



南京大學

研究生畢業論文  
(申請碩士學位)

論文題目	<u>基于机器学习的水下信号识别研究</u>
作者姓名	<u>刘恒</u>
专业名称	<u>计算机科学与技术</u>
研究方向	<u>机器学习</u>
指导教师	<u>申富饶 教授</u>

2022 年 5 月 22 日

学 号： MG1937016

论文答辩日期： 2022 年 5 月 17 日

指导教师：  (签字)

# Research on underwater signal recognition based on machine learning

by

**Liu Heng**

Supervised by

**Professor Shen Fu-Rao**

A dissertation submitted to  
the graduate school of Nanjing University  
in partial fulfilment of the requirements for the degree of

MASTER

in

Computer Science and Technology



School of Artificial Intelligence  
Nanjing University

May 22, 2022



# 南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目： 基于机器学习的水下信号识别研究

计算机科学与技术 专业 2022 级硕士生姓名： 刘恒  
指导教师（姓名、职称）： 申富饶 教授

## 摘 要

我国是海洋大国，随着航海技术和海洋产业的快速发展，水下通信的应用变得愈加广泛。对采集到的水声信号进行正确分类，从中获取信号的特征信息，是水声信号系统的一个重要课题。船只辐射噪声信号是一类重要的水声信号，该类信号可以反映船只的载货量、速度以及零部件工作状况等信息，正确识别辐射噪声信号所属的船只类型，对于海洋产业有重要的实际意义。

船只辐射噪声信号是船只的航行运动所产生的一类信号，它是一种典型的时变信号，其信号特征由对应船只的船体结构、推进器工作状态以及水域特性所决定，同时还会受到船只航行速度、天气等因素的影响。船只辐射噪声信号复杂的形成机理，以及其容易受到环境影响的特性，导致了现有识别方法的识别性能大都较差。同时由于收集船只辐射噪声信号需要使用昂贵的采集设备，噪声数据采集周期长，采集实验难度大等原因，船只辐射噪声数据集的规模一般较小，这进一步提高了辐射噪声识别算法开发的难度。本文主要聚焦了上述水下辐射噪声信号识别问题，开展了如下的研究：

- 在辐射噪声信号识别任务中，我们设计了一种带帧维度注意力机制的时延网络模型，简称 TDNN-attn。该模型利用时延卷积单元来构建特征提取模块，时延卷积单元可以提取具有时频特性的深度特征，相较于传统的卷积特征，该特征更适合辐射噪声任务。同时为了加强有效信号片段对分类结果的影响，我们还设计了一个帧级别注意力层，该结构输出不同时间段对应的深度特征的重要性系数，通过向量点乘的方法来加强深度特征中的有效片段。我们在两个辐射噪声信号数据集上进行实验，实验结果表明，相较于现有的时频分析算法以及神经网络算法，TDNN-attn 模型可以更好的识别不同类别船只发出的辐射噪声信号。

2. 在辐射噪声信号的数据增强任务中，我们设计了一种基于变分自编码器的数据增强算法，简称 VAE-tdc。该模型使用我们设计的 TDNN-attn 模型的特征提取部分作为编码器，学习输入数据在特征空间的分布，然后在该分布上进行采样得到特征向量，将此特征向量通过转置时延卷积网络解码生成重构数据，用于数据增强。和另外几种基于时频分析的数据增强方法对比，我们设计的数据增强模型可以更好地提高识别算法的性能。

3. 在前面研究的基础上，我们基于 Python 设计了一个水下辐射信号识别软件，该软件实现了辐射噪声数据可视化、数据存储、识别模型训练以及辐射噪声信号识别的功能。通过在辐射噪声处理任务中使用该软件，可以一站式地完成各项操作，从而极大地提高工作效率。

**关键词：**机器学习, 水下信号识别, 信号处理

## 南京大学研究生毕业论文英文摘要首页用纸

THESIS: Research on underwater signal recognition  
based on machine learning

SPECIALIZATION: Computer Science and Technology

POSTGRADUATE: Liu Heng

MENTOR: Professor Shen Fu-Rao

### **Abstract**

China is a country with rich marine resources, and rapidly expanding maritime activities, where there is correspondingly a growing demand by practitioners in the marine sector, for more advanced technologies to explore the underwater environment, and in particular, for efficient underwater acoustic communication systems to facilitate and serve better a broader range of practical applications such as maritime transport, deep-ocean scientific exploration, energy exploitation, etc. It is an important topic of underwater acoustic signal system to correctly classify the collected underwater acoustic signals and obtain the characteristic information of signals. Ship radiated noise signal is a kind of important underwater acoustic signal. This kind of signal can reflect the information of ship's cargo volume, speed and working conditions. Correctly identifying the ship type of radiated noise signal has important practical significance for marine industry.

The ship radiated noise signal is a kind of signal generated by the ship's navigation action. It is a typical time-varying signal. Its signal characteristics are determined by the corresponding ship's hull structure, propeller working state and water area characteristics around ship. At the same time, it is also affected by the ship's navigation speed, weather and other factors. The complex formation mechanism of ship radiated noise signal and its vulnerability to environmental impact lead to the poor recognition performance of the existing recognition methods. At the same time, due to the use of expensive ship radiated noise acquisition equipment, long noise signal acquisition cycle and great difficulty in signal acquisition experiments, the scale of ship radiated noise dataset is generally small, which further improves the difficulty of radiated noise

classification. This paper mainly focuses on the underwater radiated noise signal classification, and carries out the following research:

1. In the task of radiated noise signal classification, we design a time delay neural network model with frame level attention mechanism, which is called TDNN-attn for short. The feature extraction module is constructed by time delay convolution units, which can extract the depth feature with time frequency characteristics. Compared with the traditional convolution feature, this feature is more suitable for radiated noise classification tasks. At the same time, in order to strengthen the influence of effective signal segments on classification results, we also designed a frame level attention layer, which outputs the importance coefficients of depth features corresponding to different signal segments, for strengthening the depth features of effective segments by vector point wise multiplication. We conduct experiments on two radiated noise signal datasets. The experimental results show that compared with the existing time-frequency analysis algorithms and neural network recognition algorithms, TDNN-attn model can classify the radiated noise signals of different ship types better.

2. In the task of radiated noise data augmentation, we design a data augmentation algorithm based on variational self-encoder, which is called VAE-tdc for short. The model uses the feature extraction part of TDNN-attn model as the encoder to learn the distribution of the input data in the feature space, and then obtain the feature vector by sampling the feature space distribution. The feature vector is decoded through the transpose time delay convolution network to generate the reconstructed data for data augmentation. Compared with other data augmentation methods based on time-frequency analysis, the data generated by VAE-tdc can better improve the performance of recognition algorithm.

3. Based on the above research, we designed a underwater radiation signal recognition software based on python. The software realizes the functions of radiation noise data visualization, data storage, recognition model training and radiation noise signal recognition. By using the software in the radiated noise processing task, various operations can be completed easily, so as to greatly improve the work efficiency.

**keywords:** machine learning, underwater signal recognition, signal process

# 目 录

目 录 .....	v
插图清单 .....	ix
附表清单 .....	xi
<b>1 绪论 .....</b>	<b>1</b>
1.1 问题背景 .....	1
1.2 现有的研究进展 .....	2
1.2.1 基于建模的方法 .....	2
1.2.2 基于时频分析的方法 .....	3
1.2.3 基于深度学习的方法 .....	4
1.3 本文研究内容 .....	6
1.4 本文组织结构 .....	7
<b>2 相关工作 .....</b>	<b>9</b>
2.1 辐射噪声组成机理 .....	9
2.1.1 机械噪声原理 .....	9
2.1.2 螺旋桨噪声原理 .....	10
2.1.3 水动力噪声原理 .....	10
2.2 辐射噪声识别问题定义 .....	11
2.3 基于时频分析的辐射噪声识别算法 .....	12
2.3.1 短时傅里叶变换方法 .....	12
2.3.2 离散小波变换方法 .....	15
2.4 基于深度学习的辐射噪声识别算法 .....	19
2.4.1 卷积神经网络 .....	19
2.4.2 循环神经网络 .....	21
2.5 基于深度学习的辐射噪声增强算法 .....	22

2.5.1 对抗生成网络方法 .....	22
2.5.2 变分编码器方法 .....	24
2.6 本章小结 .....	26
<b>3 水下辐射噪声信号识别 .....</b>	<b>27</b>
3.1 研究动机 .....	27
3.2 基于帧维度注意力的深度时延网络 .....	27
3.2.1 模型结构 .....	28
3.2.2 时延卷积模块 .....	29
3.2.3 帧维度注意力模块 .....	32
3.2.4 统计池化模块 .....	35
3.2.5 模型训练 .....	36
3.3 实验结果和分析 .....	37
3.3.1 实验环境设置 .....	37
3.3.2 实验数据分析 .....	38
3.3.3 实验评估指标 .....	39
3.3.4 对比实验结果 .....	39
3.3.5 消融实验结果 .....	43
3.4 本章小结 .....	45
<b>4 水下辐射噪声数据增强 .....</b>	<b>47</b>
4.1 研究动机 .....	47
4.2 基于时延卷积的变分自编码器 .....	47
4.2.1 模型结构 .....	48
4.2.2 时延卷积编码器 .....	49
4.2.3 转置时延卷积解码器 .....	50
4.2.4 模型训练 .....	51
4.3 实验结果和分析 .....	53
4.3.1 数据增强结果可视化 .....	53
4.3.2 对比实验结果 .....	54
4.3.3 消融实验 .....	56
4.4 本章小结 .....	57

---

<b>5</b>	<b>水下辐射噪声识别系统</b> .....	<b>59</b>
5.1	系统开发背景 .....	59
5.2	系统设计 .....	60
5.2.1	系统需求设计 .....	60
5.2.2	系统架构设计 .....	61
5.3	系统实现 .....	63
5.3.1	系统开发环境 .....	63
5.3.2	系统功能模块 .....	63
5.3.3	系统演示 .....	65
5.4	本章小结 .....	67
<b>6</b>	<b>总结与展望</b> .....	<b>69</b>
	参考文献 .....	71
	致 谢 .....	77
	简历与科研成果 .....	79
	学位论文出版授权书 .....	81



# 插图清单

2-1	辐射噪声识别	11
2-2	不同信号傅里叶变换结果	13
2-3	不同信号的时频图	14
2-4	离散小波变换结果展示	16
2-5	梅尔滤波器组	17
2-6	梅尔倒谱系数流程图	18
2-7	循环神经网络结构图	21
2-8	长短时记忆网络结构图 <sup>[1]</sup>	22
2-9	对抗生成网络模型	23
2-10	自编码器模型	25
2-11	差分自编码器对比编码器模型	25
3-1	辐射噪声识别模型结构图	28
3-2	时延卷积数学原理	30
3-3	残差连接简易结构	31
3-4	普通卷积 vs. 空洞卷积 <sup>[2]</sup>	32
3-5	小波变换阈值去噪算法	33
3-6	帧维度注意力结构图	34
3-7	最大池化原理	35
3-8	统计池化原理	36
3-9	不同数据集的混淆矩阵结果	42
3-10	MFCC 特征可视化 vs. 时延卷积特征可视化	42
4-1	基于注意力的 VAE 模型	48
4-2	转置卷积原理图 <sup>[3]</sup>	51
4-3	不同数据增强方法结果图	54
4-4	不同数据集的混淆矩阵结果	56

---

5-1	辐射噪声信号识别用途 .....	59
5-2	数据可视化和存储需求原型图 .....	60
5-3	训练需求原型图.....	61
5-4	测试需求原型图.....	62
5-5	系统架构设计图.....	62
5-6	数据存储表结构.....	64
5-7	训练模块 UML 时序图 .....	65
5-8	数据可视化页面截图 .....	66
5-9	识别系统训练模式截图 .....	66
5-10	识别系统测试模式截图 .....	67

# 附表清单

3-1	TDNN-attn 模型参数细节 .....	37
3-2	实验服务器配置 .....	38
3-3	数据集 1 辐射噪声分布 .....	38
3-4	数据集 2 辐射噪声分布 .....	39
3-5	辐射噪声分类实验识别率 .....	41
3-6	时域波形 vs.MFCC 输入对比实验 .....	43
3-7	空洞卷积和残差连接结构对比实验 .....	44
3-8	不同池化层对比实验 .....	44
3-9	帧维度注意力层对比实验 .....	45
4-1	VAE-tdc 数据增强模型参数 .....	52
4-2	辐射噪声数据增强实验 .....	55
4-3	帧维度注意力层对比实验 .....	57



# 第一章 绪论

## 1.1 问题背景

我国是海洋资源大国，随着近现代航海工业，渔业等行业的快速发展，水声信号的研究越来越受到人们的重视。水下辐射噪声是一种常见的水声信号，此类信号是由船只在水中行驶所产生的。船只的辐射噪声可以反映船只的运行速度、船只载货量、零部件工作状况等船体信息，因此辐射噪声信号经常被用作识别不同船只的重要信息源。目前大多数船只辐射噪声分类工作都是由经过训练的专业人员完成的，受到专业人员培训的成本以及人工识别效率的限制，该项技术目前的应用范围仍然非常有限。为了降低辐射噪声识别技术的成本，扩大该项技术的应用范围，越来越多的科研工作者投入到辐射噪声自动化识别的研究中。

舰船辐射噪声识别是一类典型的时间序列识别任务，然而和其他时间序列分类问题相比，辐射噪声信号的识别更加困难，这是由该类信号的数据特点所导致的。下面我们总结了该问题的几个关键技术难点：

1. 辐射噪声信号是复杂的时变信号。和在空气中传播的信号相比，在水下传播的信号更加容易受到水下信道和变幻莫测的水生环境的影响。因为海面的随机运动、海底的不平整以及随时间变化、水体的不均匀分布等原因，水声信道不仅在空间上不是均匀的，在时域上也是随机可变的。再加上水声信号传播速度慢，码元周期长，使得水声信号更容易受到复杂的水声信道的影响，这进一步导致了水声信号时变、非平稳的特点。同时水声信号中还会包含由于各种水下生物的活动、海水的运动以及船只的行驶等行为产生的噪声，这些干扰信号会影响识别算法对水下信号的正确识别。因此即使是由同种类型的船只发出的辐射噪声信号，它们之间也存在着各种各样的差异，这种复杂的类内数据差异提高了辐射噪声信号分类的难度。

2. 辐射噪声信号数据规模小。辐射噪声的采集需要通过相关类别的船只进行海试来获得，而船只海试的时间周期一般较长，同时用于采集信号的设备也较为昂贵，这导致了一次测试中可以采集到的辐射噪声训练数据是有限的。训

练数据的规模也在一定程度上限制了机器学习算法的识别性能。

现有的水下辐射噪声识别研究大都是基于经典的信号处理算法来开展的，该类算法通过构造复杂的信号处理模型来完成信号降噪、信号特征提取的工作，然后通过比对不同信号之间特征的相似度来完成分类。此类方法具有较好的可解释性和理论依据，但是容易受到环境因素的影响，当实验场景发生变化时，需要依靠相关领域知识对模型参数进行调整。为了解决上述问题，本文基于人工智能技术展开了辐射噪声识别算法研究，提出了深度神经网络模型来完成辐射噪声识别任务，同时还设计了自动化的深度学习数据增强模块，来改善辐射噪声信号数据匮乏的问题，充实了水下辐射噪声识别领域的研究。

## 1.2 现有的研究进展

按照辐射噪声信号识别的研究方法，可以将其分为三类：

1. 基于信号建模的方法。此类研究大都会把辐射噪声信号视作一个复杂的信号系统的输出，利用相关领域知识和信号系统的方法，来建立辐射噪声信号的数学模型，然后基于此模型来计算辐射噪声信号的特征参数，最后使用这些特征参数来区分不同类别的辐射噪声。

2. 基于时频分析的方法。此类研究主要是通过信号处理技术来分析信号在时域、频域或者时频联合域的一些特性，比如信号能量分布特性、信号波形变化特性，来区分不同类别的辐射噪声。该类方法会通过信号去噪、信号分解、特征分解等手段对原始信号进行处理，然后将上述时频特性转变为低维的信号特征向量，基于这些信号特征向量构建识别模型。

3. 基于深度学习的方法。此类研究通过设计深度神经网络模型，通过大规模数据训练的方法，来学习辐射噪声数据背后中的复杂非线性特征，从而基于这些非线性特征构建识别模型。

下面我们将分别介绍这三类识别方法中，不同的研究者对于辐射噪声识别问题的研究。

### 1.2.1 基于建模的方法

水下辐射噪声可以看做是一个复杂的信号系统的输出，系统的输入是船只的机械零部件的振动，比如螺旋桨转动、船体振动，该输入如何通过一个个信号变换模块，进入接收装置得到最终的辐射噪声信号，正是该类方法研究者所

关心的内容。

Ross 等人<sup>[4]</sup> 通过研究商船的辐射噪声特点，最早提出了一种源模型-传播模型的建模方式，来对水下辐射噪声信号进行建模和分析。其中源模型描述了船舶产生辐射噪声的过程，以及船舶行驶的轨迹。它是通过将船只的发声源频谱，和水中某一固定深度的单级子声源进行联合计算得到的，其中船只的发声源频谱主要是由螺旋桨空化以及船只机械结构工作所产生的，它既包含了固定频率的窄带信号，也包含了一定范围频率的宽带信号。传播模型描述了辐射噪声传入信号收集系统的过程，它是通过计算发声源到接收装置之间的格林函数得到的。该辐射噪声模型对后来的研究起到了非常重要的启发作用。但是此类模型的识别效果很大程度上依赖船只发声源频谱的建模效果，因此应用于实际的识别任务时，需要进行复杂的理论分析和细致的模型设计才能保证模型的识别性能。

Wales 等人<sup>[5]</sup> 设计了一种新的发声源频谱建模方法，用来改进经典的源模型-传播模型建模算法。该算法提出了一种统一的源谱估计模型，用来建模不同船只的发声源频谱，并将其命名为有理谱模型（rational spectrum model）。有理谱模型通过一系列近似函数的线性组合来逼近原始声源谱中的线谱成分，该设计较好地建模了发声源频谱和船只的行驶速度，以及发声源频谱和船只的尺寸之间的相关性。在不同频段的辐射噪声信号上进行实验的结果，验证了有理谱模型刻画发声源频谱的有效性，基于有理谱模型构造的特征进行船只辐射噪声识别时，可以比以往的模型获得更小的识别误差。

尽管对辐射噪声进行建模，研究其背后的发声机理有利于从理论角度分析辐射噪声的特性，但是现实水域中充满了各种各样的干扰信号，这极大地提高了辐射噪声建模的复杂程度。因此基于建模的方法难以应用于真实的生产环境，仅适合作为一种理论分析手段。

### 1.2.2 基于时频分析的方法

随着人们对辐射噪声建模研究的深入，人们发现船只辐射噪声的一些时频特征和船只特性之间存在一定的对应关系，比如辐射噪声声谱中的线谱部分和螺旋桨的特性有关，声谱的连续谱和船只的吨位、航速、航行水域的深度有关。因此越来越多的研究者开始从辐射噪声的时频特性入手，研究如何从中挖掘出具有丰富表达能力的特征，从而构建优秀的辐射噪声识别模型。

由于辐射噪声声谱中的窄带线谱和船只的推动装置特性息息相关，提取声

谱中的线谱特征便成为了此类方法中的经典方法。Liu 等人<sup>[6]</sup>通过对舰船辐射信号进行功率谱估计和非线性最小二乘多项式拟合,得到连续谱,然后通过选择峰值和去除旁瓣的操作,从时频图中得到线谱,从中提取线谱的数目和强度特征,经过归一化处理来建立辐射噪声的线谱特征,最后利用支持向量机<sup>[7]</sup>作为分类器,完成信号分类。Zhang 等人<sup>[8]</sup>提出了一种基于卡尔曼滤波器<sup>[9]</sup>的线谱增强方法,该方法首先通过傅里叶变换确定了原始信号的频谱分布,然后通过卡尔曼滤波器降低了无用信号的权重,从而得到了更加清晰的线谱特征,最后同样是采用了支持向量机来完成信号分类。

另外一部分学者直接从信号的整体时频特征入手,通过数据降维的手段,从中提取出辐射噪声信号的特征。Li 等人<sup>[10]</sup>提出了一种基于小波包分形 (wavelet packets-fractal) 的方法来提取辐射噪声信号的时频特征。该算法首先通过小波包分解来计算辐射噪声信号的时频图,然后通过对不同尺度的小波分解序列构造分形维数特征 (fractal dimension),完成对时频图特征的压缩。在得到基于小波包分形的信号特征后,该方法使用支持向量机来对信号特征进行分类。在三类噪声信号上的实验结果表明,该算法具有优秀的识别性能。廖等人<sup>[11]</sup>利用辐射噪声信号在离散余弦变换 (Discrete Cosine Transform, DCT) 域上的近似稀疏特性,提出了一种基于感知压缩思路<sup>[12]</sup>的特征提取方法,在 DCT 域构造了具有稀疏特性的信号特征向量,并采用了支持向量机分类器来完成信号特征分类。实验结果表明该算法在提高了目标分类识别率的同时,还降低了水声信号的数据传输量,具有更加优秀的实用性能。

基于时频特征的方法具有优秀的可解释性和完备的理论依据,这类方法目前仍然是辐射噪声信号识别的主流方法。但是随着任务场景越来越复杂,人们发现此类方法需要基于数据特点精心设计特征,而这些特征在一些数据上表现得好,在另外的数据上就有可能效果很糟糕。因此设计一种可以自动提取辐射噪声信号特征,并且具有优秀的泛化能力的识别算法对于辐射噪声识别任务来说显得愈加重要。

### 1.2.3 基于深度学习的方法

近年来随着深度学习理论的发展以及数据计算设备的更新换代,深度神经网络技术得到了快速的发展,深度神经网络模型已经在图像识别、语音识别、自然语言处理等领域取得了巨大的成果。目前越来越多的学者提出了基于深度学习模型构建的辐射噪声信号识别模型<sup>[13]</sup>。相较于以往的时频分析方法,深度

学习模型不需要精心设计手工特征，通过大规模的数据训练即可提取到更加有表达力的非线性特征，同时该类模型在识别性能上也有了较大程度的提升。

Shen 等人<sup>[14]</sup> 根据人耳的听觉特性，设计了一种具有听觉机制的卷积神经网络用于辐射噪声分类任务。该模型主要包括耳蜗模型和听觉中枢模型，其中耳蜗模型通过时间卷积层和听觉滤波器来模拟人耳基膜上的声信号分解功能；听觉中枢模型通过注意力机制来模仿人耳对不同听觉特征的感知效果。在四种船型和海洋背景噪声进行的分类实验结果表明，该模型在辐射噪声识别任务中具有较高的准确度。

Tian 等人<sup>[15]</sup> 提出了一种基于多尺度残差单元的深度卷积神经网络 (Multi Scale Residual Deep Neural network, MSRDN) 来解决辐射噪声识别的问题。MSRDN 模型将不同尺寸的卷积核提取到的卷积特征在通道维度上进行融合，利用多尺度的卷积特征来提取更加丰富的信号特征；同时该模型还设计了一种自适应阈值的软阈值算子，对深度特征中的无效特征分量进行了过滤，提高了模型的鲁棒性。实验结果表明，该模型可以较好地识别各类船只的辐射噪声。

Hu 等人<sup>[16]</sup> 认为目前的神经网络模型的识别性能依赖模型的全连接层，该层待优化参数多的同时，泛化性能也较受限。因此他们首先训练一个完整的卷积神经网络模型，学习鲁棒并且具有丰富表达能力的深度特征，然后将模型中的全连接层替换为极限学习机模型<sup>[17]</sup> 来完成分类。辐射噪声识别实验证明，该方法在原有模型的基础上，极大地提高了模型的识别准确率。

尽管深度神经网络模型可以构造更加复杂的非线性特征，提高辐射噪声信号识别任务的识别性能，但是此类方法对训练数据的数量和质量要求较高。辐射噪声信号由于信号采集周期长，采集设备价格昂贵，所以在大多数任务场景下，实际可用的训练数据是有限的，这极大程度地限制了深度学习模型的识别能力，因此也有部分学者针对辐射噪声的数据增强开展研究。Jin 等人<sup>[13]</sup> 提出了一种基于对抗生成网络<sup>[18]</sup> 的数据增强方法，该算法基于卷积结构设计了对抗生成网络，然后将信号的时频谱图作为输入，通过对抗生成网络模型产生的生成数据来对辐射噪声数据进行数据扩充。实验结果表明，该模型生成的数据具有较好的分类质量，显著提高了原有识别算法的性能。

目前深度学习技术已经在辐射噪声信号问题中取得了优秀的成果，但是这些方法大都是将原始信号转换为二维图像的格式作为模型的输入，较少关注到辐射噪声信号本身的时频特性，因此容易丢失部分原始信号的特征。同时辐射噪声数据规模一直是限制深度神经网络模型性能的瓶颈，但是却少有人开展这

部分研究。综上所述，辐射噪声信号识别问题仍然有很大的改进空间和研究价值。

### 1.3 本文研究内容

本文基于水下辐射噪声识别问题展开研究，在分析了水下辐射噪声信号数据的特点之后，提出了一种基于深度神经网络的识别算法，该算法模型可以挖掘辐射噪声的复杂信号特点，从而有效地完成分类任务。此外本文还针对水下辐射噪声采集难度大，有效数据匮乏的问题，设计了一种辐射噪声信号数据增强模型，通过生成新的辐射噪声数据来扩充训练数据集，来提高识别模型的识别性能。最后本文还设计了一种水下辐射噪声信号识别系统，完成数据存储、模型训练和信号预测的工作，从而将我们开发的算法模型应用于实际的生产生活中。本文的主要研究内容如下：

1. 在辐射噪声信号识别任务中，我们设计了一种基于注意力机制的时延网络模型，简称 TDNN-attn。该模型利用时延卷积单元构建了特征提取模块，时延卷积单元可以提取具有时频特性的深度特征，相较于传统的卷积特征，该特征更适合辐射噪声任务。同时为了加强有效信号片段对分类结果的影响，我们还设计了一个帧级别注意力层，该结构输出不同时间段对应的深度特征的重要性系数，通过向量点乘的方法来加强深度特征中的有效片段。我们在两个辐射噪声信号数据集上进行实验，实验结果表明我们设计的模型可以较好的识别不同类别船只发出的辐射噪声信号，同时我们还和现有的主流的神经网络模型进行了对比实验，发现我们设计的模型具有最优的识别性能，最后我们还针对模型中不同的模块展开消融实验，证明了每个模块的有效性。

2. 在辐射噪声信号的数据增强任务中，我们设计了一种基于自编码器的数据增强算法，简称 VAE-tdc。该模型使用我们设计的 TDNN-attn 模型的特征提取部分作为编码器，学习输入数据在特征空间的分布，然后在该分布上进行采样得到特征向量，将此特征向量通过转置时延卷积网络解码生成增强数据。经过实验证明，和另外几种基于手工特征的数据增强方法对比，我们设计的数据增强模型可以更好地提高识别算法的性能，同时我们还针对模型中不同的模块展开消融实验，证明了每个模块的有效性。

3. 在前面研究的基础上，我们基于 Python 设计了一套水下辐射信号识别软件，该软件实现了辐射噪声数据可视化、数据存储、识别模型训练以及辐射噪

声信号识别的功能。

## 1.4 本文组织结构

本文的第一章节对研究课题的背景以及研究现状做了介绍；第二章对辐射噪声的信号特点，以及常用的辐射噪声识别算法方法展开了介绍；第三章介绍了本文的第一个研究课题，即辐射噪声信号识别任务的研究内容；第四章介绍了辐射噪声信号数据增强的研究内容；第五章介绍了我们设计的一套水下辐射噪声信号识别软件的相关内容；最后一章将会对全文的内容进行总结，并指出该研究后续的发展方向。



## 第二章 相关工作

辐射噪声信号是船只航行过程中发出的一类信号，通过对辐射噪声信号进行识别，我们可以了解其所属的船只类别，进而了解船只的工作状况等信息。本章我们将介绍辐射噪声识别的一些基础内容，其中包括了辐射噪声信号的组成原理，辐射噪声识别的问题定义，以及几种目前常用的识别算法。

### 2.1 辐射噪声组成机理

舰船辐射噪声是一类复杂的水声信号，它是由于舰船中多种不同的噪声传入水体中，与水体发生相互作用而产生的<sup>[19]</sup>。舰船上不同的噪声是由结构各异的噪声源所产生的，按照噪声源的类型进行分类，辐射噪声可以分为机械噪声、螺旋桨噪声和水动力噪声，其中机械噪声和螺旋桨噪声是辐射噪声最主要的两个成分。机械噪声是船上的机械结构所产生的噪声；螺旋桨噪声是一种混合型的噪声，它与机械噪声和水动力噪声有相似的特征和的发声源；水动力噪声是当不规则的水流流过舰船时，由于水动力过程的变化所引起的噪声。

在研究具体的识别算法之前，我们有必要了解一下辐射噪声的主要声音成分的作用机理，下面我们将会详细介绍上述三种主要噪声成分的特点。

#### 2.1.1 机械噪声原理

机械噪声是通过舰船机械结构的振动而形成的一类噪声。船舶机械振动的方式包括两类，一类是机械设备振动，比如电机转动产生的振动，齿轮旋转产生的振动等；另一类是推进系统振动，比如凝汽器中气体流动产生的振动，推进器部件之间的摩擦等。其中，机械设备振动产生线谱噪声，此类辐射噪声具有较为固定的频率；推进系统振动产生连续谱噪声，此类辐射噪声的频率范围分布较广，并且能量分布和频率之间没有明显的关系。船只的机械噪声是由线谱噪声和连续谱噪声组合而成的，其中的线谱噪声通常由多种船只结构的规律性振动所产生，它们的频率一般较低，是舰船辐射噪声中主要的低频分量，并

且这些低频线谱噪声和舰船运行的速度、机械装置的作业状态等信息有关，因此机械噪声一般是复杂多变的。

### 2.1.2 螺旋桨噪声原理

螺旋桨是舰船的主要推进装置，它产生的噪声与机械噪声有不同的源和不同的频谱。机械噪声基本都是由舰船内部的噪声源产生，然后通过船体传播进入水体；而螺旋桨噪声在船体外部产生，由螺旋桨转动及船在水中航行引起。螺旋桨噪声主要来源于螺旋桨的空化现象，船舶前进时会把大量空气卷进海水，此时螺旋桨在水中转动就会产生大量的空气泡，这些气泡在螺旋桨周围会发生空化现象。螺旋桨空化可分为以下两类：

1. 叶尖涡流空化。螺旋桨在海水中高速转动时，螺旋桨周围会形成低压水体，当水压下降达到某一阈值时，水中就会产生气泡，这类气泡的出现和消失导致了螺旋桨叶尖、叶面出现强烈的空化现象。叶尖涡流空化是螺旋桨的重要噪声源，它们的幅度不稳定，产生了包含了丰富频率分量的频率谱。

2. 叶片表面空化。在螺旋桨转动过程中，螺旋桨叶片周围水体的运动状态呈现周期性变化，此时水体产生的周期性空化即为叶片表面空化现象。此类空化现象具有相对固定的频率，因此所产生是离散的线谱。

总结来说，螺旋桨空化产生的噪声具有如下特点：叶尖涡流空化噪声是由大量随机小气泡破裂引起的，因此此类空化产生的噪声声谱为高频段的连续谱；叶片表面空化噪声是由规律性的叶片旋转产生的，此类空化噪声声谱为高频段的线谱。由于螺旋桨运动速度较快，螺旋桨空化噪声的频率较高，占据了舰船辐射噪声的高频部分。

### 2.1.3 水动力噪声原理

水动力噪声是由船只所处水域的水体和船体结构共同作用而形成的。水流和船体结构共振所产生的噪声，以及航船循环系统和海水作用产生的噪声都是水动力噪声的主要组成成分。

在正常的情况下，水动力噪声在船只辐射噪声中的占比不大，此时其他两种辐射噪声是舰船辐射噪声信号的主要组成成分。但当船体结构和海水产生较强的共振现象时，水动力噪声会演变成具有较高能量的线谱噪声，成为辐射噪声中重要的信号组成成分。

综上所述，辐射噪声信号是一类由多种舰船噪声组成的复合信号，其频谱中即包含了固定频率的窄带线谱，也包含了大量的宽带连续谱。宽带连续谱噪声的形成机理较为复杂，通常认为其能级和频率结构与船舶吨位、类型、航速、航行区域的水深等诸多因素有关。窄带线谱噪声的形成与舰船的机械零件、推进器、螺旋桨等部件有关，其中机械零件产生的辐射噪声声谱与航速无关；推进器与螺旋桨产生的辐射噪声声谱，其幅度、频率均与航速有关，且带宽比前者要宽，有周期性的变化规律。当螺旋桨未产生空化时，辅机与推进系统所产生的线谱是很强的，当航行速度增加导致空化发生时，其所产生的宽带噪声强度就会增加并掩盖某些线谱成分，此时从辐射噪声声谱中提取特征就会变得更加困难。

## 2.2 辐射噪声识别问题定义

辐射噪声识别问题是时间序列分类问题的一种，其目标是通过提取辐射噪声信号的关键特征，来区分不同类别的信号。辐射噪声识别问题的输入为一段船只行驶时发出的辐射噪声信号，该信号为一维的时间信号；输出为该信号所属船只类别，比如渔船、游轮等。我们使用图 2-1 描述了解决辐射噪声识别问题中的几个关键要素之间的关系。

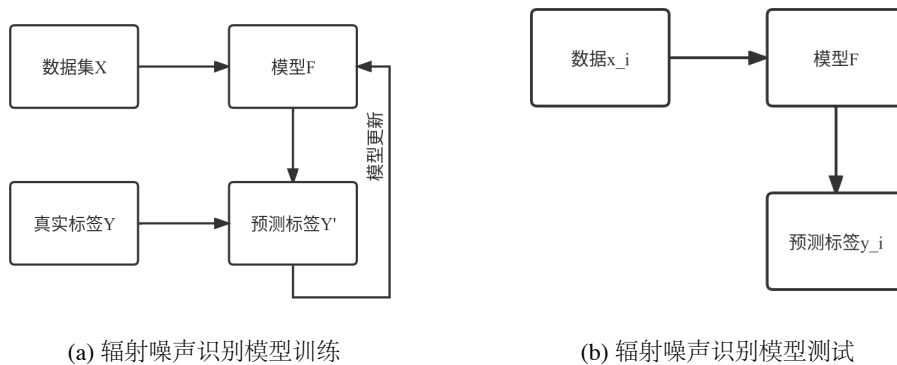


图 2-1: 辐射噪声识别

子图 2-1a 描述了识别模型的训练过程，其中  $X$  代表了带有标签的训练辐射噪声信号数据集， $F$  代表了用于辐射噪声识别的分类模型， $Y$  代表了辐射噪声识别模型输出的标签。在该部分中，通过计算模型输出标签和真实标签的偏差，来对模型  $F$  进行参数更新。子图 2-1b 描述了识别模型的测试过程，其中  $x_i$  代表了待预测的辐射噪声信号， $y_i$  代表了模型输出的预测类别。如何通过训

练辐射噪声数据来训练一个具有优秀的识别模型，是辐射噪声识别问题的研究重点。

## 2.3 基于时频分析的辐射噪声识别算法

在介绍完了辐射噪声的发生机理之后，我们可以看到辐射噪声的时频特性和船只特征之间是息息相关的。通过分析辐射噪声的时频特征，设计对应的手工特征来完成辐射噪声分类仍然是目前主流的识别算法，下面我们将会针对几种经典的时频分析算法展开介绍。

### 2.3.1 短时傅里叶变换方法

离散傅里叶变换（Discrete Fourier Transform, DFT）是频谱分析的基本手段，但是该算法在分析时变信号的时候会出现严重的频谱混淆现象，即两个完全不同的信号，通过离散傅里叶变换会产生相似的频谱图。图 2-2 的例子中对三种完全不同的信号做了频谱分析，可以看到它们具有几乎一致的频谱图。

图 2-2 中，左边一列是原始的信号波形，右边一列是通过 DFT 操作得到的信号频域分布。其中第一行的信号是 200hz、400hz 和 600hz 的正弦波叠加产生的，后两行信号是在不同的时间段改变正弦波的频率生成的，频率成分和第一行信号一样。可以看到这三个完全不同的信号具有相似的频谱特征，因为它们频率成分相同，只是各种频率的信号出现的时机不同。

短时傅里叶变换（Short-Time Fourier Transform, STFT）的提出较好地解决了上述问题。短时傅里叶变换在离散傅里叶变换的基础上引入了加窗的操作，将上述时变信号分为若干小段，将每一小段看作平稳信号，然后在每一小段内进行傅里叶变换，这样就可以得到原始信号的时频特征，其中横轴代表信号的时间跨度特征，纵轴代表信号的频率跨度特征。图 2-3 给出上面三个信号对应的由短时傅里叶变换产生的时频图。

上图左边展示的是原始信号的波形，右边是时域信号做短时傅里叶变换得到的。可以看到时频图除了刻画了信号的频域特征之外，额外引入的时间轴还描述了时变信号的频谱随时间变化的样貌，因此时频图较好地描绘了信号的时频特征，克服了上述傅里叶变换的频谱混淆的问题。

在短时傅里叶变换进行时频分析的基础上，陆等人<sup>[20]</sup>提出了一种基于有效能量过滤的信号特征构建方法，该方法对信号的能量谱进行处理，得出原始

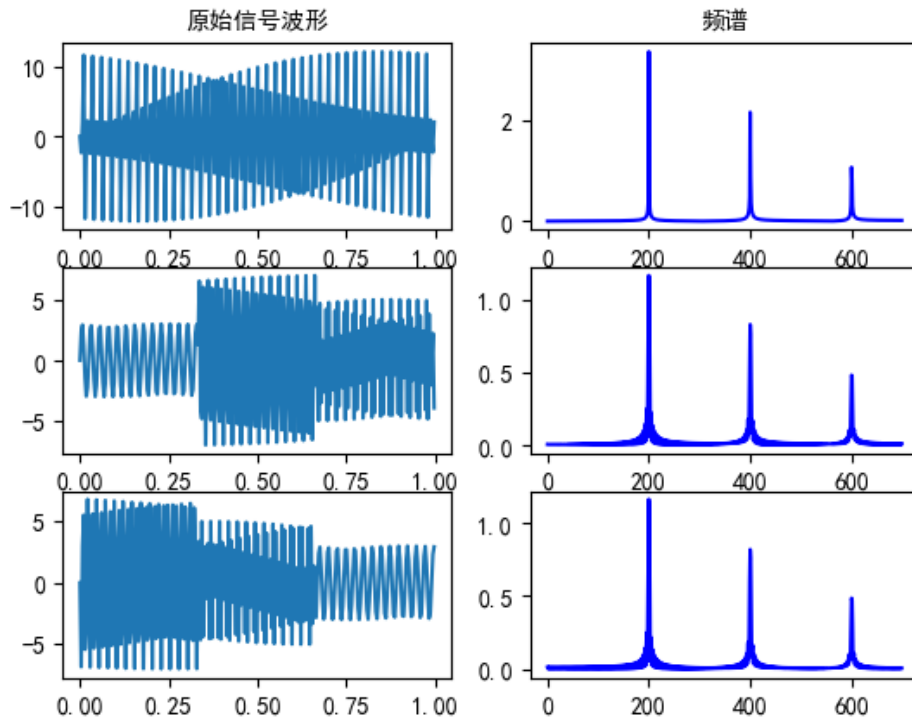


图 2-2: 不同信号傅里叶变换结果

信号中有效信号的频带范围，然后根据频带范围滤除原始时频图中的无效信号成分。

能量谱中能量较高的区间代表了有效信号存在的频带，为了从能量谱中得到有效信号的频率范围，需要根据能量阈值过滤掉那些能量较弱的频带，这个过程需要用到两个滤波器，第一个滤波器根据局部频带的统计特性，对能量谱进行局部平滑，去除噪声的影响，第二个滤波器滤除能量谱中的脉冲干扰。

统计平滑滤波器通过能量谱的局部统计特性来完成滤波，对于能量谱  $E(k)$  中任意位置  $n$ ，该滤波器首先选取一个以该位置为中心，长度为  $2m + 1$  的窗口，计算窗口内能量的均值和方差，如公式 2-1 所示。

$$\begin{aligned}\mu(n) &= \frac{1}{2m+1} \sum_{k=n-m}^{k=n+m} E(k) \\ \sigma^2(n) &= \frac{1}{2m+1} \sum_{k=n-m}^{k=n+m} E^2(k) - \mu^2(k)\end{aligned}\quad (2-1)$$

然后根据计算出的局部均值和局部方差，滤波器将原始能量谱  $E(k)$  转换为

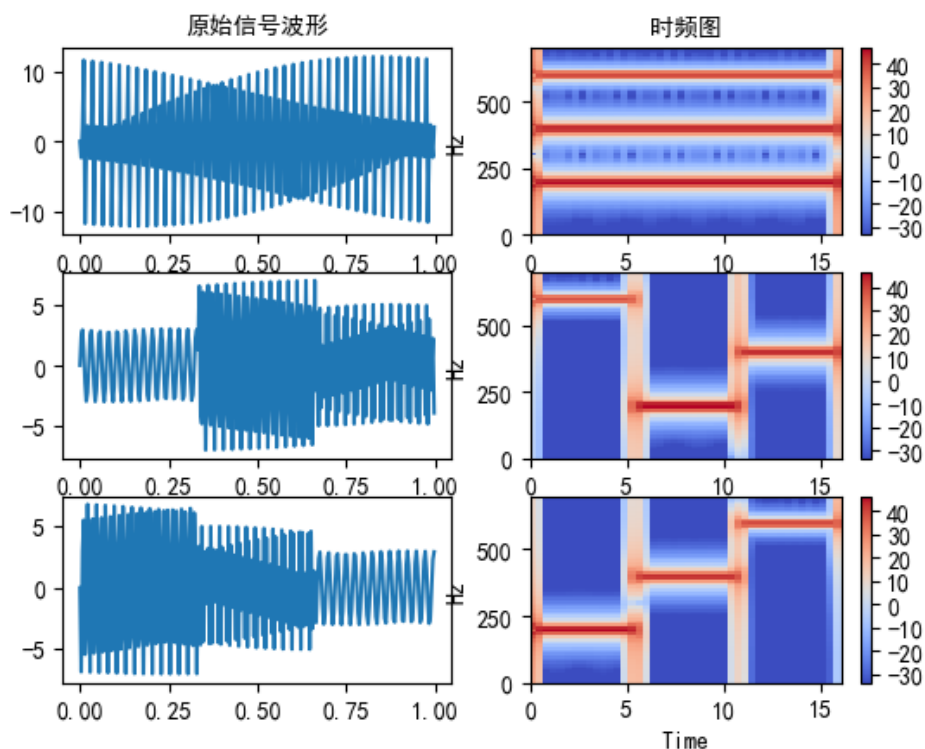


图 2-3: 不同信号的时频图

去除了局部噪声的能量谱向量  $E'(k)$ , 如公式 2-2 所示, 其中  $V^2$  是所有的局部方差的均值, 这里进行操作的原理是, 当局部方差  $\sigma^2(k)$  比较大时, 滤波器进行平滑的程度较强, 反之则较弱。

$$E'(k) = \mu(k) + \max\left\{1, \frac{V^2}{1 + \sigma^2(k)}\right\}(E(k) - \mu(k)) \quad (2-2)$$

统计滤波器有效地减少了噪声, 但是此时的能量谱中仍然存在一些类似尖刺的脉冲干扰, 这些大多是由实际的水声环境干扰产生的, 平滑滤波器并不能有效地滤去这部分干扰。因此在平滑滤波器之后串联一个滑动平均滤波器, 原理如公式 2-3 所示。滑动平均的思路可以较好地处理能量谱中的毛刺, 使得这部分能量被均匀地分配到周围的频带空间中, 从而削减这部分干扰的影响。

$$E_{avg}(n) = \frac{1}{m + p + 1} \sum_{k=-m}^p E'(n + k) \quad (2-3)$$

在经过滤波器完成能量谱的去噪和削减尖刺干扰后, 该方法进一步获得了

有效信号的频率下限和上限，利用这两个临界值可以滤除原始时频图中的无用信号，接着将时频图展开成一维特征，训练分类器模型。基于短时傅里叶变换的方法是辐射噪声信号时频分析的经典方法，但是此类方法的效果依赖窗口值的选取，同时抗噪声能力较差。

### 2.3.2 离散小波变换方法

小波变换 (Wavelet Transform) [21]，顾名思义它的基函数不再是充斥整个时域的正弦信号，而是一段有限长度的信号，这使得将原始信号变换到小波基函数对应的小波域之后，每个点都描述的是一段频率窗口内的信息，而不是傅里叶变换那样的某一个频率点的信息。同时小波基函数不是时域连续的，和正弦函数相比，它需要尺度和位移两个变量来控制它的形态和位置，从而描述整个原始信号在小波域的信息。公式 2-4 给出小波变换的数学表达式。

$$CWT_x^\psi(\tau, s) = \Psi_x^\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \psi^* \left( \frac{t - \tau}{s} \right) dt \quad (2-4)$$

其中变量  $\tau$  和  $s$  分别对应了小波基函数的位移和缩放尺度。上述公式可以理解为，首先通过参数  $s$  固定小波的形状，即此时变换对应的小波值，将小波在原始信号的时域上进行平移以及计算，得到信号在某一小波值下的时间变化；然后连续地改变  $s$  来进行多轮上述计算，得到信号在不同小波值下的小波变换结果，最终达到时频分析的效果。由于不同的小波值对应了不同的频率范围，因此改变上述两个参数，计算信号在不同时刻和小波值下的小波变换，该方法就可以像短时傅里叶变换一样，获得信号的时频特征，同时小波变换和短时傅里叶变换相比还具有下面的优点：

1. 改变了基函数的形式，使得基函数天然具备了刻画信号时频特征的性质，此时小波变换的量纲不再是时频空间，而是小波基空间；
2. 小波基空间描绘了不同窗口尺度和不同信号位置的时频特征，计算更加简单，不需要加窗然后再傅里叶变换；

离散小波变换 (Discrete Wavelet Transform, DWT) 是小波变换的离散算法实现，它采用几组固定的位移值  $\tau$  和缩放尺度  $s$  来计算原始信号的小波变换。Mallet 算法 [21] 是一种经典的离散小波变换算法，它采用了一种半分解的思路，将原始信号逐层分解为低频成分和高频成分，然后对其中的低频部分进行进一步的分解。

为了更加形象地说明离散小波变换所构造特征的结构，我们对公式2-5所示的信号波形进行5层离散小波变换，该信号是频率时变的正弦波信号，其频率  $x$  随着时间的增大而变大。

$$y = \sin(2\pi x^2) \quad (2-5)$$

我们在图2-4中展示了该信号的原始波形，并使用离散小波变换对其进行了5层的小波分解，所采用的小波基为常见的 haar 小波基。

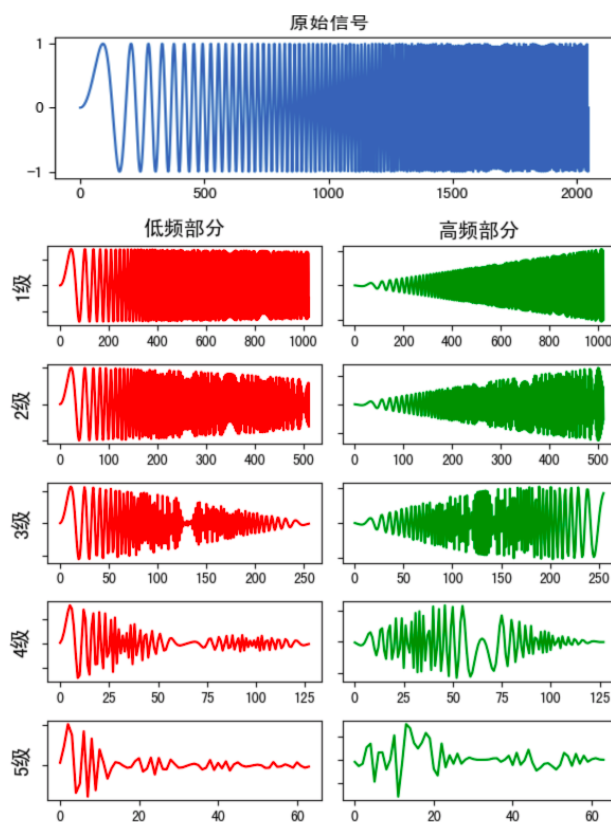


图 2-4: 离散小波变换结果展示

图2-4中，通过离散小波变换，原始信号进行了5层分解，每一层分解都是对上一层的低频部分（或者是原始信号）进行小波分解，从中得到对应的低频部分和高频部分。因为给出的样例信号的频率是随着时间逐步提高的，所以可以看到高频部分的能量也是随着时间逐步提高的，而经过逐步分解之后，低频部分的能量也逐渐收敛在前半段时间范围内。

在通过小波分解获得信号的时频特征之后，Li 等人<sup>[10]</sup>通过对不同尺度的小波分解序列构造分形维数特征 (fractal dimension)，完成对时频图特征的压

缩。在得到基于小波包分形的信号特征后，该方法使用支持向量机来对信号特征进行分类。基于小波变换的方法克服了短时傅里叶变换方法在窗口分辨率选择上的问题，但是这两种方法都没有考虑到辐射噪声自身的特点，不能很好地提取辐射噪声中低频线谱部分的特征。

### 2.3.2.1 梅尔倒谱系数方法

梅尔频率倒谱系数 (Mel Frequency Cepstrum Coefficients, MFCC) 是在 1980 年由 Davis 和 Mermelstein 等人<sup>[22]</sup> 提出来的声学特征，该方法被广泛应用于声信号识别问题中。梅尔频率分析和倒谱分析是计算声音信号 MFCC 特征的两个基本算法，下面我们将会依次介绍这两个核心算法。

人类听觉感知的实验表明，人类听觉的感知只聚焦在频谱中某些特定的区域，梅尔频率分析就是基于该特点而设计的。通过观察不同频率的声音信号在人类听觉系统中的响应情况，人们发现听觉系统会对某些频率的信号产生较强的响应，而对于另外一些频率的信号则只会有微弱的响应。Davis 等人受到人类听觉系统的启发，设计了梅尔滤波器组。梅尔滤波器组是由频率坐标轴上的一系列滤波器组成的，其中大量的滤波器在低频段密集分布，少量的滤波器在高频区域稀疏分布。梅尔滤波器在频域上的分布如图 2-5 所示，其中横轴代表了信号的频率轴，纵轴代表了滤波器组的系数，每个三角形形状的信号代表了梅尔滤波器，它对频率范围内的信号的加强作用呈现三角形的变化趋势，同时随着频率升高，梅尔滤波器变得越来越宽，滤波器的系数也越来越小。

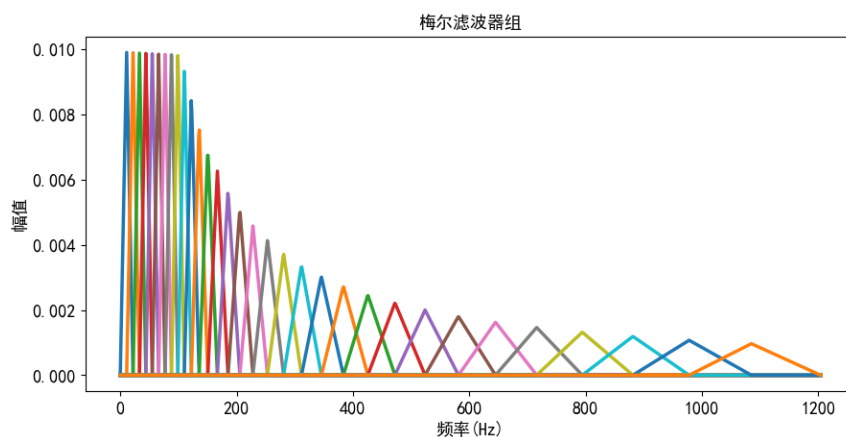


图 2-5: 梅尔滤波器组

倒谱分析的作用是在经过梅尔滤波器处理后的频谱中提取出包络。对于一段语音的频谱图，峰值表示语音的主要频率成分，这些峰值称为共振峰。共振峰是具有辨识度的声音特征，频谱图的包络描述了共振峰的位置以及它们转变的过程。倒谱分析算法使用傅里叶变换方法来计算频谱的包络。对一段序列进行傅里叶变换后，傅里叶变换序列中的低维部分对应了原序列的包络，高维部分对应了原信号的细节，倒谱分析便是利用了这一点，该算法对频谱图进行傅里叶变换，然后从中过滤出低维度的序列，这部分序列便对应了频谱图中包络部分的信息。

综上所述，梅尔倒谱系数的构建流程如图 2-6 所示。首先通过预加重和分帧操作来对输入信号进行梅尔频率处理，得到经过梅尔滤波器处理过的分帧频谱；然后通过对数操作和离散余弦变换来完成倒谱分析操作，提取到分帧频谱的包络信息；最后计算分帧包络信息的一阶差分和二阶差分，得到分帧包络的动态特征，动态特征便是最终的 MFCC 特征。

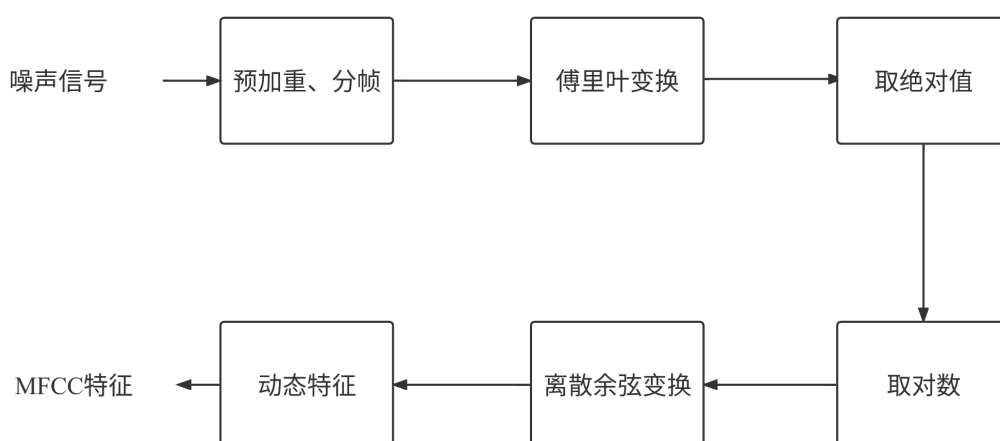


图 2-6: 梅尔倒谱系数流程图

梅尔倒谱系数作为辐射噪声信号的特征，已经被广泛地应用于噪声识别工作中。Zhang 等人<sup>[23]</sup>通过相关距离 Fisher 准则，将 MFCC 特征和对应的高阶差分 MFCC 特征进行融合得到辐射噪声信号的特征，基于该特征训练得到了具有较高识别准确率的分类模型。Tong 等人<sup>[24]</sup>利用 MFCC 特征提取辐射噪声信号的声学特征，然后基于最近邻分类器完成辐射噪声分类任务，该方法在实测数据上取得了优秀的识别结果。尽管 MFCC 特征可以模仿人耳的听觉原理，从辐射噪声中提取具有更好辨识能力的特征，但是此类识别模型的性能非常依赖领域专家对模型参数的设置，模型泛化性能有限。

## 2.4 基于深度学习的辐射噪声识别算法

近年来随着人工智能技术的快速发展，以深度神经网络为代表的深度学习算法已经在图像识别、语音识别、自然语言处理等领域取得了巨大的成果。现代深度神经网络模型受到了生物神经结构的启发<sup>[25]</sup>，设计了具有激活、传递以及更新功能的人工神经元，通过改变人工神经元的工作机制以及连接结构，可以衍生出丰富多样的神经网络模型。目前越来越多的基于深度学习的辐射噪声识别算法被开发出来，相较于以往的时频分析方法，深度学习模型可以提取到更加有表达力的非线性特征，模型的识别性能和泛化性能有了较大程度的提升。

### 2.4.1 卷积神经网络

卷积神经网络模型 (Convolutional Neural Network, CNN)<sup>[26]</sup> 已经在模式识别领域已经取得了巨大的进展。通过多级的卷积运算和池化操作，卷积神经网络可以构建各种复杂的深度特征，从而在人脸识别<sup>[27]</sup>、目标检测<sup>[28]</sup>、时间序列分类<sup>[29]</sup> 等任务中都取得了优秀的识别性能。

卷积单元是卷积神经网络的基本组成部件，每个卷积单元又由卷积层、批归一化层和激活函数组成。通过多层卷积单元的作用，CNN 网络可以提取到一些复杂的深度特征，最后通过池化层对这些特征进行综合，输入 softmax 激活函数得到最终的分类向量。

卷积层是由一些参数可学习的卷积核集合构成的。卷积层的一个重要特性是权值共享，卷积层的每个卷积核在输入序列上滑动时，其参数是不变的。这是因为信号的底层特征和其在信号中出现的位置无关，这种设计可以节约大量的模型参数。局部连接是卷积层的另一个特点，这是因为信号中的某一块区域中，距离较近的序列具有较强的相关性，采用这个结构可以让模型在浅层时学习数据的局部特征，同时随着网络结构的加深，逐渐将这些局部特征联系起来，得到信号的全局特征。

批归一化层<sup>[30]</sup> 是深度神经网络的常见组件，它具有加快训练速度、防止模型过拟合的特性。随着深度神经网络的层数加深，模型训练会变得越来越难收敛，这是由于随着网络每一层对输入数据的非线性变换，导致了在网络上流通的特征的分布发生了偏移，从而导致每次训练后网络中各层的输入变量分布都有变化，从而导致了深度网络的训练困难。批归一化层对其输入向量进行了

标准化操作，将每层的数据都控制在稳定的分布内，缓解了模型训练困难的问题。算法 2.1 给出了批归一化算法的算法原理，其中  $\gamma$  和  $\beta$  是可学习参数，用来控制经过规范化后的数据的尺度变换和偏移。

---

**Algorithm 2.1** 批归一化算法
 

---

输入：批数据  $\mathcal{B}_i = x_1, \dots, x_m$

输出：经过批归一化的批数据  $\mathcal{B}_o = y_1, \dots, y_m$

- 1: 计算批处理数据的均值  $\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$ ;
  - 2: 计算批处理数据的方差  $\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2$ ;
  - 3: 对输入数据进行规范化  $\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$ ;
  - 4: 对规范化数据进行尺度变换和偏移  $y_i \leftarrow \gamma \hat{x}_i + \beta$ ;
- 

池化层在卷积神经网络中的主要作用是压缩数据和参数数量，防止过拟合，同时也可以减少噪声对特征的影响，增强特征的健壮性，因此该层大都是一些简单的综合性操作，比如求平均，求最大值等。目前有较多的辐射噪声识别模型都是基于卷积神经网络结构搭建的。Doan 等人<sup>[31]</sup>提出了一种使用密集卷积神经网络模型 (Dense-CNN) 进行水下目标识别的方法，该网络结构通过残差连接 (skip-connection) 的设计巧妙地重用了网络的浅层特征，在克服了深层网络梯度不稳定的问题的同时，提高了模型的识别准确率。Cao 等人<sup>[32]</sup>提出了一种基于二阶池化操作的 CNN 模型 (CNN-SOP)，该模型的使用二阶池化层 (Second Order Pooling, SOP) 替换传统 CNN 模型中的最大池化层，使用模型可以学习到不同输入信号之间的时间相关性，使得在特征空间中，具有较强时间相关性的数据特征逐渐汇聚到一起，从而改善模型的识别性能。

卷积神经网络采用误差反向传播算法 (Back propagation, BP)<sup>[33]</sup>来进行模型的参数更新。BP 算法包括两个步骤，前向推理和反向传播。

1. 在前向推理时，网络计算输入数据对应的输出，然后通过损失函数计算输入数据和对应标签之间的误差值。

2. 反向传播时，网络首先计算误差值对网络输出的梯度，然后通过求导链式法则，反向计算出上游网络中所有神经元参数的梯度更新值，最后通过梯度下降算法 (Gradient descent, GD) 完成参数更新。

### 2.4.2 循环神经网络

循环神经网络<sup>[34]</sup> (Recurrent Neural Network, RNN) 可以从输入数据中捕捉到序列化的信息, 在 RNN 模型中, 神经元不但可以接受其他神经元的信  
息, 也可以接受自身的信  
息, 形成具有环路的网络结构。和卷积神经网络相比, 循环神经网络可以显式地对噪声数据的波形变换规律等时域特性进行编码。RNN 的基础结构如图 2-7 所示, 可以看到模型的隐藏层输入不但有该时刻的输入数据, 还有前一时刻的隐藏层输出。

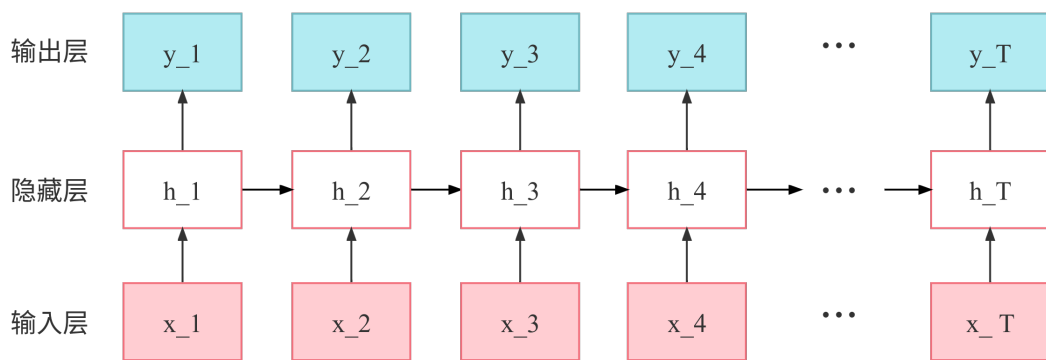


图 2-7: 循环神经网络结构图

上述循环神经网络的计算原理可以通过公式 2-6 进行概括, 其中  $h_t$  是时刻  $t$  输入对应的隐状态,  $x_t$  是模型在时刻  $t$  的外部输入,  $\sigma(\cdot)$  是激活函数,  $W_{xt}$  是外部输入对应的系数,  $W_{ht}$  是上一时刻的隐藏层输出对应的系数。循环神经网络模型可以看做是增加了一层按照时间维度传递的隐藏层, 因此在 RNN 的参数更新路线中需要增加一条沿着时间链路的路线, 来更新对应的参数。

$$h_t = \sigma(W_{xt}x_t + W_{ht}h_{t-1}) \quad (2-6)$$

然而 RNN 模型在处理长度较长的数据时, 由于时间维度的计算链路较长, 导致了在模型参数更新的时候, 多级的矩阵乘法运算会使得时间链路上游参数矩阵的更新值越来越小。因此随着时间维度的网络计算, 模型会逐渐遗忘历史的信息, 这一缺点极大地限制了 RNN 模型捕捉长时间范围时序特性的能力。为了解决上述问题, Hochreiter 等人<sup>[1]</sup> 提出了长短时记忆网络 (Long Short Term Memory, LSTM) 模型。和传统的 RNN 模型相比, LSTM 模型引入了门机制来控制时间维度特征的流通和传递, 图 2-8 给出了 LSTM 单元的结构示意

图，图中从左往右的三个  $\sigma(\cdot)$  算子，分别对应了 LSTM 单元的遗忘门机制、输入门机制和输出门机制，通过这三个门机制，LSTM 可以控制无效信息的遗忘，有效信息的传递，从而解决 RNN 遗忘历史信息以及训练难的问题。

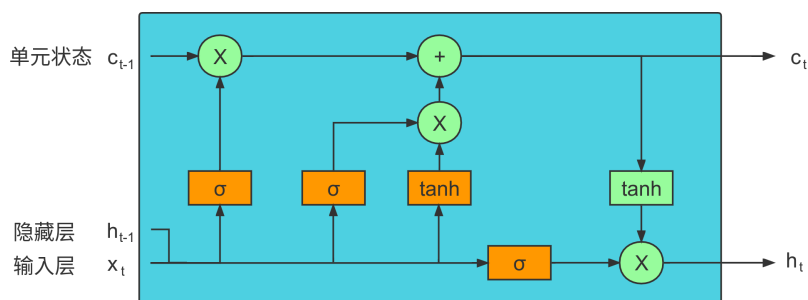


图 2-8: 长短时记忆网络结构图<sup>[1]</sup>

也有不少的学者基于 LSTM 模型开展辐射噪声信号识别研究。Zhang 等人<sup>[35]</sup>提出了一种基于 LSTM 的目标识别方法，该模型将信号的时域、频域和梅尔域数据的融合作为输入，训练一个 LSTM 模型来提取信号的深度特征，最后基于 softmax 分类器完成分类任务。

## 2.5 基于深度学习的辐射噪声增强算法

大数据是深度学习方法的重要支撑，神经网络模型可以在图像、自然语言处理、时间序列分类等领域取得巨大成就，离不开 ImageNet<sup>[26]</sup>、SQuAD<sup>[36]</sup>、UCR 时间序列档案<sup>[37]</sup> 等大规模高质量数据集。但是辐射噪声数据由于采集设备成本高、采集时间周期长、实验设备成本高等原因，目前少有公开的大规模辐射噪声数据集。因此开展辐射噪声数据增强研究，可以弥补辐射噪声数据不足的问题，从而提高深度识别模型的鲁棒性。

### 2.5.1 对抗生成网络方法

对抗生成网络模型 (Generative Adversarial Network, GAN) 由两部分组成，一个用于生成模拟数据的生成器，另外一个用于判别生成数据和真实数据的判别器，二者共同训练，使得生成器产生的模拟数据可以逐渐逼近真实数据，判别器越来越难以区分二者，GAN 网络的通用模型如图 2-9 所示。

GAN 网络的目的是为了训练出可以让判别器无法分辨真假的生成数据，

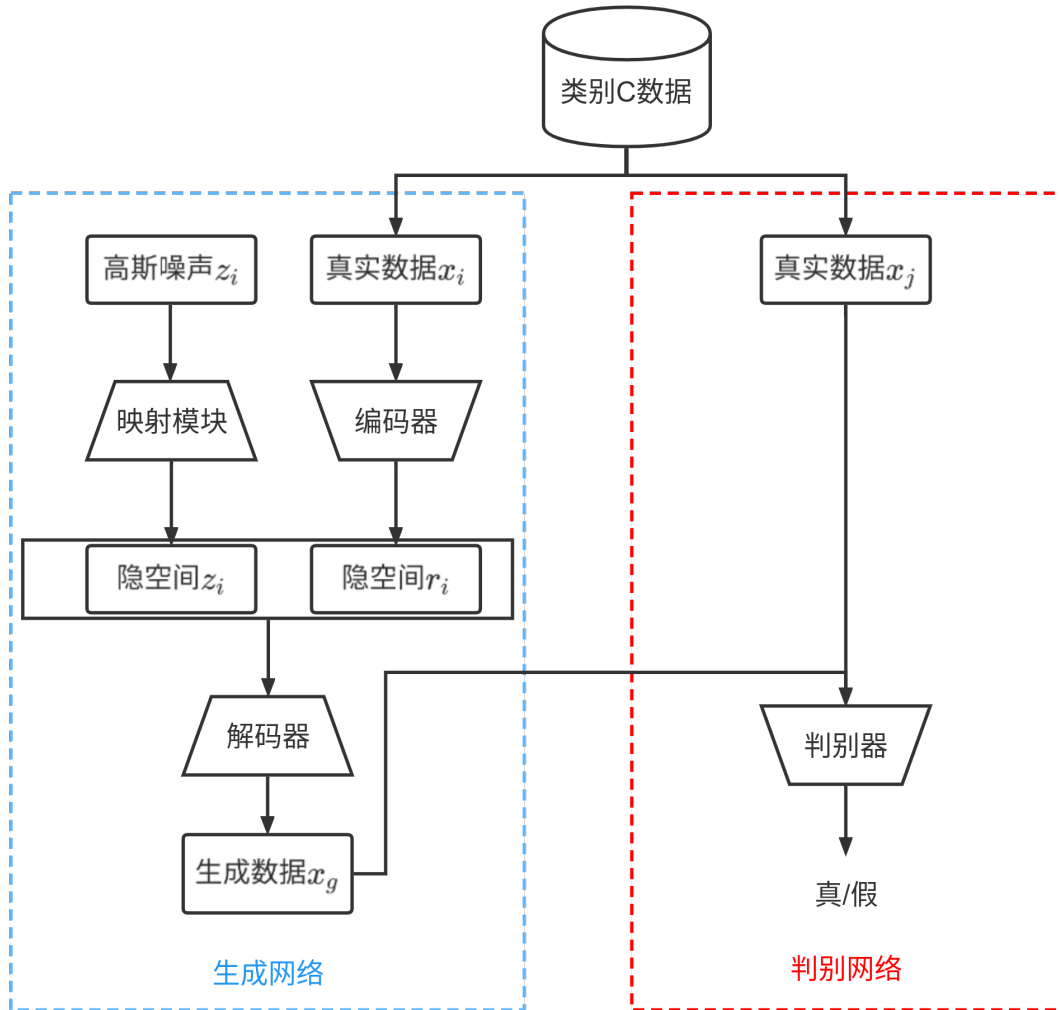


图 2-9: 对抗生成网络模型

其目标函数如公式 2-7 所示。GAN 网络的目标函数由两个期望函数组成，第一项期望函数中的分布  $\mu$  代表真实数据的分布，该项期望的目标是最大化判别器  $D(x)$  在真实数据上的识别性能；另一项期望函数中的分布  $\gamma$  代表随机噪声  $z$  的分布，该项期望的目标是最小化生成器产生的模拟数据  $G(z)$  在判别器  $D(G(z))$  中的识别性能，使得判别器无法正确分辨此时的生成数据。

$$Loss(D, G) := E_{x \sim \mu}[\log D(x)] + E_{z \sim \gamma}[\log(1 - D(G(z)))] \quad (2-7)$$

通过上面的公式 2-7 可以看到，GAN 网络的训练处于一种对抗进行的过程，如果对判别器  $D(x)$  进行过度优化，则会导致生成器  $G(z)$  生成的数据和真实数据  $x$  之间差异过小，起不到数据增强的效果，反之如果对生成器  $G(z)$  过度

优化, 则会导致判别器  $D(x)$  无法正确识别真实数据和生成数据, 从而无法将生成数据的数据增强效果控制在一定范围内, 有可能导致生成数据和原始数据差异过大, 同样起不到数据增强的效果。

Goodfellow 等人<sup>[18]</sup> 为了解决上述优化问题, 使用极大极小值算法来优化 GAN。极大极小值算法<sup>[38]</sup> 是一种适用于围棋等零和博弈问题的优化算法, 该算法通过两步迭代的方式, 使得博弈的双方不断寻找最适合自己的优化策略。具体而言, Goodfellow 等人将 GAN 的判别器和生成器的优化过程拆分开来, 通过迭代式的方法让二者不断朝着自己的优化目标前进:

1. 首先在给定生成器  $G(z)$  参数的前提下, 对判别器模型进行参数优化, 其损失函数如公式 2-8 所示。

$$Loss(D) = \log D(x) + \log(1 - D(G(z))) \quad (2-8)$$

2. 在给定判别器  $D(z)$  参数的前提下, 对生成器模型进行参数优化, 其损失函数如公式 2-9 所示。

$$Loss(G) = \log(1 - D(G(z))) \quad (2-9)$$

通过上述两轮迭代式的优化过程, GAN 网络就可以逐渐学习得到能够混淆判别器的生成数据了, 通过将生成数据加入训练集中, 可以有效地提高原有数据集的多样性, 改善识别模型的性能。在辐射噪声数据增强问题上, Yang 等人<sup>[39]</sup> 提出了一种基于 GAN 的数据增强方法, 该算法基于实验数据生成了大量辐射噪声数据的低频向量谱, 经过实验验证, 使用生成数据扩充训练集, 可以显著提高原有的辐射噪声识别模型的分类准确率。

## 2.5.2 变分编码器方法

自编码器模型 (Auto Encoder, AE)<sup>[40]</sup> 是一类用来学习输入数据的特征映射的无监督学习模型, 其结构如图 2-10 所示, 该类模型的编码器对输入数据进行编码, 得到对应的特征向量; 解码器对特征向量进行解码, 得到重构数据。AE 模型通过最小化重构数据和输入数据的均方误差来进行参数训练。

变分自编码器 (Variational Auto Encoder, VAE)<sup>[41]</sup> 是一类用于数据生成的自编码器模型。它和普通的自编码器模型最大的区别在于, VAE 通过编码器学习输入数据在特征空间上的分布, 然后通过特征空间上进行采样来得到

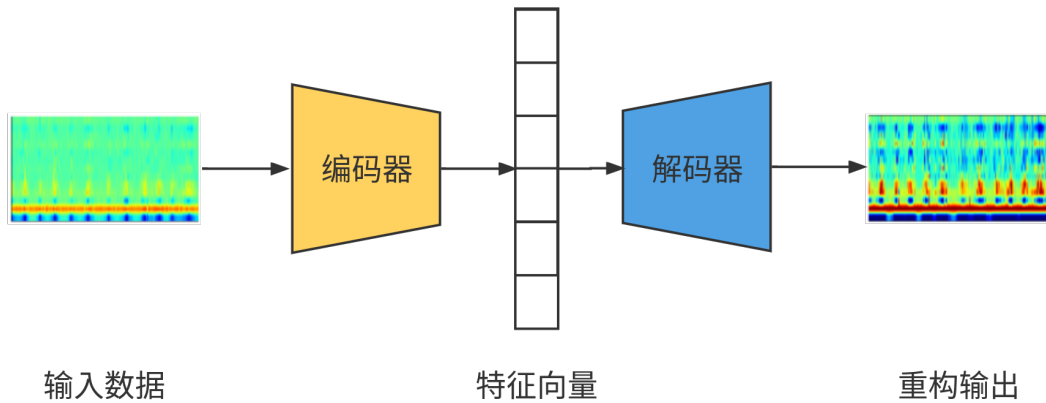


图 2-10: 自编码器模型

新的特征向量，而 AE 则是将输入数据变换到特征空间上一个具体的向量，如图 2-11 所示。VAE 模型通过学习输入数据在特征空间的分布，使得采样得到的特征向量和输入数据的特征映射服从相同的分布，此时解码器生成的重构数据在保留了原始数据特征的同时，还可以引入更多非线性变化，从而增加训练数据的多样性，提高识别模型的性能。

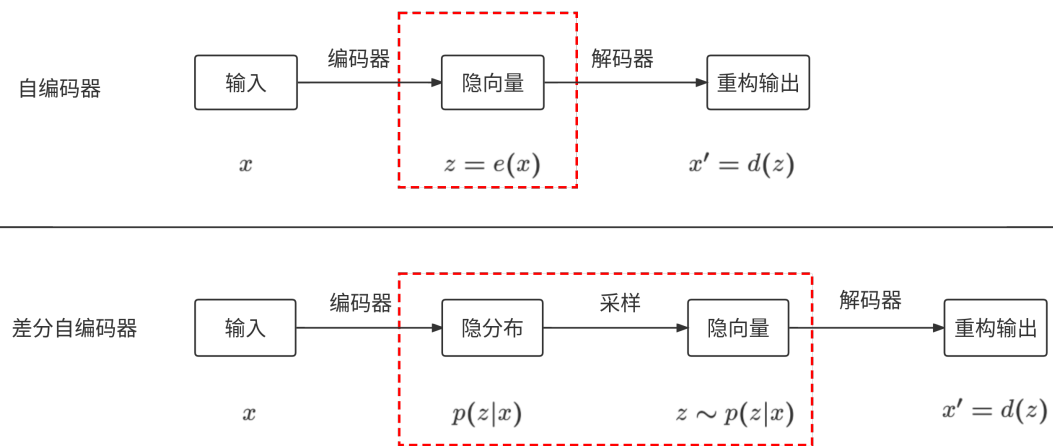


图 2-11: 差分自编码器对比编码器模型

在辐射噪声数据增强任务中，Chen 等人<sup>[42]</sup>提出了一种基于 VAE 的舰船辐射噪声分类方法。该方法首先将真实的舰船辐射噪声信号输入到 VAE 模型中，生成大量的 VAE 仿真信号；然后利用 VAE 生成的仿真数据预训练卷积神经网络分类模型；最后利用真实数据微调预训练的分类模型。实验结果表明，经过预训练的 CNN 模型的分精度优于直接训练的 CNN 模型。

## 2.6 本章小结

本章对辐射噪声信号识别的背景知识和一些经典算法展开介绍。首先我们介绍了辐射噪声信号的组成原理，然后我们介绍了常用的几种基于时频分析的辐射噪声特征构建方法，接着我们介绍了基于深度学习的辐射噪声识别算法，最后我们对用于辐射噪声数据增强的算法模型进行了介绍。

## 第三章 水下辐射噪声信号识别

本章主要研究辐射噪声信号识别问题。我们首先分析了现有方法的不足，然后针对这些问题提出了一种基于帧维度注意力的深度时延网络结构，以往的方法相比，我们设计的模型可以提取更加具有表达力的辐射噪声深度特征，从而获得更加优秀的识别性能。

### 3.1 研究动机

在相关工作部分，我们介绍了几种基于传统手工特征的辐射噪声识别算法，这些基于时频分析的手工特征方法具有较为完备的理论基础和可解释性，配合上支持向量机，最近邻分类器等机器学习算法，便可以获得不错的识别性能。但是海洋环境是复杂多变的，噪声信号的特点随着船只航行速度、天气、海洋气候等因素也是不断变化的，有的时候用某个特征可以较好地进行识别，但是场景变化之后，可能就需要设计新的特征了，这在一定程度上降低了此类方法的实用性。

随着深度学习技术的发展，涌现了大量的基于深度学习的辐射噪声识别方法。但是这些方法在模型设计上大都参考了一些通用的网络结构，少有从辐射噪声自身特性出发，设计对应的模型结构。因此基于辐射噪声的信号特点，设计一种基于深度学习的自动化特征提取模型，是辐射噪声识别领域急需解决的问题。

### 3.2 基于帧维度注意力的深度时延网络

针对现有方法的不足，我们提出了一种带帧维度注意力机制的深度时延网络模型（Time Delay Neural Network with frame level attention, TDNN-attn）。在该模型中，我们设计了一种可以较好地捕捉辐射噪声时频特性的时延卷积模块，该模块使用时延卷积操作来提取具有时频特性的深度特征，同时我们利用了空洞卷积和残差连接的结构设计来降低模型训练的难度。为了解决真实辐射

噪声数据有效信息稀疏的问题，我们还提出了一种帧维度注意力模块，该模块通过学习不同帧维度特征之间的差异，输出对应的注意力权重，从而提高包含丰富噪声信息的帧特征对分类结果的影响，削弱包含无效信息的帧特征对分类结果的影响。经过实验证明，我们提出的模型相较几种经典的辐射噪声识别算法，具有最优的识别性能。

本节我们将会对 TDNN-attn 模型的设计原理进行介绍。我们首先介绍模型的整体结构，然后对时延卷积层，帧维度注意力层以及统计池化层这三个主要模块展开详细介绍，最后我们介绍模型的参数训练过程。

### 3.2.1 模型结构

TDNN-attn 模型的整体结构如图 3-1 所示。

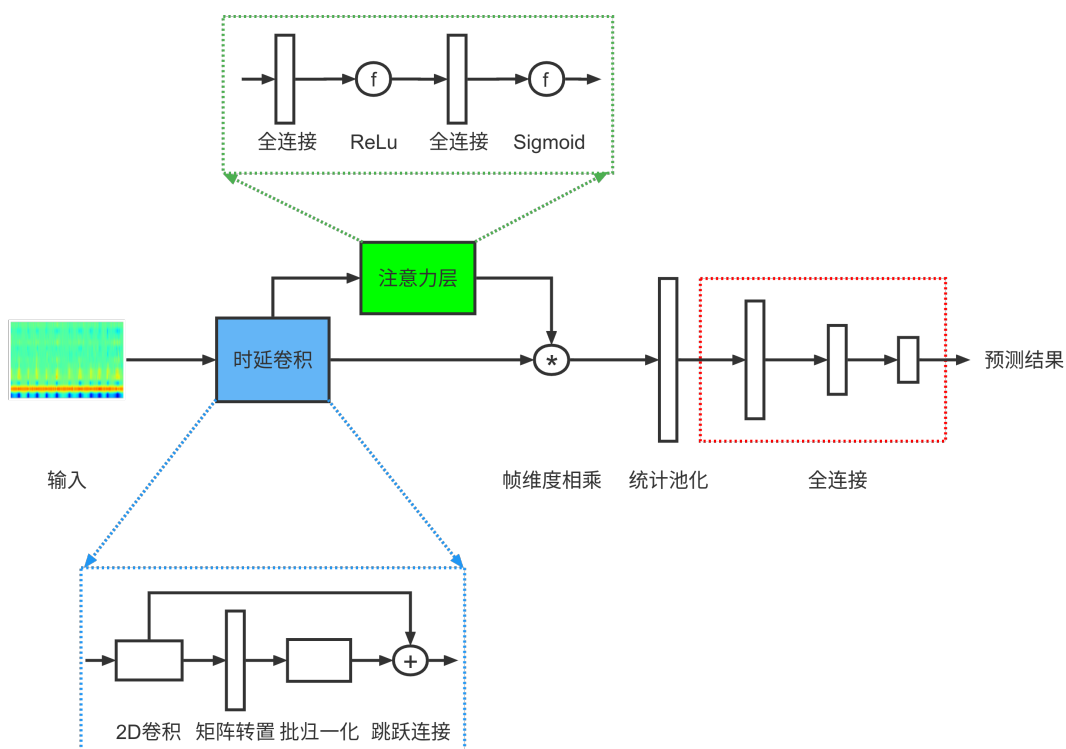


图 3-1: 辐射噪声识别模型结构图

上述模型的主要模块如下：

1. 模型输入。该模型的输入为辐射噪声信号的 MFCC 特征，因为船只的很多特性都隐藏在辐射噪声的低频声谱中，而 MFCC 特征通过梅尔滤波器的作用，可以加强辐射噪声中的低频窄带声谱，削弱高频连续谱，因此该特征比辐

射噪声的时域波形更适合作为识别模型的输入。我们在实验部分也会对模型输入对识别效果的影响进行分析。

2. 时延卷积模块。时域卷积模块的主要作用是提取辐射噪声的帧维度深度特征。和传统的卷积结构不同，我们的采用了更大尺寸的卷积核，来捕捉更大感受野的辐射噪声信息。同时我们在每次卷积之后添加了一个维度转置操作，将卷积输出的特征图通道维度和宽维度进行交换。此时特征图仍然保持了时频维度，方便我们根据数据的时频特性添加优化模块。同时为了防止由于大卷积核导致的优化问题，我们引入了空洞卷积和残差连接的优化设计。经过实验证明，这两个设计较好地缓解了模型优化困难的问题，提高了最终的识别性能。

3. 帧维度注意力模块。通过对辐射噪声的信号特点分析，我们发现由于采集时间长、采集设备位置固定等原因，一段辐射噪声信号中往往会包含一些无效的声音片段。为了削弱这部分信号对最终分类结果的影响，我们设计了一个帧维度注意力模块。该模块通过学习不同帧维度特征之间的差异，输出对应的注意力权重，从而提高了包含丰富辐射噪声信息的帧特征对分类结果的影响，削弱了包含无效信息的帧特征对结果的影响。经过实验证明，帧维度注意力模块可以较好地改善模型的识别性能。

4. 统计池化层。该模块对经过注意力模块处理过的时延卷积特征进行综合，从帧维度的特征中得到辐射噪声整体的深度特征。考虑到辐射噪声信号的复杂时变特性，我们采用了声纹识别算法中常用的统计池化层<sup>[43]</sup>。统计池化层在平均池化的基础上，还引入了方差计算来捕捉特征偏离平均值的程度。经过实验对比，使用统计池化层可以提高模型的识别准确率。

5. 全连接模块。该部分将经过池化层处理得到的深度特征经过几个连续的全连接层，进一步提取辐射噪声的非线性特征，然后经过一个 `softmax` 分类器得到不同类别的预测概率，预测概率最高的类别即为模型的预测类别。

### 3.2.2 时延卷积模块

众所周知，用于图像识别领域的卷积神经网络模型，都会采用大量的  $3 \times 3$  尺寸的卷积核，因为这个尺寸的卷积核可以较好地捕获图片中的细节纹理。但是对于船舶辐射噪声信号而言， $3 \times 3$  尺寸的卷积核却不太合适，因为辐射噪声信号的采样频率都集中在  $8\text{kHz}$  到  $16\text{kHz}$  之间，此时为了获得有效的感受野大小，需要堆叠相当深度的卷积层才能实现。这一方面为模型设计增强了难度，同时也加大了模型训练的难度。

一些研究学者选择将卷积神经网络模型压扁，即采用浅层、大卷积核尺寸的模型设计来解决辐射噪声识别问题。比如 Hu 等人<sup>[16]</sup> 使用了  $204 \times 1$ 、 $14 \times 1$  和  $12 \times 1$  大小的卷积核来提取信号特征；Shen 等人<sup>[14]</sup> 也采用了大小为  $100 \times 1$ 、 $200 \times 1$ 、 $400 \times 1$  和  $800 \times 1$  的卷积核来构建卷积神经网络模型。通过使用大尺寸的卷积核，可以扩大模型的感受野，从而更加容易捕捉到信号中波形的细节信息。但是大尺寸的卷积层产生了大量的训练参数，加大了模型训练的难度。

同时也有另外一部分学者选择使用和图像识别模型相同的小卷积核结构，通过加深模型来提高模型的感受野。比如 Tian 等人<sup>[15]</sup> 将不同尺寸的小卷积核提取到的卷积特征在通道维度融合，利用多尺度的卷积特征来提高深度特征的丰富度。

我们在卷积尺寸上沿用了大卷积核的设计理念。时延卷积的结构借鉴了 Waibel 等人<sup>[44]</sup> 设计的时延神经网络结构，该网络最早提出时是用于识别人声中不同的音素。和 Waibel 等人设计的时延模块不同，我们的时延卷积本质上属于二维卷积。图 3-2 描述了时延卷积的基本工作原理，其中符号  $W$  代表了输入向量的宽度， $H$  代表了输入向量的高度，通道数目为 1；符号  $L$  代表了卷积核的宽， $H'$  代表了卷积核的通道数， $W'$  代表了输出向量的宽。可以看到时延卷积的本质仍然是卷积核，但是卷积核的宽度是固定的，和输入向量的宽维度相同。

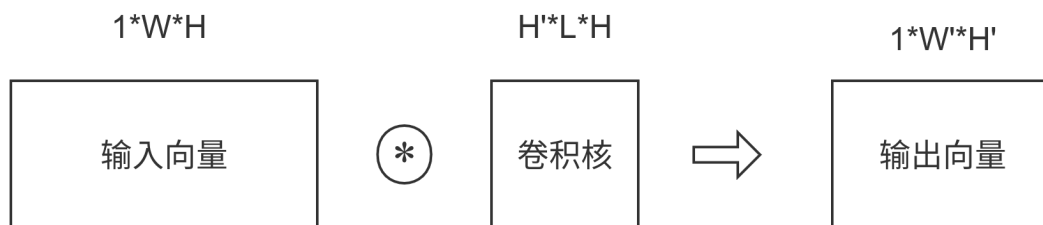


图 3-2: 时延卷积数学原理

时延卷积和普通卷积核的计算方式几乎一致，但是在输出格式上有所不同。我们在卷积运算的下游添加了一个矩阵转置运算，交换了卷积输出特征图的通道维度和高维度，使得输出特征图的通道数目仍然为 1，卷积核的数目成为输出特征图的高维度。

通过这样的设计，时延卷积得到了类似时频图的深度特征。作为时延卷积的第一级输入，MFCC 特征的宽方向描述了信号时间轴上的变化，高方向描述了信号频率轴上的变化。经过时延卷积运算后，宽方向的特征进行了局部综

合，而高方向的特征则进行了类似矩阵乘法的变换过程，因此输出特征图仍然保有原始信号的时频特性。这种计算逻辑也对应了时延卷积的命名，即模型通过延迟相加的方法，将信号的时域上的特征进行局部的综合。这样设计使得时延卷积提取到的特征的宽方向是和信号的时域上的变化对应的，高方向是和信号的频域上的变化对应的，这也方便了基于信号的特点来设计后续模块。

但是时延卷积由于卷积核尺寸过大，使得模型优化变得困难起来。为了解决这个问题，我们在时延卷积结构中加入了残差连接，来缓解了由大卷积核带来的参数优化困难；同时还引入了空洞卷积的设计，在保证卷积核感受野大小的同时，降低模型参数。

残差连接的结构在大多数神经网络模型中都有使用，它是由 He 等人<sup>[45]</sup>提出的深度残差神经网络中的代表性设计，其原理图如图 3-3 所示。残差连接模块的主要作用是缓解反向传播过程中梯度消散的问题。该结构的核心思想是在原有时延卷积基础上添加一条支路，将时延卷积的输入直接和模型的主干路提取到的特征进行相加。

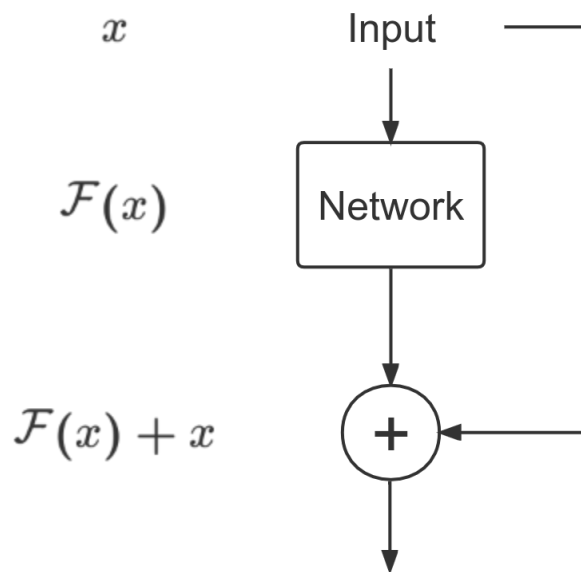


图 3-3: 残差连接简易结构

在神经网络模型训练的过程中，会通过反向传播的方法来计算需要更新的梯度。随着网络层数的加深，由网络下游的输出层产出的损失函数在对网络上游的参数计算更新梯度时，会通过链式法则来将逐层的输入输出引入进来，可想而知，如果链式法则每一项的偏导值都很小，那么最终导出的更新值同

样会变得很小，这样多次循环之后，参数的更新就会停滞，导致模型无法继续训练。残差连接的好处在于，通过将模块的输入和输出建联起来，这样链式法则的中间项就会多出一个常数项，从而缓解多层反向传播导致的梯度不稳定问题。

空洞卷积<sup>[2]</sup>是图像语义分割和目标检测中常用的一种卷积核设计手段，它在标准的卷积核中引入了空洞间隔，使得卷积运算涉及到的输入范围更大，如图3-4所示。

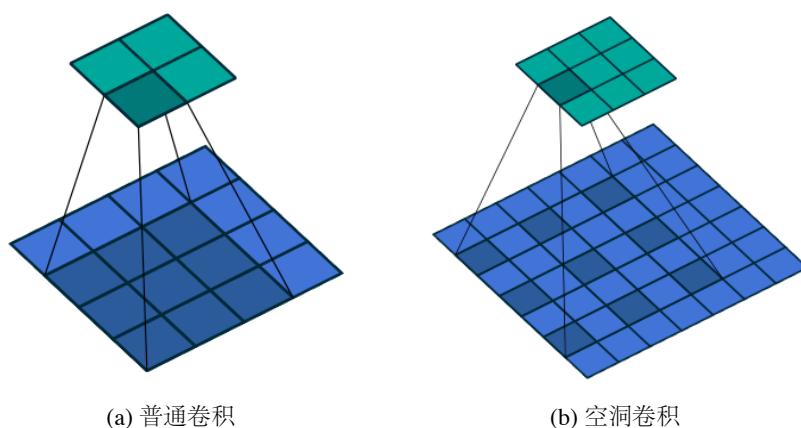


图 3-4: 普通卷积 vs. 空洞卷积<sup>[2]</sup>

空洞卷积的引入，使得相同感受野下所需卷积核参数变少，降低模型训练难度，同时通过在多层时延卷积中使用不同核间隔（dilation）的卷积核，可以获取多尺度的上下文信息，有利于我们捕捉信号中的短时特征和包络特征。

### 3.2.3 帧维度注意力模块

通过对实验数据进行分析，我们发现一段辐射噪声数据中，船只发出的噪声强度是不固定的。这是因为船只不是以固定的速度进行航行的，有的时候船只会全速航行，此时发出的辐射噪声强度较大，能听到清晰的螺旋桨等噪声；有的时候船只会进行减速，此时只能听到断断续续的机械制动装置的声音；有的时间段内船只未运动，此时只能听到类似海浪翻涌的海洋噪声。因此加强辐射噪声中的有效信号片段，削弱其中的海洋噪声片段，将提高模型的分类能力。

在传统的时频分析方法中，可以通过去噪算法来解决上述问题。去噪算法通过滤除辐射噪声中的海洋噪声，可以提高整体辐射噪声信号的信噪比，从而

提升有效信号对分类结果的影响。小波变换阈值去噪算法是一种典型的去噪算法，该算法流程图如图 3-5 所示。

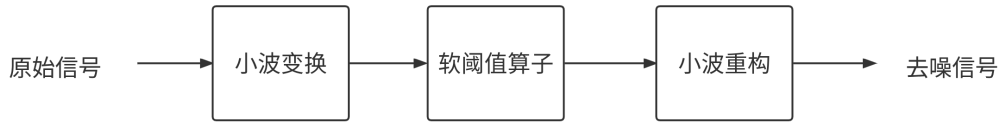


图 3-5: 小波变换阈值去噪算法

通常情况下，实际采集到的信号都是由纯净的信号和噪声构成的，而自然界的噪声中又是以高斯白噪声为主，这些高斯白噪声大都是以随机分布的状态充斥在信号中。当将信号转换到小波域之后，这些白噪声对应的小波系数一般都是较小的，因此如果在小波域通过过滤掉这一小部分系数，就可以令纯净信号的占比进一步增加，从而在重构之后得到信噪比更高的信号。因此可以通过给小波系数施加一个区间滤波的操作来完成小波变换去噪，即当小波系数低于某一阈值时，便将其置为 0，就可以滤除原始数据中噪声的部分。

经过上面的介绍，我们可以看到传统的去噪算法大都假设背景噪声是高斯白噪声，同时这类算法大都使用固定的阈值完成去噪，不能根据不同的信号数据动态调整滤波阈值或者权重。注意力机制结构是神经网络模型中的一类常见结构，它经常被用来建模模型中输入特征对输出特征的贡献度，使得模型更加关注某一部分的特征，从而提高该部分特征对最终识别结果的影响。注意力机制结构主要可以分为通道维度注意力<sup>[46]</sup>，空间维度注意力<sup>[47]</sup>，混合注意力<sup>[48]</sup>等结构。为了解决传统去噪算法存在的问题，我们提出了一种帧维度注意力机制，来加强辐射信号中有效信号对分类结果的影响。该结构和一般的空间维度注意力结构相比，只保留了时间维度的注意力计算。由于辐射噪声信号在时间维度上存在明显的信息分布不均匀问题，因此这样的结构设计既符合我们的数据特点，同时也可以节约大量的计算资源。该部分的网络结构如图 3-6 所示。

图 3-6 中左侧虚线框内的部分就是统计池化层的部分，它计算了时延卷积模块提取到的特征在时间维度上的平均值以及方差，并将这两个统计特征拼接起来，作为信号的整体特征。

图 3-6 中右侧虚线框内的部分我们设计了一个小的子网络结构，该模块的输入是经过时延卷积模块之后提取到的深度特征，输出是不同帧特征的影响因子。该模块首先将输入经过一层全连接层和 ReLu 激活函数，全连接层中的参数经过多次数据的训练，将会具有自动计算输入特征重要程度的能力。全

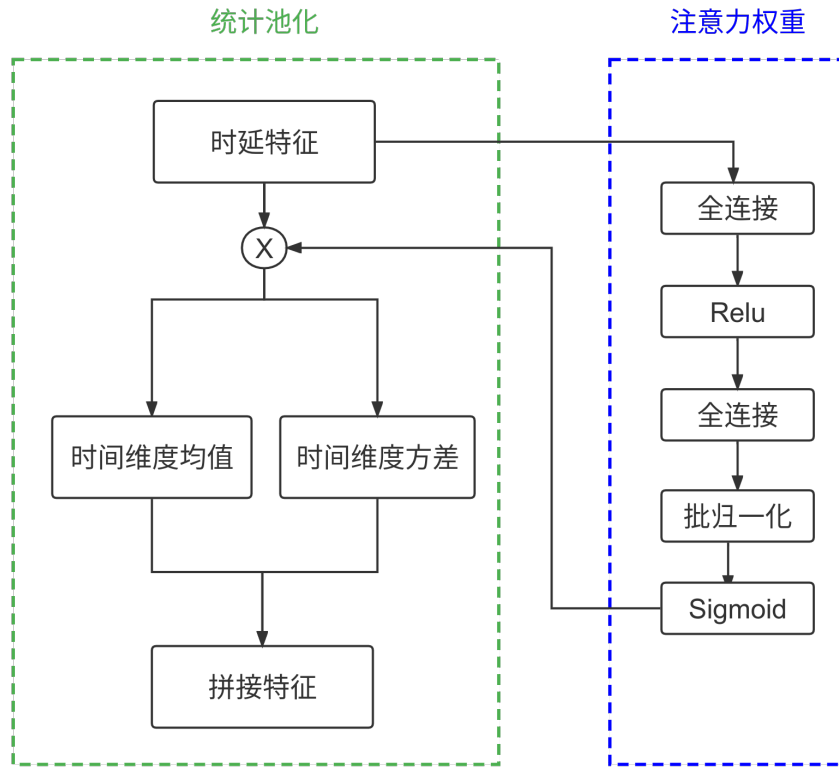


图 3-6: 帧维度注意力结构图

连接层的输出紧接着经过批归一化层，来消除批训练数据引入的额外的偏差和方差，在经过第二个全连接层和 **sigmoid** 激活函数之后，模块输出和输入特征向量宽维度相同的权重向量  $\alpha$ ，它用来描述输入特征向量在时间维度上的重要性。

因为在模型训练的初期，权重向量  $\alpha$  几乎是随机输出的，如果直接将其作为不同帧特征的影响因子和特征向量相乘，将会影响模型的训练效果。因此我们通过分布在  $(0, 1)$  区间内的超参数  $\beta$  来控制  $\alpha$  的影响程度，如公式 3-1 所示，通过  $\beta$  我们控制了注意力模块输出对时延卷积特征图的影响，使得在模型训练的初期，随机输出的注意力模块不会过度影响模型更新的走向。

$$\alpha' = 1 + \beta * \alpha \quad (3-1)$$

最后我们将经过调整的  $\alpha'$  值和时延卷积特征图在时间维度上进行乘积，便得到了经过注意力机制加强的帧维度特征。

### 3.2.4 统计池化模块

池化模块是卷积神经网络中的一个重要结构，它通过对卷积模块提取到的特征图进行下采样操作，得到具有浓缩分辨率的新特征图。通过使用池化模块，可以减少模型中参数的数量，从而降低计算成本；另一方面，池化层也可以从卷积模块输出的特征图中进一步提取有用的信息，丢弃其中无用的细节，防止模型过拟合。

在相关工作部分，我们介绍了卷积神经网络中常用的最大池化模型。最大池化的核心思想是从特征图的每个区域中提取具有最强激活值的部分，去除掉其余特征值对后续分类模型的影响。因此最大池化操作可以看做是一个阶跃滤波器，它将池化区域内低于最大值的所有特征值置为 0，如图 3-7 所示。

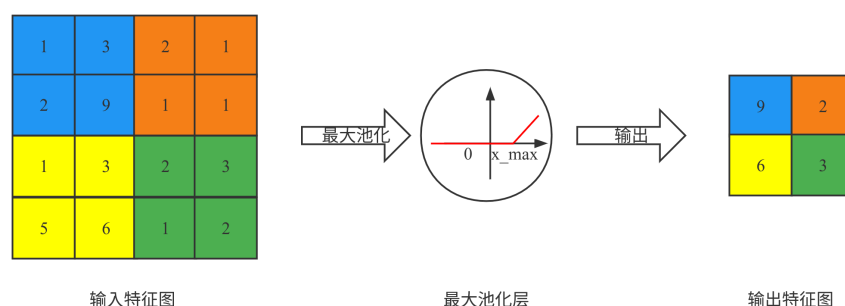


图 3-7: 最大池化原理

最大池化层结构在图像识别模型中已经被广泛使用，但是对于辐射噪声识别任务来说，最大池化设计存在两个问题：

1. 池化层尺寸难确定。因为和图像数据相比，辐射噪声数据规模较大，如果采用传统 CNN 结构中的池化层结构，将难以发挥最大池化提取有效信息的作用，而如果使用大尺寸的池化层，则容易丢失掉一些有用信息。同时由于辐射噪声数据的采样率各异，因此统一尺寸的池化结构难满足识别不同的辐射噪声信号的需求。

2. 最大池化操作容易丢失信号的统计特征。和图像数据相比，辐射噪声的一些统计特征，比如均值、方差、高阶统计量往往蕴含着较多的信号特性，采用最大池化容易丢失掉这部分信息。

因此为了解决上述两种问题，我们选用了说话人识别任务中常用的统计池化模型来作为 TDNN-attn 模型的池化层。统计池化模型对时延卷积提取到的特征在时间维度上进行综合，得到特征图在时间维度上的均值和方差，将这两个

向量进行拼接，来作为下游全连接模型的输入。和图像任务中常用的最大池化、平均池化等结构不同，统计池化是一种单维度操作的池化层，它只对输入特征图的时域维度进行统计池化操作。这种设计使得对于不同长度、不同采样率的输入数据，统计池化层都可以输出统一长度的特征向量，提高了模型的泛用性；同时在统计池化层中，我们可以根据信号特性加入不同的统计量计算，方便了模型的拓展。统计池化层原理如图 3-8 所示，其中输出特征图中的黑色箭头代表向量拼接操作。

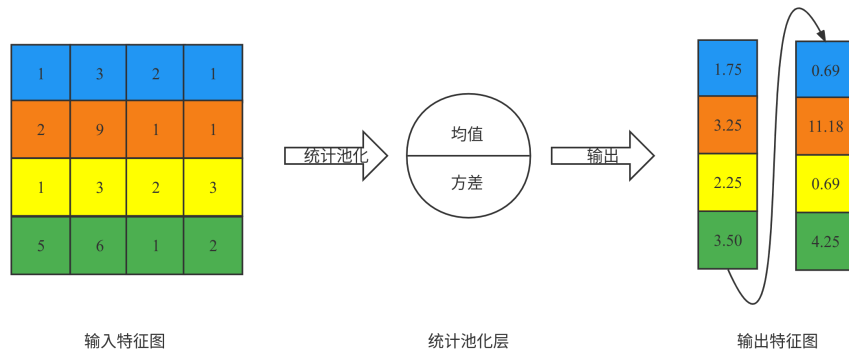


图 3-8: 统计池化原理

### 3.2.5 模型训练

我们使用损失函数 (Loss Function, loss) 来度量模型预测结果和真实标签的偏差，从而计算每轮模型更新的参数梯度，完成模型参数更新。在 TDNN-attn 模型的优化过程中，我们使用了交叉熵损失函数，其数学公式如 3-2 所示，其中符号  $N_b$  代表模型每轮参数更新的批次大小； $N_l$  代表分类任务中类别的数目； $I(\cdot)$  代表指示函数，该函数的输入值为布尔值，当输入值为真时，指示函数输出为 1，反之为 0； $p_{ic}$  代表输出预测向量中，模型预测样本  $i$  属于类别  $c$  的概率。

$$loss = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{N_l} I(y_i == c) \log(p_{ic}) \quad (3-2)$$

在模型训练过程中，我们使用大小为 8 的批次尺寸进行小批量梯度更新，更新轮数为 50 轮；使用 Adam 优化器<sup>[49]</sup>进行参数更新，初始学习率设置为 0.0003，使用阶梯式学习率调整策略进行学习率更新，其原理如公式 3-3 所示，其中  $lr$  代表学习率， $\gamma$  代表学习率衰减因子， $e$  代表当前训练轮次， $e_1, e_2, \dots, e_n$  代表预设定的阶梯列表值，函数  $C(\cdot)$  代表当前训练轮次  $e$  超过阶梯列表值中元素的个数，实验中我们将  $\gamma$  值设置为 0.1。

$$lr = lr * \gamma^{C(e, e_1, e_2, \dots, e_n)} \quad (3-3)$$

模型参数细节如表格 3-1 所示，其中  $n_m = 13$  代表 MFCC 特征的阶数； $n_1$  代表数据集中船只类别的数目；时延卷积层 TDC 的参数  $(k, d, c)$  中， $k$  代表卷积核的宽度， $d$  代表空洞卷积中卷积核元素的间隔， $c$  代表卷积核的个数；FC 代表全连接层，其参数  $(i, o)$  中， $i$  代表全连接输入维度， $o$  代表全连接输出维度。

表 3-1: TDNN-attn 模型参数细节

层名称	输入维度	参数	输出维度
TDC1	(1,2498, $n_m$ )	(128,5, $n_m$ )	(1,2498,128)
TDC2	(1,2498,128)	(128, 3, 128)	(1,2498,128)
TDC3	(1,2498,128)	(128,1,128)	(1,2498,128)
ATTN-FC1	(1,2498,128)	(128,128)	(1,2498,128)
ATTN-FC2	(1,2498,128)	(128,1)	(1,2498,1)
STAT-POOLING	(1,2498,128)	-	(1,256)
FC1	(1,256)	(256,128)	(1,128)
FC2	(1,128)	(128, $n_1$ )	(1, $n_1$ )

## 3.3 实验结果和分析

### 3.3.1 实验环境设置

本文中的实验均在课题组的 GPU 服务器上进行，使用 Python 语言进行代码编写，其中深度学习部分代码使用了 Pytorch 框架<sup>[50]</sup>，辐射噪声处理使用了 librosa 框架<sup>[51]</sup>。具体的服务器配置如下表 3-2 所示。

表 3-2: 实验服务器配置

配置	参数
CPU	Inter Xeon E5-2678
内存	94Gb
GPU	GeForce RTX 2080TI
操作系统	Ubuntu 16.04 LTS

### 3.3.2 实验数据分析

本次实验中我们选取了两个实测数据集:

1. 数据集 1 包含了四种船只在进出港口时发出的辐射噪声信号; 每类信号的数据较少, 且数目分布不均匀。由于不同的噪声数据的大小和格式不是统一, 并且采集到的数据规模较小, 我们在进行实验前首先基于窗口切片法做了实验数据的格式统一以及数据增强, 该方法首先对原始时间序列按固定长度切片, 然后将切片后的数据来作为实验数据, 进行训练和预测。表 3-3 给出了数据集 1 经过窗口切片后的各类别数据分布, 实验中我们按照 2: 1 的比例划分训练集和测试集。

表 3-3: 数据集 1 辐射噪声分布

信号类别	原始数据	切片后	训练集	测试集
货船	10	220	140	80
拖轮	4	88	56	32
渔船	11	242	154	88
游艇	6	127	81	46

2. 数据集 2 包含了两种船只的辐射噪声信号; 每类数据的数据量较多, 且分布比数据集 1 更为均匀; 由于环境因素, 所采集的实验数据质量较差, 包含了较多的背景噪声。我们同样基于窗口切片法对实验数据进行了处理, 并按照 2: 1 的比例划分了训练集和测试集, 实验数据分布如表格 3-4 所示。

表 3-4: 数据集 2 辐射噪声分布

信号类别	原始数据	切片后	训练集	测试集
小渔船	102	2124	1349	775
大货船	125	2676	1700	976

### 3.3.3 实验评估指标

实验中我们使用如下两个关键指标来评估模型的识别性能:

1. 识别准确率 (Accuracy)。准确率描述的是测试集中, 预测正确的样本数目占有所有测试集样本的比例, 如公式 3-4 所示, 其中  $N_{correct}$  代表被正确分类的测试集样本数目,  $N_{total}$  代表所有的测试集样本数目。该指标是分类模型中最常用的评估指标之一。

$$accuracy = \frac{N_{correct}}{N_{total}} \quad (3-4)$$

2. ROC 曲线面积 (Area Under Curve, AUC)。因为 AUC 指标是基于两分类任务提出的, 本文中通过一对一 (one vs one) 的方法来计算所有的类别对之间的 AUC 值, 然后进行平均得到对应的 AUC 值。AUC 值是评估分类模型的重要指标, 如果两个分类模型的准确率相同, AUC 值不同, AUC 值大的模型在预测的过程中, 会倾向于产生更高的正确类别预测概率。在去识别一些包含噪声的数据时, AUC 值高的模型往往会有更好的识别效果。对于二分类任务来说, AUC 值的计算公式如 3-5 所示。其中  $i$  代表测试集中第  $i$  个样本;  $C_{positive}$  代表正类, 即二分类中标签值为 1 的类别;  $rank_i$  代表按照正类预测概率进行从小到大排序, 样本  $i$  所处的队列位置;  $M$  代表正类的数目;  $N$  代表负类的数目。

$$auc = \frac{\sum_{i \in C_{positive}} rank_i - \frac{M*(M+1)}{2}}{M * N} \quad (3-5)$$

### 3.3.4 对比实验结果

表格 3-5 给出了不同的模型在两个辐射噪声信号数据集上的识别效果, 除了我们设计的模型之外, 我们还使用了一些常用的时频分析算法以及深度学习算法来作为对比。为了保证对比实验的公平性, 我们采用了相同的模型输入以

及学习策略，来训练这些识别模型。

1. MFCC+SVM。MFCC 特征作为输入，支持向量机 (Support Vector Machine, SVM) 作为分类器的分类模型。因为 SVM 在高维特征空间难以寻找分类超平面，实际输入分类器的是 MFCC 特征沿着时域的平均值。

2. MFCC+KNN。MFCC 特征作为输入，最近邻分类器 (K-Nearest Neighbors, KNN) 作为分类器的分类模型。因为 KNN 模型在高维特征空间效果较差，实际输入分类器的是 MFCC 特征沿着时域的平均值。

3. MLP。多层感知器 (Multilayer Perceptron, MLP) 是一类经典的全连接神经网络模型。这里我们使用了 Wang 等人<sup>[29]</sup>的论文中设计的 MLP 结构，该结构的隐藏层由三个全连接层组成，每个全连接层神经元个数均为 512。

4. FCN。全卷积神经网络 (Fully Convolutional Networks, FCN) 是一类只由卷积层和全连接层组成的神经网络模型，它在语义分割任务中表现出了极高的识别质量和效率<sup>[52]</sup>，这里我们利用 FCN 模型的多层全卷积设计来构建辐射噪声识别模型。我们使用了 Wang 等人<sup>[29]</sup>的论文中设计的 FCN 结构，该结构的隐藏层由三个卷积模块组成的，每层卷积对应的卷积核个数分别为 128, 256 和 128，每个卷积模块都由卷积层、批归一化层和 Relu 激活函数组成。

5. ResNet。残差连接神经网络 (Residual Network, ResNet) 在目标检测和其他视觉任务中已经取得了巨大的成功<sup>[45]</sup>，通过在卷积模块中引入残差连接的机制，ResNet 模型可以有效地扩展模型的深度，提高识别模型的表达能力。这里我们使用了 Wang 等人<sup>[29]</sup>的论文中设计的 ResNet 结构，该结构的隐藏层包含了三个残差卷积模块，每个卷积模块内都包含了三个通过残差结构连接起来的卷积层，它们具有相同的卷积核个数；三个残差卷积模块对应的卷积核个数分别是 64, 128, 128。

表 3-5: 辐射噪声分类实验识别率

数据集	算法	Accuracy	AUC
数据集 1	MFCC+SVM	45.53%	78.25%
	MFCC+KNN	69.92%	90.81%
	MLP	71.14%	91.31%
	FCN	73.98%	93.14%
	ResNet	72.36%	93.10%
	TDNN-attn	<b>76.83%</b>	<b>94.51%</b>
数据集 2	MFCC+SVM	60.99%	63.62%
	MFCC+KNN	72.24%	78.90%
	MLP	72.47%	81.54%
	FCN	73.33%	83.51%
	ResNet	75.04%	84.61%
	TDNN-attn	<b>76.81%</b>	<b>85.79%</b>

通过表格 3-5 的实验结果可以看到，相较于传统的时频分析方法，几种深度神经网络模型获得了更好的识别性能，这是因为深度神经网络方法可以从复杂的辐射噪声数据中学习到的表达力更丰富的深度特征。同时，通过和其他三种时间序列分类任务常用的深度学习方法对比，可以看到我们设计的 TDNN-attn 模型获得了最好的识别性能。其中我们的模型结构和 FCN 最为相似，但是我们采用了时延卷积层来替换传统的卷积层，这说明了时延卷积可以提取到更好反映辐射噪声特性的深度特征。

图 3-9 给出了我们设计的模型在两个辐射噪声数据集上识别结果对应的混淆矩阵，其中左侧的图片是任务 1 测试集识别结果对应的混淆矩阵，右侧是任务 2 测试结果对应的混淆矩阵。可以看到我们的方法可以较好地识别不同辐射噪声所属船只类别，在任务 1 中，识别效果最好的是渔船，任务 2 中识别效果最好的是货船。这两类船只的辐射噪声数据都有一个共同的特点，就是其对应的实验数据较多，我们的模型可以对这两类辐射噪声特点展开充分的学习。可以看到，实验数据的规模直接影响了模型的识别性能，因此我们将在下一章中开展辐射噪声数据增强研究，通过生成辅助训练数据来缓解数据不足的问题。

我们将数据集 1 中数据的 MFCC 特征和经过 TDNN 模型提取到的深度特征通过 t-SNE 可视化算法<sup>[53]</sup>进行了降维可视化。t-SNE 是一种经典的数据降维算

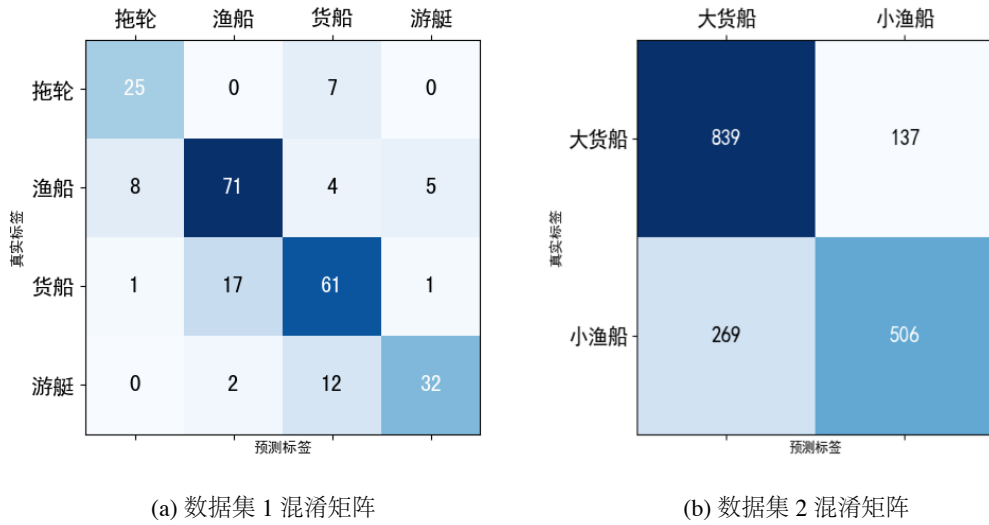


图 3-9: 不同数据集的混淆矩阵结果

法，该算法可以在降维过程中保持原有数据的局部特征，因此经常被用来对高维特征进行降维可视化。可视化结果如图 3-10 所示。可以看到经过我们设计的深度时延网络模型的特征提取之后，在新的特征空间中各个类别的类内距离和类间距离都相较 MFCC 特征有了很大的提升，这也进一步证明了我们所设计的模型分类能力的有效性和泛化性。

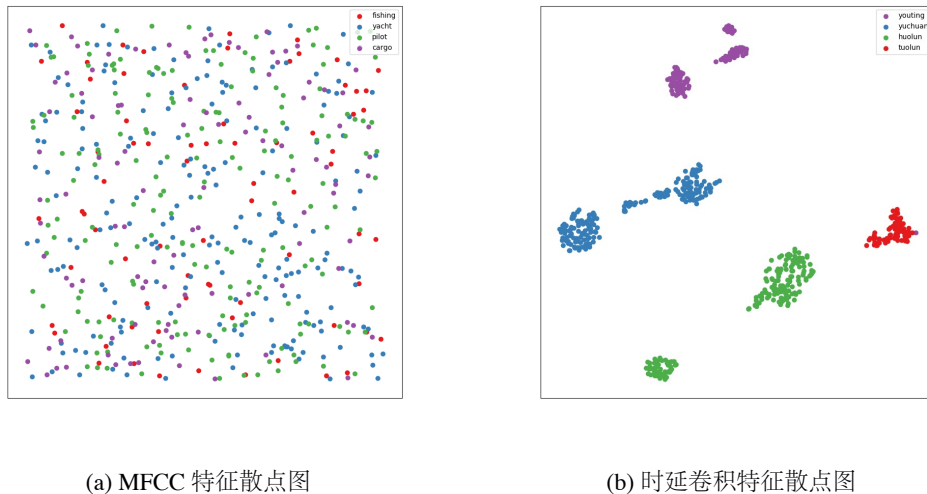


图 3-10: MFCC 特征可视化 vs. 时延卷积特征可视化

### 3.3.5 消融实验结果

这一部分我们对模型中的一些细节设计进行消融对比实验，来验证这些模块的设计是否合理。

#### 1. 模型输入的影响

我们将模型输入从 MFCC 特征替换为时域波形，重新在 TDNN-attn 模型上进行了实验，识别结果如表格 3-6 所示。

表 3-6: 时域波形 vs.MFCC 输入对比实验

数据集	输入	Accuracy	AUC
数据集 1	time	69.51%	90.62%
	MFCC	<b>76.83%</b>	<b>94.51%</b>
数据集 2	time	69.67%	79.48%
	MFCC	<b>76.81%</b>	<b>85.79%</b>

从表 3-6 中的实验结果可以看到，将输入替换为时域波形之后，TDNN-attn 的识别性能在两个数据集上都有所下降，这说明在基于 MFCC 特征构建的时频图上进行训练，有利于模型从时域和频域两个角度来捕捉不同数据之间的差异，替换为时域波形之后模型缺乏了频域角度的信息，从而影响了识别性能。

#### 2. 残差连接和空洞卷积的影响

为了研究时延卷积结构中，残差连接和空洞卷积这两个机制对模型的影响，我们在 TDNN-attn 模型中依次去除这两个结构设计来进行消融实验，识别结果如表 3-7 所示。

表 3-7: 空洞卷积和残差连接结构对比实验

数据集	结构	Accuracy	AUC
数据集 1	none	73.67%	91.80%
	only dilated	74.88%	92.29%
	only skip connection	75.50%	93.17%
	dilated+skip connection	<b>76.83%</b>	<b>94.51%</b>
数据集 2	none	74.21%	83.77%
	only dilated	75.39%	84.91%
	only skip connection	74.41%	84.01%
	dilated+skip connection	<b>76.81%</b>	<b>85.79%</b>

从表 3-7 的实验结果可以看到，空洞卷积和残差连接都可以有效地提高模型的识别性能。其中空洞卷积对模型的改进效果更明显，这是因为我们使用的卷积核尺寸较大，空洞卷积结构可以在不影响感受野的同时，较大幅度地降低模型参数，从而有效地降低了模型训练的难度。由于我们使用的模型结构相对较浅，梯度不稳定的问题可能并不明显，因此残差连接的优化效果一般。

### 3. 池化层类型对模型的影响

为了研究使用不同的池化层对模型识别性能的影响，我们将 TDNN-attn 模型中的统计池化层 (stat-pooling) 依次替换为了两种常见的池化结构：最大池化 (max-pooling) 和平均池化 (avg-pooling) 进行实验，来对比不同池化层的实验效果，识别结果如表 3-8 所示。

表 3-8: 不同池化层对比实验

数据集	池化层	Accuracy	AUC
数据集 1	max-pooling	67.07%	87.93%
	avg-pooling	75.61%	94.14%
	stat-pooling	<b>76.83%</b>	<b>94.51%</b>
数据集 2	max-pooling	72.82%	80.30%
	avg-pooling	75.16%	84.41%
	stat-pooling	<b>76.81%</b>	<b>85.79%</b>

从表 3-8 的实验结果可以看到，相较于图像识别中常用的最大池化和平均

池化层，统计池化层获得了最优的识别性能。这和我们前面的理论分析是一致的，因为统计池化层提取到的统计信息中包含了丰富的信号特征，因此基于该模块构造出的辐射噪声整体特征具有更高的辨识度，从而可以获得更优的识别性能。

#### 4. 帧维度注意力层对模型的影响

为了研究我们提出的帧维度注意力层对模型识别性能的影响，我们从 TDNN-attn 模型中删除了对应的注意力层结构进行了实验，识别结果如表 3-9 所示。

表 3-9: 帧维度注意力层对比实验

数据集	注意力层	Accuracy	AUC
数据集 1	no	73.58%	93.10%
	yes	<b>76.83%</b>	<b>94.51%</b>
数据集 2	no	75.50%	85.47%
	yes	<b>76.81%</b>	<b>85.79%</b>

从表 3-9 的实验结果可以看到，帧维度注意力层的引入可以有效地提高模型的识别性能。这是因为帧维度注意力层可以提高输入数据中有效信号对应的帧维度特征，削弱无效信号对应的帧维度特征，从而提高了 TDNN-attn 提取到的深度特征在特征空间的类内聚合程度，降低了误分类的情况。

### 3.4 本章小结

本章我们对辐射噪声信号识别问题展开了深入的研究。我们首先分析了现有方法的不足，然后针对这些缺点提出了一种基于帧维度注意力的深度时延网络 TDNN-attn。在该模型中，我们设计了一种可以较好地捕捉辐射噪声时频特征的时延卷积模块；同时为了有效信息稀疏分布的问题，我们还提出了一种帧维度注意力模块，该模块可以有效地提高包含丰富噪声信息的帧特征对分类结果的影响。最后我们在两个辐射噪声数据集上展开了实验，结果显示，我们提出的模型相较于其他辐射噪声识别算法，具有最优的识别性能。同时我们对模型中的各个模块展开了消融实验，充分证明了这些模块对模型性能的影响。



## 第四章 水下辐射噪声数据增强

尽管 TDNN-attn 等深度学习方法已经可以较好地识别各类辐射噪声信号了，但是其识别性能仍然受到实验数据规模的限制。然而由于采集成本高、采集周期长、采集方法复杂等原因，使得收集大量的船只辐射噪声数据是非常困难的。因此开展数据增强研究，根据真实数据的特性产生更多的增强数据，对于扩充噪声信号数据集以及提高识别算法的性能都有非常重要的研究价值。

### 4.1 研究动机

在辐射噪声识别任务的研究中，我们发现对于数据比较充足的辐射噪声信号类型，识别模型可以较好地识别，但是那些实验数据较少的噪声信号类别，模型则难以区分。为了解决该问题，我们考虑利用数据增强的手段，对原有的辐射噪声训练集进行扩充，来缓解训练数据不足对模型识别性能的影响。

目前关于辐射噪声数据增强的研究相对较少，主流的方法大都是通过一些时频分析方法，来手动在原始数据中加入可控的噪声，从而完成数据增强<sup>[54-57]</sup>。这类基于手工设计的数据增强方法难以在原有数据的基础上引入更多的多样性，因此对识别模型的识别性能提升有限。针对上述方法的不足，本章的研究旨在基于深度学习算法开发一种自动化的辐射噪声数据增强模型，通过生成更加丰富的扩充数据，来帮助识别模型改善性能。

### 4.2 基于时延卷积的变分自编码器

本章我们构建了一种基于时延卷积的变分自编码器模型（Variational Auto Encoder with time delay convolution, VAE-tdc），通过生成新的辐射噪声数据来扩充训练数据。本节我们将会对 VAE-tdc 的设计原理进行介绍，包括模型的整体结构，时延卷积编码器和转置时延卷积解码器的结构设计，以及模型参数训练过程。

### 4.2.1 模型结构

我们使用上一章提出的时延卷积单元来构建了编码器结构，同时设计了对应的转置时延卷积解码器结构，共同组成了 VAE-tdc 数据增强模型，如图 4-1 所示。

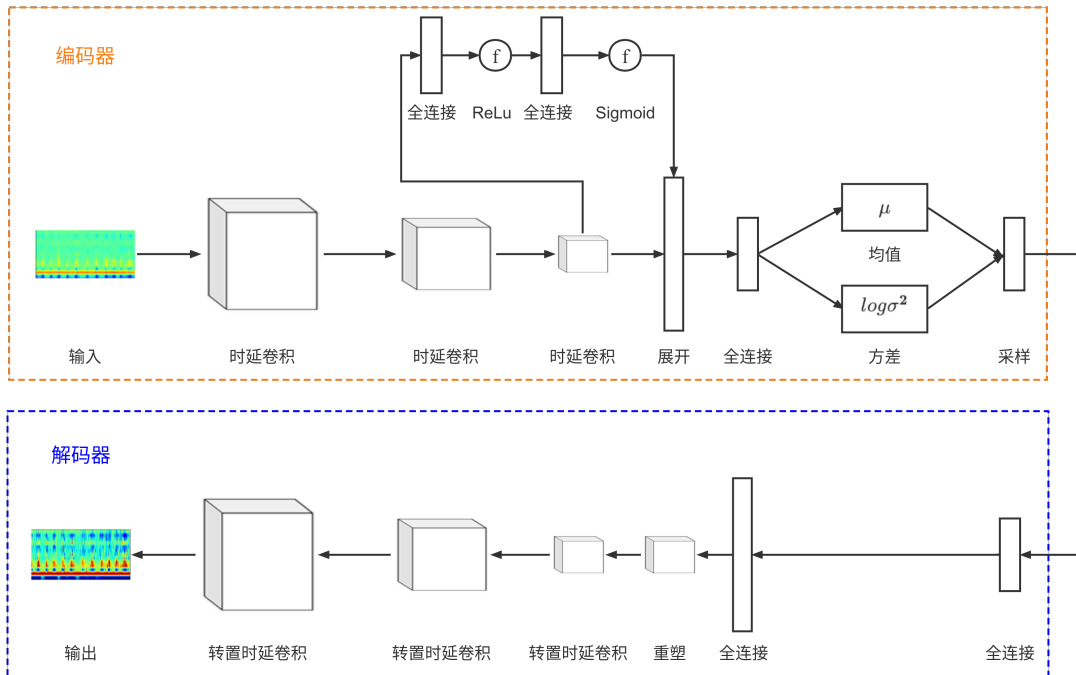


图 4-1: 基于注意力的 VAE 模型

该模型主要由下面几部分组成：

1. 模型输入。和辐射噪声识别任务相同，我们采用了辐射噪声信号的 MFCC 声学特征作为 VAE-tdc 模型的输入。在上一章的实验中我们发现，MFCC 特征作为模型输入特征，比起时域波形输入可以获得更加优秀的识别性能；同时基于 MFCC 特征进行数据增强，可以产生具有丰富时频差异的增强数据，有利于提升训练样本的多样性。

2. 时延卷积编码器。该模块首先利用时延卷积模块来提取辐射噪声的帧维度特征，接着将特征经过帧维度注意力层进行加强。然后将帧维度特征分别输入两个全连接模块，得到特征向量概率分布的均值  $\mu$  和方差的对数  $\log \sigma^2$ ，最后在特征向量概率分布  $N(\mu, \sigma^2)$  中采样得到带扰动的特征向量。

3. 转置时延卷积解码器。在该部分，我们根据时延卷积的特性设计了对应的转置时延卷积单元，将编码器输出的特征向量输入解码器，经过逐层解码

得到和输入数据尺寸相同的重构输出，经过训练后的重构输出便可以用于数据增强。

### 4.2.2 时延卷积编码器

在上一章实验分析的图 3-10 中我们可以看到，时延卷积模块可以将不同类别的辐射噪声数据映射到一个具有较好分类间隔的特征空间，在该特征空间中，同一类别的辐射噪声对应的特征向量之间彼此近邻，不同类别的特征向量则相对疏远。时延卷积的这一特性可以降低在数据增强过程中，由一个类别的数据生成具有其他类别特性数据的可能性，从而保证了数据增强模型的可用性。

这里我们使用了和上一章相同的帧维度注意力层，用来加强有效信号对应的时延卷积帧特征，降低那些较弱信号对应的帧特征。数学原理如公式 4-1 所示，其中  $x_{\text{input}}$  代表输入注意力层的特征向量， $W_1$  和  $W_2$  代表注意力层中二个全连接层的参数， $x_{\text{attn}}$  代表注意力层的输出， $\gamma$  代表注意力权重的缩放因子，用来控制  $x_{\text{attn}}$  对帧维度特征的影响程度， $*$  代表向量乘法， $\odot$  代表帧维度乘法， $x_{\text{output}}$  代表经过帧维度注意力层处理过的时延卷积特征。通过引入帧维度的注意力层设计，我们希望有效信号片段的特征可以更好地影响编码器的输出，减少无效信号片段引起的特征空间分布漂移现象，从而防止采样特征向量偏离真实的特征空间分布。

$$\begin{aligned} x_{\text{fc1}} &= \text{relu}(W_1 * x_{\text{input}}) \\ x_{\text{attn}} &= \text{sigmoid}(W_2 * x_{\text{fc1}}) \\ x_{\text{output}} &= (1 + \gamma x_{\text{attn}}) \odot x_{\text{input}} \end{aligned} \quad (4-1)$$

和 TDNN-tdnn 模型不同的是，在编码器结构中我们没有使用统计池化层来对帧维度特征进行综合。这是因对帧维度特征进行时间维度的平均和方差计算，会导致编码器得到的特征向量丢失较多的时域信息，使得最终生成的重构数据和原始信号在时域上产生较大的偏差。

在 VAE-tdc 模型中，我们假设编码器得到的特征向量概率分布为正态分布，编码器将时延卷积特征通过两个独立的全连接网络结构来计算正态分布的两个参数：均值  $\mu$  和方差  $\log\sigma^2$ 。这里输出方差的对数而不是方差本身，是因为正态分布的方差是正数，而全连接模型的输出是分布在  $(-\infty, +\infty)$  上的。此

时通过对数操作就可以将方差映射到  $(-\infty, +\infty)$  的取值空间上。

在得到特征空间分布的均值和方差后，我们通过在正态分布  $\mathcal{N}(\mu, \sigma^2)$  上进行采样来得到一个新的特征向量  $z$ ，用于输入解码器产生重构数据。由于采样操作不是可导的，因此我们采用了 Kingma 等人<sup>[41]</sup>提出的重参数操作来完成采样，其原理如公式 4-2 所示。重采样利用正态分布的乘积特性，通过标准正态分布  $\mathcal{N}(0, 1)$  中采样得到的向量  $\epsilon$  来构建服从正态分布  $\mathcal{N}(\mu, \sigma^2)$  的特征向量，该过程是可导的，因此可以通过梯度下降来进行参数优化。

$$\begin{aligned}\epsilon &\sim \mathcal{N}(0, 1) \\ z &= \mu + \sigma * \epsilon\end{aligned}\tag{4-2}$$

### 4.2.3 转置时延卷积解码器

因为时延卷积和传统卷积的计算方式不同，因此在解码器结构中，我们根据时延卷积的特点，设计了对应的转置时延卷积，将编码器输出的特征向量逐渐解码得到重构数据。下面我们对转置时延卷积中的关键步骤进行介绍：

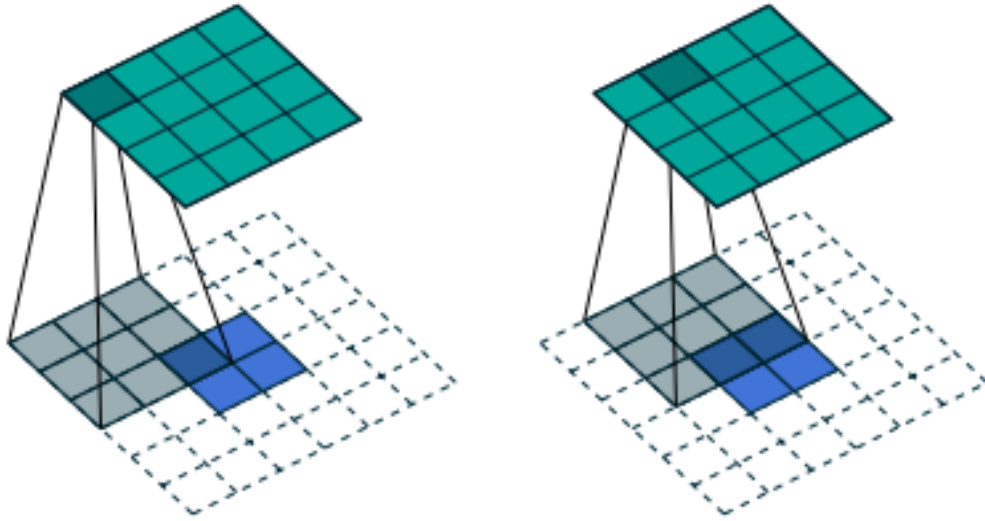
1. 转置操作。首先对输入特征向量的通道维度和高维度进行交换，得到转置特征向量。

2. 转置卷积。转置卷积操作首先将转置特征向量扩散为一个更大尺寸的特征矩阵，然后依次将转置特征向量中的元素和转置卷积核相乘，得到一系列转置卷积运算的中间结果。我们的转置卷积操作参考了 Dumonlin 等人<sup>[3]</sup>的做法，如图 4-2 所示。转置卷积是转置时延卷积的关键步骤，每个转置卷积的输出都和编码器中相应位置的时延卷积的输入具有相同的尺寸，这样多次运算之后，就可以得到和编码器输入数据相同尺寸的重构输出。

3. 批归一化。VAE-tdc 模型包含了两个深度相同的编码器和解码器结构，因此模型深度较深，为了缓解由于多层非线性运算导致的向量分布漂移问题，我们在每个转置卷积之后接入一个批归一化层，来对批次训练数据的特征分布进行动态调整。

4. 残差连接。同样为了防止过拟合问题，我们在转置时延卷积的输入和批归一化层的输出之间加入了一条残差连接支路，来缓解深层网络模型在参数更新时出现的梯度不稳定问题。

经过转置时延卷积，我们得到了和输入 MFCC 特征相同尺度的重构输出，但是此时重构输出和输入之间还存在着较大的偏差，并不能直接用于数据增

图 4-2: 转置卷积原理图<sup>[3]</sup>

强。通过模型训练，才能使 VAE-tdc 逐渐产生和原始输入相近，同时又具有一定多样性的增强数据。

#### 4.2.4 模型训练

VAE-tdc 模型的损失函数如公式 4-3 所示，其中符号  $N_b$  代表模型每轮参数更新的批次大小； $\|\cdot\|_2$  代表向量的 1-2 范数； $\text{KL}(\cdot|\cdot)$  代表计算两个不同分布的 Kullback-Leibler 散度，简称 KL 散度，该指标反映了两个分布之间差异程度。

$$\begin{aligned} \text{loss} &= \text{loss}_{\text{recon}} + \text{loss}_{\text{reg}} \\ &= -\frac{1}{N_b} \sum_{i=1}^{N_b} \|\vec{x}_{\text{recon}} - \vec{x}_{\text{input}}\|_2^2 + \text{KL}(\mathcal{N}(\mu, \sigma^2) | \mathcal{N}(0, 1)) \end{aligned} \quad (4-3)$$

上述损失函数的第一部分  $\text{loss}_{\text{recon}}$  是重构误差，主要是为了惩罚模型重构输出和输入数据之间的差异；第二部分  $\text{loss}_{\text{reg}}$  是正则项，通过将特征空间分布  $\mathcal{N}(\mu, \sigma^2)$  和标准正态分布  $\mathcal{N}(0, 1)$  之间的 KL 散度来作为正则项，使得编码器输出的特征空间分布不会偏离标准正态分布过远。

在模型训练之初，损失函数中重构误差的影响较大，模型主要朝着缩小输出和输入之间的重构误差的方向进行，此时重构误差减小，KL 散度正则项处于较大的水平；而随着模型逐渐训练，重构误差变小，此时 KL 散度正则项的

影响逐渐增大，从而加大了特征空间采样向量的噪声。此时的解码器的重构生成数据开始逐渐变得多样化，但是受到重构误差的影响，重构输出和输入之间的差异仍然保持在一定范围的。当模型收敛后，VAE-tdc 就可以产生具有丰富多样性的重构数据，来进行数据增强。

在模型训练过程中，我们使用大小为 8 的批次尺寸进行小批量梯度更新，更新轮数为 50 轮；使用 Adam 优化器<sup>[49]</sup>进行参数更新，初始学习率设置为 0.008，使用阶梯式学习率调整策略进行学习率更新，阶梯式学习率调整参数为 0.95；损失函数中，KL 散度正则项的权重为 0.00025。

下面我们给出 VAE-tdc 模型的参数细节，如表格 4-1 所示，其中  $n_m$  代表输入 MFCC 特征的阶数；时延卷积层 TDC 和转置时延卷积层 TRANS-TDC 的参数  $(k, d, c)$  中， $k$  代表卷积核的宽度， $d$  代表空洞卷积中卷积核元素的间隔， $c$  代表卷积核的个数；FC 代表全连接层，其参数  $(i, o)$  中， $i$  代表全连接输入维度， $o$  代表全连接输出维度。

表 4-1: VAE-tdc 数据增强模型参数

层名称	输入维度	参数	输出维度
TDC1	$(1, 2498, n_m)$	$(3, 1, 128)$	$(1, 2498, 128)$
TDC2	$(1, 2498, 128)$	$(3, 1, 128)$	$(1, 2498, 128)$
ATTN-FC1	$(1, 2498, 128)$	$(128, 128)$	$(1, 2498, 128)$
ATTN-FC2	$(1, 2498, 128)$	$(128, 1)$	$(1, 2498, 1)$
FLATTEN	$(1, 2498, 128)$	-	$(1, 319744)$
ENCODER-FC	$(1, 319744)$	$(319744, 256)$	$(1, 256)$
MEAN-FC	$(1, 256)$	$(256, 256)$	$(1, 256)$
NORM-FC	$(1, 256)$	$(256, 256)$	$(1, 256)$
DECODER-FC	$(1, 256)$	$(256, 319744)$	$(1, 319744)$
RESIZE	$(1, 319744)$	-	$(1, 2498, 128)$
TRANS-TDC1	$(1, 2498, 128)$	$(1, 1, 128)$	$(1, 2498, 128)$
TRANS-TDC2	$(1, 2498, 128)$	$(1, 3, n_m)$	$(1, 2498, n_m)$

## 4.3 实验结果和分析

上面我们介绍了 VAE-tdc 的模型设计原理，在这一部分我们将会基于 VAE-tdc 以及其他数据增强方法产生的增强数据进行对比实验，研究增强数据对辐射噪声识别性能的影响。

### 4.3.1 数据增强结果可视化

我们使用了下面两个基于时频分析的数据增强算法来作为对比方法：

1. 基于 MFCC 特征域的数据增强。该方法参考了刘等人<sup>[54]</sup>的研究，他们使用随机裁剪的方法来对辐射噪声的 MFCC 特征进行数据增强。该方法通过从图片中随机选中若干个小区域范围，然后将这些区域内的像素值置为 0，来完成 MFCC 时频图的信息随机裁剪。因为对于人眼来说，少量的遮挡或者图片缺失并不影响人类视觉系统对图片类型的判断，这种随机裁剪的设计利用了人眼的感知原理来生成增强数据。

2. 基于离散小波域的数据增强。该方法参考了 Priyadharsini 等人<sup>[55]</sup>的研究，他们使用了一种基于离散小波变换的方法来对水下信号进行去噪、信号增强等操作。利用该方法的思路进行数据增强时，首先通过离散小波变换对原始信号进行逐级分解，然后对每一步分解得到的高频信号成分添加时域随机扰动，最后通过反向小波分解操作，将所有的小波分解信号合成得到增强信号。因为小波分解的低频部分属于信号的粗粒度信息，对其添加扰动容易影响原始信号的样貌，而对高频部分对应的细粒度信息添加扰动，可以在不对原始信号造成大幅度改变的同时，引入更多的数据多样性，从而达到数据增强的目的。

图 4-3 展示了上述两种方法以及 VAE-tdc 模型产生的增强数据，其中第一列是原始的辐射噪声 MFCC 特征，第二列是基于 MFCC 特征域的数据增强算法产生的数据，第三列是基于离散小波域的数据增强算法产生的数据，最后一列是 VAE-tdc 模型产生的增强数据。从数据增强结果可以看到，对比另外两种增强方法，VAE-tdc 模型产生的增强数据和原始数据相比有了更丰富的变化，同时增强数据中又保留了原始数据的分布特点。

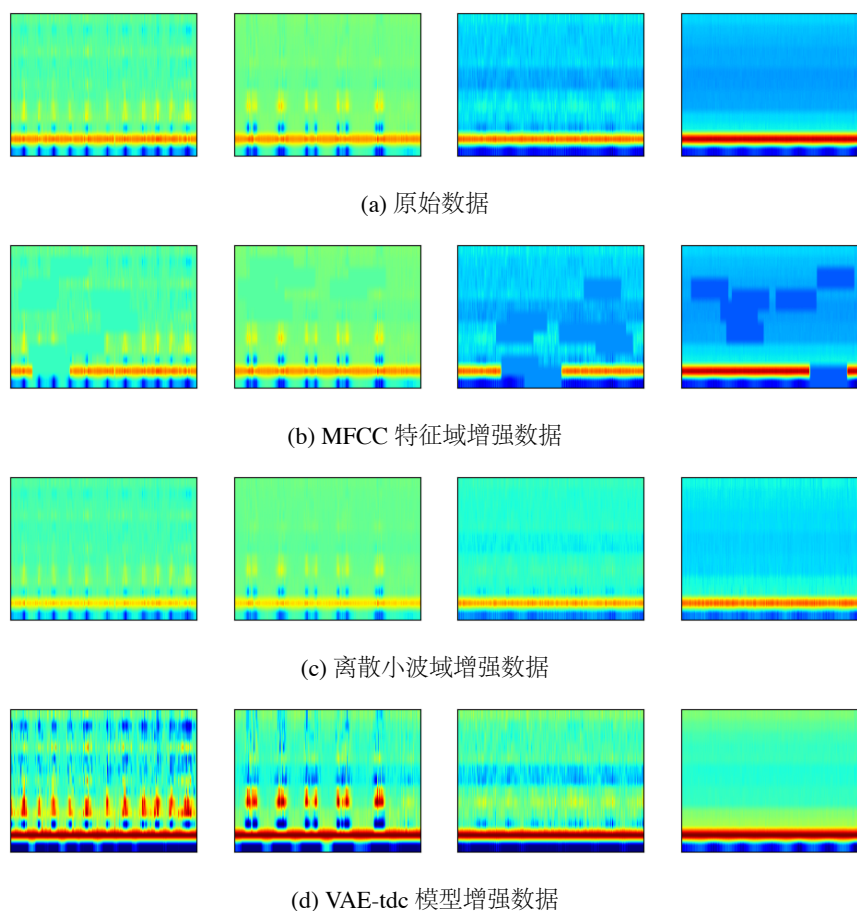


图 4-3: 不同数据增强方法结果图

### 4.3.2 对比实验结果

我们在这一部分采用了上一章节中所使用了两个辐射噪声数据集进行数据增强实验。对比实验主要分为两个阶段，数据增强阶段和模型评估阶段。在数据增强阶段，我们分别使用三种数据增强方法对两个辐射噪声数据集的训练数据进行数据增强，其中每个原始训练数据产生一个对应的增强数据，增强数据的标签和原始数据相同，然后将生成的增强数据加入训练集，完成训练数据扩充。在模型评估阶段，我们选取了上一章实验中识别效果最好的三种算法模型：FCN、ResNet 以及 TDNN-attn 来进行实验，每个模型在经过扩充后的训练集上进行训练，然后在测试集上进行评估，计算 Accuracy 和 AUC 两项指标。数据增强实验结果如表格 4-2 所示。

表 4-2: 辐射噪声数据增强实验

数据集	算法	数据增强	Accuracy	AUC
数据集 1	FCN	none	73.98%	93.14%
		mfcc	74.39%	<b>94.06%</b>
		wavelet	74.39%	93.67%
		VAE-tdc	<b>75.20%</b>	93.44%
	ResNet	none	72.36%	93.10%
		mfcc	73.17%	91.81%
		wavelet	74.39%	92.57%
		VAE-tdc	<b>75.61%</b>	<b>94.20%</b>
	TDNN-attn	none	76.83%	94.51%
		mfcc	78.05%	94.80%
		wavelet	78.46%	94.75%
		VAE-tdc	<b>80.90%</b>	<b>95.13%</b>
数据集 2	FCN	none	73.33%	83.51%
		mfcc	<b>75.67%</b>	83.54%
		wavelet	75.39%	84.01%
		VAE-tdc	75.39%	<b>84.27%</b>
	ResNet	none	75.04%	84.61%
		mfcc	75.50%	84.72%
		wavelet	<b>77.16%</b>	<b>85.91%</b>
		VAE-tdc	76.19%	84.68%
	TDNN-attn	none	76.81%	85.79%
		mfcc	78.24%	86.86%
		wavelet	78.58%	87.26%
		VAE-tdc	<b>79.44%</b>	<b>87.88%</b>

从对比实验结果可以看到，三种数据增强方法都不同程度地提高了辐射噪声识别模型的性能，其中基于 MFCC 域的数据增强方法对 FCN 模型性能的提升更明显，VAE-tdc 算法对 ResNet 和 TDNN-attn 模型的提升效果更好。通过进行数据增强实验，证明了基于数据增强来扩充训练数据有利于缓解辐射噪声数据不足的问题；同时通过对比不同的数据增强方法的实验效果，可以看到

VAE-tdc 产生的增强数据在大多数模型上都可以更好地改善模型的识别性能，这是因为 VAE-tdc 的重构数据和真实数据之间的变化受到了重构损失函数的约束，同时由于 VAE-tdc 模型引入了非线性变换，使得生成数据相较原始数据的变化更加自然和多样化。

为了进一步说明基于 VAE-tdc 方法生成的增强扩充数据对辐射噪声分类模型的影响，我们绘制了经过数据增强改善的 TDNN-attn 模型在测试集上的混淆矩阵，如图 4-4 所示。和上一章节中的混淆矩阵结果 3-9 进行对比可以发现，对于实验数据较少的辐射噪声类别，分类模型的识别准确率有了一定的提升。这说明通过 VAE-tdc 方法来扩充辐射噪声训练数据，可以有效地缓解训练数据不足对分类模型的影响。

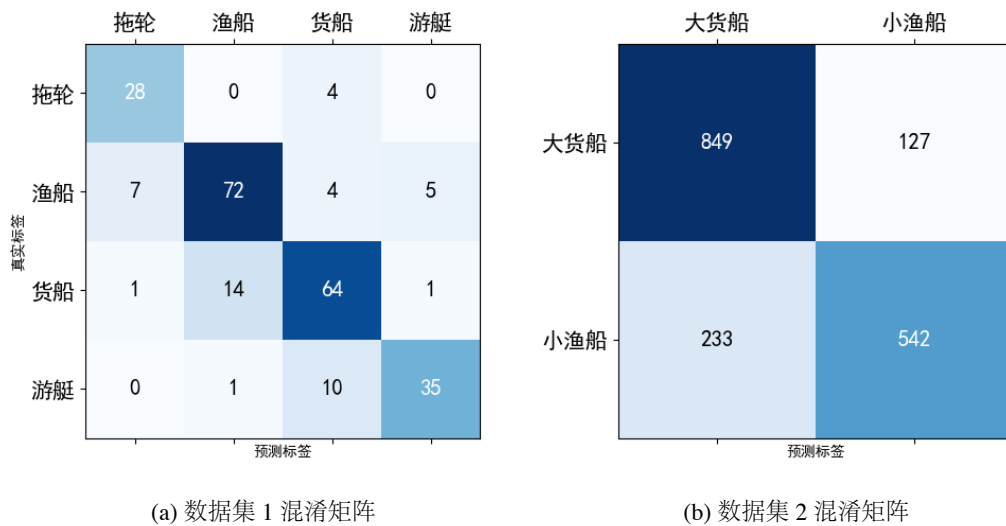


图 4-4: 不同数据集的混淆矩阵结果

### 4.3.3 消融实验

在这一部分，我们将会对 VAE-tdc 中的部分模块展开了消融实验，来研究其对整体模型的影响。

1. 证明帧维度注意力层对模型识别性能的影响，实验结果如表 4-3 所示，该实验中我们使用 TDNN-attn 作为基础的识别模型。

表 4-3: 帧维度注意力层对比实验

数据集	注意力层	Accuracy	AUC
数据集 1	no	80.14%	94.89%
	yes	<b>80.90%</b>	<b>95.13%</b>
数据集 2	no	79.01%	87.23%
	yes	<b>79.44%</b>	<b>87.88%</b>

从表 3-9 的实验结果可以看到，帧维度注意力层的引入可以帮助 VAE-tdc 模型生成更高质量的增强数据，改善识别模型的性能。这是因为帧维度注意力层提高了有效信号片段特征对编码器输出的影响，降低了由于无效信号片段导致特征空间分布发生漂移的可能性。不过实验结果中帧维度注意力层的提升效果并不大，这有可能是因为帧维度注意力层过度削弱了部分信号的特征，导致解码器的重构输出中丢失了部分原始 MFCC 特征的时域特性。

## 4.4 本章小结

本章我们对辐射噪声信号数据增强展开了研究。我们首先分析了现有方法的不足，然后针对这些缺点提出了一种基于时延卷积的变分自编码器模型 VAE-tdc。在该模型中，我们使用时延卷积层来构建编码器部分，从输入 MFCC 特征中提取具有表达力的深层信息；使用转置时延卷积层来构建解码器部分，将在特征空间采样得到的特征向量解码得到重构 MFCC 特征，完成数据增强。我们在两个辐射噪声数据集上展开了实验，经过实验证明，我们提出的模型相较于其他辐射噪声数据增强算法，可以更好地提升识别模型的性能。同时我们还对模型中的部分模块展开了消融实验，证明了这些模块对模型性能的影响。



# 第五章 水下辐射噪声识别系统

在前面的章节中，我们介绍了辐射噪声信号识别的相关问题和研究方案，为了让我们的算法可以转变为具体的工作设备，投入到生产生活中使用，在这一章节中，我们将会讨论我们基于上述算法开发的一套辐射噪声信号识别系统。

## 5.1 系统开发背景

随着各类海洋产业的快速发展，辐射噪声信号识别技术的应用范围越来越广泛，渔业生产、海底勘测、以及各种调查和科学研究都涉及到对辐射噪声的识别，如图5-1所示。在不同的任务场景中，采集并存储各类型的辐射噪声信号，对其完成正确识别，将会极大地提高我们日常的生产生活质量。



(a) 探测鱼群规模



(b) 判断货船载货量

图 5-1: 辐射噪声信号识别用途

目前市面上尚未出现集数据分析、可视化以及模型训练和预测为一体的辐射噪声识别软件系统，为了解决上述问题，我们设计了一套包含了上述功能的识别系统。对于输入的每一条辐射噪声信号，我们的系统都可以对其进行波形可视化，然后完成类别标记，并将其加入对应的任务数据库中，同时我们的系统还提供了一站式的模型训练、评估功能。下面我们将会对所设计的软件系统展开详细的介绍。

## 5.2 系统设计

这一部分我们首先明确了系统的需求，并基于这些需求设计了系统的整体架构，为后续的具体模块的设计打下基础。

### 5.2.1 系统需求设计

在开发具体的软件系统之前，我们需要首先明确系统的功能需求，下面我们梳理出了系统开发的几个关键需求，并针对每个需求给出了简单的原型设计图。

1. 数据可视化和存储。在后续的认可功能开发之前，我们的系统需要具备数据加载和可视化的能力，在输入新数据之后，用户可以通过屏幕看数据的波形，通过扬声器来听数据的声音，从而决定这是属于哪个识别任务的哪类数据，同时我们需要提供一个存储数据库，来保存输入系统的辐射噪声数据，该需求的原型图如 5-2 所示。

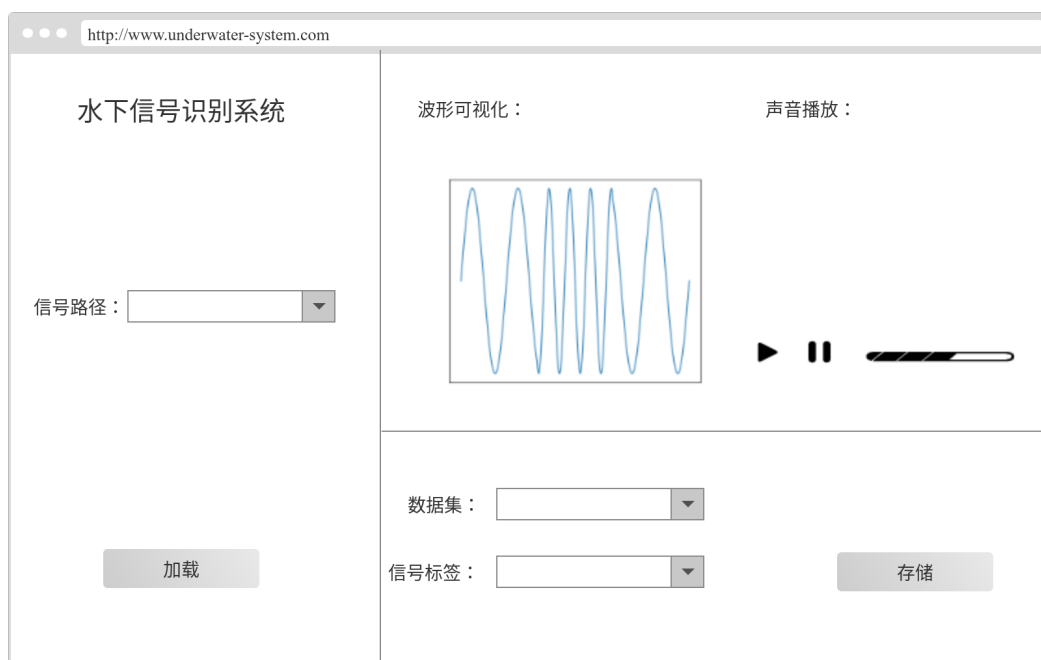


图 5-2: 数据可视化和存储需求原型图

2. 训练功能。这部分是我们的识别系统的一项重要功能，模型训练开始时，我们需要为用户提供数据集选择、识别算法选择、预训练模型选择的选

项，同时识别算法的一些超参数，比如训练轮数、批次大小、学习率等参数也需要提供给用户。模型训练过程中，我们需要为用户展示损失下降曲线，测试集上的精度变化曲线，以及模型保存的按钮，同时我们还需要为用户预留一个文字输出窗口，方便输出一些调试信息，该需求的原型图如 5-3 所示。

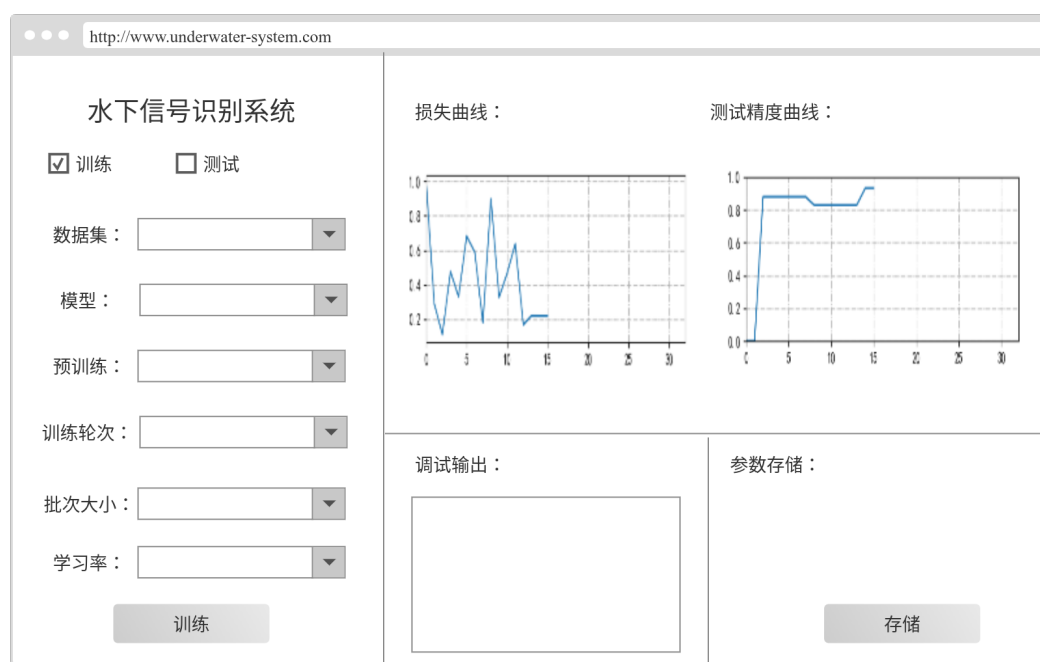


图 5-3: 训练需求原型图

3. 测试功能。这部分是我们的识别系统的另一项重要功能，在完成模型训练之后，我们在测试阶段需要为用户提供测试新数据所属类别的功能，同时我们还需要为用户预留一个文字输出窗口，方便输出一些调试信息，该需求的原型图如 5-4 所示。

### 5.2.2 系统架构设计

在明确了软件系统开发的需求之后，我们对系统架构进行了设计，梳理了需要开发的功能模块，以及不同模块之间的交互，如图 5-5 所示。通过架构图可以看到，我们的系统主要需要完成三个主要的功能模块：

1. 数据存储。该模块需要设计合理的存储结构，来保存采集数据的信息；
2. 模型训练。该模块需要在后端服务器上完成模型训练，同时设计接口协议，将训练过程中的各种信息传输给前端；

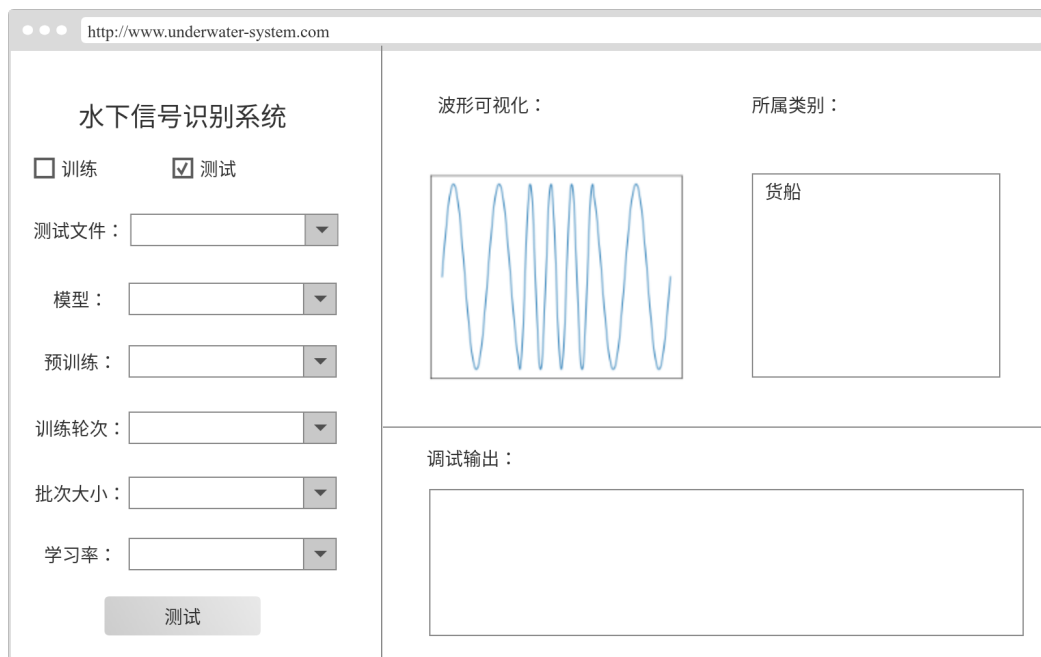


图 5-4: 测试需求原型图

3. 数据测试。该模块需要处理任意输入长度的信号，并对其进行数据处理和识别。

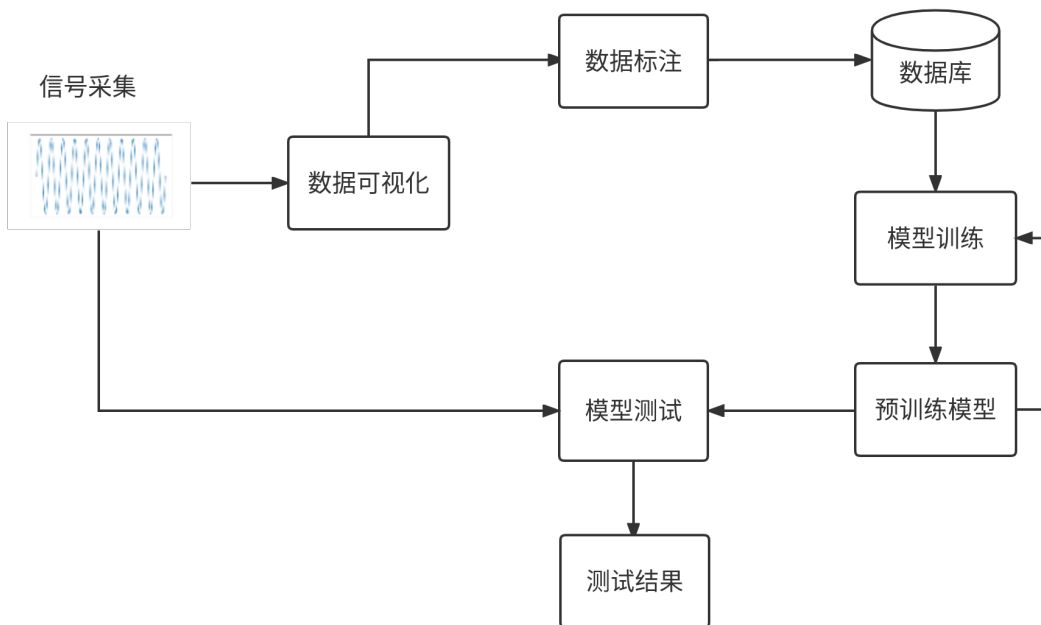


图 5-5: 系统架构设计图

## 5.3 系统实现

### 5.3.1 系统开发环境

我们采用前后端分离的设计思路，通过 `http` 协议接口来传输数据，`http` 协议部分我们使用了开源的 `Flask` 网络框架<sup>[58]</sup> 进行接口的封装。`Flask` 是基于 `Python` 开发的轻量级网页应用框架，我们可以将后端的功能模块封装成对应的 `http` 协议服务，这样使得前端可以部署在一些方便展示的设备上，比如平板电脑、树莓派等，后端可以部署在一些计算能力比较强的设备上，比如计算服务器上。

前端我们使用 `Python` 语言和对应的图形界面库 `PyQt5` 进行开发。`PyQt5` 是 `Python` 的一个开源图形界面开发库，它包含了大量的界面组件，同时为开发者提供了信号机制，方便了基于 `Python` 的图形界面开发。该部分占用系统资源较少，因此对部署设备没有限制，可以部署在树莓派、平板电脑等展示设备上，本文中我们基于笔记本电脑进行演示。

后端我们通过 `Python` 语言和 `Pytorch` 深度学习框架进行开发，利用 `MySQL` 进行数据存储。`Pytorch` 是一个开源的机器学习库，它为开发者提供了基于动态图的神经网络模型搭建接口，同时还封装好了一系列常用的网络模型、损失函数和可视化工具，便于开发和调试深度学习算法。`MySQL` 是常用的关系型数据库，我们这里主要是为了存储不同任务的数据集，以及标签、长度、采样率等信息。该部分对系统的计算能力和存储能力有要求，因此需要部署在 `GPU` 服务器上，本文中我们将后端部署在课题组的 `NVIDIA GTX 2080ti` 深度学习服务器上。

### 5.3.2 系统功能模块

通过上面的技术架构部分，我们将整个系统划分出了三个关键功能模块，下面我们对这三个模块的实现原理展开介绍。

1. 数据存储模块。为了方便后端在训练的时候可以快速检索到不同任务的数据集，我们对每个任务单独建立一张 `MySQL` 数据表，并将一些数据的关键信息存储在表中，我们将本文中所使用的两个数据集对应的 `MySQL` 表结构进行可视化，如图 5-6 所示。

该表结构和前后端均有交互，前端在读入新的辐射噪声数据，并完成数据

NoiseData1		NoiseData2	
DataID	int	DataID	int
FileName	string	FileName	string
SampleRate	int	SampleRate	int
Length	int	Length	int
Label	int	Label	int
Trainable	boolean	Trainable	boolean

图 5-6: 数据存储表结构

标注之后，点击存储按钮就会将该数据的信息插入对应任务数据集的表格中，后端在进行模型训练和模型评估的时候，就会从对应的任务数据集中将数据读出，并加载入内存，完成相应的操作。

2. 模型训练模块。当该模块工作时，后端需要完成读取数据库中的数据、调用深度学习框架进行训练和评估、通过 **https** 协议接口回传训练过程的损失值以及测试精度值的任务，前端需要完成损失函数曲线和测试精度曲线展示的任务。在该模块中，我们采用了一种前端轮询的交互方式，即前端每间隔固定时间对 **https** 接口进行一次访问，拿到最新的训练过程中间数据，并在图形页面上完成渲染。前端进行一次轮询访问时，对应的 **UML** 时序图如 5-7 所示。

3. 数据测试模块。该模块需要对输入数据的类别进行预测，因为输入数据的格式有可能和实验数据不一致，比如数据长度、采样率等，该模块会首先对原始数据进行降采样或者升采样，使得输入数据采样率和实验数据一致，然后将输入信号划分为若干段和实验数据相同长度的数据  $x_1, x_2, \dots, x_n$ 。该模块首先对上述过程产生的分段数据进行类别预测，然后通过投票法来决定测试数据的类别，原理如 5-1 所示，其中  $f(\cdot)$  代表用于预测数据类别的实验模型， $I(\cdot)$  代表指示函数，该函数的输入变量为布尔值，当输入布尔值为真时函数输出 1，反之输出 0， $m$  为所有的标签的数量。

$$y_{\text{pred}} = \max_{j \in [1, m]} \sum_{i=1}^n I(f(x_i) == j) \quad (5-1)$$

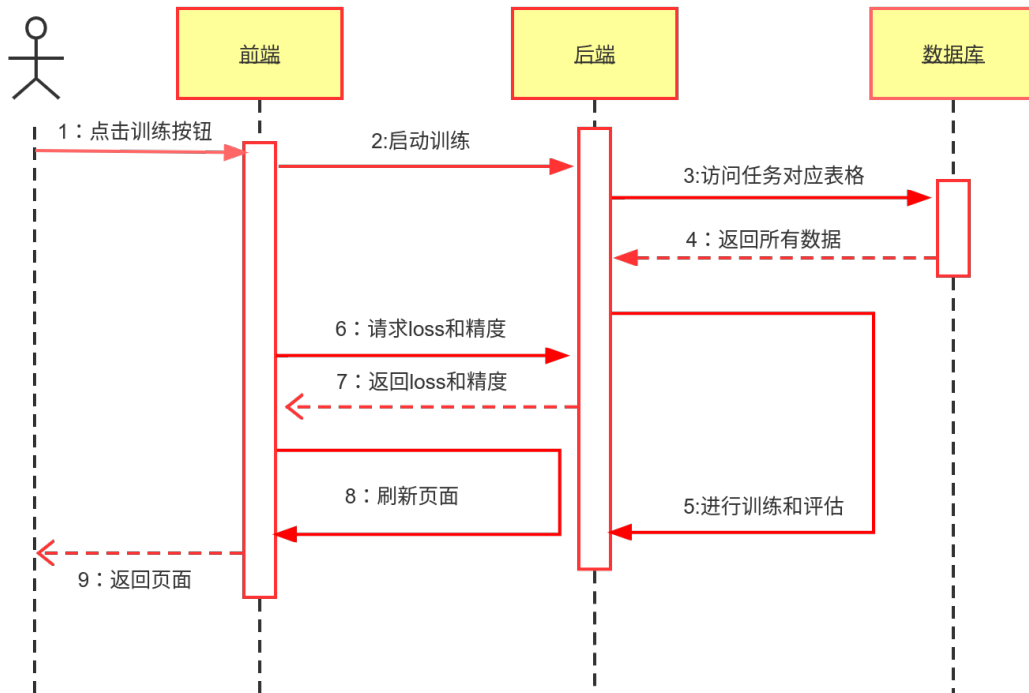


图 5-7: 训练模块 UML 时序图

### 5.3.3 系统演示

这一部分我们会演示水下辐射噪声识别软件的几个关键功能。图 5-8 展示了数据加载时的软件界面。通过点击界面右侧的播放按钮，软件将会播放辐射噪声的音频，同时波形部分也会滚动展示当前音频对应的时域波形，通过专业人员进行辨别之后，选择该数据的真实标签，点击存储，就可以将该数据存入数据库中。

图 5-9 展示了模型训练时的软件界面。使用模型训练功能时，我们需要在界面左上角选中训练模式，然后在数据集的下拉菜单中选择使用的任务数据集。接着在界面左侧的模型下拉框中选择所使用的算法模型结构，如果需要加载预训练模型，还可以在模型加载的地方选择对应的预训练模型参数文件路径。选中识别模型之后，可以在训练轮数、批次大小以及学习率的下拉框中选择具体的参数值。最后点击左下角的训练按钮，软件就会开始训练识别模型。训练过程中，软件界面的右侧还会展示模型的损失函数变化曲线，以及当前模型在测试集上的准确度。

图 5-10 展示了测试新数据时的软件界面。在对新的信号数据进行测试时，

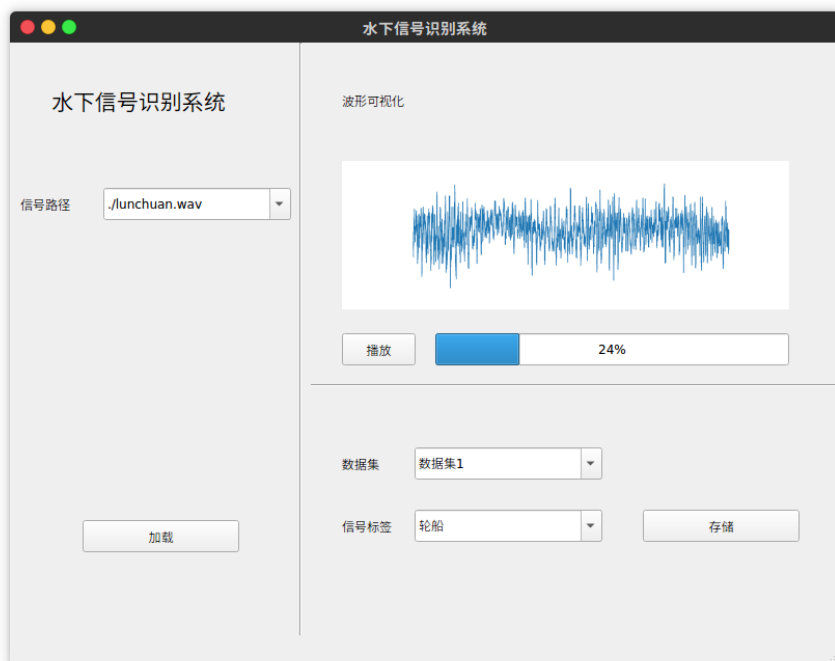


图 5-8: 数据可视化页面截图



图 5-9: 识别系统训练模式截图

我们首先在左侧选中测试数据的路径，然后选中对应的实验任务，以及所使用的识别模型和对应的预训练模型路径，点击测试就可以开始进行预测了。预测过程中右侧会展示待预测信号的波形，同时右侧的所属类别下方的文字框中会显示该信号的预测类别。



图 5-10: 识别系统测试模式截图

## 5.4 本章小结

本章我们详细介绍了水下辐射噪声识别系统的设计方案和实现细节。我们首先介绍了识别系统的开发背景，然后对系统的需求进行了汇总，详细介绍了三个需求场景以及对应的原型图。在明确了系统需求之后，我们设计了识别系统的系统架构，对具体的功能模块进行了划分。接着我们详细介绍了几个主要功能模块的实现细节，最后我们使用真实的辐射噪声数据进行了系统演示，给出了相应的系统截图。



## 第六章 总结与展望

水下信号识别是海洋信息技术发展的关键技术之一,该任务旨在通过构建水下目标信号的关键特征,来完成不同类别的水下信号的分类任务。但由于水下信号的传播容易受到复杂水体环境的影响,因此该课题一直是水声信号领域的难题。目前,海洋信号识别大多基于时频探测特征,但该特征容易受到环境噪声、信道衰落的影响,从而导致海洋目标信号的特征提取和识别性能受损。本文中我们从辐射噪声信号识别问题出发,基于现代深度学习算法开发了识别模型,并针对海洋数据规模小,难以采集的问题,提出了对应的数据增强方法。最后我们还设计了一套识别软件系统,可以一站式地完成数据存储、模型训练和信号识别的功能。

在辐射噪声信号识别任务中,我们详细介绍了辐射噪声信号的组成成分,基于辐射噪声的信号特点,提出了一种带帧维度注意力机制的深度时延网络模型 TDNN-attn。在该模型中,我们设计了一种可以较好地捕捉辐射噪声特性的时延特征提取模块,该模块使用大尺度卷积核来提取具有时频特性的深度特征,同时利用了空洞卷积和残差连接的结构设计来缓解大卷积核带来的优化问题。我们还设计了一种帧维度注意力模块,该模块通过学习不同帧维度特征之间的差异,输出对应的注意力权重,提高了包含丰富噪声信息的帧特征对分类结果的影响,削弱了包含无效信息的帧特征对分类结果的影响,从而较好地缓解了真实辐射噪声数据的有效信息稀疏的问题。经过在真实的辐射噪声数据集上进行实验,我们发现 TDNN-attn 算法提取到的深度特征在特征域具有较好的区分度,同时和其他识别模型对比,其识别性能也是最优的。

在辐射噪声信号数据增强任务中,我们首先分析了现有方法的优点和不足,然后提出了一种基于时延卷积的变分自编码器模型 VAE-tdc,通过生成新的辐射噪声数据来扩充训练数据。该模型利用了我们在识别任务阶段开发出来的时延卷积单元作为自编码器模型的编码器结构,通过时延卷积编码器来提取具有丰富表达能力的特征向量。然后我们还设计了与之对应的转置时延卷积解码器结构,该模块将特征空间采样得到的特征向量映射到数据域,产生新数据用于数据增强。经过在数据集上进行对比实验,我们发现我们设计的 VAE-tdc

模型可以有效地改善算法的识别性能。

在前面研究的基础上，我们设计了一套水下辐射噪声识别系统，该系统实现了数据存储，在实验数据上进行模型训练，以及对新数据进行测试的功能。我们针对系统的需求和架构展开了分析和设计，最终设计出来的系统可以方便地部署在各类任务场景中，对采集到的信号类型进行存储和识别。

本文从实际的任务场景出发，针对水下信号识别问题展开了相关研究。在本文的研究基础上，我们认为有如下几个研究方向可以继续深入：

1. 增量式的水下信号分类研究。由于水声数据的采集难度较大，在实验环境下可以采集到的信号数据，其类别以及数据规模都有限。设计增量式的、可持续式的水声信号识别模型，可以在实际的生产环境中不断更新和优化，具有更好的实用价值。

2. 水下信号数据压缩研究。水声信号数据的采样点数较多，这为信号处理和存储都带来了不小的难度。数据压缩算法的研究可以帮助识别系统降低信号存储占用，从而提高系统响应速度，增加工作时限。

## 参考文献

- [1] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural computation*, 1997, 9(8): 1735 – 1780.
- [2] ZHANG Z, WANG X, JUNG C. DCSR: Dilated Convolutions for Single Image Super-Resolution[J]. *IEEE Transactions on Image Processing*, 2019, 28: 1625 – 1635.
- [3] DUMOULIN V, VISIN F. A guide to convolution arithmetic for deep learning[J]. *ArXiv*, 2016, abs/1603.07285.
- [4] ROSS D. Mechanics of underwater noise[J]. *The Journal of the Acoustical Society of America*, 1989, 86(4): 1626 – 1626.
- [5] WALES S C, HEITMEYER R M. An ensemble source spectra model for merchant ship-radiated noise[J]. *The Journal of the Acoustical Society of America*, 2002, 111(3): 1211 – 1231.
- [6] JIAN L, YANG H, ZHONG L, et al. Underwater target recognition based on line spectrum and support vector machine[C] // *Proceedings of the International Conference on Mechatronics, Control and Electronic Engineering (MCE2014)*, Shenyang, China. 2014: 29 – 31.
- [7] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine learning*, 1995, 20(3): 273 – 297.
- [8] ZHANG J, LI Y, ALI W, et al. Line Spectrum Enhancement of Underwater Acoustic Signals Using Kalman Filter[J]. *Journal of Marine Science and Application*, 2020, 19(1): 148 – 154.
- [9] KALMAN R E. A new approach to linear filtering and prediction problems[J], 1960.

- [10] LI H, CHENG Y, DAI W, et al. A method based on wavelet packets-fractal and SVM for underwater acoustic signals recognition[C] // IEEE. 2014 12th International Conference on Signal Processing (ICSP). 2014 : 2169 – 2173.
- [11] 廖明熙, 张小蓓, 张歆. 基于稀疏表示的水声信号分类识别 [J]. 探测与控制学报, 2014, 36(4): 67 – 70.
- [12] DONOHO D L. Compressed sensing[J]. IEEE Transactions on information theory, 2006, 52(4): 1289 – 1306.
- [13] JIN G, LIU F, WU H, et al. Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal[J]. Journal of Experimental & Theoretical Artificial Intelligence, 2020, 32(2): 205 – 218.
- [14] SHEN S, YANG H, YAO X, et al. Ship type classification by convolutional neural networks with auditory-like mechanisms[J]. Sensors, 2020, 20(1): 253.
- [15] TIAN S, CHEN D, WANG H, et al. Deep convolution stack for waveform in underwater acoustic target recognition[J]. Scientific reports, 2021, 11(1): 1 – 14.
- [16] HU G, WANG K, PENG Y, et al. Deep learning methods for underwater target feature extraction and recognition[J]. Computational intelligence and neuroscience, 2018, 2018.
- [17] HUANG G-B, ZHU Q-Y, SIEW C-K. Extreme learning machine: Theory and applications[J/OL]. Neurocomputing, 2006, 70(1): 489 – 501.  
<https://www.sciencedirect.com/science/article/pii/S0925231206000385>.
- [18] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [19] 李林, 冯燕尔. 一种基于小波分析的船舶振动故障检测方法 [J]. 船舶, 2009, 20(4): 12 – 14.
- [20] 陆扬, 王雪松, 赵鹏远, et al. 基于时频分析和神经网络的水声通信信号识别技术 [J]. 科技导报, 2011, 29(28): 33 – 36.
- [21] SHENSA M J. The discrete wavelet transform: wedding the a trous and Mallat algorithms[J]. IEEE Transactions on signal processing, 1992, 40(10): 2464 – 2482.

- 
- [22] DAVIS S, MERMELSTEIN P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences[J]. IEEE transactions on acoustics, speech, and signal processing, 1980, 28(4): 357–366.
- [23] ZHANG L, WU D, HAN X, et al. Feature extraction of underwater target signal using mel frequency cepstrum coefficients based on acoustic vector sensor[J]. Journal of Sensors, 2016, 2016.
- [24] TONG Y, ZHANG X, GE Y. Classification and recognition of underwater target based on MFCC feature extraction[C] // 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). 2020: 1–4.
- [25] LIU W, WANG Z, LIU X, et al. A survey of deep neural network architectures and their applications[J]. Neurocomputing, 2017, 234: 11–26.
- [26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [27] TAIGMAN Y, YANG M, RANZATO M, et al. Deepface: Closing the gap to human-level performance in face verification[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701–1708.
- [28] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
- [29] WANG Z, YAN W, OATES T. Time series classification from scratch with deep neural networks: A strong baseline[C] // 2017 International joint conference on neural networks (IJCNN). 2017: 1578–1585.
- [30] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C] // International conference on machine learning. 2015: 448–456.
- [31] DOAN V-S, HUYNH-THE T, KIM D-S. Underwater acoustic target classification based on dense convolutional neural network[J]. IEEE Geoscience and Remote Sensing Letters, 2020.

- [32] CAO X, TOGNERI R, ZHANG X, et al. Convolutional neural network with second-order pooling for underwater target classification[J]. *IEEE Sensors Journal*, 2018, 19(8): 3058–3066.
- [33] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. *nature*, 1986, 323(6088): 533–536.
- [34] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[J]. *arXiv preprint arXiv:1409.2329*, 2014.
- [35] ZHANG S, XING S. Intelligent recognition of underwater acoustic target noise by multi-feature fusion[C] // *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*: Vol 1. 2018: 212–215.
- [36] RAJPURKAR P, ZHANG J, LOPYREV K, et al. Squad: 100,000+ questions for machine comprehension of text[J]. *arXiv preprint arXiv:1606.05250*, 2016.
- [37] DAU H A, BAGNALL A, KAMGAR K, et al. The UCR time series archive[J]. *IEEE/CAA Journal of Automatica Sinica*, 2019, 6(6): 1293–1305.
- [38] SHANNON C E. Programming a computer for playing chess[C] // first presented at the National IRE Convention, March 9, 1949, and also in Claude Elwood Shannon Collected Papers. 1993: 637–656.
- [39] YANG H, GU H, YIN J, et al. GAN-based Sample Expansion for Underwater Acoustic Signal[C] // *Journal of Physics: Conference Series*: Vol 1544. 2020: 012104.
- [40] KRAMER M A. Nonlinear principal component analysis using autoassociative neural networks[J]. *AIChE journal*, 1991, 37(2): 233–243.
- [41] KINGMA D P, WELLING M. Auto-encoding variational bayes[J]. *arXiv preprint arXiv:1312.6114*, 2013.
- [42] CHEN C, SHEN X, MA S, et al. A Classification Method of Ship Radiated Noise Based on Simulation Signal of Variational Auto-Encoder[C] // *Journal of Physics: Conference Series*: Vol 2026. 2021: 012039.

- [43] SNYDER D, GARCIA-ROMERO D, POVEY D, et al. Deep Neural Network Embeddings for Text-Independent Speaker Verification[C] // Interspeech. 2017: 999 – 1003.
- [44] WAIBEL A, HANAZAWA T, HINTON G, et al. Phoneme recognition using time-delay neural networks[J]. IEEE transactions on acoustics, speech, and signal processing, 1989, 37(3): 328 – 339.
- [45] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770 – 778.
- [46] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
- [47] HU J, SHEN L, ALBANIE S, et al. Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks[C/OL] // BENGIO S, WALLACH H, LAROCHELLE H, et al. Advances in Neural Information Processing Systems: Vol 31. [S.l.]: Curran Associates, Inc., 2018.  
<https://proceedings.neurips.cc/paper/2018/file/dc363817786ff182b7bc59565d864523-Paper.pdf>.
- [48] WOO S, PARK J, LEE J-Y, et al. CBAM: Convolutional Block Attention Module[C] // FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 3 – 19.
- [49] ZHANG D, YIN J, ZHU X, et al. Network representation learning: A survey[J]. IEEE transactions on Big Data, 2018, 6(1): 3 – 28.
- [50] PASZKE A, GROSS S, CHINTALA S, et al. Automatic differentiation in pytorch[J], 2017.
- [51] MCFEE B, RAFFEL C, LIANG D, et al. librosa: Audio and music signal analysis in python[C] // Proceedings of the 14th python in science conference: Vol 8. 2015: 18 – 25.

- 
- [52] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015 : 3431 – 3440.
- [53] Van der MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008, 9(11).
- [54] 刘峰, 罗再磊, 沈同圣, et al. 时频谱图和数据增强的水声信号深度学习目标识别方法 [J]. 应用声学, 2021, 40(4): 518 – 524.
- [55] PRIYADHARSINI R, SREE SHARMILA T, RAJENDRAN V. A wavelet transform based contrast enhancement method for underwater acoustic images[J]. Multidimensional Systems and Signal Processing, 2018, 29(4) : 1845 – 1859.
- [56] SHARMA S, VARMA T. Graph signal processing based underwater image enhancement techniques[J]. Engineering Science and Technology, an International Journal, 2022, 32 : 101059.
- [57] WEN Q, SUN L, YANG F, et al. Time series data augmentation for deep learning: A survey[J]. arXiv preprint arXiv:2002.12478, 2020.
- [58] GRINBERG M. Flask web development: developing web applications with python[M]. [S.l.] : ” O’Reilly Media, Inc.”, 2018.

# 致 谢

毕业论文工作接近尾声，也意味着研究生的生活即将告一段落。在南京大学三年的读研时光里，我在人工智能学院 RINC 课题组的大家庭里学习到了很多宝贵的知识与经验。

感谢申富饶教授在三年里对我的教诲和指导。申老师治学严谨、平易近人，深受我们大家的爱戴。刚来到南大时，我对许多领域知识、专业技能都不能熟练运用，申老师总是鼓励我不能急于求成，要打好基础，一步一个脚印。有申老师的耐心指导，我的研究工作才能步入正轨。在科研中申老师总是教导我们不要迷信学术权威，多注入自己的思考。在生活中，申老师经常和我们谈心，关心大家的身心健康，帮我们解决学习生活上的问题。

同时还要感谢赵健老师。赵老师定期为我们做组会报告，为我们介绍科研工作和论文写作的经验，给我们的科研提供了极大的帮助。

感谢 RINC 课题组的同学们，他们在学习和生活中都为我提供了许多帮助，感谢在自己迷茫时愿意和我谈心的好兄弟们，感谢在找工作时耐心指导我的师兄师姐们。感谢我的女朋友，平时督促我运动，和我一起学习，你的陪伴让我变得更加优秀。

最后我要感谢我的父母，他们数十多年含辛茹苦的付出才能有我的今天。如今即将走向社会，我要通过自己的劳动来回报他们的付出，成为一个可以让他们依靠的顶梁柱。

在南京大学的三年里我遇到了许多小伙伴，他们中有的人专注于科学技术事业，有的人立志于从政，也有的人身怀才艺爱好表演。每个人兴趣、特长不同，但都心怀自己的理想。毕业之后，我会时刻铭记自己南大人的身份，努力做人，认真做事，在社会的熔炉中继续成长，怀揣南大的精神一路前行。



# 简历与科研成果

## 基本信息

刘恒，男，汉族，1997年11月出生，河南省南阳人。

## 教育背景

2019年9月 — 2022年6月 南京大学人工智能学院 硕士

2015年9月 — 2019年6月 西北工业大学电子信息学院 本科

## 攻读硕士学位期间完成的学术成果

1. 葛轶洲, 刘恒, 王言, 徐百乐, 周青, 申富饶. 小样本困境下的深度学习图像识别综述. 软件学报. 2021 Apr 21;33(1):193-210.

## 攻读硕士学位期间参与的科研课题

1. 国家自然科学基金“基于深度感知增量式联想记忆神经网络的信息融合系统研究”（课题年限2019年1月 — 2022年12月），负责时间序列识别相关问题的研究。



# 《学位论文出版授权书》

本人完全同意《中国优秀博硕士学位论文全文数据库出版章程》(以下简称“章程”),愿意将本人的学位论文提交“中国学术期刊(光盘版)电子杂志社”在《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》中全文发表。《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》可以以电子、网络及其他数字媒体形式公开出版,并同意编入《中国知识资源总库》,在《中国博硕士学位论文评价数据库》中使用和在互联网上传播,同意按“章程”规定享受相关权益。

作者签名: 刘恒  
2022年 5月 22日

论文题名	基于机器学习的水下信号识别研究				
研究生学号	MG1937016	所在院系	人工智能学院	学位年度	2022
论文级别	<input checked="" type="checkbox"/> 学术学位硕士 <input type="checkbox"/> 学术学位博士		<input type="checkbox"/> 专业学位硕士 <input type="checkbox"/> 专业学位博士 (请在方框内画钩)		
作者 Email	mg1937016@smail.nju.edu.cn				
导师姓名	申富饶 教授				

论文涉密情况:

不保密

保密, 保密期(\_\_\_\_年\_\_\_\_月\_\_\_\_日至\_\_\_\_年\_\_\_\_月\_\_\_\_日)

注: 请将该授权书填写后装订在学位论文最后一页(南大封面)。

