


图 4-9 超参数网格搜索与随机搜索方法

Figure 4-9 Grid search and random search for hyperparameters

网格线性搜索的时间复杂度是 $O(n^k)$, 即超参数数量的 k 阶这个数量级上, 同等网格密度上的随机搜索其所需要的运行时间则要小的多。随机搜索比同等网格密度的所消耗的计算资源要小的多。所以网格线性搜索方法和随机搜索各具优点。因此, 考虑将网格线性搜索和随机搜索方法结合, 以期进一步减小计算量, 加快最优超参数的寻找。

表 4-3 层次网格超参数搜索

Table 4-3 Hierarchical grid scaling hyperparameter random search

Algorithm: HGSRs (Hierarchical Grid Scaling Hyperparameter Random Search)

Input: Pretrained-NNM (Neural network model)

PEF (Performance estimation function)

G(Hypermeter Grid initialization)

S_0 (Hypermeters initialization)

I_0 (Iterator initialization)

R (Grid size scaling factor)

Output: NNM (Neural network model)

S_n (Hypermeters)

1: model:= Pretrained-NNM

2: F:=PEF

3: S:= S_0

4: I:= I_0

5: n:=the number of Pretrained-NNM layers

6: while $I < n$

7: I = I+1

8: S = random_search(G)

9: train_acc, train_loss, test_acc, test_loss= model_train (S, model)

10: performance estimation value = F (train_acc, train_loss, test_acc, test_loss)

11: Smax = (max (performance estimation value))⁻¹

12: model = model_frozen layer(I)

13: G = G * R

14: end

表格 4-3 是本文中提出的层次网格超参数搜索方法的算法伪码描述。其输入初始化包括一个基本的预训练模型和一组基本的超参数, 输入则是经过整个算法优化过后得到的针对当前数据集有限资源内的最优模型与最优超参数。图 4-10 是该

算法的另一种表达方法，其中的子程序结构利用了任意选定的基本深度学习模型的训练方法。

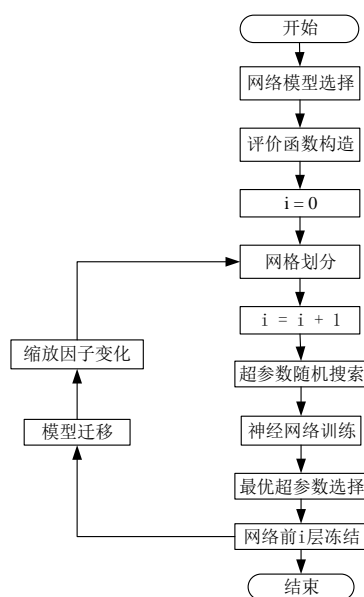


图 4-10 层次网格超参数搜索算法流程图

Figure 4-10 Hierarchical grid scaling hyperparameter random search flowchart

4.3 数值实验结果分析与讨论

数值实验以经典得 LeNet5 为基础的网络结构，将其的网络层次划分为三个主要部分。第一和第二部分均由卷积层和池化层构成，第三层为了节省计算资源将全部的全连接层作为同一个层级。本文中提及到的神经网络都是由 Pytorch 框架完成，测试的主机是主频 2.60GHz，GPU GTX840m，实验结果都在该配置上完成。层级网格缩放超参数随机搜索算法的应用基础是需要一个基础的网格作为初始化的网格分布，具有一定的任意性。算法本身对网格进行了密度的逐级调整，使搜索范围逐步的缩小，期望在每一个阶段的网格中都找到局部最优点。将这些局部最优点进行比较，进一步得到在有限时间和计算资源范围内的近似全局最优。并将获得的最优超参数组和其对应的模型保存下来，方便进行迁移加速或者用于其他的具有不同目标的任务。

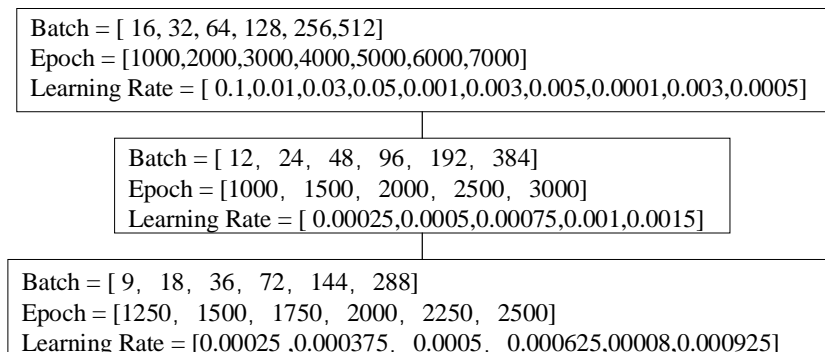


图 4-11 三个阶段搜索空间的网格划分
Figure 4-11 Three-stage meshing of search space

图 4-11 中展示了进行的三个阶段的随机寻优，限于计算条件，主要选择了小批次、迭代次数和学习率三个参数。可以观察到每个批次网格的设定都是两倍的增量，迭代次数的间隔由 1000 到 5000 再到 250，是网格密度不断增大的过程。学习率总体呈现出较小的特点，并且逐渐降低。Pytorch 库的默认数值精度是浮点 32 位，即小数点之后 8 位，理论上来说，学习率在这个范围内均是合理的并且在计算上是有效的。在三个阶段的超参数寻优过程中，所有的结果如图 4-12 所示。该平行线图展示了所探究的超参数和其对应模型性能之间的关系。第一个阶段因为初始的超参数组是随机选择的，所以进行了比较多的训练和测试，并且其拟合的误差和测试的准确度都较低，同时在不同超参数组的选择下明显准确率具有方差比较大稳定性差的特点。第二第三阶段相比第一阶段的其拟合误差较小同时准确率也更高。此外，在平行线图中偶然观察了一个现象，训练集拟合误差和测试集的准确率的所有折线的交点在准确率 0.55 左右。

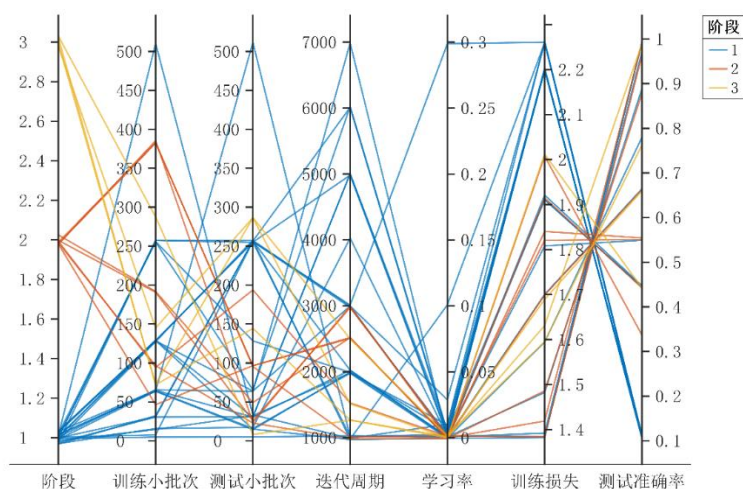


图 4-12 三个阶段超参搜索的结果
Figure 4-12 Three-stage hyperparameter search results

上述算法选择了出了最优的超参数组，然而在 Adam 优化器作为求解算法的条件下得出的结论。为了避免优化算法对神经网络训练过程的影响，忽略超参数组的作用，所以选择了目前使用最广泛的七种优化器对最优超参数组进行测试，其结果在图 4-13 和图 4-14 中进行了图示。图 4-13 的四张子图分别表达了训练过程中的训练损失、训练准确率、测试损失和测试准确率的变化情况。训练过程中的误差和测试准确率波动较大并且极不稳定，但是测试误差和准确率则相对变化较小。从图 4-14 中的条形图中发现对于特定的超参数组来说，除了 Adam 和 SGD 优化算法之外，该组参数均不会出现过拟合的现象。此外，所选取的剩余几种优化算法的测试准确率都比较高，并且几乎所有优化算法对于该组超参数的优化误差都保持在 1.4~1.6 之间的范围内，说明选定的超参数组的稳定性比较高。没有出现测试准确率在 0.5 以下的情况，且无论是测试误差曲线和训练误差曲线都处于降低的方向，不存在发散的现象，其稳定性较高。

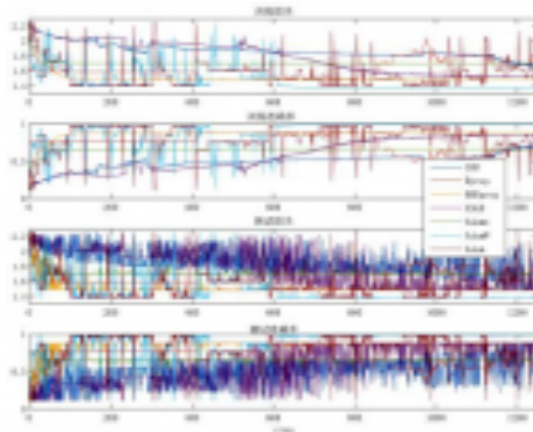


图 4-13 七种不同的优化器的训练过程和测试过程的可视化

Figure 4-13 The training process and testing process visualization of seven optimizers

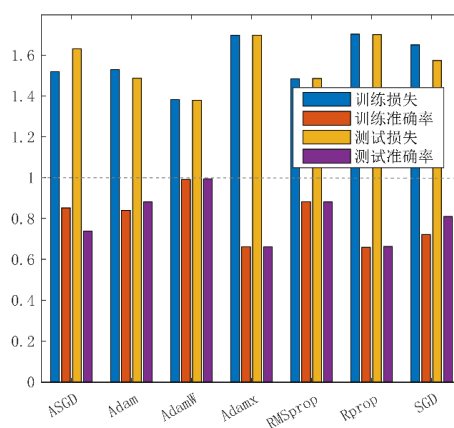


图 4-14 七种优化算法的损失与准确率示意图

Figure 4-14 The loss and accuracy of seven optimizers

AdamW 优化算法在给定的超参数组的表现最佳，误差在 1.4 左右且准确率在 0.990 浮动。其中几种优化算法的准确率在 0.6~0.8 范围内，表现出比较一般的性能。通过上述的实验结果和对不同优化器的结果分析，可以找到在有限计算资源的范围内使网络结构性能最优的超参数。验证了论文中提出的层级网格缩放超参数随机搜索算法的有效性。

4.4 小结

本章主要针对风电机组齿轮箱中的断齿和点蚀故障进行了仿真分析，并利用仿真获得的数据整理成数据集。将一维仿真的模态数据转化为二维数据，生成新的特征图。试验了一种将经典网络模型进行复用的方法，在此基础上提出了一种新的超参数搜索方法，对于提高风电机组故障诊断的准确率有非常重要的意义。

5 滚动轴承稀疏特征提取与迁移故障诊断方法

风电机组齿轮箱滚动轴承信号具有时变非平稳的特征，且在振动信号测量和数据获取的过程中，常常伴随着一定强度的高斯白噪声。这样具有高随机性特征的信号，直接进行常规的信号处理或者进行特征提取得不到比较理想的效果，因此在进行智能故障诊断分析前需要对数据进行变换或者预处理。自然界中可以观测到的数据大多数在选取合理的“基”的情形下都具有其稀疏表达的形式。换句话说，如果选择的“基”比较好，那么复杂的随机信号，在这种坐标下可以仅有较少的参数来表达。信号的可压缩性质在于是否可以稀疏表达，典型的方法比如傅里叶变换与小波变换，其核心思想就是寻找时间信号在固定的基底和可变基底的参数量最小化。除此之外，在数学领域存在已久的压缩感知理论已经比较成熟，可以看作是稀疏特征提取的一个近年来比较重要的进展^[78]。这与滚动轴承故障信号的特点相似，因此将压缩感知的方法用于滚动轴承信号特征提取是合理的。

5.1 滚动轴承信号的稀疏特征提取

滚动轴承故障信号是一种典型具有周期性的冲击信号，从某种意义上来说，一定存在某种“基”，使得冲击信号在这种基下的表达具有最小的参数量。因此，寻找到适合冲击信号的基底坐标，获得冲击信号的稀疏表达成为一种顺其自然的想法。

5.1.1 压缩感知的基本理论与方法

大部分的自然信号，例如声音和图像信号具有高度可压缩的特征。这种可高度压缩的信息意味着该信号存在着合适基下的稀疏表达，仅有少量的参数不为零的状态同样保持了信号的信息量。也就是说，一个可压缩信号 $x \in R^n$ 可以由一个稀疏向量 $s \in R^n$ （大部分的向量分量值是零）在一组变换基底坐标 $\phi \in R^{n \times n}$ 下进

行表示:

$$x = \phi s \quad (5-1)$$

如果其中存在着 K 个非零元素, 这种向量被称为 K 稀疏向量。常见的这类基, 例如傅里叶基和小波基, 如果想要重构原始信号, 只需要构造合适的有规律的基底类就可以完成信号恢复与重构的任务。同时可以观察到信号在时间域与频率域之间进行转换, 是不损失信号的特征的, 在于基底是一个方阵^[79]。

图像与声音信号在傅里叶基或者小波基下是可压缩的, 因此可将它们进行相应的数学变换。大多数的稀疏数值很小, 并且是很接近零的, 在这种情形下, 信号的信息损失是很小的, 失真程度很低^[80]。这些少量的非零稀疏数据与压缩之前的高维信号进行比较, 无论是进行传输或者存储均有很大的优势。想要恢复和重构这些原始信号, 只需要进行对应的逆变换即可。傅里叶基与小波基是应用最为广泛的基底变换方法, 几乎所有的图像与声音信号在这类基底下都是可以进行稀疏表达的。因此, 一旦确定某些信号是可压缩的, 只需要传输和存储对应的稀疏信号即可。至于其基底矩阵 ϕ , 可以由机器即时的生成。尽管在声音, 图像和视频压缩上压缩理论有了非常广泛的应用, 然而在工程领域进行相关的理论和应用研究仍然需要进一步思考。对于高维微分方程描述的系统与其连接的低维流形, 常常伴随着确定性的低维稀疏表达^[81]。即使是宽频带的振动信号, 其瞬时特性也同样可以进行稀疏表达。

尽管压缩理论在现实世界的应用已经取得了巨大的成功, 然而其中对于高维数据特征的依赖仍然没有改变。相较于传统压缩理论中获取高维数据进行压缩, 丢弃大部分信息的做法, 压缩感知利用随机矩阵理论进行了极少的信息压缩, 然后在变换的基底组中获得了信号的稀疏表达形式^[82]。压缩感知理论的思想用数学表达是相当简单的, 然而计算出测度矩阵对应的稀疏向量确是一个有难度的问题。从待求解问题的可解性和复杂性角度来看, 属于一类 NP-hard 问题。压缩感知理论在工程和应用科学的快速应用依赖于在一定条件下该问题条件的松弛进而使用凸优化求解方法进行完整信号的重构^[83]。

从数学理论上来进行解释, 压缩感知目的是获得信号在较少测度下稀疏数值基的信号重构。假设信号 x 在基 ϕ 的表达下是 K 稀疏的, 通过在设计的随机测度矩阵, 然后在这个变换下求解稀疏向量 s 中的非零元素。实际数据 $y \in R^p$, 其中 $K \prec p \ll n$

$$y = Cx \quad (5-2)$$

测度矩阵 $C \in R^{p \times n}$ 表示 x 的 p 个线性测度。比较简单的测度矩阵元素选取方式是从高斯分布或者伯努利分布中取得。当然，另外一种设计方法是测度矩阵的每一行都单独进行设计。

至此，想要重构信号 x 可以通过稀疏信号 s 及其测量矩阵获得公式(5-1)。压缩感知方法的目标已经非常明确。找出信号 y 在给定测量矩阵下的稀疏向量 s 。

$$y = C\phi s = \theta s \quad (5-3)$$

式(5-3)表达的线性等式显然是欠定系统，因此解 s 的数目是无穷的。其中最稀疏的解可以由下列公式(5-4):

$$\begin{aligned} \hat{s} &= \arg \min_s \|s\|_0 \\ \text{s.t. } & y = \theta s \end{aligned} \quad (5-4)$$

5.1.2 信号的 L1 与 L2 正则化

压缩感知方法的稀疏解主要在于求解公式(35)，问题在于目标函数的零范数非常难以用解析表达式进行表达。因此，常规的求解约束优化的方法在此失效了。公式(35)中的目标函数的 L0 范数与目标函数的 L1 范数在稀疏问题的求解上是等价的，而且使用 L1 范数进行替代并不影响稀疏问题的求解精度。至此稀疏信号的求解目标已经转化为 L1 范数的求解问题，然而这个目标的求解仍然效率不够高。利用其对应的 L2 范数可以提高求解速度但带来的是精确度的损失和求解目标稀疏性的下降。图 5-1 分别对应 L0~L2 的二维平面图形，L1 范数对应的菱形与 L2 对应的圆形相比较明显表达的信号分量具有更强的稀疏性质。L1 范数有着良好的信号稀疏性表达和精确性的优势，而 L2 范数具有求解速度快且求解算法比较成熟的便捷性，所以做出一种假设提出本文的方法。

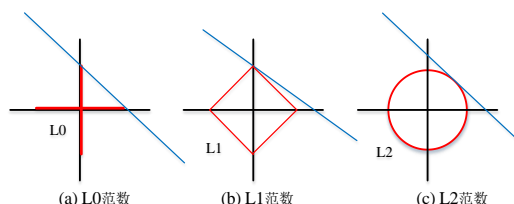


图 5-1 L0~L2 范数的二维平面解析示意图
Figure 5-1 L0~L2 norm 2D plane diagram

考虑公式(5-5)中的优化目标，其中 $0.5 < \lambda < 1$ 。设计这样的优化函数其实非

常易于理解想要保持 L1 和 L2 范数特性的同时，尽可能不影响待求解信号的稀疏性同时兼顾求解效率。参数 λ 的选择可以根据信号的特点在范围内自由选取。

$$\begin{aligned} \hat{s} = \arg \min_s (\lambda \|s\|_1 + (1-\lambda)\|s\|_2) \\ \text{s.t. } y = \theta s \end{aligned} \quad (5-5)$$

5.1.3 振动信号的获取与数据描述

实验数据的来源是 JSNU-WT-1 型风电机组故障模拟实验台，如图 5-2 所示。该实验台利用对传动件进行电火花加工来模拟齿轮和轴承等多种故障形式。实验台的主要组件包括：控制柜、调速电机、行星齿轮箱、同步带、风力发电机等。在实验中，INV3062T0 型信号采集系统用于收集振动信号。使用 4 个 INV9822 型加速度传感器收集振动信号，并使用磁性底座将传感器放置在行星齿轮箱的壳体上，振动信号由数据记录仪以 25.6kHz 的采样频率获取。轴承损伤是用电火花线切割加工的，在滚动轴承外圈、内圈滚道加工裂缝用以模拟外圈、内圈滚道损伤所产生的冲击。滚珠切割出凹坑用以模拟滚珠的损伤所生的冲击。



图 5-2 风力发电机组实验台
Figure 5-2 Test-bed of wind turbine

在不同的工况下进行数据采集，构建样本集中的五种类型的滚动轴承故障数据，包括内圈故障、外圈故障、滚珠故障、内外圈组合故障与正常运转的情形。数据采集设备如图 5-3 在 0.16s 内取得 4096 个样本点作为一个故障样本 (4096/25600=0.16s)。



图5-3 振动信号的采集
Figure 5-2 Acquisition of vibration signals

直接进行振动信号的时域表达往往受到噪声的干扰会因为影响信号的典型特征，所以使用 Hilbert 变换来获得解析信号，进一步突出信号的包络特征。对于一个时间域信号 $x(t)$ 来说通常这样计算。

第一步，获得 $x(t)$ 的 Hilbert 变换(5-6):

$$\tilde{x}(t) = x(t) * \frac{1}{\pi t} \quad (5-6)$$

第二步，构造解析函数 $z(t)$ (5-7):

$$z(t) = x(t) + j\tilde{x}(t) \quad (5-7)$$

第三步，得到解析信号的模(5-8)，

$$|z(t)| = \sqrt{x(t)^2 + \tilde{x}(t)^2} \quad (5-8)$$

图 5-4 中的(a)(b)(c)(d)(e)分别表达了五种下的滚动轴承信号的时域信号及其解析信号的时域图。可以明显的看到解析曲线与原时域获得信号相比，表达能力更强，突出了滚动轴承的冲击信号特征。

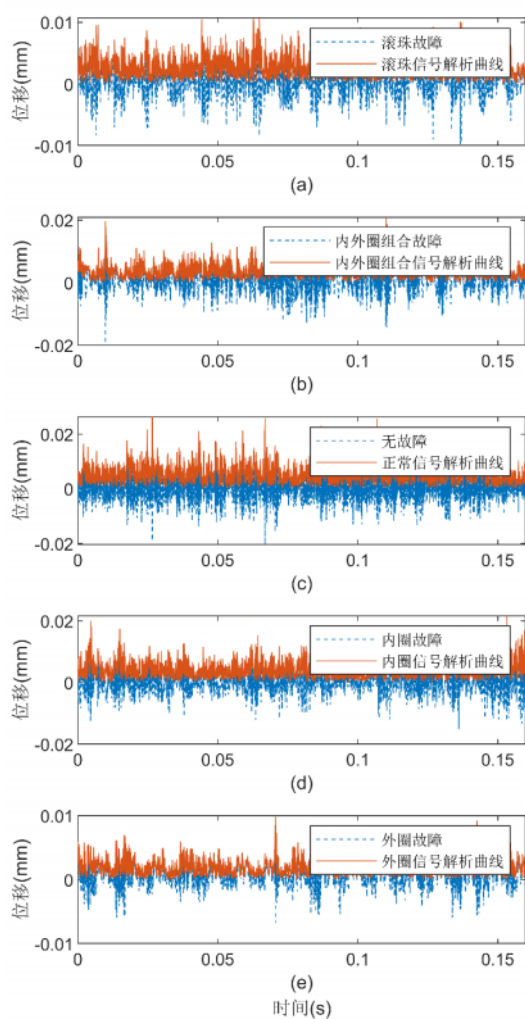


图 5-4 五类信号的时域与其解析信号的时域表达

Figure 5-4 Time domain of five types of signals and the time domain expression of their analytical signals

5.1.4 振动信号的稀疏特征提取与试验数据结果分析

信号的稀疏特征求解实际上转化为了一组凸优化问题。本研究中采用了 CVX 软件中凸优化求解包 SDPT3 算法的 4.0 版本(Infeasible path-following algorithms)。图 5-6 中包含了 15 个子图。每一行中的(a)(b)(c)(d)与(e)分别代表了滚珠故障、内外圈组合故障、正常信号、内圈故障与外圈故障信号。每一列中的从左至右分别代表了求解稀疏信号的不同范数是 L1、L2 和 L1 与 L2 的组合。本文中的 λ 取 0.65。

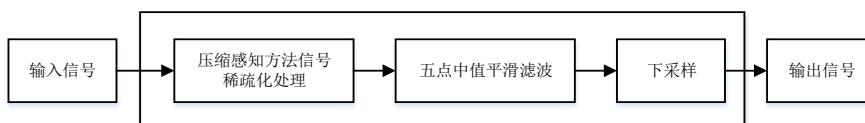


图 5-5 L0~L2 滚动轴承信号的信号处理设计流程

Figure 5-5 L0~L2 Signal processing design process of rolling bearing signal

将输入信号进行稀疏化求解之后的信号，仍然需要进一步进行平滑处理。之后进行信号的下采样处理，下采样的采样率和压缩感知进行处理的稀疏信号的放大倍数相同。

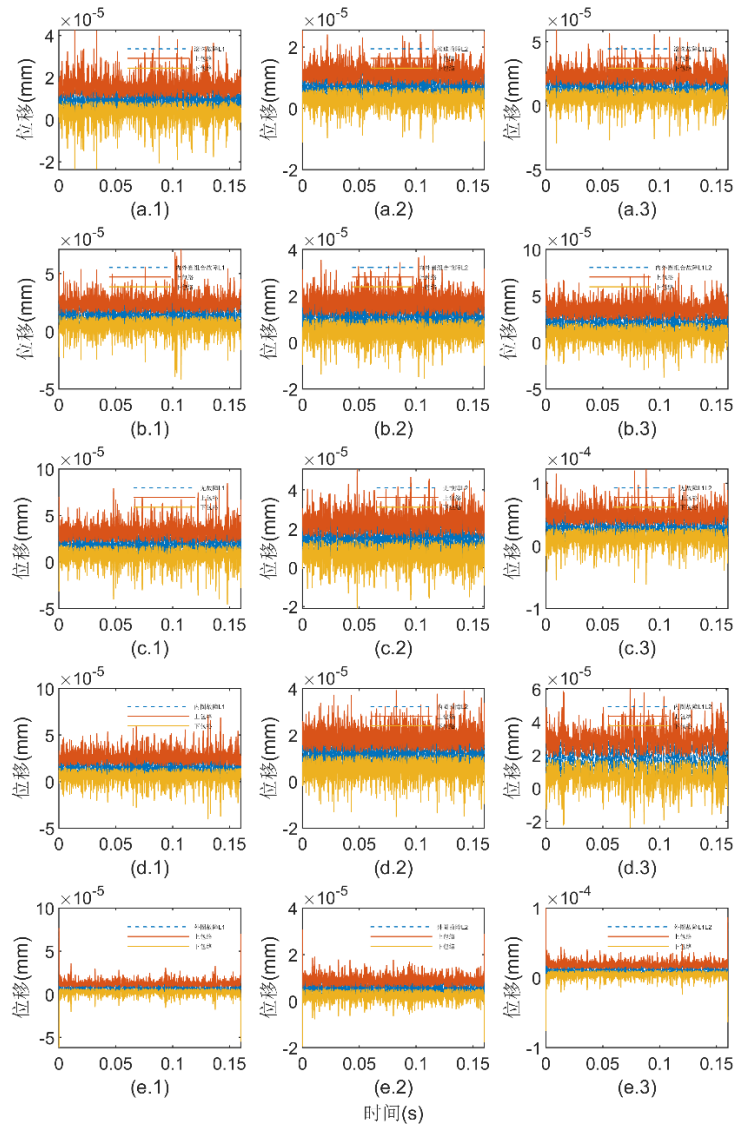


图 5-6 L0~L2 滚动轴承信号的上包络和下包络曲线

Figure 5-6 L0~L2 Upper envelope and lower envelope curve of rolling bearing signal

从图 5-6 子图中第一列可以观察到 L1 稀疏化方法获得的信号，冲击特征最明显。子图第二列 L2 稀疏化方法求解的速度最快，在正常状态、内圈故障及外圈故障上的表现比较优良。子图第三列 L1 与 L2 的组合稀疏的表现比较综合。滚珠故障、正常状态及内圈故障信号的冲击特征比较明显，周期性的波动在滚珠故障信号曲线上具有突出的特点。

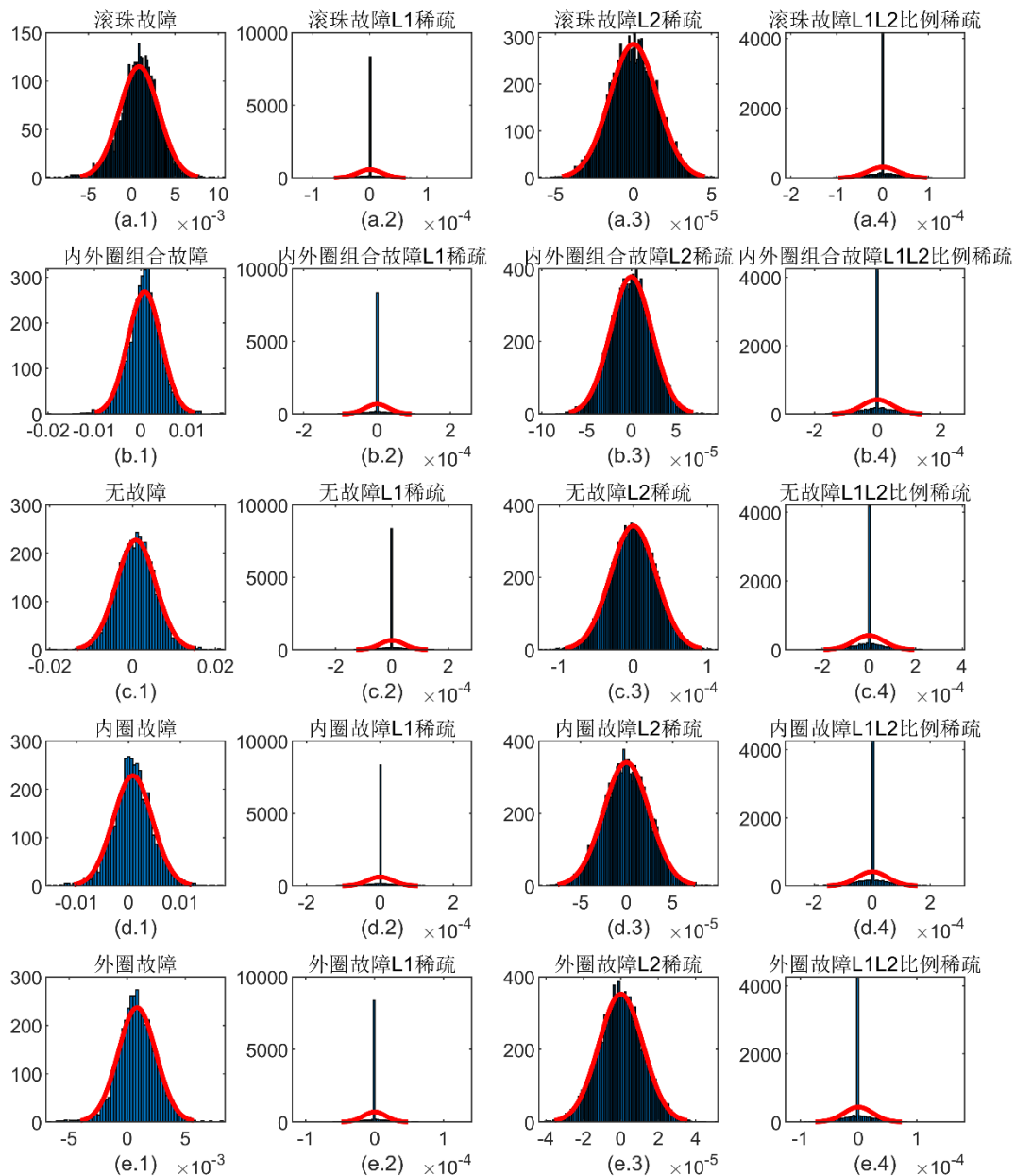


图 5-7 不同故障状态下的密度分布曲线
Figure 5-7 Density distribution under different fault states

图 5-7 从随机信号的概率统计特性来分析稀疏化处理之后的振动信号特点。绘制五类信号的概率直方图并且进行正态拟合，进行了 L2 稀疏化处理的信号与原信号一样，保持了基本的正态分布密度。而 L1 稀疏化处理的信号与 L1L2 组合稀疏处理的信号，在某一频率特别突出，说明了非零元素在时间信号中较少，达到了信号的稀疏化要求。另外，注意到 L1L2 组合稀疏信号优化在提高了计算速度的基础上，保持了 L1 稀疏的单一频率的特点。

除了信号的包络曲线与概率密度的角度，从信号的信息熵的角度也可以用来

度量信号的信息量和随机程度。信号的随机性与不确定性大，其对应的信息熵也就越大。反正，熵值越小，表现出信号的规律性强，被预测的可能性也就越大。式(5-9)给了信息熵的一个定义。

$$H(X) = -\sum_{i=1}^n p(x_i) \log p(x_i) \quad (5-9)$$

其中, $p(x_i)$ 代表了某个信号数值出现的可能概率。表 5-1 分别对获得五类信号的信息熵进行了计算。L2 稀疏信号的复杂度比原始信号还要高, 而 L1 稀疏与 L1L2 组合稀疏均小于原始信号。可见 L1 稀疏对于原始信号的稀疏特征提取是有效的。同时 L1L2 组合稀疏的香农熵处于 L1 稀疏与 L2 稀疏之间, 这也是符合直观感受的。保证了追求信息的稀疏性同时保持高计算特性的目标一致。

表 5-1 故障信号不同类型的香农熵计算结果
Table 5-1 Shannon entropy calculation results of different types of fault signals

	原始信号	L1 稀疏	L2 稀疏	L1L2 组合稀疏
滚珠故障	3.8692	1.9983	4.1533	2.7311
内外圈故障	3.6453	1.9859	3.9646	2.7400
正常状态	3.9030	1.9969	4.1041	2.7552
内圈故障	3.8870	2.0327	4.1106	2.7458
外圈故障	3.8230	1.9367	4.0606	2.6914

5.2 深度迁移方法的设计和实验验证

深度迁移相关的方法相关研究大多数是采用的从机器视觉领域引申出来的结构, 通常会采用基于图像数据的源域。这样的方法固然有它的优越之处, 不用去手工制作预训练模型, 并且预训练模型中的参数也进行了复用。跨领域的知识在这个过程中被学习和记忆, 然而其中的机理并不十分的清晰, 同时其能力和效率也很难以被评估。考虑这上述这些问题, 所以深度迁移的全流程需要被重新讨论和思考。在下面几个小节中, 从深度网络结构设计与迁移策略两个方面分别进行研究和分析其能力和效果。

5.2.1 深度迁移结构的参数设计与结构解释

利用其它领域中的模型直接进行迁移学习是一个不错的思路，然而更加完备的想法应该是依据数据的特征进行网络的设计和预训练。图 5-8 所示的 CNN 网络模型是设计的第一个模型(下面称为 V0)，具有全卷积的设计思路，采用了四路并行的通道，除最后一层之外主要由卷积层和池化层构成。

V0 卷积网络的作为网络结构设计的第一版主要特点是三个层次。第一层次是四个双卷积核的并行，不同的特征图大小获得不同的感受野，多个差异化的特征图能获得对数据较多的理解和认知。第二层采用了池化算子进行操作将特征图大小进行统一，将特征图进行了融合。第三层次主要是常规的全连接层作为分类器这样一个配置。

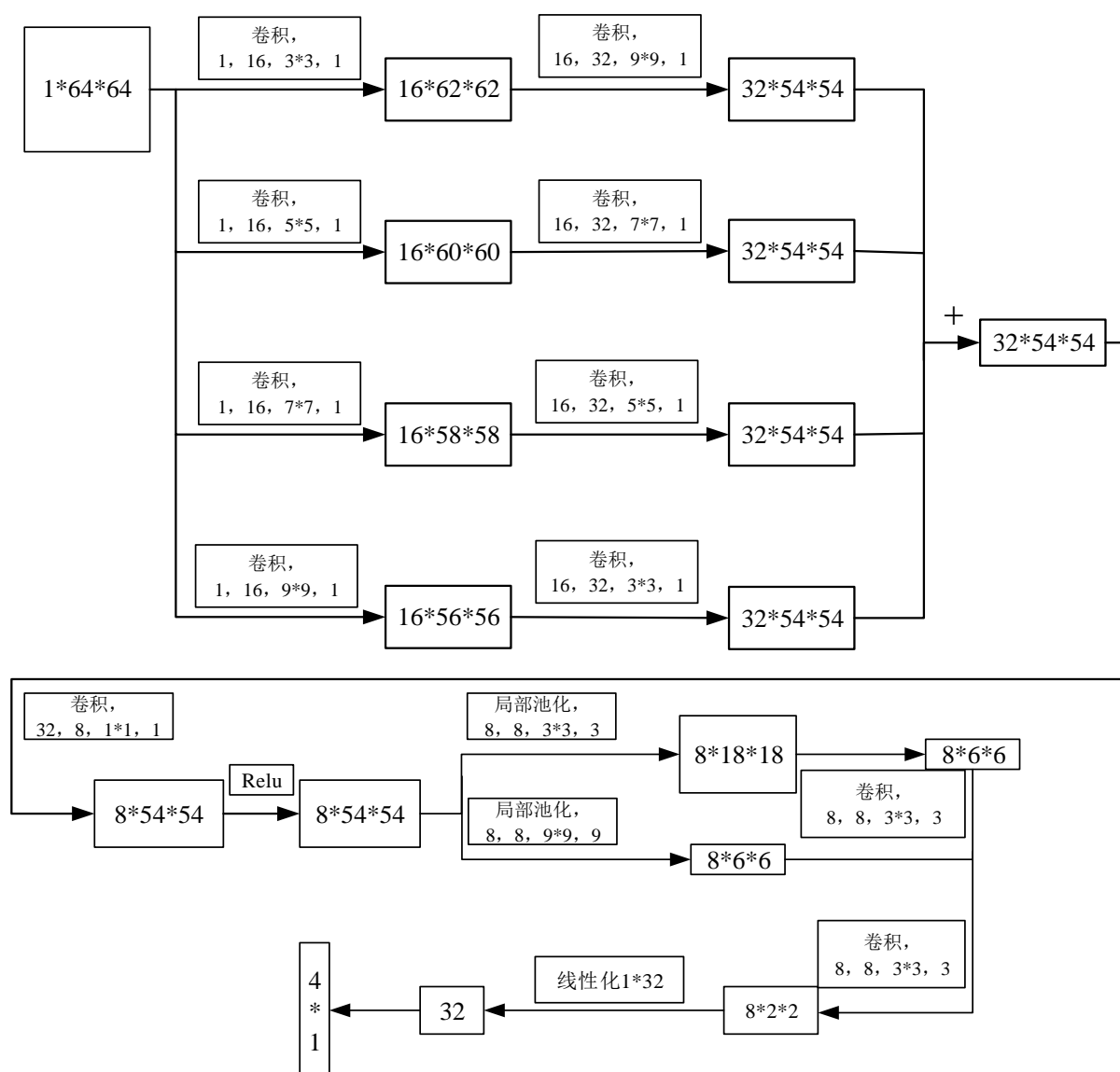


图 5-8 V0 卷积神经网络结构示意图

Figure 5-8 V0 convolutional neural network structure

V0 卷积网络的参数并不多，总计 92245 个可变参数。V0 是以卷积核为主的

计算模型，且全部卷积核尺寸均是方的。同一个通道中进行串联的卷积核，输出核的个数核与输入核的个数相同。V0 网络的设计参数在数量上属于小型网络，计算速度相对来说比较快。然而参数规模的小型化，在某种程度上意味着特征提取的能力不足，某些数据信息密度较高的数据上可能会出现效率较低的情况。同时，在设计 V0 时，刻意将全连接层的参数设计的比较少，使其在整个网络参数中的比例稍小，突出卷积计算的特征。在相同的参数规模下，卷积计算的特征提取能力比全连接计算模式高，所以尽可能多的卷积计算可以作为一个设计的准则和参考。在表 5-2 将 V0 中的全部结构和参数列举了出来。因为多层卷积的存在所以在最后几层加入了特征图数据的归一化，缩减了多层卷积造成的整个网络可能不收敛的风险。

表 5-2 V0 卷积神经网络结构具体参数
Table 5-2 V0 design parameters of convolutional neural network structure

层定义	计算方法	输入核	输出核	核尺寸	核步长
branch1conv1	Conv2d	1	16	3*3	1
branch1conv2	Conv2d	16	32	9*9	1
branch2conv1	Conv2d	1	16	5*5	1
branch2conv2	Conv2d	16	32	7*7	1
branch3conv1	Conv2d	1	16	7*7	1
branch3conv2	Conv2d	16	32	5*5	1
branch4conv1	Conv2d	1	16	9*9	1
branch4conv2	Conv2d	16	32	3*3	1
mainconv1	Conv2d	32	8	1*1	1
mainconv2	Conv2d	8	8	3*3	3
mainconv3	Conv2d	8	8	3*3	3
partpool1	MaxPool2d	8	8	3*3	3
partpool2	MaxPool2d	8	8	9*9	9
bnmain	BatchNorm2d	8			
bnbranch	BatchNorm2d	8			
classifier	Full-connect			32*120*5=19200	

在 V0 网络的基础上，又设计了参数和结构更复杂的 V1 卷积结构。可变参数的总数达到了 358037 个，加入了两个双层的并行卷积设计。同时在串联卷积层之后，进行了归一化，这是因为网络的宽度和深度逐渐增加卷积特征图的奇异性增大，为了避免这种效应必须采用归一化来抵消。

表 5-3 V1 卷积神经网络结构具体参数
Table 5-3 V1 design parameters of convolutional neural network structure

层定义	计算方法	输入核	输出核	核尺寸	核步长
stage1branch1	Conv2d	1	16	3*3	1
	Conv2d	16	32	9*9	1
	BatchNorm2d	32	32		
stage1branch2	Conv2d	1	16	5*5	1
	Conv2d	16	32	7*7	1
	BatchNorm2d	32	32		
stage1branch3	Conv2d	1	16	7*7	1
	Conv2d	16	32	5*5	1
	BatchNorm2d	32	32		
stage1branch4	Conv2d	1	16	9*9	1
	Conv2d	16	32	3*3	1
	BatchNorm2d	32	32		
stage2branch1	Conv2d	32	64	3*3	3
	Conv2d	64	64	3*3	3
	BatchNorm2d	64	64		
stage2branch2	Conv2d	32	64	2*2	4
	Conv2d	64	64	4*4	2
	BatchNorm2d	64	64		
stage2branch3	Conv2d	32	64	5*5	7
	Conv2d	64	64	1*1	1
	BatchNorm2d	64	64		
	Conv2d	64	64	3*3	1
mainbranch	Conv2d	64	128	2*2	2
	AdaptiveMaxPool2d				1
classifier		128*120*0.5(Dropout)*5			

V1 结构的分类器部分的参数占网络总参数的约 10%，比 V0 网络中的占比降低了约 7%。双层的并行卷积结构比单层并行卷积的特征提取能力再一次提高。卷积核尺寸在整体上奇数小卷积的特点中，加入了偶数小卷积核。在卷积计算模块中卷积核的移动步长也不再局限于一步，进行多步设计，增加一部分泛化性能。

V1 结构的双层并行结构之间存在多个特征图融合的过程，提取的特征信息更加丰富。图 5-9 是 V1 结构模型的示意图，可以非常清晰的看到它的各种特点和设计思路。多并行卷积的数目在逐层减少，高维特征的表达也更明确和清晰。

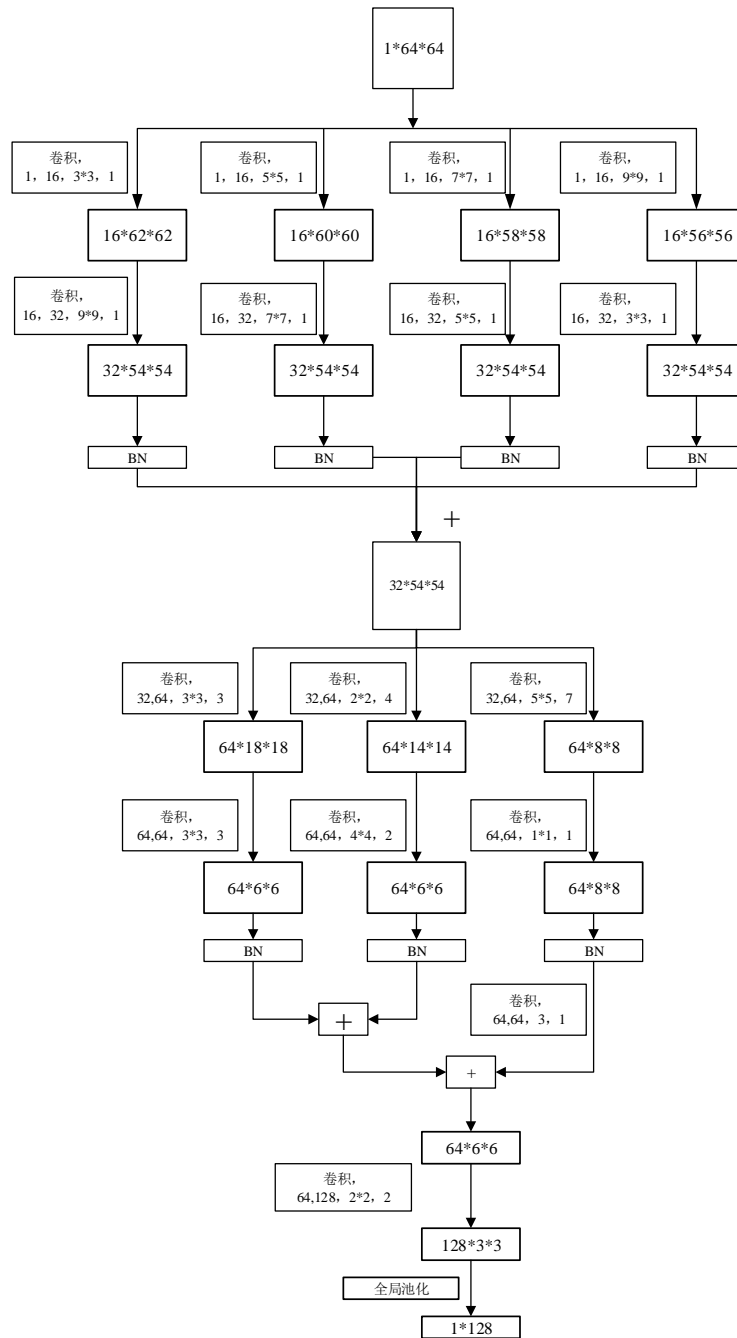


图 5-9 V1 卷积神经网络结构示意图
Figure 5-9 V1 convolutional neural network structure

在 V1 的基础上在两个并行层中加入一个即插即用的注意力模块（V2 本文中称网络），进一步提高整个网络结构的复杂度和计算速度，以期对数据集的认识能力和高阶特征的提取能力进行增强。实际上注意力结构的插入，增加的可变参数量并不多，但是对于神经网络结构的改变有影响。在后续的迁移数值实验中将讨论和证实这一个观点。

5.2.2 神经网络的迁移策略

神经网络的迁移方法主要分为两个方向的研究。其一，是在足量的源域情况下进行网络的训练和测试，达到一定的准确率之后，保存模型。然后数据集小的目标域中进行部分层级网络的微调。通常是用于特征提取的层级进行参数冻结，而作为分类层的全连接层进行训练，达到迁移的目标。其二，是在目标域中进行迁移结构的全部参数进行微调或者训练。后者的待训练参数显然更多，所需的计算力也更高。为了实证相关的思想和方法，制作了两个数据集。在图 5-10 中，将实验台测量的滚动轴承和齿轮箱中的振动数据进行 T-SNE 降维，并进行主成分提取，作出了两个主成分分量的数据分布密度和分布图。

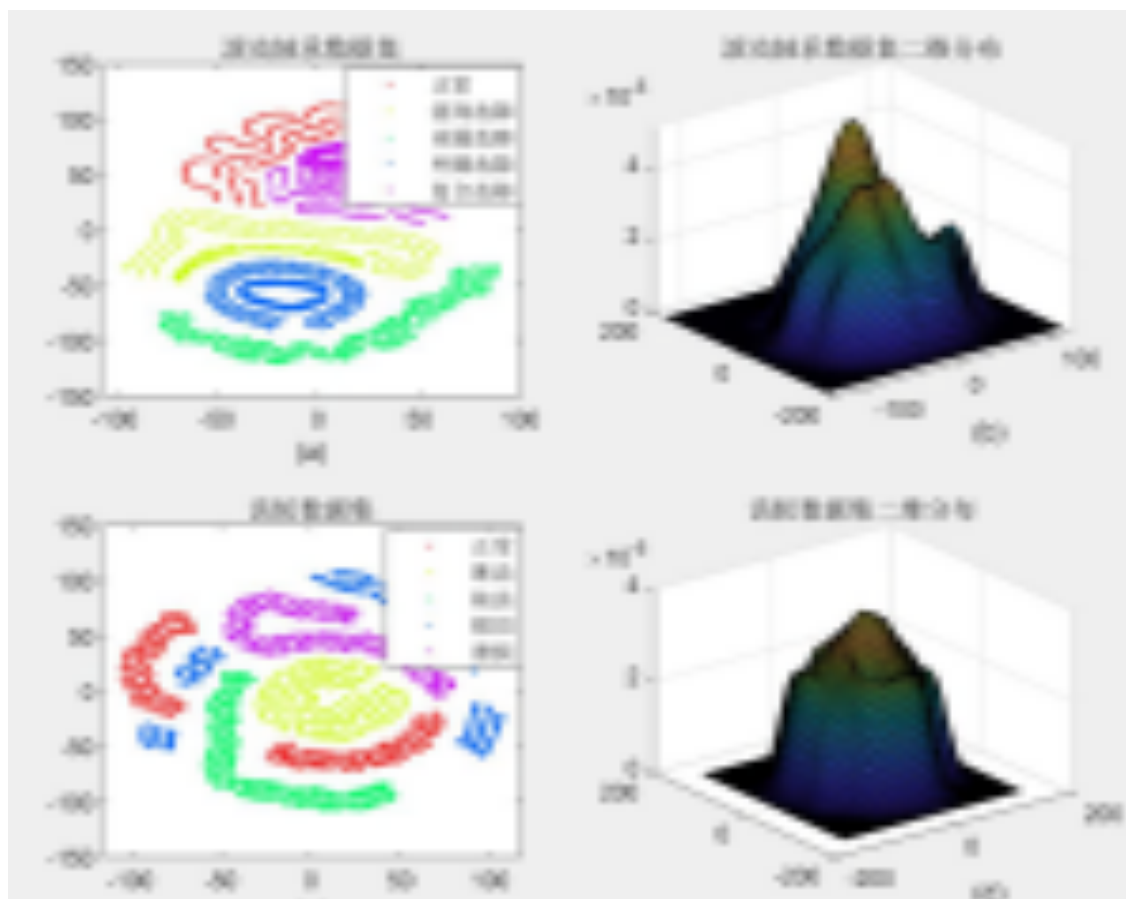


图 5-10 滚动轴承数据集作为源域，齿轮数据集作为目标域

Figure 5-9 Rolling bearing dataset is used as the source domain and gear dataset is used as the target domain

图 5-10 中可以观察到滚动轴承和齿轮数据集的分布特征并不完全相同，存在一定的差异性。尽管是在同一工况下测量得到的数据，但是因为测点的不同分布差异性比较明显。滚动轴承的样本集分布比较集中，而齿轮样本集的分布呈现非

连续性的特点。滚动轴承样本的密度分布曲线具有多峰的特性，齿轮样本的密度分布曲线除了中心之外都比较平缓。

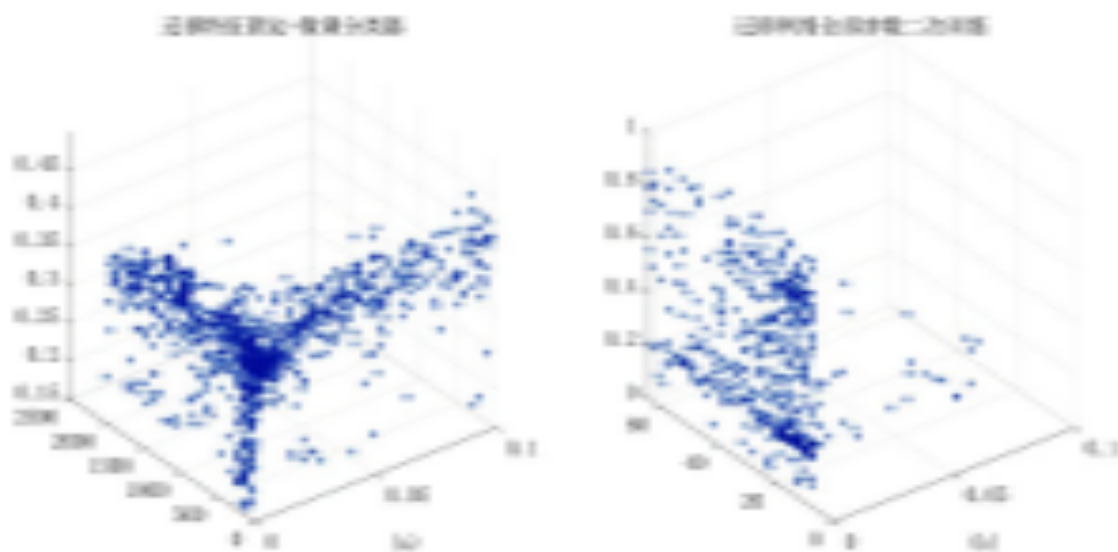


图 5-11 三种卷积神经网络在随机超参数分布下的两种迁移策略效果

Figure 5-11 Results of two transfer strategies of three kinds of convolutional neural networks under random hyper-parameter distributions

以轴承数据集作为源域，齿轮数据集作为目标域。以 V0、V1 和 V2 三种 CNN 模型作为深度迁移中的迁移结构，分别进行 20-500，50-200，100-100（随机抽样实验，20-500 代表进行了 20 次网络训练 500 次）的数值实验。将三种网络结构在微调分类层和训练全部参数两种情况下分别进行大量数值实验，得到图 5-11 所示的结果。其中图 5-11 中的子图坐标分别是学习率、小批次数目和测试准确率。可以发现在源域与目标域样本数量规模相同的情况下，采用 Adam 学习算法在学习率在 0.1 至 0.00001 之间与不同的小批次搜索空间下，两种迁移策略的测试准确率存在明显的差异。特征固定（卷积层参数冻结）加微调的方法，对于样本数据规模变化不大的情形下，效果比较差，其准确率甚至不超过 50%。对于迁移全部网络参数并进行再次训练的方法来说，源域的数据量对于目标域中的测试准确率没有任何影响，并且在某些情况下，可以对目标域中的模型调节起到加速作用。因此，训练迁移结构的全部参数的方法适用性更强，参数训练的效率 and 准确率也更高。

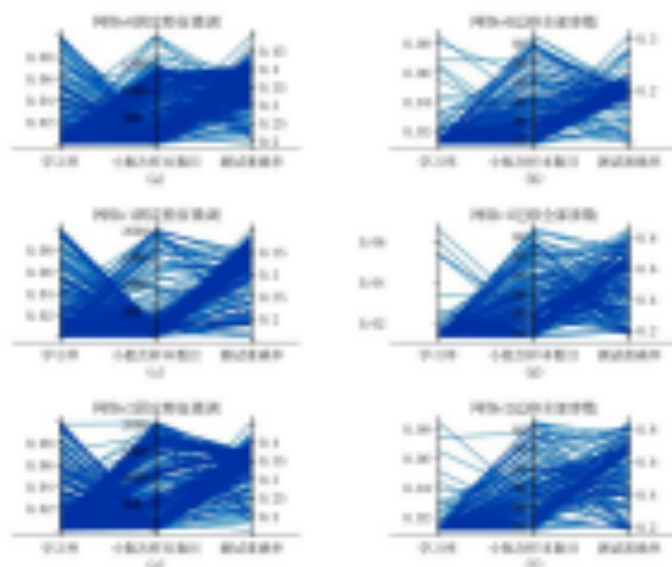


图 5-12 三种网络结构在不同的超参数下的迁移结果对比的平行线图
Figure 5-12 Parallel line graphs comparing the transfer results of three network structures under different hyper-parameters

将 V0、V1 与 V2 进行 10000 步的迭代，并保存作为预训练模型，其参数与网络结构均被存储下来，在三个网络的性能测试阶段重复使用。图 5-12 表示了这三种网络结构在两种迁移策略下的性能表现。图 5-12 的子图(a)(c)(e)的测试准确率均不超过 0.5，与图 5-11 中进行的大规模测试结果是统一的。在网络结构的全部参数迁移方法中 V1 与 V2 网络的准确率可以到达故障识别的要求。V0 没有完成测试目标准确率是可变参数的数量过小造成的，V1 与 V2 的可变参数总量差异不大，尽管 V2 中加入了注意力结构。同时，还发现了学习率对于网络性能的影响，较小的学习率对于找到在计算资源有限情形下最佳的深度模型有倾向性。

5.3 小结

本章节利用了压缩感知的相关理论进行了滚动轴承故障信号的去噪和特征提取，提出了 L1 L2 组合的信号稀疏化方法，并且在不同的故障类型下进行了数值实验从信息量、密度分布、时域曲线几个角度进行了结果比较。在本章节的后半部分，利用了滚动轴承和齿轮的故障数据进行了特征微调方法和参数全迁移方法的数值实验验证和结果分析。针对获取的振动数据，设计了特定的深度学习模型并且进行了迁移策略的方法验证。验证了超参数的变化对不同迁移策略的准确率影响。

6 结论与展望

6.1 结论

6.1.1 主要结论

风力发电作为新能源的一种，对于未来世界的发展有着重要的作用。而用于产生电力的风电机组系统的稳定性和可靠性也就自然很关键。针对风电机组的工况判断与运行监测任务也越来越受到重视，因此提出了一系列的故障诊断和运行监测的方法。针对风电机组运行过程中的振动信号中含有大量噪声的的干扰和齿轮箱中故障情形下振动的问题，结合深度迁移学习方法，做了一些研究和提出了对应的解决方案。重要的研究内容和结论如下：

(1)对现有的针对风电机组运行工况中的状态监测的方法进行了总结和分类，对基信号处理方法与传统机器学习的机械故障诊断方法进行研究，分析其优缺点。研究了将深度学习与迁移理论应用在故障诊断领域的可行性。

(2)针对风电机组齿轮箱中的单齿可能出现的故障进行了有限元的振动模式分析，利用仿真取得的数据结合自编码器的降维能力，提取故障数据的特征进行故障模式的识别和准确率改进。在单一模型可能出现准确率过低的情形下，提出用来评估多个模型效率与性能的方法。

(3)针对齿轮箱中的行星轮系各个部件可能出现的故障类型进行仿真模拟，以断齿和点蚀两种典型齿轮故障为例，提出一种将一维振动特征转化为二维图像特征的方法，并在此基础上利用经典的卷积网络进行了故障类型的分类。考虑到超参数对模型性能的影响，在此过程中为了提高识别准确率，提出了一种逐层冻结参数的有限计算资源的最优超参搜索方法。

(4)将压缩感知方法中的稀疏特征提取方法用在滚动轴承振动信号中，并且应用了L1与L2范数结合的正则化方法，将其中的稀疏化与下采样流程进行了重新设计。在信号进行了稀疏特征提取的基础上，设计了针对滚动轴承数据的模型，

并且制作了预训练模型。实现了从滚动轴承数据到齿轮故障数据的深度迁移学习和模型性能的评价分析。

6.1.2 主要创新点

本文的主要创新点如下：

(1)提出了利用单齿仿真数据进行故障诊断数据集制作，利用自编码器进行降维的特征提取方法，探索了用于评估多个模型效率与性能的方法。

(2)提出了利用行星轮系中各个零件仿真数据实现振动数据升维同时复用经典网络进行故障诊断的方法。同时，提出了逐层冻结网络参数的最优超参数搜索方法。

(3)提出了从滚动轴承到齿轮数据的深度迁移方法。在源域中利用压缩感知方法进行稀疏特征提取，制作预训练模型加速超参数搜索过程，并且比较了特征微调和全局参数调节两种方法在迁移学习中的适用性。

6.2 展望

风电机组的运行状态监测与故障诊断技术在我国仍然处于一个长期需要发展的阶段，与欧美发达国家相比存在技术上不成熟的问题。许多研究方法和研究思路，需要转化成实实在在的工程技术。所以在相当一段时间内，还是需要继续开展研究和发展。

本文中提及到的仿真数据与实验采集数据的融合发展，对于许多工业数据不完整或者无法直接获得工业数据的情形下，是一个很好的补充思路。所以多源数据融合、虚拟仿真数据与工业采集数据融合有很高的应用前景。

将深度学习理论应用在风电机组故障诊断领域，这种模式值得进一步深入挖掘和探索。深度学习领域实际上包含了很多的子领域，例如迁移学习、少样本学习、增量学习（终身学习）、元学习等，将其中的几种方法的优势进行吸收，重新设计模型与参数，这会是一个可以发展和研究的方向。

参考文献

- [1] W.Y. Liu, B.P. Tang, J.G. Han, et al. The structure healthy condition monitoring and fault diagnosis methods in wind turbines:A review[J]. *Renewable and Sustainable Energy Reviews*, 2015 (44)C: 466-472.
- [2] 赵少伟. 云平台下风电机组故障诊断技术研究[D]. 天津理工大学, 2018.
- [3] 鲁尔洁. 面向风电系统的分布式云存储关键技术研究[D]. 电子科技大学, 2017.
- [4] 王稳,李新星. 装配误差引发风电机组齿轮箱失效分析[J]. *自动化应用*, 2018(11):119-121.
- [5] Liang X , Zuo M J , Feng Z. Dynamic modeling of gearbox faults:A review[J]. *Mechanical Systems and Signal Processing*, 2018, 98(1):852-876.
- [6] 陈雪峰, 李继猛, 程航等. 风力发电机状态监测和故障诊断技术的研究与进展[J]. *机械工程学报*, 2011, 47(09):45-52.
- [7] 张霆. 风力发电机叶片状态监测与故障诊断系统设计和实现[D]. 上海电机学院, 2016.
- [8] 郭鼎. 双馈式风电机组发电机故障分析与诊断研究[D]. 华北电力大学(北京), 2016.
- [9] 王世全, 赵志森, 李兴见. 浅谈风电机组发电机对中重要性与对中原理及应用[A]. 中国农业机械工业协会风力机械分会:中国农业机械工业协会风力机械分会, 2018:5.
- [10] 张善秀. 双馈风电机组不对中试验台与故障诊断研究[D]. 北京交通大学, 2019.
- [11] 邵沛泽, 张万军, 冀满忠等. 风力发电机组气动噪声监测与预测的发展现状[J]. *船舶工程*, 2019, 41(S1):295-298+301.
- [12] 覃楚东, 贺石中, 庞晋山等. 基于有限元分析和油液监测的轴承疲劳磨损寿命研究[J]. *润滑与密封*, 2018, 43(07):121-125+156.

-
- [13] 李志农, 张芬, 肖尧先. 基于 CVA-ICA 的机械故障源动态盲分离方法[J]. 机械工程学报, 2015, 51(12):24-29.
- [14] 雷亚国, 杨彬, 杜兆钧等. 大数据下机械装备故障的深度迁移诊断方法[J]. 机械工程学报, 2019, 55(07):1-8.
- [15] 雷亚国, 贾峰, 孔德同等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报, 2018, 54(05):94-104.
- [16] 张贤达. 现代信号处理(第二版)[M]. 清华大学出版社, 2002.
- [17] Tang Y, Blincoe K, Kempa-Liehr A W. Enriching feature engineering for short text samples by language time series analysis[J]. EPJ Data Science, 2020, 9(1):1-59.
- [18] Rik Das. Content-based image classification:efficient machine learning using robust feature extraction techniques[M].CRC Press:2020-10-22.
- [19] Hyun Kwon, Hyunsoo Yoon,Ki-Woong Park. Acoustic-decoy:Detection of adversarial examples through audio modification on speech recognition system[J]. Neurocomputing, 2020, 417:357-370.
- [20] Y. Lin, D. Guo, J. Zhang, et al. A Unified Framework for Multilingual Speech Recognition in Air Traffic Control Systems[J]. IEEE transactions on neural networks and learning systems, 2020:1-13.
- [21] Zhu Y, Mottaghi R, Kolve E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017.
- [22] G.Dalpiaz, A. RivolaI, R. Rubini. Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears[J]. Mechanical Systems and Signal Processing, 2000, 14(3):387-412.
- [23] I. Antoniadou, G. Manson, W.J., et al. A time-frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions[J]. Mechanical Systems and Signal Processing, 2015, (64–65):188–216.
- [24] Z. Feng, M. Liang, Y. Zhang, et al. Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation[J]. Renewable Energy, 2012, (47):112–126.
- [25] Feng Z, Zuo M J, Qu J, et al. Joint amplitude and frequency demodulation analysis based on local mean decomposition for fault diagnosis of planetary

- gearboxes[J]. *Mechanical Systems and Signal Processing*, 2013, 40(1):56-75.
- [26] Liu W Y, Zhang W H , Han J G , et al. A new wind turbine fault diagnosis method based on the local mean decomposition[J]. *Renewable Energy*, 2012, 48:411-415.
- [27] Wang J, Gao R X , Yan R. Integration of EEMD and ICA for wind turbine gearbox diagnosis[J]. *Wind Energy*, 2014, 17(5):757-773.
- [28] Liao S H. Expert system methodologies and applications-a decade review from 1995 to 2004[J]. *Expert Systems with Applications*, 2005, 28(1):93-103.
- [29] Sahin S, Tolun M R, Hassanpour R. Hybrid expert systems: A survey of current approaches and applications[J]. *Expert Systems with Applications*, 2012, 39(4):4609-4617.
- [30] Gelgele H L, Wang K. An expert system for engine fault diagnosis: development and application[J]. *Journal of Intelligent Manufacturing*, 1998, 9(6):539-545.
- [31] Lee H J, Park D Y. A fuzzy expert system for the integrated fault diagnosis[J]. *IEEE Transactions on Power Delivery*, 2000, 15(2):833-838.
- [32] Berredjem T, Benidir M. Bearing faults diagnosis using fuzzy expert system relying on an Improved Range Overlaps and Similarity method[J]. *Expert Systems with Applications*, 2018, 108(OCT.):134-142.
- [33] Jayaswal P, Verma S, Wadhvani A. Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis[J]. *Journal of Vibration & Control*, 2011, 17(8):1131-1148.
- [34] Merainani, Boualem, Rahmoune, et al. A novel gearbox fault feature extraction and classification using Hilbert empirical wavelet transform, singular value decomposition, and SOM neural network[J]. *Journal of vibration and control: JVC*, 2018:2512-2531.
- [35] Chen X, Zhou J, Xiao J, et al. Fault diagnosis based on dependent feature vector and probability neural network for rolling element bearings[J]. *Applied Mathematics and Computation*, 2014, 247:835-847.
- [36] Zhong B, Macintyre J, He Y, et al. High Order Neural Networks for Simultaneous Diagnosis of Multiple Faults in Rotating Machines[J]. *Neural Computing and Applications*, 1999, 8(3):189-195.
- [37] Pang B, Tang G, Zhou C, et al. Rotor Fault Diagnosis Based on Characteristic

-
- Frequency Band Energy Entropy and Support Vector Machine[J]. *Entropy*, 2018, 20(12).
- [38] Yuan S F, Chu F L. Support vector machines-based fault diagnosis for turbo-pump rotor[J]. *Mechanical Systems and Signal Processing*, 2006, 20(4):939-952.
- [39] Zhang X L, Chen W, Wang B J, et al. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization[J]. *Neurocomputing*, 2015, 167:260-279.
- [40] He D, Li R, Zhu J. Plastic Bearing Fault Diagnosis Based on a Two-Step Data Mining Approach[J]. *IEEE Transactions on Industrial Electronics*, 2013, 60(8):3429-3440.
- [41] Jiang L, Xuan J, Shi T. Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2013, 41(1-2):113-126.
- [42] Su H, Chong K T. Induction Machine Condition Monitoring Using Neural Network Modeling[J]. *IEEE Transactions on Industrial Electronics*, 2007, 54:241-249.
- [43] Li B, Chow M Y, Tipsuwan Y, et al. Neural-network-based motor rolling bearing fault diagnosis[J]. *IEEE Transactions on Industrial Electronics*, 2002, 47(5):1060-1069.
- [44] Samanta B, Al-Balushi K R. Artificial neural network-based fault diagnostics of rolling element bearings using time-domain features [J]. *Mechanical Systems and Signal Processing*, 2003, 17(2):317-328.
- [45] Aminian M, Aminian F. Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor[J]. *IEEE Transactions on Circuits and Systems II Analog and Digital Signal Processing*, 2000, 47(2):151-156.
- [46] Sun W, Shao S, Zhao R, et al. A sparse auto-encoder-based deep neural network approach for induction motor faults classification[J]. *Measurement*, 2016:171-178.
- [47] Lu C, Wang Z Y, Qin W L, et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification[J]. *Signal*

- Processing, 2017, 130(1):377-388.
- [48] Tao S, Zhang T, Yang J, et al. Bearing fault diagnosis method based on stacked autoencoder and softmax regression[C]. IEEE Control Conference, 2015.
- [49] F. Jia, Y. Lei, J. Lin, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data[J]. Mechanical Systems and Signal Processing, 2016, 72:303–315.
- [50] Sun J, Yan C, Wen J. Intelligent Bearing Fault Diagnosis Method Combining Compressed Data Acquisition and Deep Learning[J]. IEEE Transactions on Instrumentation and Measurement, 2017, 67(99):185-195.
- [51] F. Zhou, Y. Gao, C. Wen. A novel multimode fault classification method based on deep learning[J]. Journal of Control Science and Engineering, 2017:1-14.
- [52] Junbo T, Weining L, Juneng A, et al. Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder[C]. The 27th Chinese Control and Decision Conference (2015 CCDC), IEEE, 2015.
- [53] L. Guo, H. Gao, H. Huang, et al. Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring[J], Shock and Vibration, 2016:10.
- [54] Cheng F, Wang J, Qu L, et al. Rotor current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier[C]. Industry Applications Society Meeting, IEEE, 2017.
- [55] Chen R, Chen S, He M, et al. Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion[J]. Journal of Risk and Reliability, 2017, 231(6): 666-679.
- [56] R. Liu, G. Meng, B. Yang, et al. Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine[J]. IEEE Transactions on Industrial Informatics, 2017, 13(2):1310–1320.
- [57] Wang P, Ananya, Yan R, et al. Virtualization and deep recognition for system fault classification[J]. Journal of Manufacturing Systems, 2017, 44(c):310-316.
- [58] C. Lu, Z. Wang, B. Zhou. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network-based health state classification[J]. Advanced Engineering Informatics, 2017, 32:139–151.
- [59] Ding X, He Q. Energy-Fluctuated Multiscale Feature Learning With Deep

-
- ConvNet for Intelligent Spindle Bearing Fault Diagnosis[J]. *IEEE Transactions on Instrumentation and Measurement*, 2017, 8:1-10.
- [60] X. Guo, L. Chen, C. Shen. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis[J]. *Measurement*, 2016, 93:490–502.
- [61] W. Sun, B. Yao, N. Zeng, et al. An intelligent gear fault diagnosis methodology using a complex wavelet enhanced convolutional neural network[J]. *Materials*, 2017, 10 (7):790.
- [62] Z. Chen, C. Li, R.V. Sanchez. Gearbox fault identification and classification with convolutional neural networks[J]. *Shock and Vibration*, 2015, 10:390134.
- [63] Weimer D, Scholz-Reiter B, Shpitalni M. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection[J]. *Cirp Annals Manufacturing Technology*, 2016, 65(1):417-420.
- [64] H. Y. Dong, L. X. Yang, H. W. Li. Small fault diagnosis of front-end speed-controlled wind generator based on deep learning[J]. *Wseas Transactions on Circuits and Systems*, 2016, 15:64–72.
- [65] W. You, C. Shen, X. Guo, et al. A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery[J]. *Advances in Mechanical Engineering*, 2017,9(6):168–176.
- [66] Lee K B, Cheon S, Kim C O. A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes[J]. *IEEE Transactions on Semiconductor Manufacturing*, 2017, 30(2):135-142.
- [67] M. Yuan, Y. Wu, L. Lin. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network[C]. *IEEE International Conference on Aircraft Utility Systems(AUS)*, 2016, 135–140.
- [68] R. Zhao, R. Yan, J. Wang, et al. Learning to monitor machine health with convolutional bi-directional lstm networks[J]. *Sensors*. 2017, 17(2):273.
- [69] R. Zhao, D. Wang, R. Yan, et al. Machine health monitoring using local feature-based gated recurrent unit networks[J]. *IEEE Transactions on Industrial Electronics*, 2017, 65(2):1539–1548.
- [70] Jiang N, Ning L I. Wind turbine frequent principal fault detection based on a self-attentive LSTM encoder-decoder model[C]//2020 39th Chinese Control Conference (CCC). 2020.

- [71] 石祥滨, 房雪键, 张德园等. 基于深度学习混合模型迁移学习的图像分类[J]. 系统仿真学报, 2016, 28(01):167-173+182.
- [72] 张博, 史忠植, 赵晓非等. 一种基于跨领域典型相关性分析的迁移学习方法[J]. 计算机学报, 2015, 38(07):1326-1336.
- [73] 于重重, 田蕊, 谭励等. 非平衡样本分类的集成迁移学习算法[J]. 电子学报, 2012, 40(07):1358-1363.
- [74] 张倩, 李明, 王雪松等. 一种面向多源领域的实例迁移学习[J]. 自动化学报, 2014, 40(06):1176-1183.
- [75] 洪佳明, 印鉴, 黄云等. TrSVM:一种基于领域相似性的迁移学习算法[J]. 计算机研究与发展, 2011, 48(10):1823-1830.
- [76] 沈飞, 陈超, 严如强. 奇异值分解与迁移学习在电机故障诊断中的应用[J]. 振动工程学报, 2017, 30(01):118-126.
- [77] 陈仁祥, 杨星, 胡小林等. 深度置信网络迁移学习的行星齿轮箱故障诊断方法[J]. 振动与冲击, 2021,40(01):127-133+150.
- [78] 李玉博, 张景景, 韩承桓等. 基于分圆类构造卷积压缩感知测量矩阵[J]. 电子与信息学报, 2021, 43(02):419-425.
- [79] 郭俊锋, 王茁. 一种基于多测量向量模型的机械振动信号联合稀疏重构方法[J]. 振动与冲击, 2021, 40(01):254-263.
- [80] 丁剑飞, 孙德春, 李兆刚. 基于压缩感知的电力线脉冲噪声抑制改进算法[J]. 计算机应用研究, 2020, 37(S2):289-290.
- [81] 赵晨, 罗自炎, 修乃华. 稀疏优化理论与算法若干新进展[J]. 运筹学学报, 2020, 24(04):1-24.
- [82] 王鹤, 李石强, 于华楠等. 基于分布式压缩感知和边缘计算的配电网电能质量数据压缩存储方法[J]. 电工技术学报, 2020, 35(21):4553-4564.
- [83] 王雪琼, 郭静波. 动态压缩感知理论研究综述[J]. 电子测量与仪器学报, 2020, 34(10):1-16.



致谢

时间转瞬即逝，原来以为很漫长的三年时间没想到会这么快。至 2021 年，在学校的日子接近二十载，总计约为平均人生长度的四分之一。在大学里的七年时光是我求学路途之中最快乐的一段。在这里，学到了很多有用的知识和技能，也让我明白了我需要的是什么。感谢这所大学，让我知道了我所要抵达的海岸。从某种程度上来说，学业生涯给与我最大的意义就在于此。在本科和研究生阶段遇到的各种特点和专长的老师给了我很大的帮助，在这里非常感谢老师们的帮助。

研究生三年，更要感谢我的导师刘文艺老师。刘老师对我的毕业论文进行了非常耐心的指导与交流。对于我的课题和相关的研究方向每次都能探讨和交流，表明了导师对于当前学术课题和潮流的把握。师生本是偶然，与导师年龄相差十岁，许多思路和方法都能互相交流，极其难得。在这里，向导师表达敬意。

能完成学业，父亲母亲的支持与家人的鼓励是其中比较重要的原因。读书之于我，更多的是一种运气。外婆一直对于我的学业非常关注，每次想要放弃或者退出外婆总能给出很多理由和建议。我的家庭的心态对我的学业也很重要，因为对我没有过高的期望，所以我的学业生涯一直比较轻松快乐，这也是我保持对知识的兴趣、对自由的无限追求、对公平的执着的重要原因。希望自己努力的成为合格的现代社会的公民。

还要感谢任贺师兄、同门单梦晨、师弟顾恒和张岩，大家之间相处的极其融洽。此外，18 级的其他小伙伴也是对我的学习和生活有很大的帮助。室友李大权、王伟以及高帅，同样给与生活增添了无限的乐趣与魅力。最后，感谢 Dr.Z 一年的陪伴，让我对学业还有职业的选择更加坚定。感谢季岸大学的伙伴们，他们同样也很重要，辅导员、白鱼（大黄鱼）、燕子、类脑哥、故障哥、筑波、张北大、小混沌、曾博、火哥、诺哥、猫哥、龟龟、自闭哥、熊猫哥、弟中弟、171、圆圆、车神、Lis、明斯特等一批热心人士给与了我无限的力量。

最后特别感谢参加论文评审和答辩的各位老师和专家们。感谢江苏师范大学。

作者简介

一、基本情况

姓名：王欣 性别：男 民族：汉 出生年月：1994-10 籍贯：江苏省徐州市
2014.09-2018.06 江苏师范大学科文学院工学学士
2018.09-2021.06 江苏师范大学机电工程学院工学硕士

二、学术论文及专利

1. **Wang X.**, Liu W.Y., Zhang Y, Gu H. A novel wind turbine fault diagnosis method based on generator current analysis[J]. Journal of Vibroengineering. 2020, 22 (8): 1758-1769.(EI 检索: 20210309788780)

2. **X. Wang**, W. Liu and M. Shan, Fault Diagnosis Analysis of Wind Turbine Gear Based on Transfer Function Model[C]. 2019 Prognostics and System Health Management Conference (PHM-Qingdao), 2019:1-5. (EI 检索: 20200408061025)

3. Shan M, Liu W, **Wang X**, et al. A novel Robot Joint Bearing fault diagnosis method based on VMD in BP neural network[C]. International conference on ubiquitous robots and ambient intelligence, 2019: 760-764.(EI 检索: 20193307314020)

4. He Ren, Wenyi Liu, Mengchen Shan, **Xin Wang**. A new wind turbine health condition monitoring method based on VMD-MPE and feature-based transfer learning[J]. Measurement, 2019, 148. (SCI 检索: 000487930000022)

三、获奖情况

获 2018-2021 学年度研究生二等学业奖学金

四、研究项目

江苏师范大学研究生科研创新计划项目：风电机组故障诊断仿真研究，项目编号：2019XKT161，主持。

国家自然科学基金：基于多维核域谱的风机传动链系统微征兆健康状态识别方法，项目编号：51505202，参与。

江苏省基础研究计划（自然科学基金）项目：基于深度迁移学习的风电机组故障诊断方法研究，项目编号：BK20201463，参与。