



上海大学  
SHANGHAI UNIVERSITY

# 本科毕业论文（设计）

UNDERGRADUATE THESIS (PROJECT)

题 目 基于深度迁移学习的故障诊断研究

学 院 机电工程与自动化

专 业 自动化

学 号 19121802

学生姓名 丁一平

指导教师 修贤超

起讫日期 2023.2.15 ~ 2023.6.1



姓名: 丁一平  
论文题目: 基于深度迁移学习的故障诊断研究

学号: 19121802

## 原创性声明

本人声明: 所呈交的论文是本人在指导教师指导下进行的研究工作。除了文中特别加以标注和致谢的地方外, 论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签名: \_\_\_\_\_ 日期: \_\_\_\_\_

## 本论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定。即: 学校有权保留论文及送交论文复印件, 允许论文被查阅和借阅; 学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

签名: \_\_\_\_\_ 指导教师签名: \_\_\_\_\_ 日期: \_\_\_\_\_



# 上海大学

SHANGHAI UNIVERSITY

## 本科毕业论文（设计）

UNDERGRADUATE THESIS (PROJECT)

### 基于深度迁移学习的故障诊断研 究

学 院 \_\_\_\_\_ 机电工程与自动化  
专 业 \_\_\_\_\_ 自动化  
学 号 \_\_\_\_\_ 19121802  
学生姓名 \_\_\_\_\_ 丁一平  
指导教师 \_\_\_\_\_ 修贤超  
起讫日期 \_\_\_\_\_ 2023.2.15 ~ 2023.6.1

## 摘要

随着经济的高速发展和科学技术的不断提升，工业系统越来越大型化、复杂化、智能化，使得工业故障的发生几率大幅上升，并可能导致严重后果。因此，准确有效地进行故障诊断对于工业生产的安全和效率至关重要，已经成为学术界和工业界的研究热点。本文针对故障诊断实际工程问题，探讨了基于深度迁移学习的故障诊断方法，以正确分类和识别所选数据集为导向，设计了多个神经网络模型，主要工作如下：

(1) 针对深度迁移学习中的特征提取问题，采用了卷积神经网络作为故障特征提取工具。以 VGG11 为基础，构建了基于 ImageNet 预训练的迁移模型，并根据模型调整故障数据的形式，对模型进行故障分类验证实验，实现了有效的故障分类。同时为了进一步提高故障分类效率，讨论了模型不同参数对故障分类结果的影响。

(2) 针对深度迁移学习中的数据不均衡问题，采用了生成对抗网络作为数据增强办法，对原始的故障数据进行增强。根据故障数据的形式，分别以人工神经网络和卷积神经网络为基础，构建了两种生成对抗网络，用训练好的生成器对少数样本进行增强，并通过使用同一模型将原始数据与增强后的数据进行测试对比，验证了生成对抗学习能够提高故障分类少数样本中的特征提取效果，从而实现了更准确的故障分类。

(3) 针对故障诊断的变工况问题，采用深度域适应将不同工况的数据进行迁移。根据故障数据的形式，构建了基于 VGG11 和 ResNet50 两种卷积神经网络作为骨干模型，对 3 种轴承状态互相进行领域对齐，得到了高精度的迁移结果，同时讨论不同模型骨干和不同领域对于迁移学习结果的影响。

**关键词:** 故障诊断；深度学习；迁移学习；神经网络

## ABSTRACT

With the rapid development of the economy and continuous improvement of science and technology, industrial systems are becoming larger, more complex, and more intelligent, resulting in a significant increase in the occurrence probability of industrial faults and potential serious consequences. Therefore, accurate and effective fault diagnosis is crucial for ensuring the safety and efficiency of industrial production and has become a research hotspot in both academia and industry. This paper addresses practical engineering problems in fault diagnosis and explores fault diagnosis methods based on deep transfer learning. Multiple neural network models are designed with a focus on correctly classifying and identifying the selected datasets. The main contributions are as follows:

(1) In order to tackle the feature extraction issue in deep transfer learning, convolutional neural networks (CNNs) are adopted as tools for fault feature extraction. Based on the VGG11 architecture, a transfer model pretrained on ImageNet is constructed. The form of the fault data is adjusted accordingly, and experiments are conducted to validate the effectiveness of fault classification. Additionally, the impact of different model parameters on fault classification results is discussed to enhance efficiency.

(2) To address the problem of data imbalance in deep transfer learning, generative adversarial networks (GANs) are employed as a data augmentation method to enhance the original fault data. Two types of GANs are constructed based on artificial neural networks and CNNs, respectively. The trained generators are used to augment minority samples. By comparing the original data with the augmented data using the same model, it is verified that generative adversarial learning improves feature extraction in minority samples, thereby achieving more accurate fault classification.

(3) To tackle the issue of varying operating conditions in fault diagnosis, deep domain adaptation is utilized to transfer data from different operating conditions. Based on the VGG11 and ResNet50 architectures as backbone models, three different bearing states are aligned to obtain highly accurate transfer results. Furthermore, the influence of different backbone models and domains on the results of transfer learning is discussed.

**Keywords:** Fault diagnosis; Deep learning; Transfer learning; Neural networks

# 目 录

摘要 .....	I
ABSTRACT .....	i
<b>第一章 绪论 .....</b>	<b>1</b>
1.1 研究背景与意义 .....	1
1.2 故障诊断技术研究现状 .....	1
1.2.1 基于信号处理的方法 .....	2
1.2.2 基于解析模型的方法 .....	3
1.2.3 基于机器学习的方法 .....	3
1.3 深度学习研究现状 .....	4
1.3.1 前馈神经网络 .....	4
1.3.2 卷积神经网络 .....	5
1.3.3 递归神经网络 .....	5
1.3.4 生成对抗网络 .....	5
1.4 迁移学习研究现状 .....	6
1.4.1 基于特征的迁移学习 .....	8
1.4.2 基于样本实例的迁移学习 .....	8
1.4.3 基于模型的迁移学习 .....	9
1.5 论文研究内容和整体结构 .....	9
<b>第二章 基于卷积神经网络的模型迁移方法 .....</b>	<b>11</b>
2.1 卷积神经网络 .....	11
2.1.1 卷积层 .....	12
2.1.2 池化层 .....	13
2.1.3 全连接层 .....	14
2.1.4 损失函数 .....	14
2.1.5 梯度下降 .....	15
2.2 模型迁移 .....	15
2.2.1 预训练模型 .....	15
2.2.2 微调 .....	16
2.3 实验设计 .....	17
2.3.1 实验数据集简介 .....	17

2.3.2 数据预处理 .....	18
2.3.3 网络模型的选取 .....	19
2.4 实验结果与分析 .....	20
2.5 本章小结 .....	23
<b>第三章 基于生成对抗网络的不均衡数据增强方法 .....</b>	<b>24</b>
3.1 数据增强 .....	24
3.2 生成对抗网络 .....	25
3.2.1 网络框架 .....	25
3.2.2 目标损失参数 .....	25
3.3 基于 GAN 的数据增强能力分析 .....	27
3.3.1 模型结构 .....	27
3.3.2 实验结果与分析 .....	28
3.4 基于 DCGAN 的数据增强能力分析 .....	31
3.4.1 模型结构 .....	31
3.4.2 实验结果与分析 .....	31
3.5 本章小结 .....	34
<b>第四章 基于深度域适应的变工况故障诊断方法 .....</b>	<b>35</b>
4.1 变工况故障诊断 .....	35
4.2 领域自适应 .....	36
4.3 深度适配网络 .....	37
4.3.1 基于分布差异的方法 .....	38
4.3.2 基于对抗学习的方法 .....	40
4.3.3 基于重构和基于生成的方法 .....	41
4.4 实验设计与分析 .....	42
4.4.1 数据预处理及划分 .....	42
4.4.2 模型结构 .....	42
4.5 本章小结 .....	43
<b>第五章 总结与展望 .....</b>	<b>44</b>
5.1 工作总结 .....	44
5.2 未来展望 .....	44
<b>插图索引 .....</b>	<b>46</b>
<b>表格索引 .....</b>	<b>47</b>
<b>参考文献 .....</b>	<b>48</b>

致 谢 .....	51
-----------	----

# 第一章 绪论

本章介绍了本课题的研究背景、意义与研究现状，并对本文的主要研究内容和整体结构进行了概述。

## 1.1 研究背景与意义

随着经济的高速发展以及科学技术的提升，工业系统朝着大型化、复杂化、智能化的方向不断升级。为了提高了工业水平满足更加智能和精细的工业需求，装备工业系统的集成组织规模不断扩大，单一装备的结构和组成变得更加复杂，不同装备之间的关联也变得更加紧密，系统分布也更加密集。由于复杂的机械结构和多变的运行工况，设备中的核心部件，如齿轮、轴承、和主轴等，可能会随着时间的推移，发生不同程度的损伤甚至断裂，从而引发故障<sup>[1,2]</sup>。特别是在实际生产中，一些工业设备往往需要工作在重载、高温、腐蚀以及高作业率的恶劣环境中，这极大增加了工业故障发生的几率。工业故障发生时，轻则导致机械加工偏离指定指标，造成生产进度降低、良品率下降；重则损坏设备，造成设备停产，造成重大的经济损失，甚至会威胁到人员的生命安全。

2015年8月12日晚，位于天津市滨海新区的瑞海公司危险品仓库运抵区起火，并接连发生两次爆炸。事故现场形成6处大火点及数十个小火点。事故造成165人遇难，8人失踪，直接经济损失68.66亿元人民币；2018年11月28日，张家口发生爆燃事故，造成24人死亡、22人受伤，直接经济损失4148.86万元人民币；2021年2月8日，辽宁康缘华威药业有限公司发生爆炸事故，造成5人受伤，其中2人经抢救无效死亡。

正是由于工业流程中设备故障会造成严重的影响，开展工业设备状态监测和故障诊断研究，对设备可能发生的故障进行检测、定位、诊断和预测，以“防范于未然”，对保证机械设备的安全和稳定运行，减少经济损失和运行成本，避免重大事故的发生，是保证设备运行安全稳定的基础，具有重要的实际需求和现实意义。

## 1.2 故障诊断技术研究现状

故障诊断技术根据设备运行中采集和存储的数据，通过对于设备进行建模、专家知识以及人工智能技术等手段，对于运行设备进行状态监测、故障诊断以及综合管理。良好的状态监测可以保证设备正常稳定的运行，而通过对设备各关键变量的

预测，能够对设备及其各组件的预期寿命进行估计，使得对设备的维护更有针对性，降低整体运维的成本，故障诊断则对于设备的受损和运行的影响情况进行评估，进行相应的处置，使设备恢复正常工作状态。

作为当前控制领域的国际研究热点之一，故障诊断是保证工业生产安全的必要手段。故障诊断的目的是通过对传感器获取的数据进行分析，对系统中存在的故障进行检测、判别，确定故障位置以及故障的严重程度；同时对可能发生的故障情况进行预测及预防。

故障是指系统中的零件或部件丢失或性能指标低于标准的事件或现象，包括执行器堵塞、传感器信号丢失，如传感器卡在特定值或传感器恒定参数发生变化，或系统组件断开等情况。这些故障可能会中断控制器对对象的控制动作，或者产生超出限定的测量误差，抑或是直接改变系统的动态输入或输出特性，导致系统性能下降，甚至驱使整个系统受到损坏、崩溃。为了提高系统的可靠性，通常采用故障诊断技术对系统出现的故障进行监测、定位和识别。

故障诊断技术是一门正在高速发展的新兴学科，自 20 世纪 60 年代美国为其国家宇航局专门创立故障预防小组以来，故障诊断技术逐渐成为结合了数学、物理、现代控制理论、模式识别、人工智能、神经网络等技术的综合学科。故障诊断技术在我国的化工、冶金、电力、铁路等行业得到广泛应用，取得了良好效果，尤其是在电子计算机普及后，故障诊断技术已经进入了智能化应用的阶段<sup>[3]</sup>。

早期的故障检测法大多凭借着工人积累的经验进行故障的诊断，如音棒法、观察法等，但随着传感器技术和故障诊断技术的不断发展改进，仅仅依靠人工经验已无法再满足故障的诊断<sup>[4]</sup>。目前较为流行的故障诊断方法多是通过传感器对设备工作状态进行分析，常见的有：基于信号处理的方法、基于解析模型的方法和基于机器学习的故障诊断方法，如图 1.1 所示<sup>[5]</sup>。

### 1.2.1 基于信号处理的方法

基于信号处理的方法是故障诊断中的传统方法之一，它主要通过分析特定数据和提取信号特征的方法来实现故障诊断。当传感器采集到振动信号后，由于其特征并不明显且可能受到噪声的影响，所以需要对信号进行处理并提取特征，以便进一步对故障进行分类。这种方法通常利用谱分析法、相关函数、小波变换等信号模型，从时域、频域、时频域等对信号进行特征提取，分析相关信号并判断是否发生故障。这种方法对数学模型的依赖较少，能在信号较少的情况下得到较优的诊断结果，因此在一些特定领域中仍然在使用这种方法，如 Ge 等人提出了一种新的多通道轴承信号故障诊断方法，结合了振动信号的峭度指标和峰值频率并与理论故障相关频率进行了比较，最后对轴承的故障进行了诊断<sup>[6]</sup>。

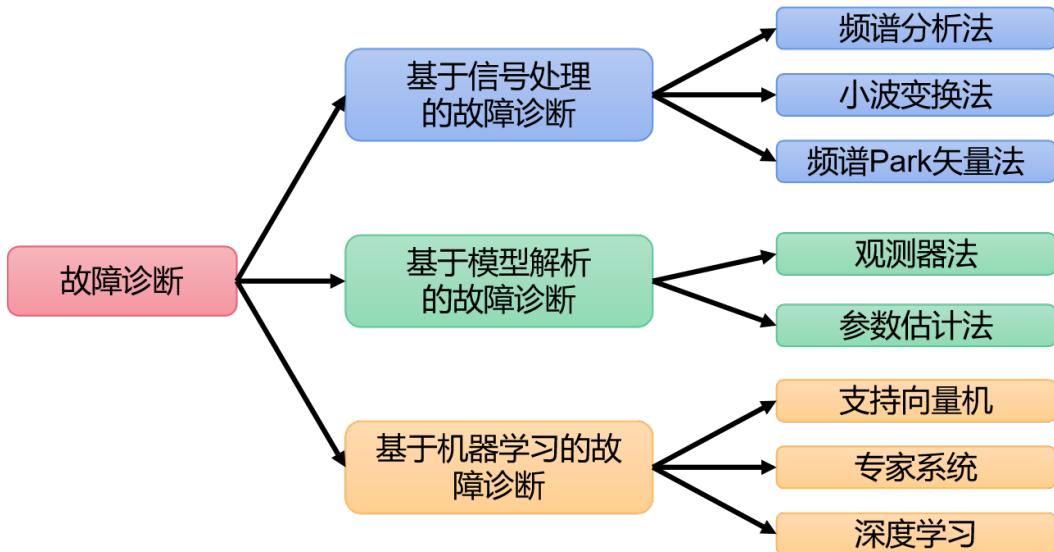


图 1.1 故障诊断分类

然而，基于信号的方法存在着许多不足。一方面其没有精确的数学模型，使得它对于早期潜在故障的诊断能力不足，更依赖于专家的经验，另一方面，在一些噪声干扰较大的情况下，基于信号的方法的诊断准确度也会受到一定的影响。

### 1.2.2 基于解析模型的方法

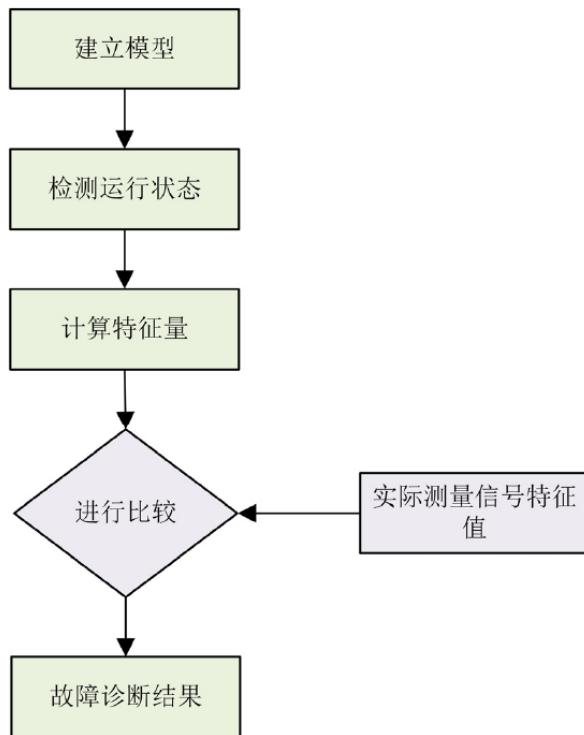
基于解析模型的故障诊断方法，它通过建立数学模型来进行故障诊断的方法，诊断过程如图 1.2 所示。它基于物理模型，通过对模型的状态和行为进行建模和分析，来检测和诊断系统的故障。这种方法具有较高的可靠性和准确性，能够有效地识别和分析潜在的故障，并且可以在早期阶段发现故障。

基于解析模型的故障诊断方法一般需要较为精确的物理模型，并且需要对模型进行状态观测和控制。这种方法的局限性在于需要较为精确的数学模型，并且对模型的参数进行估计和校准也需要较高的精度。此外，对于复杂的系统，建立数学模型也会比较困难，因此在实际应用中需要针对具体情况选择。

### 1.2.3 基于机器学习的方法

基于机器学习的故障诊断方法主要是通过机器学习算法，学习故障样本的特征并对故障进行分类。它与传统的基于信号和解析模型的方法相比，具有更强的自适应性和泛化能力，能够自动学习故障的特征，从而更好地适应不同的故障场景。

这种方法的主要优点是可以处理多变量、高维度的数据，并能够自适应地适应不同的故障场景。与传统的基于信号和解析模型的方法相比，它需要更少的先验知识和手工特征工程，并且可以更好地应对故障模式的变化。另外，它还可以实现在线学习，随着时间的推移不断提高故障诊断的准确率和可靠性。但是与基于信号和



**图 1.2 基于解析模型的故障诊断过程**

解析模型的方法相比，基于机器学习的方法需要大量的训练数据和时间来训练模型，并且对于新的故障类型，需要重新训练模型。此外，在实际应用中，机器学习模型的解释性较差，因此对于一些关键应用场景的故障诊断，需要与基于解析模型和基于信号的方法相结合，以保证故障诊断的可靠性和准确性。

机器学习作为故障诊断的一个重要手段，无论是在理论体系的构建方面，还是在实际问题的应用方面，都有很大的发展空间。

### 1.3 深度学习研究现状

深度学习通过模拟神经元的结构和功能，将多个神经元之间进行连接，构建深层神经网络，并模拟人类大脑处理信息的方式进行模式识别和数据分析，从而获得从输入数据中自动学习复杂特征的能力。深度学习方法包括前馈神经网络、卷积神经网络、递归神经网络和生成对抗网络等。

#### 1.3.1 前馈神经网络

前馈神经网络 (Feedforward Neural Networks, FNN) 是最基本的深度学习模型，也被称为多层感知器 (Multilayer Perceptron, MLP)。它由一个或多个全连接层组成，其中每个神经元通过激活函数将输入传递给下一层。FNN 中的信息流向是单向的，从

输入层经过隐藏层最终到达输出层。

FNN 的目的是学习从输入到输出之间的非线性映射关系。通过在训练过程中调节网络的权重和偏置，FNN 能够自动学习输入数据中的特征表示，从而实现对数据的分类、回归等任务。目前，FNN 广泛应用于图像分类、语音识别、自然语言处理等领域。

### 1.3.2 卷积神经网络

卷积神经网络 (Convolutional Neural Networks, CNN) 是一种特殊的神经网络结构，主要应用于图像和视觉任务，其核心是卷积层 (Convolutional Layer) 和池化层 (Pooling Layer)。卷积层通过卷积操作提取输入数据中的局部特征，每个卷积核对输入进行滑动并计算与之对应的特征图。池化层则通过取最大值或平均值等操作来减小特征图的空间尺寸。

CNN 的卷积和池化操作能够捕捉到图像中的局部模式和特征，而通过堆叠多个卷积层，网络能够学习到更高级别的抽象特征表示。最后，全连接层将这些特征结合起来进行分类或回归。目前，CNN 在图像分类、目标检测、图像分割等任务中取得了巨大的成功。

### 1.3.3 递归神经网络

循环神经网络 (Recurrent Neural Networks, RNN) 是一种特殊的神经网络结构，其专用于处理序列数据的深度学习模型。与前馈神经网络不同，RNN 具有循环连接，允许信息在网络中沿时间步传递，这使得 RNN 能够处理具有时序依赖性的数据，如自然语言、音频等。

RNN 中的每个神经元都有一个隐藏状态，它在每个时间步接收当前输入和前一个时间步的隐藏状态，并产生输出和新的隐藏状态。通过循环连接，RNN 能够捕捉到序列中的上下文信息。RNN 能够对序列数据进行建模和预测，它在机器翻译、语言建模、情感分析等领域取得了显著的成果。

### 1.3.4 生成对抗网络

生成对抗网络 (Generative Adversarial Networks, GAN) 是由生成器 (Generator) 和判别器 (Discriminator) 组成的一种特殊的深度学习模型，其核心思想是通过对抗训练来提高生成器和判别器的性能。

生成对抗网络的生成器通过学习生成真实数据分布的能力来提高生成数据的质量，而判别器则通过学习区分真实数据和生成数据的能力来提高辨别能力。两个网

络通过对抗训练的方式相互竞争学习，最终生成器能够生成与真实数据相似的合成数据。GAN 在图像生成、图像转换、图像修复等任务中取得了显著的成果。

深度学习能够自动地从数据中学习特征，无需手动提取，这与传统的机器学习方法不同。深度学习的优势在于其能够处理非常大、高维度的数据集，并通过大规模的训练数据提高模型的准确性。随着计算机性能的提高和大数据的出现，深度学习的应用越来越广泛，它已经成为机器学习领域的主流方法之一，并在图像识别、自然语言处理、语音识别等多个领域取得了丰硕成果。

## 1.4 迁移学习研究现状

虽然深度学习在许多领域中取得了重大的突破和成功，但也存在一些缺点。其中一个主要问题是其往往需要大量的标记数据进行训练，而标记数据的获取和标注是一个非常费时费力的过程；另外，深度学习的训练过程需要大量的计算资源，特别是在数据集较大时，使用 CPU 训练时间过慢，而使用 GPU 则成本较高。为了克服这些问题，研究者开始关注迁移学习，例如 Shen 等人在 2016 年已经在深度网络中引入了迁移学习，用以改善不同工况下轴承诊断的性能<sup>[7]</sup>。

一般学习的过程，都是利用充足数量和质量的训练数据，训练出分类或回归模型，使其在测试数据上可以进行较为良好的分类或预测。机器学习的训练效果主要依靠两个前提：一是需要有足够的、具有明确标签的样本；二是测试数据的分布需要与训练样本基本一致。然而上述两个前提在实际应用中难以实现：就训练样本方面来说，为每一条训练数据打上准确的标签需要耗费大量人工和计算成本；而对于测试数据来看，一旦与训练数据的获取方式不一致，其两者的分布也很难相似，这就容易导致模型对于训练数据很完美在测试数据上的表现较差。

在 2005 年，杨强教授引入了迁移学习的思想到计算机领域，此后这一概念引起了广泛的研究讨论。迁移学习的核心思想是将从一个数据域或任务域中通过数据积累、挖掘和分析获得的有用信息应用到另一个数据库或任务中，从而解决标注数据不足、测试数据分布不一致等问题。相对于传统的机器学习方法，迁移学习的思想在数据分布、数据标签和学习模型上都有着显著的不同之处，具体如表 2.1 所示。

**表 1.1 机器学习与迁移学习间的区别**

比较项目	机器学习	迁移学习
数据标签	训练数据需要足够标签	训练数据无需大量标签
数据分布	训练和测试数据分布一致	训练和测试数据分布无需一致
学习模型	每个任务都需要建模	可在不同任务间迁移

如果在一个领域中训练好的模型能够被迁移到另一个领域中进行训练，那么就可以减少在新领域中训练模型的时间和精力，这就是迁移学习的优点之一。同时，迁

移学习还可以使用一个大型数据集中的信息来帮助训练小型数据集中的模型，以帮助解决数据集不平衡的问题。

迁移学习最初是教育心理学领域的一个概念。人类在学习新知识时，会将已掌握的知识和技能应用于新的学习任务中，例如学过中文拼音的人在学习英语发音就会更容易一些；而会驾驶小汽车的人在学习驾驶卡车时也会更加轻松。这种积累和应用已有知识的能力被称为迁移学习。随着机器学习技术的不断发展，研究者们开始将迁移学习的思想与传统机器学习方法相结合。通过赋予机器对已有知识的继承和拓展能力，让其可以更好地利用它们来帮助目标知识的学习。

在机器学习中，迁移学习是指利用已有领域的经验来帮助新领域的学习。与传统的机器学习方法相比，迁移学习可以更快地获得更高的准确性。迁移学习是一种解决数据稀缺和标注缺失的方法，可以将一个领域的知识迁移到另一个领域，从而提高新领域的学习效果。

迁移学习问题通过领域 (Domain) 和任务 (Task) 这两个基本概念得以具体化。在这个框架下，学习的主体被称为领域，通常用  $D$  表示，领域  $D = \{X, P(X)\}$  由两部分组成，即数据的特征空间和数据的概率分布；学习的目标被称为任务，通常用  $T$  表示，任务  $T = \{Y, f(\cdot)\}$  也由两部分组成，即标签空间和目标学习函数。通常领域被分为源域 (Source Domain) 和目标域 (Target Domain)，这对应着机器学习中的训练集和测试集。

概率分布  $P$  是对数据的一种描述，可分为边缘概率分布和条件概率分布，分别对应着数据的整体分布情况和类间分布情况。如图 1.2 所示，图 (a) 和图 (b) 相比，数据整体分布差异较大，因此需要重点关注边缘概率分布；图 (a) 和图 (c) 相比，数据都分布于左右两侧，代表数据的整体相似，则更需要关注条件概率分布，了解数据的类间情况。

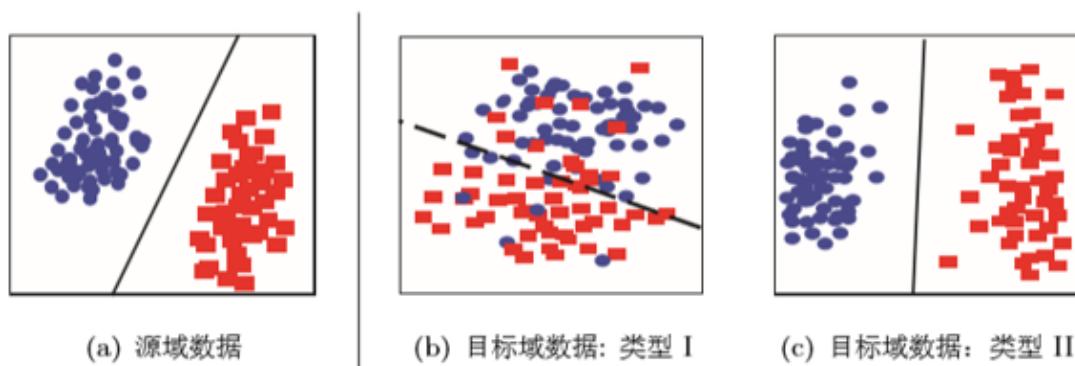


图 1.3 不同数据分布表示

迁移学习的主要工作是找到新任务和原任务之间的关系，迁移机器学习过的知

识，其优势在于它能够通过辅助域积累的大数据来训练模型和结构，提取相关信息，从而解决标注数据不足、测试数据与训练数据不一致等问题。同时，它也放宽了传统机器学习中的训练样本需求，不再要求测试数据与训练数据符合相同的分布，但仍需要具有一定的相关性。在实际应用中，如何对有限的含标签数据进行充分利用，从而提高新任务的准确率，成为了迁移学习研究的重要方向。

迁移学习可以根据特征空间  $X$  和标签空间  $Y$  的特点分为以下几个类别：特征空间  $X$  和标签空间  $Y$  相同的为同构迁移学习 (Homogeneous Transfer Learning)，其中根据条件分布和边缘分布的不同又分为数据偏移、领域自适应和多任务学习；特征空间  $X$  和标签空间  $Y$  有一个不同的为异构迁移学习 (Heterogeneous Transfer Learning)， $X$  不同为异构特征空间， $Y$  不同为异构类别空间。

迁移学习的方法亦可以按学习方法的分类形式，根据数据、特征、模型的机器学习逻辑进行划分。现有研究的方法可以分为基于特征的迁移学习、基于样本实例的迁移学习、基于模型的迁移学习三个类别。

#### 1.4.1 基于特征的迁移学习

基于特征的迁移学习 (Feature Based Transfer Learning) 就是更进一步对特征进行变换。意即假设源域和目标域的特征原来不在一个空间，或者说两个域在原有空间上不相似，通过变换将其变换到一个空间里面，使得在该空间中，两个域内样本的特征尽可能相似。该思路也非常直接，目前基于特征的迁移学习方法最为常用，也是最热的研究方向。

Paul 等人利用在 ImageNet 上预训练过得卷积神经网络模型进行肺癌的 X 射线图像进行特征提取，在对特征进行迁移学习，通过迁移后的图像特征进行肺腺癌的早期检测，方法取得了较好的效果<sup>[8]</sup>。康守强等人在对滚动轴承数据进行模态分解的基础上，提出了一种半监督的迁移成分迁移学习方法，即将不同类别的样本数据映射到同一个样本空间后，在提取其中的故障特征，从而对滚动轴承的故障诊断问题进行了解决<sup>[9]</sup>。

#### 1.4.2 基于样本实例的迁移学习

基于样本实例的迁移学习 (Instance Based Transfer Learning) 是指通过权重重用，对源域和目标域的样例进行迁移。可以直接对不同的样本赋予不同权重，例如对于相似的样本，给其一个较高权重，这样就完成了迁移。

Alothman 等人针对基于相似性的样本实例迁移学习方法进行了改进，保证了迁移后的样本数据集的类别均衡条件，文中将提出的方法与现有方法进行了比较，并在网络拥堵数据上进行了实验验证<sup>[10]</sup>。Hossain 等人针对脑电接口运动图像的标注

缺失问题提出了一种多信息的样本迁移方法，利用主动学习的方法对目标域的 EEG 图像数据进行标注，有效减少了对于图像校准和标记的工作量<sup>[11]</sup>。

### 1.4.3 基于模型的迁移学习

基于模型的迁移学习 (Model Based Transfer Learning) 也就是构建参数共享的模型。该方法主要在神经网络中应用最为广泛，因为神经网络的结构和参数可以直接进行迁移。神经网络最经典的微调方法 (Finetune) 就是模型参数迁移的很好的体现。

Segev 等人利用在源域上训练好的决策树模型迁移到目标域中，文中用两种方法对迁移的过程进行优化，第一种通过不断搜索局部最优解来相应用于树结构进行调整，第二种则不进行结构的调整而仅仅进行参数的优化，最后尝试结合两种方法进行集成的学习，文章在多类数据集上进行了实验验证<sup>[12]</sup>。Si 等人针对自然语言处理提出了一种不需要目标域语料信息的模型迁移方法，通过对于多个源域的语料数据进行模型的训练，提取出通用的分析模型并在目标域上进行应用<sup>[13]</sup>。

总之，迁移学习是一种非常有前途的方法，可以大大提高机器学习的效率和准确性。在故障诊断领域，它可以帮助我们更快地找到问题的根源，加快维修速度，提高设备的可靠性和可用性。

## 1.5 论文研究内容和整体结构

本文主要针对工业生产中故障诊断问题，研究了故障诊断分类的难点，并且针对模型迁移、不均衡数据和变工况诊断三个角度进行了深入研究。

第一章为绪论部分，主要阐述了故障诊断在实际生产中的重要意义，同时介绍了深度学习和迁移学习在故障诊断领域的发展现状，强调了它们在提高诊断准确性和适应多样化故障场景方面的优势。

第二章研究了故障诊断的特征提取问题，采用了基于卷积神经网络的模型迁移方法，并通过修改冻结层数、学习率和单个样本大小等参数，对轴承故障数据进行了诊断分类测试。通过这些实验，验证了不同模型在故障诊断方面的效果，并对比了它们在故障分类能力上的差异。

第三章针对样本不均衡的情况下故障诊断精度低的问题，使用生成对抗网络进行数据增强。分别基于 GAN 和 DCGAN 两种网络构建生成器和判别器后，对少量样本进行对抗训练，得到具有一定相似程度的伪造图像，并将两种方法增强后的数据和未增强的数据分别进行训练，并将训练的结果进行对比。该方法有效地提高了故障分类精度，验证了生成的伪造数据能够增强对原数据集中少数数据的特征提取能力。

第四章针对变工况下模型训练效率慢的问题，引入了领域自适应方法将不同工况下的数据进行迁移。首先对深度预适应的原理和分类做了详实的介绍和概述，并使用 DAN 方法，将 VGG11 和 ResNet50 分别作为 DAN 的骨干网络，随后使用最大均值差异，将源领域的数据和目标领域的数据进行特征转换，使它们在特征空间上更加接近。通过这种方法，我们可以减少针对每个工况进行独立训练的时间和资源消耗，同时在不同工况下实现高效的模型训练。

第五章是对论文整体内容的总结，以及对未来研究工作的展望。

## 第二章 基于卷积神经网络的模型迁移方法

通过深度学习建立数据分析模型时，往往需要使用海量带有准确标签的训练数据分布，同时需要耗费大量时间进行训练，使其得以准确分类或预测。本章针对模型迁移方法，首先对深度迁移学习方法和微调方法进行介绍，其次针对实际问题进行理论分析，然后设计算法在卷积神经网络上进行实验，最终对所得实验结果进行评估和讨论，给出调参建议以及对此方法的评价。

### 2.1 卷积神经网络

人工神经网络是一种模仿动物神经网络行为特征的复杂网络系统，能够对各个节点进行参数调节，使得整个系统能够完成各种各样的任务，并具有记忆和学习功能。人工神经网络最早在 1943 年被提出，主要包含输入层，隐藏层和输出层，每一层结构皆由人工神经元组成，如图 2.1 所示。

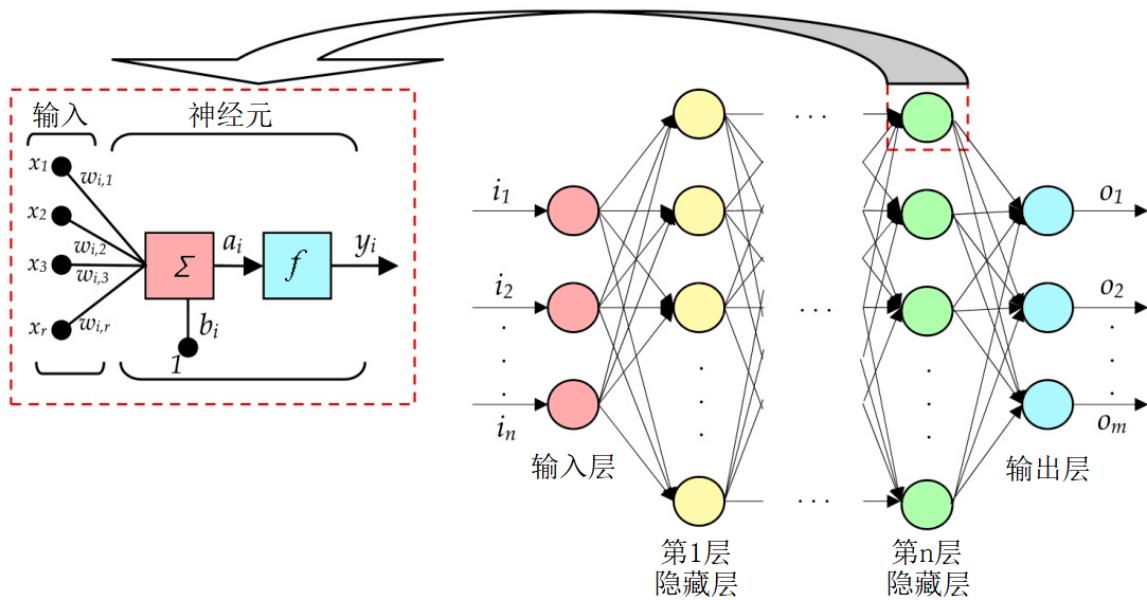


图 2.1 人工神经网络结构及其神经元结构

在人工神经元结构中， $x_i$  表示第  $i(i = 1, 2, 3, \dots, n)$  个输入值， $w_i$  表示第  $i(i = 1, 2, 3, \dots, n)$  个输入的连接权值， $b$  为该神经元的偏置， $f$  是对该神经元的一个激活函数， $Y$  是神经元的输出值。神经元输入与输出的关系可以由公式(2.1)表示：

$$Y = f \left[ \left( \sum_{i=1}^r w_i x_i \right) + b \right] \quad (2.1)$$

神经网络是由人工神经元组成的，它们可以形成层次结构，并通过将上层的输出作为下层的输入来连接。具有多个隐藏层的神经网络被称为深度学习网络。人工神经元模型将生物神经元转化为逻辑功能器件，结合了生物和数学的特点，为神经网络研究奠定了基础。

受到动物视觉皮层的启发，卷积神经网络(CNN)引入了卷积运算。CNN正在快速发展，并在图像识别、语音识别、故障诊断等领域得到应用。与经典的人工神经网络不同，卷积神经网络的隐藏层由卷积层、池化层和全连接层组成。网络的下层用于检测基本的低级视觉特征，随着网络层数的增加，能够检测到的特征变得更高级，这些特征是建立在简单低级特征的基础上的。典型的卷积神经网络如图2.2所示，下面将详细阐释。

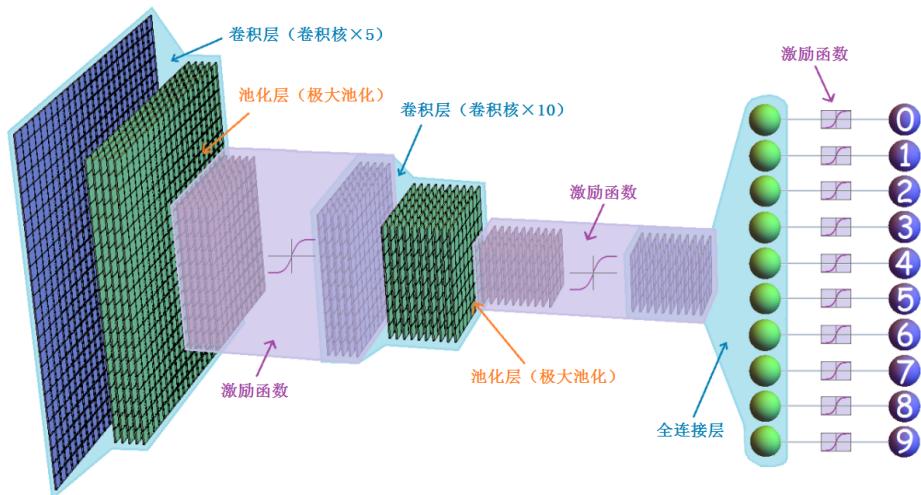


图 2.2 经典卷积神经网络模型<sup>[14]</sup>

### 2.1.1 卷积层

卷积层是CNN中最关键的一种网络层，通过卷积运算可以依次检测图像从简单到复杂的特征。卷积运算是一种数学操作，用于处理两个函数之间的关系。在卷积层中，我们使用卷积运算对输入数据和卷积核（也称为滤波器或特征提取器）之间的关系进行建模。

假设输入数据为一个三维张量（或称为特征图） $X$ ，其尺寸为 $(C_{\text{in}}, H_{\text{in}}, W_{\text{in}})$ ，其中 $C_{\text{in}}$ 表示通道数， $H_{\text{in}}$ 和 $W_{\text{in}}$ 分别表示高度和宽度。

卷积层包含一组卷积核（或称为滤波器） $K$ ，每个卷积核都是一个三维张量，尺寸为 $(C_{\text{in}}, K_{\text{out}}, H_{\text{kernel}}, W_{\text{kernel}})$ ，其中 $C_{\text{in}}$ 表示输入通道数， $K_{\text{out}}$ 表示输出通道数， $H_{\text{kernel}}$ 和 $W_{\text{kernel}}$ 分别表示核的高度和宽度。

卷积操作通过将每个卷积核与输入数据进行逐元素相乘，并对结果进行求和得到输出特征图。假设输出特征图为 $Y$ ，其尺寸为 $(K_{\text{out}}, H_{\text{out}}, W_{\text{out}})$ ，其中 $H_{\text{out}}$ 和 $W_{\text{out}}$ 分别表示输出特征图的高度和宽度。

那么，卷积层的卷积操作公式(2.2)表示。

$$Y_k = \sum_{c=1}^{C_{\text{in}}} \sum_{i=1}^{H_{\text{kernel}}} \sum_{j=1}^{W_{\text{kernel}}} X_{c,h+i,w+j} \cdot K_{c,k,i,j} \quad (2.2)$$

其中， $Y_k$ 表示输出特征图中第 $k$ 个通道的值， $X_{c,h+i,w+j}$ 表示输入特征图中第 $c$ 个通道、位置为 $(h+i, w+j)$ 的元素， $K_{c,k,i,j}$ 表示第 $c$ 个通道的卷积核中的权重， $h$ 和 $w$ 分别表示输出特征图中的高度和宽度索引。

通过卷积操作，卷积层能够在保留空间结构的同时，提取输入数据中的局部特征，并生成更有代表性的特征图作为下一层的输入，这使得卷积层成为深度学习中重要的组成部分之一。卷积核在输入数据上移动并执行卷积计算时，只针对部分输入数据进行计算，从而降低了计算的成本和复杂度，而不必对整个数据进行运算。

### 2.1.2 池化层

池化层一般紧跟在卷积层之后，能够通过减小特征图的空间尺寸，降低模型对空间位置的敏感性，同时减少参数数量，提取主要特征，从而帮助提高模型的鲁棒性和效率。

池化层是深度学习中常用的一种层级结构，用于减少特征图的尺寸并提取主要特征。池化层通常跟在卷积层之后，对卷积层提取的特征图进行下采样操作。最常见的池化操作是最大池化 (Max Pooling) 和平均池化 (Average Pooling)。它们的数学原理如下：

#### (1) 最大池化

给定输入特征图 $X$ ，池化层通过将输入特征图划分为不重叠的区域（通常是矩形区域），并在每个区域中选择最大的元素作为输出特征图的对应位置的值。假设输入特征图为 $X$ ，其尺寸为 $(C, H_{\text{in}}, W_{\text{in}})$ ，其中 $C$ 表示通道数， $H_{\text{in}}$ 和 $W_{\text{in}}$ 分别表示高度和宽度。则最大池化层的数学原理可以由公式(2.3)表示。

$$Y_{c,h,w} = \max_{i=1}^k \max_{j=1}^k X_{c,h \times s+i, w \times s+j} \quad (2.3)$$

其中,  $Y_{c,h,w}$  表示输出特征图中第  $c$  个通道、位置为  $(h, w)$  的值,  $X_{c,h \times s+i, w \times s+j}$  表示输入特征图中第  $c$  个通道、位置为  $(h \times s + i, w \times s + j)$  的元素,  $s$  表示池化窗口的大小,  $k$  表示池化窗口的边长。

## (2) 平均池化

与最大池化类似, 平均池化也将输入特征图划分为不重叠的区域, 并在每个区域中计算元素的平均值作为输出特征图的对应位置的值。假设输入特征图为  $X$ , 其尺寸为  $(C, H_{\text{in}}, W_{\text{in}})$ 。则平均池化层的数学原理可以由公式(2.4)表示。

$$Y_{c,h,w} = \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k X_{c,h \times s+i, w \times s+j} \quad (2.4)$$

其中,  $Y_{c,h,w}$  表示输出特征图中第  $c$  个通道、位置为  $(h, w)$  的值,  $X_{c,h \times s+i, w \times s+j}$  表示输入特征图中第  $c$  个通道、位置为  $(h \times s + i, w \times s + j)$  的元素,  $s$  表示池化窗口的大小,  $k$  表示池化窗口的边长。

### 2.1.3 全连接层

通常数据在经过多个卷积层和池化层的特征提取之后, 会被投入到网络结构最后的全连接层中, 此时全连接层把所局部特征结合成全局特征, 来对数据进行分类, 得到网络最终的输出。

除上述提到的网络层, 激活函数、损失函数和梯度下降是深度学习中重要的组成部分, 它们在神经网络的训练和优化过程中起着关键的作用。

激活函数是神经网络中每个神经元的非线性变换。通过引入非线性性质, 激活函数允许神经网络学习更复杂的函数。线性变换的堆叠只会得到线性结果, 而激活函数的引入使得神经网络能够建模非线性关系。激活函数的目的是引入非线性性质, 提供更大的模型表达能力。它们能够处理非线性的特征和模式, 使网络能够适应各种复杂的数据分布。此外, 激活函数还具有平滑性、抑制梯度消失等特性, 有助于提高网络的训练稳定性和性能。

### 2.1.4 损失函数

损失函数用于衡量模型预测结果与实际标签之间的差距。通过计算损失函数的值, 可以度量模型在给定数据上的拟合程度。优化的目标是最小化损失函数。损失

函数的目的是提供一个可优化的目标，以指导模型参数的调节。它衡量了模型的预测误差，使得模型能够逐渐减小预测值与实际标签之间的差异。不同的损失函数适用于不同的任务，例如回归、分类和生成等。

### 2.1.5 梯度下降

梯度下降是一种优化算法，用于调节神经网络中的参数以最小化损失函数。它基于损失函数对参数的梯度信息，沿着梯度的反方向更新参数，以逐步降低损失函数的值。梯度下降的目的是通过迭代更新参数，使模型朝着损失函数最小化的方向优化。通过计算参数的梯度，梯度下降算法可以找到损失函数的局部最小值。批量梯度下降使用所有训练样本的梯度，而随机梯度下降和小批量梯度下降使用一部分样本的梯度，从而提高计算效率。

Janssens 等人于 2016 年发表了第一篇利用 CNN 来识别轴承故障的论文，它利用分别安装在 x 方向和 y 方向上的两个单轴加速度计收集振动数据，并利用 CNN 从经过尺度离散傅里叶变换预处理的原始数据中自动学习出承载故障检测的有用特征，结果表明，基于特征学习的方法明显优于基于特征工程的传统机器学习方法；同时表明 CNN 也可以执行轴承健康预测，并识别一些没有明确的特征频率的早期错误条件，如润滑退化，这是使用经典深度学习方法无法实现的<sup>[15]</sup>。在接下来的三年里，出现了许多相关论文也逐渐运用起了这项技术，为推进故障诊断自动化、智能化发展做出了巨大贡献。

## 2.2 模型迁移

模型迁移是深度迁移学习的一种，指利用在源域中预先训练好的部分网络，包括其网络结构和连接参数，作为目标域深度神经网络的一部分进行复用。这种方法的理论基础是神经网络类似于人类大脑的信息处理机制，其是一个迭代且连续的抽象过程。网络的前面层可被视为特征提取器，这些层提取的特征是通用的。基于网络的深度迁移学习的示意图如图 2.3 所示。

### 2.2.1 预训练模型

进行模型迁移的过程中，需要有一个事先训练好的预训练模型，该模型一般是数据集较大的源域上进行训练得到的。得到预训练模型后，可以直接使用预训练模型的结构和权重，在目标域的场景中应用该模型，即将源域预训练模型迁移到目标域问题中。一般情况下，源域任务与目标域任务的要求往往不同，故可以在模型迁移之后，可以根据目标域需求，修改预训练模型的输出层，从而更好地与目标域场

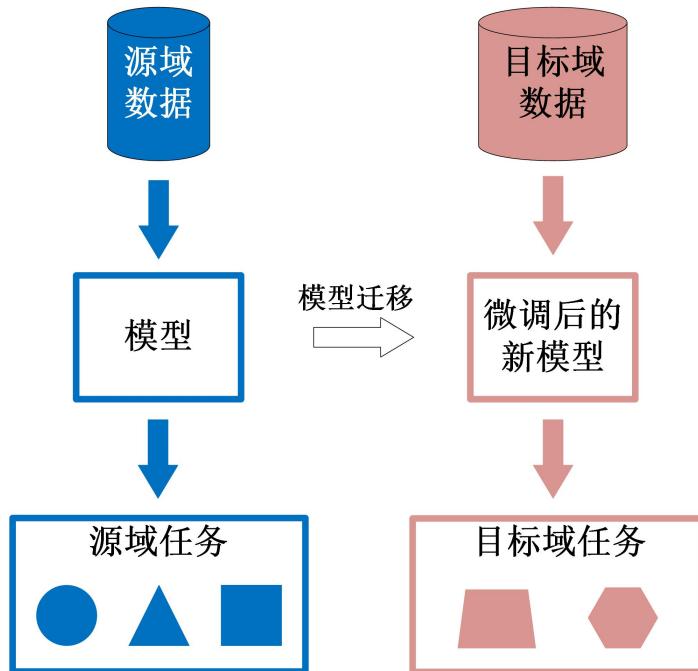


图 2.3 深度神经网络迁移过程

景匹配。

### 2.2.2 微调

在处理目标域问题时，由于与源域的概率分布可能不同，可能会导致直接迁移后的模型效果欠佳。因此，将预训练过的模型应用到目标数据集之后，还需要一个使参数适应目标数据集的过程，即微调 (fine tuning)。微调又可分为整体微调与部分微调的形式。整体微调指将整个预训练模型的权重都参与训练，并在目标域数据上进行调节，这意味着整个模型的权重都会因目标域数据的训练而被更新；部分微调则是指冻结预训练模型中前  $k$  层的权重，重新训练后面的  $n - k$  层，然后将其应用到目标场景中。

模型迁移具体是否需要用到微调，使用整体微调还是部分微调，取决于目标域规模以及源域与目标域数据的相似程度。当目标域规模较小，且两个域的相似度较高时，则无需对预训练模型进行微调，只需调节输出层的结构以适应目标域场景；当目标域规模较小，且两个域的相似度较低时，一般使用部分微调；当目标域规模较大，两个域之间存在较大差异时，多不采用模型迁移的方式；当目标域规模较大，且两个域存在一定相似度时，则可以使用整体微调。

## 2.3 实验设计

本次实验均使用 NVIDIA GeForce RTX 2060 6G 显卡，AMD Ryzen 7 4800H 处理器，32G 内存，Windows11 操作系统进行实验，所有实验均采用 pytorch 深度学习框架，通过 Python 语言进行编写和调试。本章将对模型迁移前后的损失值、准确率和运行时间进行对比，并通过改变微调时的冻结的卷积层数、优化器的学习率以及预处理数据方式来研究其对模型迁移的影响。

### 2.3.1 实验数据集简介

为方便评价所提方法的性能，本研究采用江南大学轴承故障数据集作为基准数据集进行实验<sup>[16]</sup>，如图 2.4 所示。

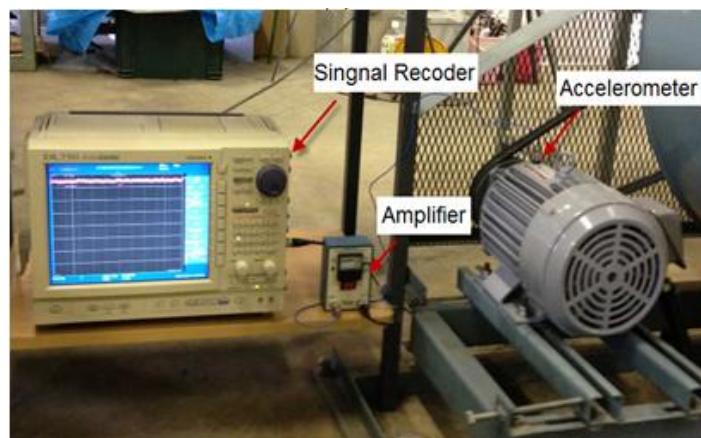


图 2.4 江南大学滚动轴承故障诊断数据采集装置

该数据集在 50kHz 采样频率下进行采集，并按照不同的转速分为三种不同的工作状态，分别是 600rpm、800rpm 和 1000rpm。每种工作状态下都有四种故障状态，包括正常 N(normal)、内圈故障 IB(inter-race defect)、外圈故障 OB(outer-race defect)、滚动体故障 TB(ball roller defect)，图 2.5 为不同故障状态下的轴承实体图。其中，每个转速下的正常状态数据有超过 150 万个采样点，其余状态数据有 50 万个左右的采样点。



图 2.5 故障轴承实体图，其中 (a) 为外圈故障，(b) 为内圈故障，(c) 为滚动体故障

本章将选取单一转速(600rpm)下不同故障状态的数据进行实验。

### 2.3.2 数据预处理

为了更加有效地从原始信号中提取特征，一般会在数据送入诊断模型前，将原始一维振动信号进行预处理，转化为一定量的二维灰度图作为诊断模型的输入。本章通过重叠截取的技术，将一维振动信号进行预处理，转换为具有更丰富信息的二维灰度图形式。

首先，在原始振动信号上进行窗口长度对1024个采样点进行重叠采样，窗口每次移动的长度固定为400个采样点，如图2.6所示。

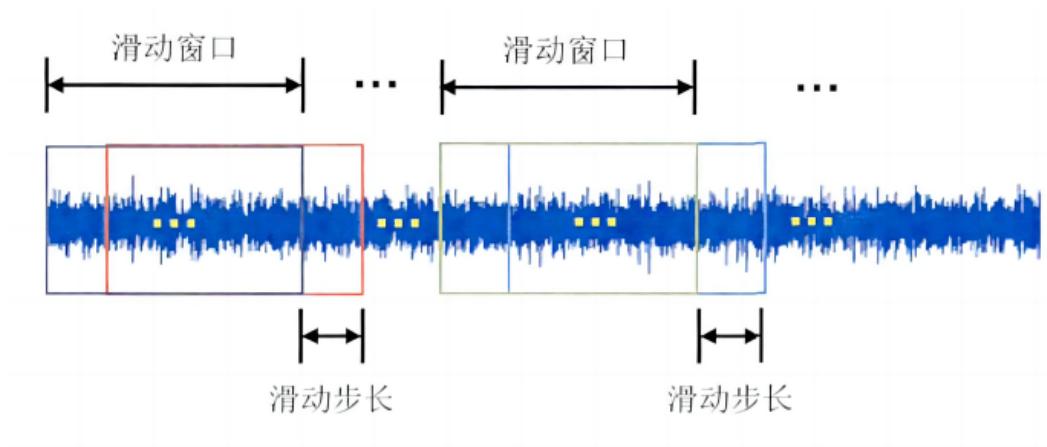


图2.6 重叠采样示意图

随后将每一个截取到的信号段无重叠地切分成长度为32个采样点的32段，并按序堆叠成 $32 \times 32$ 的二维矩阵；然后将矩阵中的每个信号点数值归一化到0-255，具体归一化方式如公式(2.5)表示。

$$x = 255.0 \times \left( \frac{i - \min(A)}{\max(A) - \min(A)} \right) \quad (2.5)$$

其中， $i$ 表示原始信号采样点的值， $A$ 表示由原始信号转换后未归一化的二维矩阵， $\max$ 表示取最大值， $\min$ 表示取最小值， $x$ 为归一化的数值。

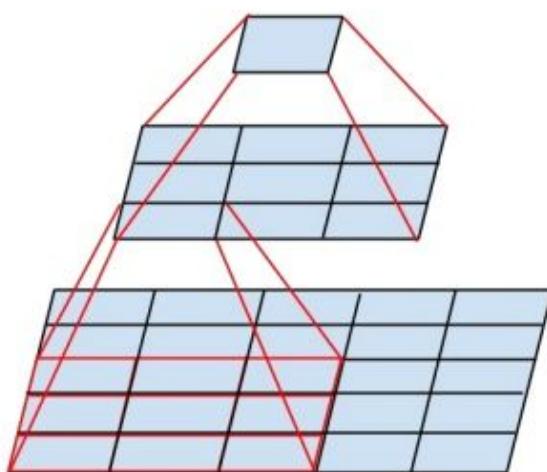
最后，将归一化后得到的数值作为灰度像素值，从而得到尺寸为 $32 \times 32$ 的二维灰度图样本。表2.1为本章实验各故障状态下使用的数据样本数量，总计3840个数据样本，其中2880个样本用于训练模型，960个样本用于测试模型的结果。

**表 2.1** 数据库样本类型及划分

编号	故障类型	样本数量	训练样本	测试样本
0	正常	960	720	240
1	内圈	960	720	240
2	外圈	960	720	240
3	滚动体	960	720	240

### 2.3.3 网络模型的选取

本章选取基于 ImageNet 数据库训练的 VGG11 模型作为待迁移的预训练模型。ImageNet 是一个用于视觉对象识别软件研究的大型可视化数据库，该数据集共有 1000 个类别，包含 128 万张训练图像、5 万张验证图像和 10 万张测试图像。VGG11 结构包含 8 个卷积层，5 个池化层和 3 个全连接层。相比于 AlexNet，VGG 的一个改进是采用连续的多个  $3 \times 3$  的卷积核来代替 AlexNet 中的较大卷积核，如使用了 3 个  $3 \times 3$  卷积核来代替  $7 \times 7$  卷积核，使用了 2 个  $3 \times 3$  卷积核来代替  $5 \times 5$  卷积核，这样做的主要目的是在保证具有相同感知野的条件下，提升了网络的深度，在一定程度上提升了神经网络的效果。对于给定的感受野，即与输出有关的输入图片的局部大小，采用多个小卷积核效率会优于使用对应的大卷积核，因为多层非线性层可以增加网络深度来保证学习更复杂的模式，且参数会更少，运算速度更快。图 2.7 展示的是 VGG 结构中将  $5 \times 5$  的卷积核转化成 2 个  $3 \times 3$  的卷积核的过程。

**图 2.7**  $5 \times 5$  卷积核转化为  $3 \times 3$  卷积核的示意图

本章实验将 VGG11 的最后一层全连接层的输出通道由 1000 个修改为 4 个，以更好地对 4 种故障状态的训练样本进行分类，以达到故障诊断的目的，具体结构参数如表 2.2 所示。

**表 2.2 VGG11 结构参数**

层序	层类型	输出通道数	卷积核大小	步幅	填充
1	卷积层	64	3x3	1	1
2	最大池化层	64	2x2	2	0
3	卷积层	128	3x3	1	1
4	最大池化层	128	2x2	2	0
5	卷积层	256	3x3	1	1
6	卷积层	256	3x3	1	1
7	最大池化层	256	2x2	2	0
8	卷积层	512	3x3	1	1
9	卷积层	512	3x3	1	1
10	最大池化层	512	2x2	2	0
11	卷积层	512	3x3	1	1
12	卷积层	512	3x3	1	1
13	最大池化层	512	2x2	2	0
14	全连接层	4096	-	-	-
15	全连接层	4096	-	-	-
16	全连接层	4	-	-	-

## 2.4 实验结果与分析

将预处理后的训练样本和 ImageNet 预训练后的 VGG11 模型载入至 GPU，为与 VGG11 模型的输入层结构相匹配，样本在进行训练前会被缩放至  $224 \times 224$  的大小，随后使用载入的模型进行 100 次迭代训练，在每次迭代之后，计算训练损失和准确率，并将结果记录下来，每迭代 10 次后使用验证集对得到的模型进行评估，得到的训练结果如图 2.8 所示。在模型迭代次数一致的情况下，使用模型迁移方法训练得到的模型，其验证损失远低于未迁移模型，说明使用模型迁移可以大幅度地缩短模型的训练周期；同时，迁移后模型的准确率相较于未迁移的模型提高了 30% 以上，说明使用模型迁移得到更好的诊断效果。

重新载入预训练模型，并对迁移后的模型由浅至深分别冻结不同层数的卷积层后，进行 100 次迭代的训练来微调模型参数，其准确率变化如表 2.3 所示。

**表 2.3 冻结不同卷积层数下的诊断效果比较**

冻结卷积层数	训练损失	验证损失	模型准确率
0	0.0133	0.1032	92.60%
1	0.0058	0.4276	91.46%
2	0.0061	0.0485	93.02%
3	0.0052	0.1597	93.23%
4	0.0169	0.1442	93.33%
5	0.0133	0.2319	89.79%
6	0.0975	0.3179	86.67%
7	0.4188	0.5394	78.84%
8	0.6205	0.5679	75.62%

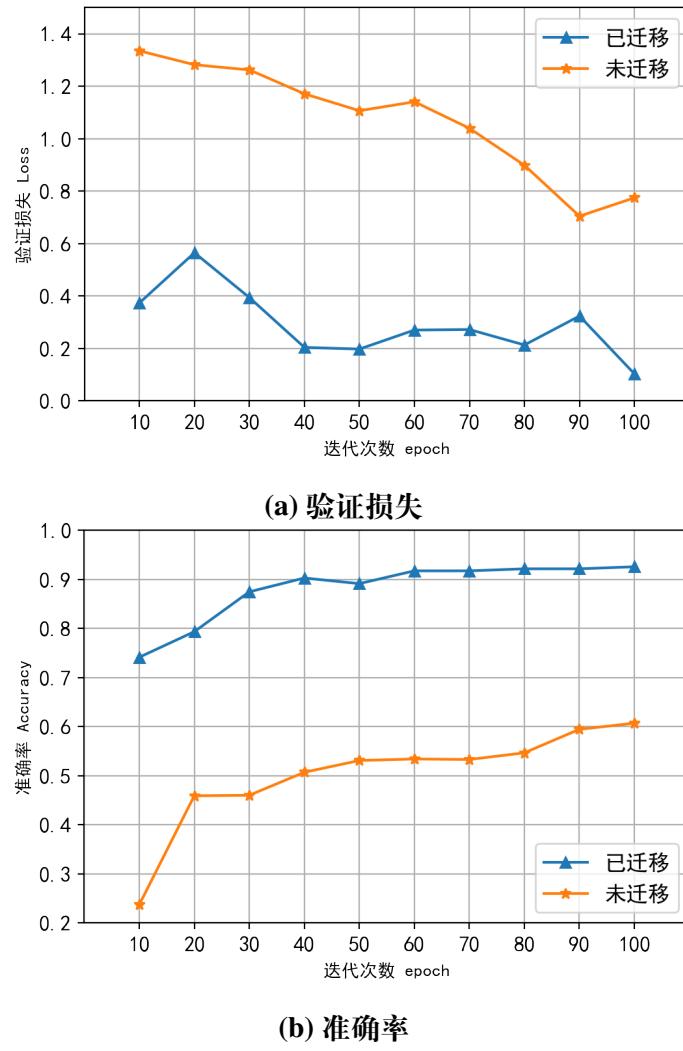


图 2.8 模型迁移前后的部分参数指标

由表 2.2 中数据可得，当冻结的卷积层数从 0 层增加到 1 层时，模型的训练损失下降，但验证损失存在异常上升的情况，导致模型的准确率有所下降。随着冻结的卷积层数由 1 层上升至 4 层期间，模型的准确率小幅上升，并维持在 93% 以上。从冻结的卷积层数上升至 5 层开始，模型的准确率大幅度下降，说明对于所使用的样本来说，前 4 层卷积层提取得到的图像特征与 ImageNet 这样的大数据提取得到的图像特征是高度相似的。因此，就二维灰度图而言，可以冻结前 4 层卷积层，这样可以在保证较好训练效果的情况下，节省算力和时间成本。

针对冻结前 4 层卷积层的模型，使用 0.0005, 0.001, 0.002, 0.005, 0.01 五种学习率分别对迁移后的模型重新进行 100 次迭代的微调，获得的准确率比较图如图 2.9 所示。其中，当学习率为 0.0005 时，训练出的模型准确率整体上低于实验其他情况；当学习率在 0.001 和 0.005 时，模型准确率在训练早期存在大量波动。当学习率设置过低时，训练初期时模型的准确率低，需要更多的迭代次数才能达到理想效果；当学习率较高时，模型的准确率上升速度更快，但可能伴随着较大的波动。在本次实

验中，在使用学习率为 0.002 的情况下，训练结果表现更优。

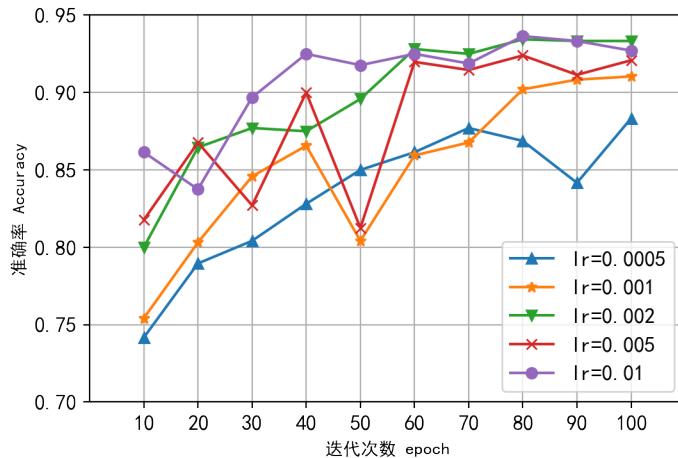


图 2.9 不同学习率对准确率的影响

改变数据预处理时的窗口长度大小，分别得到大小为  $28 \times 28$ 、 $56 \times 56$  的二维灰度图，并在冻结前 4 层卷积层，学习率为 0.002 的前提下分别重新训练，并与  $32 \times 32$  的情况进行比较，得到的准确率结果由图 2.10 所示。

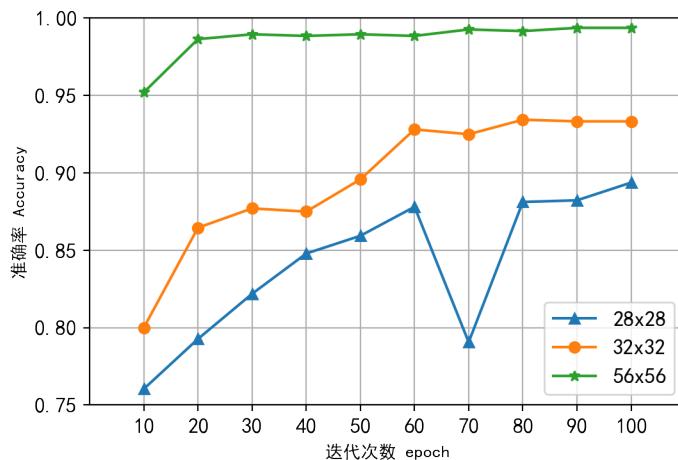


图 2.10 不同图像大小对准确率的影响

由图可知，同样的原始振动数据，预处理成更大的二维灰度图，训练得到的准确率越高。实际上，对于  $56 \times 56$  的二维灰度图像，仅需 20 次迭代训练后即可获得超过 98% 的预测准确率，之后 80 次的迭代训练对模型的优化甚微。将预处理后的样本大小设置为  $56 \times 56$  能够为单个样本提供更多的细节和特征，使得其具有更高的准确率。考虑到预处理后的样本需在正式进行训练前统一被缩放至  $224 \times 224$ ，如使用分辨率大于  $224 \times 224$  的样本可能会使得样本在缩放时出现信息丢失、特征变化等情况，导致内容这是因为较大的图像包含更多的细节和特征，使得网络更难学习到准确的

特征表示。

## 2.5 本章小结

本章针对轴承故障诊断问题，对基于 VGG11 的模型迁移方法进行了研究。首先，使用重叠窗口方法对原始振动数据进行预处理，获得适合二维卷积神经网络的二维灰度图数据，随后使用基于 ImageNet 数据集训练后的 VGG11 模型进行训练，并与为进行模型迁移的模型进行对比，得出了使用模型迁移方法可以代替模型的前期训练，缩短工期，并能有效提高故障诊断的精度。其次，对不同程度的微调进行讨论，得出从输入层开始，冻结不同数量的卷积层对模型训练速度和训练效果的影响。在此之后，针对特定的模型，改变模型损失函数的学习率，得出对于本次故障诊断问题最佳学习率。最后，通过改变数据预处理方式，获得不同大小的数据样本并分别训练，得出适当增大预处理过程生成的图像大小能有效提升训练效率的结论。

### 第三章 基于生成对抗网络的不均衡数据增强方法

在故障诊断领域，由于轴承实际工况下的故障数据往往难以采集且需要较高的造价与时间人力成本，且不同种类故障发生的频率不一致，导致了对应类别的样本数量不同，故而存在许多样本不均衡的情景。对于传统的深度学习算法，往往在样本数量较多的类别上表现较好，而在样本数量较少的类别上表现较差。这可能导致对于少数类别的故障无法有效地进行准确的识别和分类，从而影响整体的故障诊断性能。为提高故障诊断对少数故障的识别准确率，一般使用数据增强方法扩充小样本数据。本章主要讨论使用生成对抗网络方法对数据增强的实现及效果，详细描述了生成对抗网络的框架及相关理论，并通过建立生成对抗网络及其衍生模型讨论生成对抗网络的不均衡数据的增强效果。

#### 3.1 数据增强

数据增强是一种在深度学习中常用的技术，它通过对原始训练数据进行一系列随机变换和扩充来增加数据样本的多样性。数据增强的目的是提高模型的泛化能力和鲁棒性，减少过拟合的风险，并改善模型在不同场景下的性能。数据增强可以应用于各种类型的数据，包括图像、语音、文本等。传统的数据增强方法有缩放、变形、裁剪、调整色彩、添加噪声、随机遮挡等。

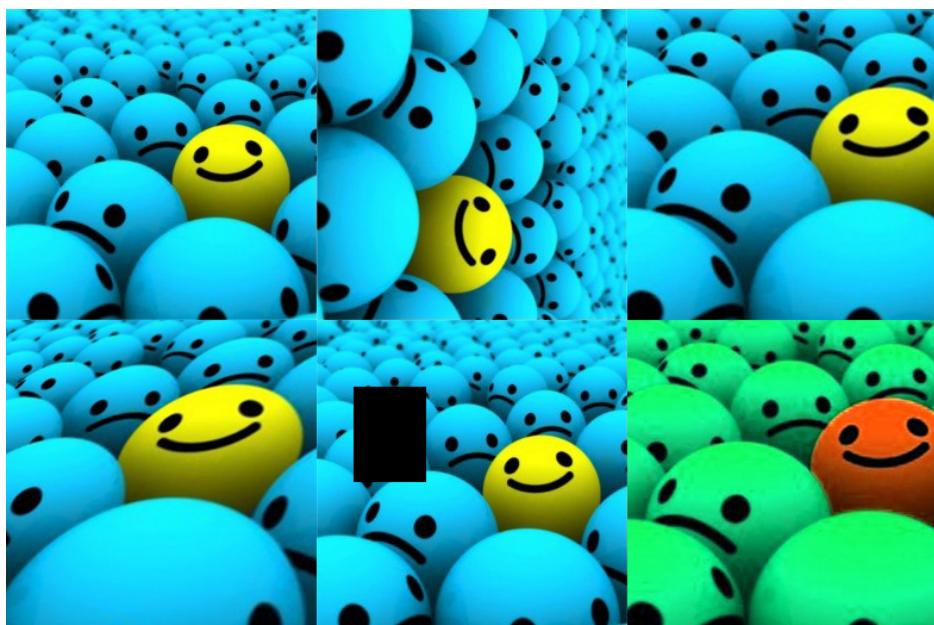


图 3.1 不同的图像增强方法

通过应用这些数据增强方法，可以扩充原始数据集，生成更多多样化的样本，从而增加训练数据的数量和质量。这样可以提高深度学习模型的鲁棒性、泛化能力和抗干扰能力，帮助模型更好地适应不同的场景和应用。同时，数据增强也可以减少过拟合的风险，提高模型的性能和稳定性。

## 3.2 生成对抗网络

### 3.2.1 网络框架

生成对抗网络（Generative Adversarial Networks, GAN）是由 Goodfellow 等人在 2014 年提出的一种深度神经网络架构<sup>[17]</sup>。作为一种无监督学习方法，GAN 具备学习复杂数据分布的能力，被认为是近年来最具发展和应用前景的方法之一。它通过一种相互对抗的方式来学习原始数据的复杂分布，并生成具有相似特定分布的数据。

GAN 的具体结构设计灵感来自于博弈论中的二人零和博弈思想<sup>[18]</sup>，包括生成器神经网络  $G$  和判别器神经网络  $D$  两个关键组件。生成器学习真实数据的分布规律并生成伪造数据，其目标是根据判别器的反馈不断改进网络结构，生成更加真实的数据以欺骗判别器。而判别器根据判断结果的误差不断提升其准确率，目的是正确地区分真实数据和生成数据。这样，生成器和判别器之间形成了一种动态的对抗过程。

GAN 的学习目标是将简单的潜在分布映射到更复杂的分布。生成器接收随机噪声  $z$  作为输入，并生成伪样本，而判别器则接收真实样本和生成器生成的伪样本作为输入。生成器的主要目标是生成逼真的伪样本，使判别器无法区分真实样本和伪样本。判别器的主要目标是通过分配更高或更低的概率来区分伪样本和真实样本。

GAN 生成伪造图像的过程中，生成器和判别器分别进行训练。第一阶段中，固定判别器，训练生成器。生成器不断生成伪造数据，判别器进行判断。初始时，生成器的能力较弱，容易被判别为伪造数据，但通过反馈和训练，生成器的能力逐渐提升，直到判别器判断结果为真。第二阶段中，固定生成器，训练判别器。判别器通过持续训练逐步提高判别能力，最终能够准确判断所有生成的伪造图像。然后循环进行第一阶段和第二阶段的训练，通过这种迭代的过程，生成器和判别器的能力都达到最佳水平。因此，可以使用具有出色效果的生成器生成足以欺骗人眼的伪造图像。

### 3.2.2 目标损失参数

GAN 通过最大最小目标函数联合求解训练生成器  $G$  和判别器  $D$ ，目标函数可以表示为公式 3.1。

$$\min_{\theta_G} \max_{\theta_D} V(D, G) = E_{x \sim P_{data}(x)} [\log(D(x))] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (3.1)$$

其中,  $\theta_D$  代表判别器的参数,  $\theta_G$  代表生成器的参数,  $V(D, G)$  代表二元交叉熵函数, 代表期望值,  $x$  代表真实数据,  $P_{data}(x)$  代表真实数据  $x$  的数据分布,  $z$  代表随机噪声,  $P_z(z)$  代表随机噪声  $z$  的数据分布。

从公式 3.1 可以得出, 生成器函数  $G$  可以被解释为将低维噪声映射到高维真实样本分布的函数, 即  $P_z$ 。判别器在此过程中扮演着评估生成器产生的伪样本质量的角色。每个模型都追求自身利益的最大化, 生成器试图生成伪样本以欺骗判别器, 而判别器则希望避免被生成器欺骗并正确认别伪样本。生成器和判别器的参数通过反向传播进行更新。公式(3.1)通过交替更新以下两个梯度来实现参数的更新:

$$\theta_D^{t+1} = \theta_D^t + \lambda^t \nabla_{\theta_D} V(D^t, G^t) \quad (3.2)$$

$$\theta_G^{t+1} = \theta_G^t + \lambda^t \nabla_{\theta_G} V(D^{t+1}, G^t) \quad (3.3)$$

其中,  $t$  为迭代次数,  $\lambda$  为学习率。在 GAN 训练中, 公式 3.1 中的  $\log(1 - D(G(z)))$  需要最大化, 这使得生成器中的梯度减小, 即梯度值变小并停止学习。为了克服梯度消失问题, 公式 3.1 中的目标函数被重新定义为两个独立的目标:

$$\max_{\theta_D} V(G, D) = E_{x \sim P_{data}(x)} [\log(D(x))] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (3.4)$$

$$\min_{\theta_G} V(G) = E_{x \sim P_z(z)} [\log(1 - D(x))] \quad (3.5)$$

此外, 生成器对这两个独立目标的梯度具的固定点相同, 并且总是在训练的方向上也一致。故在公式 3.4 和公式 3.5 中计算成本之后, 可以使用反向传播来更新生成器和判别器的参数。根据这两个不同的目标函数, 如果  $D(x)$  预测错误, 或者预测  $D(G(z))$  时, 都需要更新  $\theta_D$  和  $\theta_G$ 。

GAN 与传统生成模型相比具有两个独立的网络, 并采用相互对抗的训练方法。为了更新生成器和判别器参数, 梯度更新信息是从判别器计算的损失中获取的, 该损失衡量了两个分布之间的差异。GAN 具有多个优点: 首先, 作为一种新型生成模型, GAN 能更好地建模数据分布; 其次, GAN 可以通过学习真实样本的分布, 从而生成更清晰、真实的样本; 再者, GAN 只需使用反向传播, 无需使用马尔科夫链等方法进行重复采样; 除此之外, 生成的样本健壮性更好, 有助于提高目标模型的泛化能力。

### 3.3 基于 GAN 的数据增强能力分析

#### 3.3.1 模型结构

本章实验中所用的 GAN 模型的结构框架如图 3.2 所示，包括两个模型，即生成器  $G$  和判别器  $D$ 。在生成器中，将类别标签引入到 GAN 模型的生成器中，类别标签和随机噪声作为生成器的输入，通过多个全连接层后，将最后的一维输出值转化为二维图像，得到生成的伪样本  $x_{fake}$ 。在判别器中，将得到的伪样本  $x_{fake}$  和真实标签样本  $x_{real}$  作为判别器的输入，转换成一维输入后，同样经过多个全连接层提取时频图像特征，进而对输入的真实样本和伪样本进行判别。最后对判别结果进行反向传播，从而更新生成器和判别器的参数，循环上述步骤达到训练目的。

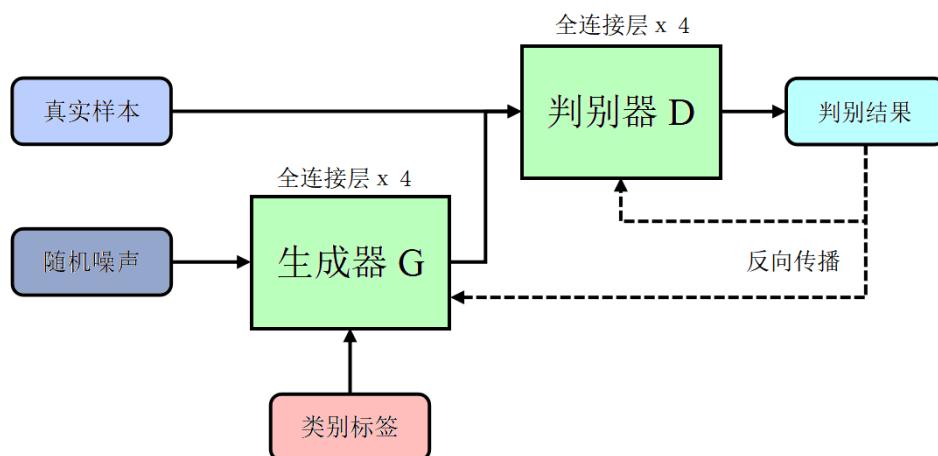


图 3.2 GAN 的模型架构

表 3.1 为本节实验中生成器和判别器的结构，其中，在生成器的结构中，输入为一个 100 维的噪声张量和一个标签张量。噪声张量和标签张量先通过次全连接层被连接起来，再经过 3 层全连接层，得到一个  $56 \times 56 = 3136$  维的神经元。最后使用 tanh 函数将神经元的权重分布在  $(-1, 1)$  的区间，从而与真实图像数据相匹配。判别器的结构可以认为是与生成器结构相对称的结构。判别器先将  $56 \times 56$  的图像数据展开成 3136 维的神经元，再经过 4 次全连接层得 1 维的标签张量，最后使用 Sigmoid 函数将标签张量的权重映射到  $(0,1)$  上，作为本次判别的概率。

表 3.1 生成器和判别器的结构

层号	生成器输入维度	生成器输出维度	判别器输入维度	判别器输出维度
1	100	256	3136	512
2	256	512	512	512
3	512	512	512	256
4	512	3136	256	1

本章实验使用二分类交叉熵损失作为判别器的损失函数。二分类交叉熵损失(BCELoss)是常用于二分类任务中一种的损失函数，其负责在每个数据样本上计算真实标签和预测标签之间的二分类交叉熵损失。该损失函数适用于预测输出为概率的情况，例如使用 Sigmoid 激活函数输出的概率值。

$$\text{BCELoss} = -(y \times \log(P) + (1 - y) \times \log(1 - P)) \quad (3.6)$$

其中， $y$  是真实标签（取值为 0 或 1）， $P$  是预测标签的概率值。

BCELoss 可以用于训练二分类模型，在例如判别器的训练过程中，可以用于衡量生成样本与真实样本之间的差异，并通过最小化该损失来优化模型参数。

### 3.3.2 实验结果与分析

本章选取江南大学滚动轴承故障诊断数据集内转速为 600rpm 的 4 种故障状态信号，对所有状态信号使用的预处理方式与第 2 章一致，其中对正常工作状态生成 1200 个  $56 \times 56$  的二维灰度图样本，对其余故障状态生成 300 个  $56 \times 56$  样本，总计 2100 个图像样本。

为验证 GAN 模型对故障样本的生成和诊断的有效性，本章先对未使用数据增强方法进行训练，并评价训练结果。使用经过 ImageNet 预训练过的模型，冻结前 4 层卷积层，设置学习率为 0.002，将数据集划分为 80% 的训练集和 20% 的验证集，对所训练样本进行 100 次迭代训练，期间每 10 次迭代训练后使用验证集进行验证，得到图 3.3。

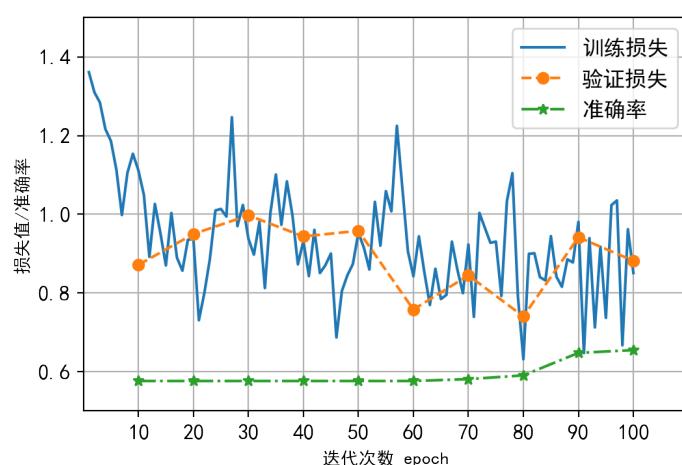


图 3.3 未使用数据增强时模型的训练过程

从上图可以观察到，数据样本不均衡导致模型训练损失的波动性较大，即使在

下降的趋势中，仍然存在大幅度的波动，这可能是由于少数类别的样本数量较少，模型难以充分学习这些类别的特征而导致的；验证损失在整个训练过程中没有明显的下降趋势，这可能是由于数据不均衡导致模型过度拟合多数类别，而在少数类别上表现较差，使验证损失无法有效降低；由于数据不均衡，模型的准确率在前 80 次迭代中都低于 60%，这表明模型在初始阶段无法有效区分少数类别，导致准确率较低。

		实际状态			
		ib	n	ob	tb
预测状态	ib	3	0	2	0
	n	281	1200	153	257
ob	16	0	145	43	
tb	0	0	0	0	

图 3.4 未使用数据增强时模型的混淆矩阵

图 3.4 为未使用数据增强时，模型对于不均衡样本训练的混淆矩阵。从图中不难发现，正常状态下的 1200 个样本全部被正确识别，但剩下的 900 个样本识别准确率不到 20%。由于正常状态样本占整个数据集的大部分，模型对于正常状态更加敏感，以至于模型提取到了更多的正常状态特征，这其中也包括其他状态下同样拥有的共性特征，使得模型在识别少数类别时，判断样本为多数样本的概率更高，导致少数样本的误判。

通过使用 GAN 模型，分别将少数的三种故障状态进行训练，得到 3 组不同生成器和判别器。图 3.5 为不同迭代次数下各故障状态通过 GAN 模型的生成器生成的伪造图像，其中第一行为内圈故障，第二行为外圈故障，第三行为滚动体故障；从左到右依次是生成器迭代 1 次、10 次、100 次、200 次和 300 次后生成的图像。图中不难看出，当迭代一次时，生成出的图像呈现出噪声的特征，但在 10 次迭代后逐渐呈现二维灰度图特有的横向纹路，100 次迭代后伪造样本与真实样本的区别较小，达到了数据增强的效果。

使用迭代 300 次训练后的生成器为样本较少的种类生成 900 个样本，使每一种故障状态的样本数量达到 1200 个。将增强后的数据库载入到使用经过 ImageNet 预训练过的模型，同样冻结前 4 层卷积层，设置学习率 0.002，用 80% 的数据集作为训练集进行 100 次迭代训练，每 10 次迭代训练后使用剩下的数据集进行验证，得到的训练结果如图 3.6 所示。

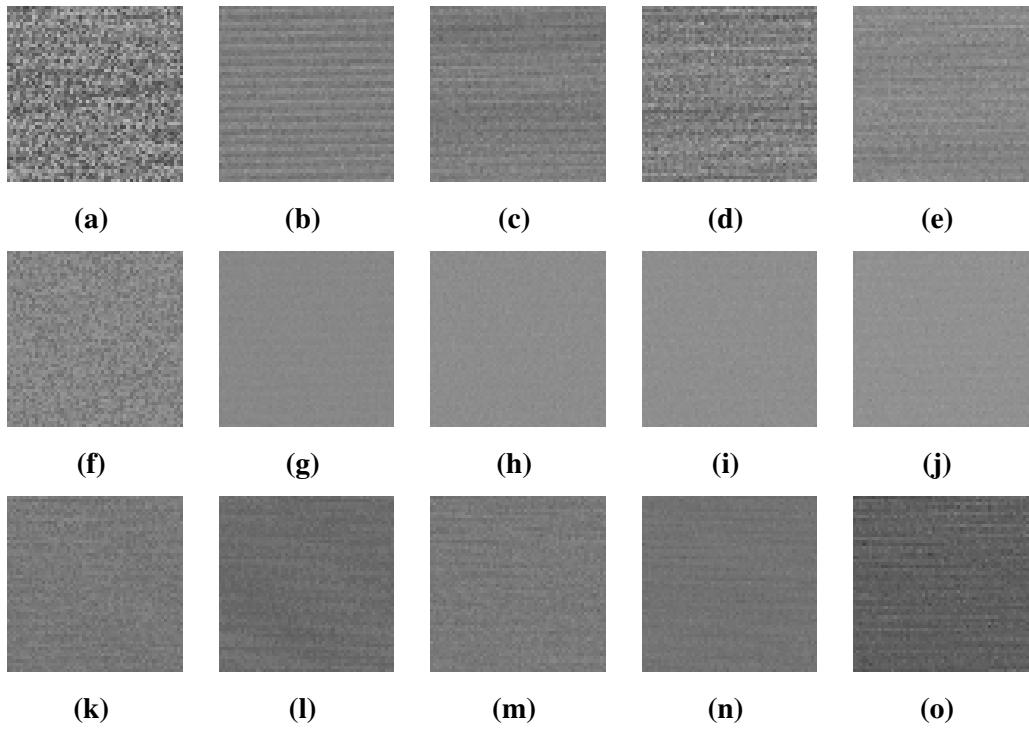


图 3.5 GAN 中生成器生成的伪造图像

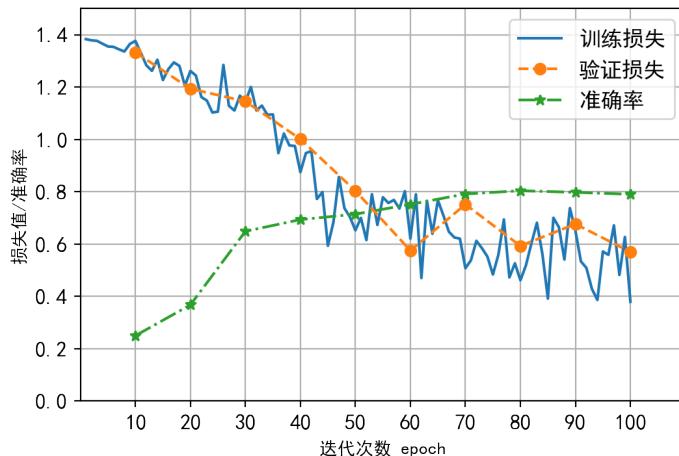


图 3.6 使用 GAN 进行数据增强后模型的训练过程

对不均衡样本使用 GAN 进行数据增强后，再次进行深度迁移学习训练，训练损失的波动幅度减小，验证损失也稳步下降，说明使用 GAN 进行数据增强能有效提升模型的特征提取能力，使原本数据较少的故障状态中，样本的特征也能被一定程度的提取。

图 3.7 是将训练好的模型重新对所有数据增强前的样本进行识别，得到的混淆矩阵。其中可以发现，虽然部分正常状态样本被判别成了内圈故障，但是其他故障状态，尤其是内圈故障与外圈故障状态的诊断能力大幅度提升。

		ib	n	ob	tb
预测状态	ib	165	156	25	102
	n	100	1044	0	59
ob	ib	16	0	264	73
	tb	19	0	11	66

图 3.7 使用 GAN 进行数据增强后模型的混淆矩阵

## 3.4 基于 DCGAN 的数据增强能力分析

### 3.4.1 模型结构

DCGAN 模型的结构与传统 GAN 模型结构类似，都是由生成器 G 与判别器 D 构成。DCGAN 模型与传统 GAN 模型的主要区别在于生成器与判别器的内部构成。本节实验中，生成器主要由 4 层反卷积层组成，激活函数为 ReLU；判别器主要由 4 层卷积层构成，激活函数为 LeakyReLU。LeakyReLU 在负输入区域引入了一个小的斜率，避免了 ReLU 的零梯度问题，增强了模型的稳定性和泛化能力。同时，两者都多次使用到 BN 层，其主要作用是通过对每个批次数据的均值和方差进行估计，将输入数据重新标准化为零均值和单位方差。这样可以使得输入数据的分布更加稳定，避免网络中间层的数据分布发生偏移或变化过大，有助于加速网络的收敛和提高模型的泛化能力。生成与判别器的具体结构参照表 3.2。

### 3.4.2 实验结果与分析

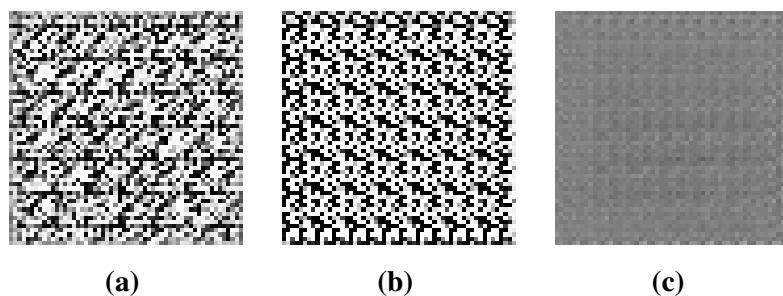


图 3.8 DCGAN 中生成器生成的伪造图像

用 DCGAN 模型，分别对样本数为 300 的三种故障状态进行迭代训练，得到不

**表 3.2 生成器和判别器的结构**

层号	层类型	输入维度	输出维度	卷积核大小	步幅	填充
1	反卷积	100	256	7×7	1	0
2	BN	256	256	-	-	-
3	ReLU	-	-	-	-	-
4	反卷积	256	128	4×4	2	1
5	BN	128	128	-	-	-
6	ReLU	-	-	-	-	-
7	反卷积	128	64	4×4	2	1
8	BN	64	64	-	-	-
9	ReLU	-	-	-	-	-
10	反卷积	64	1	4×4	2	1
11	Tanh	-	-	-	-	-
1	卷积	1	64	4×4	2	1
2	LeakyReLU	-	-	-	-	-
3	卷积	64	128	4×4	2	1
4	BN	128	128	-	-	-
5	LeakyReLU	-	-	-	-	-
6	卷积	128	256	4×4	2	1
7	BN	256	256	-	-	-
8	LeakyReLU	-	-	-	-	-
9	卷积	256	1	7×7	1	0
10	Sigmoid	-	-	-	-	-

同故障状态各自的判别器与生成器。生成器中每次迭代更新之后都生成一个批次数量的伪造图像。图 3.8 为由外圈故障状态迭代 1 次、20 次、200 次训练后的生成器所生成的批伪造图像。通过图 3.8 可以看出，DCGAN 模型在学习伪造图像的初期会产生明显的、棋盘状的纹理，这是因为网络的初步训练可能导致生成器输出的图像存在块状结构，随着学习的深入，产生的纹理会逐渐消褪。

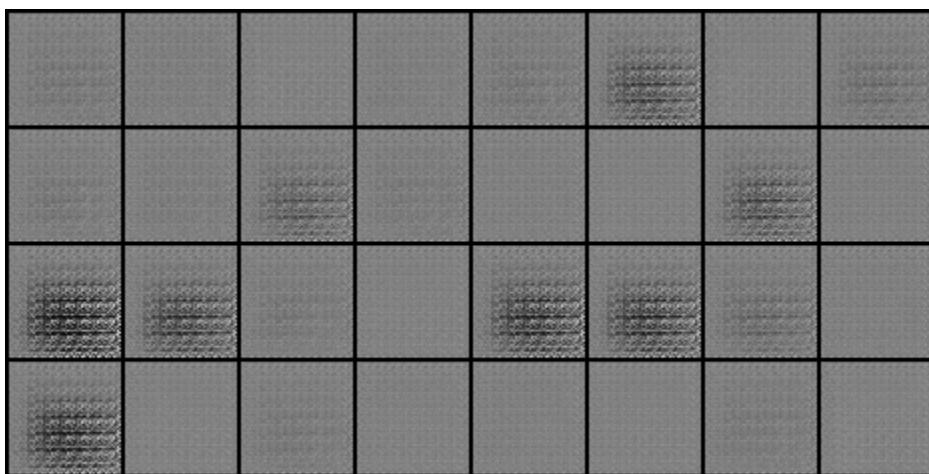
**图 3.9 迭代训练 200 次后的生成器生成的 32 张伪造图像**

图 3.9 为使用 DCGAN 模型经过 200 次迭代训练后的生成器生成的 32 张伪造图

像。从图中可以发现，虽然随着训练深入，生成器生成的图片不再具有明显的棋盘状纹理，但是部分伪造图像会出现局部泛黑的现象。

将 DCGAN 模型迭代训练 300 次后，对原始数据进行数据增强。使用增强后的数据在与上节相同的迁移学习模型下进行训练。图 3.10 展示了所得到的训练过程。

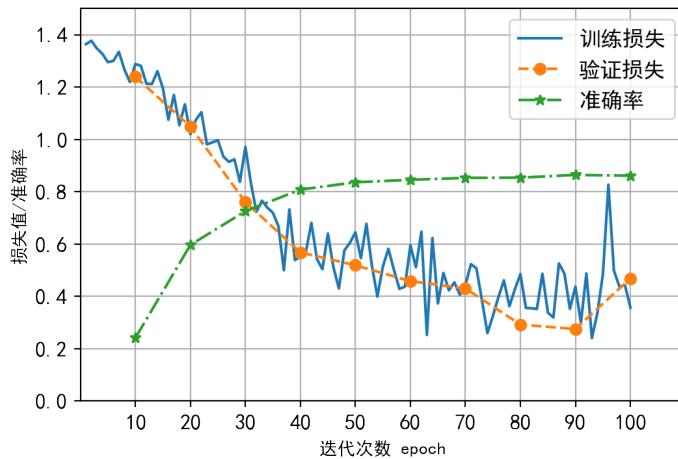


图 3.10 使用 DCGAN 进行数据增强后模型的训练过程

通过对比可以观察到，相较于使用传统的 GAN 进行数据增强，使用 DCGAN 进行数据增强后，模型的损失和准确率得到了进一步的优化和改善，损失函数的下降速度更快，准确率的提升更为显著。这说明 DCGAN 生成的合成样本具有更高的质量和代表性，能够有效地帮助模型学习到更具判别性和泛化能力的特征表示，从而提升了模型的性能和准确度。

		ib	103	32	73
		n	1075	1	120
预测状态	ib	139			
	n	107	22	267	107
tb	1	0	0	0	

图 3.11 使用 DCGAN 进行数据增强后模型的混淆矩阵

使用得到的诊断模型对所有原始样本进行识别，得到的混淆矩阵如图 3.11 所示。通过使用基于 DCGAN 方法进行数据增强后，模型的整体识别准确率进一步提升。然

而，对于某些故障状态，尤其是滚动体故障状态，DCGAN 方法并没有有效增强数据的效果。进一步分析发现，在 GAN 方法中，该故障状态的识别准确率仍然较低。这种现象可能是由于滚动体故障状态与其他状态的样本更为相似，使得模型可能更难区分不同的类别。相似的样本可能共享相似的特征，导致模型更容易将它们错误地归为同一类别。

### 3.5 本章小结

本章针对故障诊断的数据不均衡问题，运用了基于 GAN 的数据增强方法，并设计了两种不同 GAN 结构进行比较。本章首先在保持正常状态样本数量的情况下，减少其他故障数据的样本量，并直接进行训练，发现模型的准确率大幅下降，且诊断结果均向正常数据偏移。随后使用 GAN 方法对不均衡样本进行数据增强，并对同一模型进行训练，得到了数据增强可以减少对模型对多数情况的样本类型偏移的情况，从而提高模型对少数样本类型的识别准确率。最后使用 DCGAN 方法进行数据增强，并于 GAN 方法进行对比，得到了 DCGAN 方法对整体识别准确率的提升略好于 GAN 方法但对于少数样本的识别准确率较 GAN 方法更差的结论。

## 第四章 基于深度域适应的变工况故障诊断方法

常规机器学习往往针对单一工况进行研究，但现实中工业系统随时都可能需要改变工况，此时原本的模型就不能很好的诊断其他工况下的样本。本章阐释了变工况故障诊断的定义，介绍了领域自适应以及 DAN 及其分类，同时设计实验，研究 DAN 中不同骨干网络对于诊断能力的影响。

### 4.1 变工况故障诊断

工业生产中的设备类型众多，任何一个设备出现故障都有可能会导致整个工业流程的终止。例如滚动轴承，作为一种广泛应用于各种设备的基础零件，在设备运行过程中扮演着不可替代的角色，一旦轴承出现故障，就会直接影响整台设备的运行。针对运行一段时间的设备，通常可以获取到充足的故障数据，并利用这些数据对模型进行不断训练，从而得到一个成熟的诊断模型，用于设备故障诊断。

传统的机器学习模型在进行诊断时通常会在系统最常使用的一种或几种工况下进行数据采集，并对每种工况的数据进行分析，以得到相应的故障诊断模型。然而，在实际工业生产过程中，设备的工况并非一成不变。对于变工况下的故障诊断，传统的机器学习模型的诊断效果会大大下降。如果针对每种工况生成一个故障诊断模型，则需要耗费大量时间来收集、清洗和准备每个工况下的数据样本。此外，针对每个工况进行模型训练也需要耗费大量的计算资源和时间来训练和优化模型。整个过程需要处理大量的数据和执行复杂的计算任务，从而导致时间成本的大幅增加。

因此，在工程领域中，我们引入了变工况的概念。变工况指的是在不同的操作或使用情况下，设备、系统或过程所面临的不同工作条件。它涵盖了设备或系统在不同环境、负载、温度、压力、速度等方面的变化。理解和定义变工况对工程设计、测试和分析至关重要，因为不同的工作条件可能对设备的性能、可靠性、安全性和寿命产生重要影响。例如一台汽车发动机，其变工况可能包括不同的车速、负载条件、环境温度和海拔高度等。这些变工况会影响发动机的燃烧效率、冷却效果、润滑性能等。在发动机的设计和测试过程中，工程师会针对这些不同的变工况进行性能评估和优化，以确保发动机在各种工作条件下都能稳定可靠地运行。

一般情况下，同一工况内数据的均值、方差和协方差不随时间发生变化，即在特定工况下，数据的统计特性在相对稳定的状态下保持一致。例如，在特定负载下，轴承的振动幅值的均值、方差和不同传感器之间的协方差在工况持续稳定时保持相对恒定。然而，当工况发生变化时，数据的统计特性可能会发生显著变化。例如，当

轴承的负载从低到高进行变化时，轴承的振动幅值可能会逐渐增加，其均值和方差也会相应变化。此外，在不同工况下，轴承的特征之间的协方差矩阵也可能发生变化，反映了不同工况下特征之间的关系的差异。

在变工况下，某些特征可能在不同工况之间具有一定的迁移性。在某些情况下，设备在一个工况下的特征表现可以预示着其在其他工况下的表现。例如，通过观察发动机在低负载下的振动特征，可以预测其在高负载下的振动水平。迁移学习方法可以总结出不同工况之间的特征迁移关系，这种特征迁移关系可以在工况之间进行知识迁移和模型迁移，以提高设备故障诊断和预测的准确性。

综上所述，了解和考虑设备的变工况是重要的工程设计和分析因素，以确保设备或系统能够适应各种实际使用情况，提供可靠的性能和功能。同时，针对变工况下的故障诊断，需要探索更高效的方法，以克服传统机器学习模型在变工况下诊断效果下降的问题，从而提高设备故障诊断的准确性和可靠性。

## 4.2 领域自适应

领域自适应 (Domain Adaptation, DA) 是同构迁移学习中的一种方法，主要针对源域和目标域数据条件分布不同但学习任务相同这一特殊情况。在这种情况下，源域拥有大量标记好的样本，而目标域只有少量或没有标记的样本。该方法的目标是将从源域的大量标记样本中学习到的知识迁移到目标域上，以解决相同的问题。在领域自适应中，我们需要通过特征学习和领域间的特征对齐，使得源域和目标域的特征表示在某种程度上相似，从而提高在目标域上的预测性能。这样，我们可以利用源域上的标记样本来训练模型，并将学到的知识迁移到目标域上的非标记样本中。通过领域自适应的方法，我们能够充分利用源域上的标记样本和目标域上的非标记样本，从而提高在目标域上的预测准确性和性能。

在领域自适应中， $D_s$  表示源域数据集， $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ ，来自概率分布  $P$ ，其中  $x_i^s$  是源域数据集中的第  $i$  个样本， $x_i^s \in R^d$ ， $d$  对应于样本的特征维度， $y_i^s$  是源域数据中第  $i$  样本对应的类别， $n_s$  是源域数据中样本总量。 $D_t$  是目标域数据集， $D_t = \{x_i^t\}_{i=1}^{n_t}$ ， $D_t$  来自概率分布  $Q$ ， $x_i^t$  是目标域数据集中的第  $i$  个样本， $n_t$  是目标域数据中样本总量。若目标域数据集是无标签数据集，则  $D_t$  只有无标签数据，没有包含对应的标签信息  $y_t$ 。

在传统的机器学习方法中，由于源域数据集  $D_s$  和目标域数据集  $D_t$  来自相同的分布，并且具有相同的诊断任务，即满足  $P(x^s) = Q(x^t)$ ，但在领域自适应问题中，源域数据和目标域数据来自不同的分布，即  $P(x^s) \neq Q(x^t)$ 。相比传统机器学习方法，领域适应方法放宽了对测试数据和训练数据独立同分布的要求，从而更符合实际应用

场景的需求。如图 4.1 所示,当测试数据和训练数据存在领域迁移时,传统方法的分类性能会受到较大影响,而采用领域适应技术的方法旨在减少源域数据和目标域数据之间的分布差异,从而有效改善分类能力。

特征匹配和实例重加权是领域自适应中常用的两类方法。特征匹配指的是通过将源域和目标域的特征进行映射或转换,使它们在共享的特征子空间中更加接近,同时保留数据的重要属性,常用的特征匹配方法包括最大均值差异 (Maximum Mean Discrepancy, MMD)、领域对齐 (Domain Alignment) 和对抗训练 (Adversarial Training) 等。实例重加权指的是对源域的实例重新赋予权重,再利用加权后的源域实例进行训练,从而最小化源域和目标域之间的分布差异,常用的实例重加权方法包括核均值匹配 (Kernel Mean Matching, KMM)、KL 重要性估计过程 (Kullback-Leibler Importance Estimation Procedure, KLIEP) 等。

最近的研究表明,深度网络在学习更具迁移性的特征方面具有很大的潜力。因此,研究者们开始将领域自适应模块引入深度网络的特征学习中,以提取更多的特征不变量。在深度网络中,领域自适应方法主要可以分为两种类型:基于散度的域适应和基于对抗的域适应。基于散度的域适应通过将源域和目标域之间的散度最小化的方式来学习一种域不变特征,最常用的四种散度为最大均值差异、对比域差异、相关对齐和 Wasserstein 度量。基于对抗的域适应借鉴了生成对抗网络的思想,在域适应中,生成器的作用由原本的生成伪样本,转变为提取从源域和目标域中提取出可供迁移的特征,使得判别器无法区分所提取特征的来源。

### 4.3 深度适配网络

深度适配网络 (Deep Adversarial Network, DAN) 是一种用于领域自适应的深度迁移学习方法。DAN 旨在通过自适应特征学习,将源域和目标域的特征表示进行对齐,以减小域间的差异。其核心思想是通过深度神经网络,在共享的特征提取层和分类器之间引入一个领域适应层,该层可以通过优化来实现特征的域间对齐。

在 DAN 中,共享的特征提取层通常是预训练的深度神经网络 (如 AlexNet、VGG、ResNet 等),可以从大规模数据集上进行无监督或有监督的预训练。然后,通过训练领域适应层, DAN 可以自适应地调整特征表示,使得源域和目标域在特征空间中更接近,从而提高在目标域上的性能。

通过使用深度适配网络,可以在源域上学习到的知识和特征表示迁移到目标域上,从而在目标域上实现更好的泛化性能。DAN 已经在图像分类、目标检测、人脸识别等任务中取得了一定的成功,并成为领域自适应研究领域中的重要方法之一。

现有的深度适配网络方法基本上可以分为四种:基于分布差异的方法、基于对

抗学习的方法、基于生成的方法、基于重构的方法。

### 4.3.1 基于分布差异的方法

源域和目标域之间的分布差异是影响目标域任务精度的关键因素之一。为了提高目标域任务的精度，基于分布差异的深度域适应方法被广泛研究和应用。在这些方法中，设计更好的分布差异度量方法是一个重要的研究方向。以下介绍了两种最常用的分布差异度量方法。

#### (1) 最大均值差异

最大均值差异 (Maximum Mean Discrepancy, MMD) 是领域自适应中使用最广泛的一种损失函数，主要用来度量两个不同但相关的分布的距离。两个分布的距离定义为：

$$\text{MMD}(D_s, D_t) = \left\| \frac{1}{n^s} \times \sum_{i=1}^{n_s} \phi(x_i^s) - \frac{1}{n^t} \times \sum_{j=1}^{n_t} \phi(x_j^t) \right\|_{\mathcal{H}}^2 \quad (4.1)$$

其中， $\phi(\cdot)$  为特征映射函数， $n_s$  和  $n_t$  为源域和目标域的样本总量， $\mathcal{H}$  表示再生核希尔伯特空间 (Reproducing Kernel Hilbert Space, RKHS)。基于 MMD 的度量方法，通常会将源域和目标域的特征向量映射到再生核希尔伯特空间中，以计算两个域之间的均值差异。这样的映射可以通过核函数实现，例如高斯核函数或线性核函数。Tzeng 等人提出了 DDC 方法，它利用 AlexNet 作为特征提取器，MMD 作为度量准则，将提取到的特征映射到再生希尔伯特空间中进行均值对齐，是最早在领域自适应问题中应用 MMD 的方法<sup>[19]</sup>。Long 对 DDC 方法做出了改进，提出了 DAN 方法，如图 4. 所示，它将 AlexNet 网络中最后三个全连接层得到的特征均映射到再生希尔伯特空间中，分别计算对应层之间的分布差异，并提出了多核最大均值差异 (Multi-Kernel Maximum Mean Discrepancy, MK-MMD) 用于解决 MMD 中单一核无法保证均值差异最大的问题<sup>[20]</sup>。Long 等人后又认为仅将源域和目标域的边缘分布对其并不能有效缩小两个域之间的差异，相反，深度神经网络在最后几个特定任务层的激活输出可以分别表示特征分布差异和标签分布差异，于是提出了 JAN 方法，并提出了联合最大均值差异 (Joint Maximum Mean Discrepancy, JMMD)<sup>[21]</sup>。

在早期的研究中，MMD 作为一种分布度量方法，通常被用来衡量整体分布的差异。然而，仅仅对整体分布进行对齐并不能保证局部分布的对齐。因此，后续的方法通过引入类信息来构建局部对应的分布对齐。Zhang 等人提出了 DTN 方法，它利用分类器的输出对样本进行类别划分，并在对应类中计算最大均值差异<sup>[22]</sup>。Zhu 等人提出了 DSAN 方法，它不但利用分类器的输出对样本划分类别，还衡量不同样本对

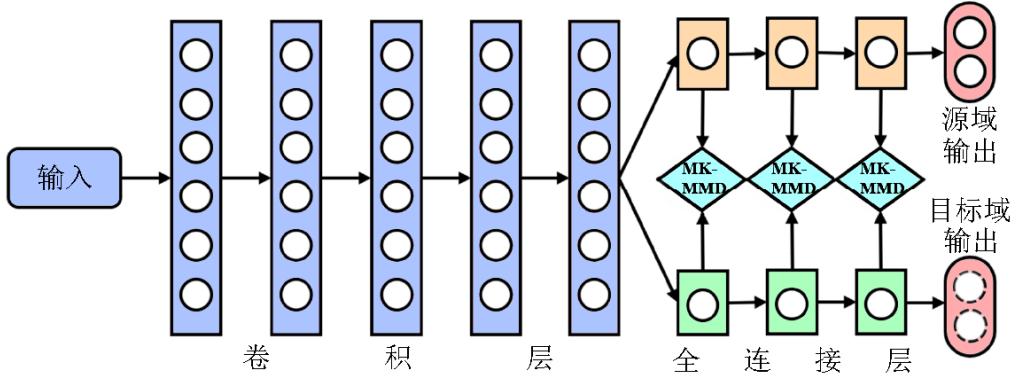


图 4.1 DAN 的网络框架

领域间分布差异损失的权重，并以此构建局部最大均值差异 (Local Maximum Mean Discrepancy, LMMD) 方法<sup>[23]</sup>，具体如下：

$$\text{LMMD}(D_s, D_t) = \frac{1}{C} \sum_{c=1}^C \left\| \sum_{x_i^s \in D_s} w_i^{sc} \phi(x_i^s) - \sum_{x_j^t \in D_t} w_j^{tc} \phi(x_j^t) \right\|_{\mathcal{H}}^2 \quad (4.2)$$

$$w_i^c = \frac{y_{ic}}{\sum_{(x_j, y_j) \in D} y_{jc}} \quad (4.3)$$

其中， $w_i^{sc}$  和  $w_j^{tc}$  由公式(4.4)可得，代表了样本  $x_i^s$  和  $x_j^t$  对第  $c$  类的贡献。 $y_{ic}$  则表示标签向量在第  $c$  类上的预测值。

## (2) 沃森斯坦距离

沃森斯坦距离的定义为：

$$W_P(P(x^s), P(x^t)) = \left( \inf_{\gamma \in \tau} \int d(x^s, x^t) d\gamma(x^s, x^t) \right)^{\frac{1}{p}} \quad (4.4)$$

其中， $\tau$  是  $P(x^s)$  和  $P(x^t)$  所有联合分布的集合， $d(x^s, x^t)$  表示  $x^s$  和  $x^t$  的距离。Damodaran 等人在深度领域自适应任务中引入了沃森斯坦距离，通过构建源域样本和目标域样本之间的相似性矩阵，对不同样本之间的沃森斯坦距离进行加权<sup>[24]</sup>。Lee 等人在领域自适应任务中借鉴了高维特征与低维流形嵌入的思想，将高维特征映射到不同的低维空间中，并分别计算每个低维空间中的沃森斯坦距离，最后进行统一的优化<sup>[25]</sup>。

### 4.3.2 基于对抗学习的方法

基于分布差异的方法旨在通过缩小源域样本和目标域样本在特征空间中的分布差异来实现领域自适应。这种方法通过优化特征提取网络，使其能够提取出具有域不变性的特征。然而，随着领域自适应研究的不断发展，研究学者们意识到需要设计更复杂的度量距离来准确衡量两个域之间的分布差异。

传统的人工设计的度量距离往往无法充分匹配真实的分布差异，因此研究者们开始探索让网络自主学习和缩小分布差异的方法。基于对抗学习的方法应运而生。这些方法引入了对抗性训练框架，通过让生成器网络和判别器网络相互博弈，实现特征的域不变性学习。

在基于对抗学习的方法中，生成器网络负责生成目标域样本，而判别器网络则负责区分生成的目标域样本和真实的目标域样本。通过这种对抗性训练，生成器网络逐渐学习到如何生成更接近目标域的样本，从而缩小源域和目标域之间的分布差异。

相比于传统的人工设计的度量距离，基于对抗学习的方法能够更有效地学习和匹配源域和目标域之间的分布差异。网络能够自主地学习如何缩小分布差异，从而实现更好的领域自适应性能。Ganin 等人提出的 DANN (Domain Adversarial Neural Network) 方法首次在域适应任务中引入对抗学习，并针对对抗训练中存在两个优化目标的对立关系从而导致的训练不稳定问题，引入了梯度反转层 (Gradient reversal layer, GRL)，使模型在反向传播过程中，梯度反转层会将判别器的反传梯度取反，以保证在判别器上进行领域对抗损失的最小化训练的同时，在特征提取器上进行领域对抗损失的最大化训练<sup>[26]</sup>，DANN 的优化目标为：

$$E(\theta_g, \theta_f, \theta_d) = \sum_{i=1}^{n_s} L_f(G_f(G_g(x_i^s)), y_i^s) + \lambda \sum_j^N L_d(G_d(G_g(x_j)), d_j) \quad (4.5)$$

其中， $G_g$  和  $\theta_g$  表示特征提取器及参数， $G_f$  和  $\theta_f$  表示分类器及参数， $G_d$  和  $\theta_d$  表示域判别器及参数， $y_i^s$  表示源域样本的真实标签， $d_j$  表示样本的领域标签， $n_s$  表示源域样本的数量， $N$  表示源域和目标域的样本总量。

Chen 等人在研究中发现，虽然域适应方法能够提高特征的可迁移性，但会对特征的可判别性产生影响，通过矩阵分析，发现在特征矩阵的第一个特征值过大，导致其他特征值对应的特征向量被过度抑制，于是提出了 BSP 方法，通过在训练过程中对最大特征值进行惩罚，来确保其他特征向量不会被过度抑制，平衡特征的权重分布，从而够提高特征的可判别性，增强域适应的性能<sup>[27]</sup>。在对抗学习领域，为了实现更细粒度的分布对齐，许多研究开始引入类别信息。例如 Pei 等人提出了 MADA

方法，该方法通过引入多个领域判别器，每个判别器专注于一个类别的分布对齐，从而实现源域和目标域的局部分布对齐<sup>[28]</sup>；Long 等人提出了 CDAN 方法，该方法通过将分类器输出的预测概率与特征向量进行外积操作，得到新的特征向量，然后将这些新特征输入到判别器中，以利用判别信息更丰富的特征来辅助进行更细致的分布对齐<sup>[29]</sup>。

### 4.3.3 基于重构和基于生成的方法

基于重构的方法利用多个自编码器来提取与领域相关和领域无关的特征。这些自编码器被设计用于保留特征的原有特性，并通过最小化重构损失来实现这一目标。通过这种方式，方法能够分离出领域无关的特征，并同时保留原始特征的信息。Ghifary 等人提出了利用编码器进行两次编码的 TLDA 方法，该方法通过对特征进行两次编码操作，即经过第一次编码后再进行第二次编码，并对对应编码结果进行两次解码操作，解码过程也类似，与传统的自编码器不同的是，该方法将第二次编码的输出直接用作预测输出，充分利用了自编码器在获取领域不变特征时认可保持原有判别信息的特点<sup>[30]</sup>。Bousmalis 等人认为简单的重构操作无法完全解耦出领域相关信息，于是提出了一种特定的编码方式即 DSN，它通过使用领域相关编码器和领域无关编码器对特征进行编码，并在编码空间引入正交约束，保证了编码器对特征的解耦能力<sup>[31]</sup>。

基于生成的方法利用生成模型来合成带标签的目标域样本，并通过比较合成样本与真实样本之间的差异来约束生成模型的学习。通常使用生成对抗网络作为生成模型，在源域样本和合成样本之间建立对应关系，以推测合成样本的标签。

尽管基于重构的方法可以抑制可迁移语义信息的损失，但由于自编码器的特征表示能力的限制，在未来的研究中，可以将自编码器作为一种辅助迁移的方式与域适应方法相结合，以提高特征的表征能力。另一方面，基于生成的方法虽然能够将域适应问题转化为监督学习问题，但同样受制于生成模型的性能。在复杂的数据集中，合成样本的难度较高，可能会导致生成模型难以生成逼真的目标域样本。因此，在进一步的研究中，需要关注提升生成模型性能的方法，以更好地应对复杂数据集中的域适应问题。综上所述，尽管基于重构的方法可以减少可迁移语义信息的损失，但其特征表示能力有限。在未来的研究中，可以将自编码器作为辅助迁移的方式与域适应方法相结合。而基于生成的方法虽然能将域适应问题转化为监督学习问题，但在复杂数据集中合成样本的难度较大。因此，进一步的研究应关注提升生成模型性能，以应对复杂数据集中的域适应问题。

## 4.4 实验设计与分析

### 4.4.1 数据预处理及划分

本章采用江南大学滚动轴承故障诊断数据集内各转速的 4 种故障状态信号，所有状态信号使用的预处理方式与第 2 章一致，得到大小为  $56 \times 56$  的灰度图像样本，总计 3600 个。其中每种领域各状态的样本总数如表 4.1 所示。

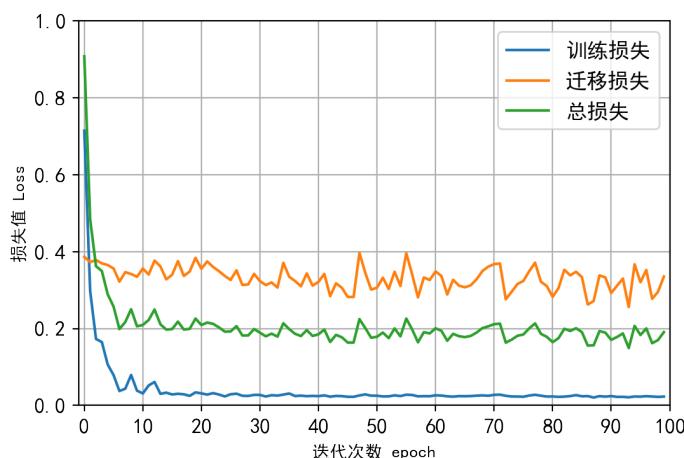
**表 4.1 每种领域各状态的样本数量**

编号	故障类别	样本数量
0	正常	300
1	内圈故障	300
2	外圈故障	300
3	滚动体故障	300

本章将 600rpm 转速下的工况称之为 A 工况，800rpm 转速下称之为 B 工况，1000rpm 下称之为 C 工况，同时定义  $A \rightarrow B$  为将 A 作为源域、B 作为目标域进行领域自适应学习。

### 4.4.2 模型结构

本章实验采用 DAN 方式来实现域适应任务，使用 MMD 作为衡量源于与目标域之间均值差异的指标，优化器选择随机梯度下降，设置动量为 0.9，同时为降低网络计算成本，设置学习率初值为 0.0003，并设置衰减率为 0.75，即在每次迭代后学习率都会是上一次的 75%。先通过 VGG11 方式进行  $A \rightarrow B$  的过程，经过 100 次迭代后得到的训练过程如图 4.2 所示。



**图 4.2 DAN 训练过程**

可以看出，VGG11 模型下的训练损失在前 10 次迭代训练中急速下滑，到 15 次

迭代训练左右稳定在最小值附近；而迁移损失波动较大，但也存在较小的下降趋势，同时该 DAN 的准确率高达 90% 以上，说明 DAN 方式可以为实验二位灰度图处理后的故障样本进行深度领域适。

分别使用 VGG11 和 ResNet50 作为骨干网络，在不同源域与目标域下各迭代 25 次，采用分类准确率来评估这些方法的性能，并迭代结果取均值，最终获得每种方法迁移任务中的平均准确率，如表 4.2 所示。

**表 4.2 模型在迁移学习中的预测准确率**

骨干网络	A→B	A→C	B→A	B→C	C→A	C→B
ResNet50	90.27%	<b>95.05%</b>	80.42%	85.62%	81.61%	92.44%
VGG11	<b>97.73%</b>	97.19%	86.96%	97.52%	82.59%	96.56%

从表中可以看出，基于 VGG11 和 ResNet50 的两种网络进行的迁移学习效果都较好，预测准确率均超过 80%；同时从骨干网络的角度来看，可以得出 VGG11 相对于 ResNet50 具有更好的域迁移性能，可能是因为 ResNet50 相对于 VGG11 来说是一个更复杂和参数更多的网络模型，复杂的模型可能容易过拟合低转速样本，并且对高转速样本的泛化能力较差；从不同迁移任务来看，将 600rpm 的低转速样本迁移到高转速领域的迁移效果要优于将高转速领域迁移到低转速领域的效果，这可能是因为机械故障中的故障信号会更多地体现在高频范围，故在低转速条件下更容易检测和区分。

## 4.5 本章小结

本章主要对变工况故障诊断进行研究，首先，总结了变工况故障诊断面临的问题，其次，给出了领域自适应定义及目的，介绍了深度适配网络的基本原理和其主流的四种方法及其衍生，然后，通过实验，将 VGG11 和 ResNet50 作为 DAN 的骨干网络，对比得出了 VGG11 作为骨干网络要优于 ResNet50 的结论，最后，将同种网络的不同迁移任务进行对比，得出使用较低转速的领域作为源域的迁移效果更好。

## 第五章 总结与展望

### 5.1 工作总结

故障诊断不仅是确保设备运行安全稳定的基础，更是保障各类工业工作人员健康安全的关键，对环境和经济的可持续发展具有积极影响。本文针对故障诊断的实际工程问题，从特征提取、不均衡样本和变工况故障诊断三个角度，探讨基于深度迁移学习的方法，设计了多个神经网络模型，主要的研究工作如下：

(1) 针对故障诊断的特征提取问题，采用了深度学习中的卷积神经网络作为故障特征提取工具。以 VGG11 为基础，构建基于 ImageNet 预训练的迁移模型，并将故障数据转化为二维灰度图，对模型进行故障分类验证实验，验证模型迁移缩短了模型的训练时间。同时为探究模型不同参数对故障分类结果的影响，对模型进行不同深度的冻结，发现当冻结 4 层卷积层时，得到的模型效果最好，效率最高。除此之外，通过改变模型的学习率和数据样本大小，发现在相同的样本下，学习率设置为 0.002 时损失函数的收敛速度较快且较为稳定；在学习率固定的情况下，预处理成  $56 \times 56$  的样本相较于  $32 \times 32$  和  $28 \times 28$  样本准确率大幅度提升。

(2) 针对故障诊断中数据不均衡问题，采用了生成对抗网络作为数据增强办法，对原始的故障数据进行增强。根据故障数据的形式，分别以人工神经网络和卷积神经网络为基础，构建了两种生成对抗网络，用训练好的生成器对少数样本进行增强，并将原始数据与增强后的数据进行测试对比，得出两种生成对抗网络均能有效提高对原数据库中少数样本的特征提取能力，从而实现了更准确的故障分类。同时将两种方式增强后的数据使用同一种模型进行分类训练，得出在层数有限的情况下，使用人工神经网络构建的生成器生成的伪造数据略优于使用卷积神经网络生成的数据。

(3) 针对故障诊断的变工况问题，采用了深度域适应将不同工况的数据进行迁移。根据故障数据的形式，构建了基于 VGG11 和 ResNet50 两种卷积神经网络作为骨干网络，对故障数据进行迁移学习，得到 VGG11 的域迁移效果比 ResNet50 更好的结果。同时，针对两种模型对 3 种不同工况的故障数据进行域迁移，得出使用较低转速的领域作为源域的迁移效果更好。

### 5.2 未来展望

本文对基于深度学习的故障诊断的方法进行了初步研究，针对故障诊断中的不同情况设计了多个有效的故障诊断模型，但是在实际应用中，大规模工业系统具有复杂性、相关性和不确定性等特点，这给故障诊断带来了许多挑战和问题，在这些

研究结论的基础上，值得进一步开展的研究工作包含：

(1) 针对模型迁移，本次研究主要运用 VGG11 和 ResNet 等主流深度卷积网络作为基本网络构建，但是近几年一些轻量级网络逐渐兴起，可能可以进一步缩短模型训练的时间。同时由于图像数据可以储存 3 个通道，可以考虑进行预处理时使用不同的方法，并将它们的结果分别保存在图像的不同通道中，形成多通道特征图像。

(2) 针对不均衡故障数据，本次研究使用生成对抗学习进行数据增强。生成对抗学习可以用于融合多模态数据，生成多模态特征表示或合成多模态样本，从而提供更全面和准确的故障诊断信息。未来的研究可以致力于提高生成模型的鲁棒性，使其对输入数据的扰动和变化具有更好的适应性。同时，还可以探索生成模型的可解释性，使其生成过程更加可解释和可控。

(3) 针对变工况故障诊断，本次研究使用的领域自适应方法往往仍需要一定数量的标记数据来训练源领域模型。未来的研究将专注于开发更有效的少标记和无标记领域自适应方法，以减少对标记数据的需求，提高模型的泛化性能。同时，深度领域自适应可以通过对不同设备数据进行域自适应，使模型能够适应不同设备的特点，实现跨设备的故障诊断。

## 插图索引

图 1.1 故障诊断分类 .....	3
图 1.2 基于解析模型的故障诊断过程 .....	4
图 1.3 不同数据分布表示 .....	7
图 2.1 人工神经网络结构及其神经元结构 .....	11
图 2.2 经典卷积神经网络模型 <sup>[14]</sup> .....	12
图 2.3 深度神经网络迁移过程 .....	16
图 2.4 江南大学滚动轴承故障诊断数据采集装置 .....	17
图 2.5 故障轴承实体图, 其中 (a) 为外圈故障, (b) 为内圈故障, (c) 为滚动体故障 .....	17
图 2.6 重叠采样示意图 .....	18
图 2.7 $5 \times 5$ 卷积核转化为 $3 \times 3$ 卷积核的示意图 .....	19
图 2.8 模型迁移前后的部分参数指标 .....	21
图 2.9 不同学习率对准确率的影响 .....	22
图 2.10 不同图像大小对准确率的影响 .....	22
图 3.1 不同的图像增强方法 .....	24
图 3.2 GAN 的模型架构 .....	27
图 3.3 未使用数据增强时模型的训练过程 .....	28
图 3.4 未使用数据增强时模型的混淆矩阵 .....	29
图 3.5 GAN 中生成器生成的伪造图像 .....	30
图 3.6 使用 GAN 进行数据增强后模型的训练过程 .....	30
图 3.7 使用 GAN 进行数据增强后模型的混淆矩阵 .....	31
图 3.8 DCGAN 中生成器生成的伪造图像 .....	31
图 3.9 迭代训练 200 次后的生成器生成的 32 张伪造图像 .....	32
图 3.10 使用 DCGAN 进行数据增强后模型的训练过程 .....	33
图 3.11 使用 DCGAN 进行数据增强后模型的混淆矩阵 .....	33
图 4.1 DAN 的网络框架 .....	39
图 4.2 DAN 训练过程 .....	42

## 表格索引

表 1.1	机器学习与迁移学习间的区别 .....	6
表 2.1	数据库样本类型及划分 .....	19
表 2.2	VGG11 结构参数 .....	20
表 2.3	冻结不同卷积层数下的诊断效果比较 .....	20
表 3.1	生成器和判别器的结构 .....	27
表 3.2	生成器和判别器的结构 .....	32
表 4.1	每种领域各状态的样本数量 .....	42
表 4.2	模型在迁移学习中的预测准确率 .....	43

## 参考文献

- [1] 虞和济, 韩庆大, 李沈, 等. 设备故障诊断工程 [M]. 冶金工业出版社, 2001.
- [2] 李国华, 张永忠. 机械故障诊断 [M]. 化学工业出版社, 1999.
- [3] 姜波, 黄捷. 机械故障诊断技术的发展现状和未来趋势 [J]. 科技创新导报, 2018, 33 (15): 65–66.
- [4] 芦奕霏. 基于深度学习的轴承故障诊断方法研究 [D]. 南京邮电大学, 2022.
- [5] 李旭东. 基于深度学习的旋转部件故障诊断研究 [D]. 中国科学院大学(中国科学院国家空间科学中心), 2021.
- [6] Ge M, Lv Y, Ma Y. Research on multichannel signals fault diagnosis for bearing via generalized non-convex tensor robust principal component analysis and tensor singular value kurtosis [J]. IEEE Access, 2020: 8:178425–178449.
- [7] 邢晓松. 基于小样本深度迁移学习的轴承智能诊断方法研究 [D]. 电子科技大学, 2022.
- [8] Paul R, Hawkins S H, Balagurunathan Y, et al. Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival among Patients with Lung Adenocarcinoma [J]. Tomography, 2016, 2 (4): 388–395.
- [9] 康守强, 胡明武, 王玉静, 等. 基于特征迁移学习的变工况下滚动轴承故障诊断方法 [J], 2019: 10.
- [10] Alothman B, Janicke H, Yerima S Y. Class Balanced Similarity-Based Instance Transfer Learning for Botnet Family Classification [C] // Soldatova L, Vanschoren J, Papadopoulos G, et al. In Discovery Science, Cham, 2018: 99–113.
- [11] Hossain I, Khosravi A, Hettiarachchi I T, et al. Informative instance transfer learning with subject specific frequency responses for motor imagery brain computer interface [C]. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017: 252–257.
- [12] Segev N, Harel M, Mannor S, et al. Learn on Source, Refine on Target: A Model Transfer Learning Framework with Random Forests [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (9): 1811–1824.
- [13] Si S, Tao D, Geng B. Bregman Divergence-Based Regularization for Transfer Subspace Learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22 (7): 929–942.
- [14] 邱锡鹏. 神经网络与深度学习 [M]. 机械工业出版社, 2020.
- [15] Janssens O, Slavkovikj V, Vervisch B, et al. Convolutional Neural Network Based Fault Detection for Rotating Machinery [J/OL]. Journal of Sound and Vibration, 2016, 377: 331–345. <https://www.sciencedirect.com/science/article/pii/S0022460X16301638>.
- [16] Li K. 江南大学滚动轴承故障数据集 [EB/OL]. <http://mad-net.org:8765/explore.HTML?t=0.5831516555847212>, accessed on August 2019.
- [17] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [J]. arXiv preprint arXiv:1406.2661, 2014.
- [18] 梁俊杰, 韦舰晶, 蒋正锋. 生成对抗网络 GAN 综述 [J]. 计算机科学与探索, 2020, 14 (1): 1–17.

- [19] Tzeng E, Hoffman J, Zhang N, et al. Deep domain confusion: Maximizing for domain invariance [J]. arXiv preprint arXiv:1412.3474, 2014.
- [20] Long M, Cao Y, Wang J, et al. Learning transferable features with deep adaptation networks [C]. In International conference on machine learning, 2015: 97–105.
- [21] Long M, Zhu H, Wang J, et al. Deep transfer learning with joint adaptation networks [C]. In International conference on machine learning, 2017: 2208–2217.
- [22] Zhang X, Yu F X, Chang S-F, et al. Deep transfer network: Unsupervised domain adaptation [J]. arXiv preprint arXiv:1503.00591, 2015.
- [23] Zhu Y, Zhuang F, Wang J, et al. Deep subdomain adaptation network for image classification [J]. IEEE transactions on neural networks and learning systems, 2020, 32 (4): 1713–1722.
- [24] Damodaran B B, Kellenberger B, Flamary R, et al. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation [C]. In Proceedings of the European conference on computer vision (ECCV), 2018: 447–463.
- [25] Lee C-Y, Batra T, Baig M H, et al. Sliced wasserstein discrepancy for unsupervised domain adaptation [C]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 10285–10295.
- [26] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation [C]. In International conference on machine learning, 2015: 1180–1189.
- [27] Chen X, Wang S, Long M, et al. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation [C]. In International conference on machine learning, 2019: 1081–1090.
- [28] Pei Z, Cao Z, Long M, et al. Multi-adversarial domain adaptation [C]. In Proceedings of the AAAI conference on artificial intelligence, 2018, 32.
- [29] Long M, Cao Z, Wang J, et al. Domain adaptation with randomized multilinear adversarial networks [J]. arXiv preprint arXiv:1705.10667, 2017.
- [30] Ghifary M, Kleijn W B, Zhang M, et al. Deep reconstruction-classification networks for unsupervised domain adaptation [C]. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14, 2016: 597–613.
- [31] Bousmalis K, Trigeorgis G, Silberman N, et al. Domain separation networks [J]. Advances in neural information processing systems, 2016, 29.
- [32] 夏天琦. 基于深度迁移学习的故障诊断算法研究 [D]. 辽宁师范大学, 2022.
- [33] Shen F, Chen C, Yan R, et al. Bearing fault diagnosis based on SVD feature extraction and transfer learning classification [C]. In 2015 Prognostics and System Health Management Conference (PHM), Beijing, 2015: 1–6.
- [34] 陈伟. 基于深度迁移学习的异步电机故障诊断方法研究 [D]. 中国矿业大学, 2022.
- [35] Zhu J, Chen N, Shen C. A New Deep Transfer Learning Method for Bearing Fault Diagnosis Under Different Working Conditions [J]. IEEE Sensors Journal, 2020, 20 (15): 8394–8402.
- [36] Yu F, Xiu X, Li Y. A Survey on Deep Transfer Learning and Beyond [J/OL]. Mathematics, 2022, 10 (19). <https://www.mdpi.com/2227-7390/10/19/3619>.
- [37] 孙自豪. 基于深度迁移学习的轴承故障诊断算法研究 [D]. 山东大学, 2022.

- [38] Maschler B, Weyrich M. Deep Transfer Learning for Industrial Automation: A Review and Discussion of New Techniques for Data-Driven Machine Learning [J]. IEEE Industrial Electronics Magazine, 2021, 15 (2): 65–75.
- [39] Zhao Z, Zhang Q, Yu X, et al. Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1–28.
- [40] 吉辉. 基于生成对抗网络的图像增强技术研究 [D]. 西安理工大学, 2022.
- [41] Yu F, Xiu X, Li Y. A Survey on Deep Transfer Learning and Beyond [J]. Mathematics, 2022, 10 (19): 3619.
- [42] 郭歌阳. 面向生成对抗网络的深度伪造图像检测算法研究 [D]. 西安理工大学, 2022.
- [43] 刘权. 基于改进生成对抗网络的旋转机械故障诊断研究 [D]. 华北理工大学, 2022.
- [44] Zhang S, Zhang S, Wang B, et al. Deep learning algorithms for bearing fault diagnostics—A comprehensive review [J]. IEEE Access, 2020, 8: 29857–29881.
- [45] Liu M-Y, Tuzel O. Coupled generative adversarial networks [J]. Advances in neural information processing systems, 2016, 29.
- [46] Sankaranarayanan S, Balaji Y, Castillo C D, et al. Generate to adapt: Aligning domains using generative adversarial networks [C]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 8503–8512.
- [47] 王梓懿. 基于样本差异性的深度领域自适应研究 [D]. 大连理工大学, 2022.
- [48] 黄志勇. 基于图像分类的领域自适应方法研究 [D]. 华北电力大学(北京), 2021.
- [49] Wang J, et al. Everything about Transfer Learning and Domain Adaption [EB/OL]. <http://transferlearning.xyz>.
- [50] Chen X, Yang R, Xue Y, et al. Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016 [J]. IEEE Transactions on Instrumentation and Measurement, 2023.

## 致 谢

时光匆匆，转瞬四年，大学本科生活即将结束，我即将迎来人生的一个重要转折点。回顾大学四年，我收获了许多宝贵的经验和知识，无论是学术上的积累、人际交往的智慧，还是身体素质的提升，都离不开那些曾经关心和帮助过我的人们。

首先，我要衷心感谢我的导师修贤超老师。他待人和蔼可亲，对待学术问题认真严谨。他为我们精心设计研究课题，频繁分享一手的学术资料，定期与我进行学术交流，并在我遇到研究难题时耐心解答，使得我的选题能够顺利进行。修老师还多次帮我点出论文中的问题与缺漏，并给予指导性的建议，最终帮助我完成了毕业论文的撰写。

同时，感谢于福超学长，在研究初期为我指点迷津，并为我的研究提出了一些建设性意见，给予了我莫大的帮助；同时他一直鼓励我，给了我完成课题的信心。

其次，我要感谢在大学四年中教授过我的每一位老师和辅导员。老师们是知识的源泉和引路人，他们教授课程不仅仅是传授知识，更锻炼了我思考问题、分析问题和解决问题的能力，他们的讲解方式或沉稳，或生动，但都让我更容易理解复杂的理论和概念。辅导员和一对一导师则是我大学生活中的导航者和关心者，他们时刻关心我的学习和生活状况，提供各种帮助和支持。

再次，我要感谢大学四年中的朋友们，是他们与我分享快乐，分担烦恼，给予各种帮助，让我感受到来自各地朋友之间纯真的友谊。这里特别感谢我的室友们，在学习上，我们经常互相探讨问题，促进对新知识的理解与思考，在生活上，我们互帮互助、相互照应，共同创造了一个温馨和谐的室友生活。

最后，我想衷心感谢我的父母和家人。感谢你们对我的理解和支持，使我拥有了向前迈进的勇气。感谢你们赐予我生命和宝贵的经历，让我成长为一个敢于面对未知、不畏挫折的人。

正如一句名言所说：“走好选择的路，别选择好走的路，你才能拥有真正的自己。”感谢过去的自己，也是你们的坚持和选择，才让我完成毕业论文的最后一个字。

