**学校编码：10384 分类号 密级**

**学号：22320221151440 UDC**



**硕 士 学 位 论 文**

**基于深度学习自监督策略的水下工程声信号降噪**

**X**

**杨承昊**

**指导教师姓名：陶毅 助理教授**

**专 业 名 称 ：海 洋 物 理**

**论文提交日期：2025 年 4 月**

**论文答辩时间：2025 年 5 月**

**学位授予日期：2025 年 月**

**答辩委员会主席：**

**评 阅 人：**

**2025 年 5 月**

**厦门大学学位论文原创性声明**

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其它个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（ ）课题（组）的研究成果，获得（ ）课题（组）经费或实验室的资助，在（ ）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

本人声明该学位论文不存在剽窃、抄袭等学术不端行为，并愿意承担因学术不端行为所带来的一切后果和法律责任。

声明人 （签名）：

指导教师（签名）：

年 月 日

**厦门大学学位论文著作权使用声明**

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ）1.经厦门大学保密委员会审查核定的涉密学位论文，于　　 年　 月 　日解密，解密后适用上述授权。

（ ）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。涉密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

随着海洋开发与建设规模的不断扩大，水下工程作业产生的声学信号种类和数量日益增加，信号特征呈现愈发复杂的变化趋势。高保真提取和重构水下工程声信号的结构特征是实现声场精确计算与声源精准定位的关键环节。这为高精度声源估计提供了坚实基础，对于提高水下施工监测精度和海洋生态环境评估的可靠性具有重要意义。然而，在实际海洋环境中，水下工程声信号的获取与处理面临多重挑战，包括多径传播效应、复杂背景噪声以及传播介质特性的变化。这些因素会导致信号的结构特征模糊，严重影响后续分析的准确性。传统降噪方法在上述复杂条件下难以同时兼顾信号结构完整性和降噪效果，已无法满足现代水声工程对高精度信号处理的需求。因此，研发针对水下工程声信号特性的精准降噪技术具有重要的科学意义和应用价值。本文围绕水下工程声信号的特性分析与精准降噪技术展开深入研究，主要工作内容如下：

（1）通过长期现场监测获取了大量水下打桩声和桥梁振动声数据，并系统分析了信号的时频特性与降噪难点。分析表明，这类信号在时域上具有明显的短时脉冲和复杂尾波干扰，频谱能量集中于100 Hz 至1 kHz 的中低频段，并伴随大量非结构性冗余成分。传统降噪方法（如维纳滤波、小波变换）难以有效抑制复杂尾波和非结构噪声；同时，由于缺乏干净参考样本，基于监督深度学习的降噪策略也难以直接应用于此类信号。

（2）针对上述难题，提出了一种基于自监督学习的水下声信号降噪方案，利用信号自身的结构特征构建训练目标，摆脱对干净标签数据的依赖。设计了自适应多重聚焦音频降噪网络（Adaptive Multi-Focus Audio Denoising Network，AMFNet），在频谱动态聚焦、多尺度特征融合和复数域结构感知等方面进行了特殊设计，有效克服了现有自监督方法易出现的频谱连续性损失和结构特征弱化问题，显著提高了模型对脉冲主能量段和结构特征的保持与重建能力。即使在无干净样本的条件下，该模型仍能够精准抑制背景噪声干扰，实现高保真的降噪效果。

（3）利用实地采集的水下打桩声、桥梁振动声和海洋背景噪声数据，构建了“累计能量段”打桩信号、原始打桩信号以及未参与训练的桥梁振动信号三类真实数据集。实验结果表明，所提出的 AMFNet 模型在上述三类数据集上均表现出色。尤其是在 0 dB 信噪比条件下，“累计能量段”打桩信号的 SNR、PSNR 和 RMSE 指标分别达到 10.25 dB、36.42 dB 和 0.0014，显著优于传统方法（维纳滤波、小波变换）以及 U-Net、N2N、NerNT、ONT 等典型自监督音频降噪模型。有力验证了 AMFNet 在信号结构重构与噪声抑制方面的鲁棒性与适用性。

**关键词：**水声信号处理；水下打桩；水下桥梁振动；海洋背景噪声；自监督神经网络

**Abstract**

With

目录

[摘要 I](#_Toc196247394)

[目录 IV](#_Toc196247395)

[第一章 引言 1](#_Toc196247396)

[1.1 水下工程声信号降噪背景及意义 1](#_Toc196247397)

[1.2 国内外研究现状 2](#_Toc196247398)

[1.3 本文主要工作与技术路线 5](#_Toc196247399)

[第二章 水下工程声信号特性分析 8](#_Toc196247400)

[2.1 水下工程声信号评估参数 8](#_Toc196247401)

[2.1.1 声压及相关参数 8](#_Toc196247402)

[2.1.2 均方根声压级 8](#_Toc196247403)

[2.1.3 声暴露级 9](#_Toc196247404)

[2.2 海洋背景噪声的来源与特性 9](#_Toc196247405)

[2.2.1 背景噪声的分类 9](#_Toc196247406)

[2.2.2 背景噪声特性分析 10](#_Toc196247407)

[2.3 水下工程声信号采集系统 11](#_Toc196247408)

[2.3.1 软硬件系统架构 11](#_Toc196247409)

[2.3.2 采集环境与布放方案 13](#_Toc196247410)

[2.3.3 数据采集情况 14](#_Toc196247411)

[2.4 打桩工程声信号特性分析 16](#_Toc196247412)

[2.4.1 采集环境与布置 16](#_Toc196247413)

[2.4.2 特性分析与讨论 17](#_Toc196247414)

[2.5 跨海桥梁振动声信号特性分析 19](#_Toc196247415)

[2.5.1 采集环境与布置 19](#_Toc196247416)

[2.5.2 特性分析与讨论 21](#_Toc196247417)

[2.6 本章小结 24](#_Toc196247418)

[第三章 自监督学习方案设计 26](#_Toc196247419)

[3.1 传统降噪方法能力分析 26](#_Toc196247420)

[3.1.1 高斯白噪声场景降噪 26](#_Toc196247421)

[3.1.2 海洋背景噪声场景降噪 27](#_Toc196247422)

[3.2 自监督学习方案动因 29](#_Toc196247423)

[3.3 自监督子采样设计 32](#_Toc196247424)

[3.4 子采样导致的频谱连续性缺失问题 34](#_Toc196247425)

[3.5 本章小结 36](#_Toc196247426)

[第四章 自监督降噪方案设计 38](#_Toc196247427)

[4.1 AMFNet网络与自监督方案 38](#_Toc196247428)

[4.2 网络架构与关键模块 41](#_Toc196247429)

[4.2.1 编码器-解码器结构 42](#_Toc196247430)

[4.2.2 频谱动态聚焦单元（SDFU） 44](#_Toc196247431)

[4.2.3 复数注意力模块（CAM） 46](#_Toc196247432)

[4.2.4 语义重校准单元（SRU） 49](#_Toc196247433)

[4.2.5 复数多尺度跳跃连接单元（CMS-SCU） 54](#_Toc196247434)

[4.3 损失函数设计 56](#_Toc196247435)

[4.4 本章小结 57](#_Toc196247436)

[第五章 降噪实验与结果分析 58](#_Toc196247437)

[5.1 实验设置 58](#_Toc196247438)

[5.1.1 实验设计内容 58](#_Toc196247439)

[5.1.2 数据集构建与来源 59](#_Toc196247440)

[5.1.3 实验环境与预处理流程 63](#_Toc196247441)

[5.1.4 评价指标 64](#_Toc196247442)

[5.2 实验结果 65](#_Toc196247443)

[5.2.1 累计能量段打桩声降噪实验 65](#_Toc196247444)

[5.2.2 原始打桩声信号降噪实验 70](#_Toc196247445)

[5.2.3 非训练类型声信号降噪实验 75](#_Toc196247446)

[5.3 模型有效性与降噪方法对比分析 77](#_Toc196247447)

[5.3.1 海洋背景噪声降噪效果对比实验 78](#_Toc196247448)

[5.3.2 非训练信号降噪效果对比实验 79](#_Toc196247449)

[5.3.3 模块消融与多方法对比分析 81](#_Toc196247450)

[5.4 本章小结 83](#_Toc196247451)

[第六章 总结与展望 85](#_Toc196247452)

[6.1 论文的主要工作与创新点 85](#_Toc196247453)

[6.2 未来工作展望 86](#_Toc196247454)

[参考文献 87](#_Toc196247455)

[攻读硕士学位期间获得的成果 92](#_Toc196247456)

[致谢 93](#_Toc196247457)

[学位论文答辩委员会名单 94](#_Toc196247458)

**Catalog**

[摘要 I](#_Toc196247459)

[目录 IV](#_Toc196247460)

[第一章 引言 1](#_Toc196247461)

[1.1 水下工程声信号降噪背景及意义 1](#_Toc196247462)

[1.2 国内外研究现状 2](#_Toc196247463)

[1.3 本文主要工作与技术路线 5](#_Toc196247464)

[第二章 水下工程声信号特性分析 8](#_Toc196247465)

[2.1 水下工程声信号评估参数 8](#_Toc196247466)

[2.1.1 声压及相关参数 8](#_Toc196247467)

[2.1.2 均方根声压级 8](#_Toc196247468)

[2.1.3 声暴露级 9](#_Toc196247469)

[2.2 海洋背景噪声的来源与特性 9](#_Toc196247470)

[2.2.1 背景噪声的分类 9](#_Toc196247471)

[2.2.2 背景噪声特性分析 10](#_Toc196247472)

[2.3 水下工程声信号采集系统 11](#_Toc196247473)

[2.3.1 软硬件系统架构 11](#_Toc196247474)

[2.3.2 采集环境与布放方案 13](#_Toc196247475)

[2.3.3 数据采集情况 14](#_Toc196247476)

[2.4 打桩工程声信号特性分析 16](#_Toc196247477)

[2.4.1 采集环境与布置 16](#_Toc196247478)

[2.4.2 特性分析与讨论 17](#_Toc196247479)

[2.5 跨海桥梁振动声信号特性分析 19](#_Toc196247480)

[2.5.1 采集环境与布置 19](#_Toc196247481)

[2.5.2 特性分析与讨论 21](#_Toc196247482)

[2.6 本章小结 24](#_Toc196247483)

[第三章 自监督学习方案设计 26](#_Toc196247484)

[3.1 传统降噪方法能力分析 26](#_Toc196247485)

[3.1.1 高斯白噪声场景降噪 26](#_Toc196247486)

[3.1.2 海洋背景噪声场景降噪 27](#_Toc196247487)

[3.2 自监督学习方案动因 29](#_Toc196247488)

[3.3 自监督子采样设计 32](#_Toc196247489)

[3.4 子采样导致的频谱连续性缺失问题 34](#_Toc196247490)

[3.5 本章小结 36](#_Toc196247491)

[第四章 自监督降噪方案设计 38](#_Toc196247492)

[4.1 AMFNet网络与自监督方案 38](#_Toc196247493)

[4.2 网络架构与关键模块 41](#_Toc196247494)

[4.2.1 编码器-解码器结构 42](#_Toc196247495)

[4.2.2 频谱动态聚焦单元（SDFU） 44](#_Toc196247496)

[4.2.3 复数注意力模块（CAM） 46](#_Toc196247497)

[4.2.4 语义重校准单元（SRU） 49](#_Toc196247498)

[4.2.5 复数多尺度跳跃连接单元（CMS-SCU） 54](#_Toc196247499)

[4.3 损失函数设计 56](#_Toc196247500)

[4.4 本章小结 57](#_Toc196247501)

[第五章 降噪实验与结果分析 58](#_Toc196247502)

[5.1 实验设置 58](#_Toc196247503)

[5.1.1 实验设计内容 58](#_Toc196247504)

[5.1.2 数据集构建与来源 59](#_Toc196247505)

[5.1.3 实验环境与预处理流程 63](#_Toc196247506)

[5.1.4 评价指标 64](#_Toc196247507)

[5.2 实验结果 65](#_Toc196247508)

[5.2.1 累计能量段打桩声降噪实验 65](#_Toc196247509)

[5.2.2 原始打桩声信号降噪实验 70](#_Toc196247510)

[5.2.3 非训练类型声信号降噪实验 75](#_Toc196247511)

[5.3 模型有效性与降噪方法对比分析 77](#_Toc196247512)

[5.3.1 海洋背景噪声降噪效果对比实验 78](#_Toc196247513)

[5.3.2 非训练信号降噪效果对比实验 79](#_Toc196247514)

[5.3.3 模块消融与多方法对比分析 81](#_Toc196247515)

[5.4 本章小结 83](#_Toc196247516)

[第六章 总结与展望 85](#_Toc196247517)

[6.1 论文的主要工作与创新点 85](#_Toc196247518)

[6.2 未来工作展望 86](#_Toc196247519)

[参考文献 87](#_Toc196247520)

[攻读硕士学位期间获得的成果 92](#_Toc196247521)

[致谢 93](#_Toc196247522)

[学位论文答辩委员会名单 94](#_Toc196247523)

[摘要 I](#_Toc196247524)

[目录 IV](#_Toc196247525)

[第一章 引言 1](#_Toc196247526)

[1.1 水下工程声信号降噪背景及意义 1](#_Toc196247527)

[1.2 国内外研究现状 2](#_Toc196247528)

[1.3 本文主要工作与技术路线 5](#_Toc196247529)

[第二章 水下工程声信号特性分析 8](#_Toc196247530)

[2.1 水下工程声信号评估参数 8](#_Toc196247531)

[2.1.1 声压及相关参数 8](#_Toc196247532)

[2.1.2 均方根声压级 8](#_Toc196247533)

[2.1.3 声暴露级 9](#_Toc196247534)

[2.2 海洋背景噪声的来源与特性 9](#_Toc196247535)

[2.2.1 背景噪声的分类 9](#_Toc196247536)

[2.2.2 背景噪声特性分析 10](#_Toc196247537)

[2.3 水下工程声信号采集系统 11](#_Toc196247538)

[2.3.1 软硬件系统架构 11](#_Toc196247539)

[2.3.2 采集环境与布放方案 13](#_Toc196247540)

[2.3.3 数据采集情况 14](#_Toc196247541)

[2.4 打桩工程声信号特性分析 16](#_Toc196247542)

[2.4.1 采集环境与布置 16](#_Toc196247543)

[2.4.2 特性分析与讨论 17](#_Toc196247544)

[2.5 跨海桥梁振动声信号特性分析 19](#_Toc196247545)

[2.5.1 采集环境与布置 19](#_Toc196247546)

[2.5.2 特性分析与讨论 21](#_Toc196247547)

[2.6 本章小结 24](#_Toc196247548)

[第三章 自监督学习方案设计 26](#_Toc196247549)

[3.1 传统降噪方法能力分析 26](#_Toc196247550)

[3.1.1 高斯白噪声场景降噪 26](#_Toc196247551)

[3.1.2 海洋背景噪声场景降噪 27](#_Toc196247552)

[3.2 自监督学习方案动因 29](#_Toc196247553)

[3.3 自监督子采样设计 32](#_Toc196247554)

[3.4 子采样导致的频谱连续性缺失问题 34](#_Toc196247555)

[3.5 本章小结 36](#_Toc196247556)

[第四章 自监督降噪方案设计 38](#_Toc196247557)

[4.1 AMFNet网络与自监督方案 38](#_Toc196247558)

[4.2 网络架构与关键模块 41](#_Toc196247559)

[4.2.1 编码器-解码器结构 42](#_Toc196247560)

[4.2.2 频谱动态聚焦单元（SDFU） 44](#_Toc196247561)

[4.2.3 复数注意力模块（CAM） 46](#_Toc196247562)

[4.2.4 语义重校准单元（SRU） 49](#_Toc196247563)

[4.2.5 复数多尺度跳跃连接单元（CMS-SCU） 54](#_Toc196247564)

[4.3 损失函数设计 56](#_Toc196247565)

[4.4 本章小结 57](#_Toc196247566)

[第五章 降噪实验与结果分析 58](#_Toc196247567)

[5.1 实验设置 58](#_Toc196247568)

[5.1.1 实验设计内容 58](#_Toc196247569)

[5.1.2 数据集构建与来源 59](#_Toc196247570)

[5.1.3 实验环境与预处理流程 63](#_Toc196247571)

[5.1.4 评价指标 64](#_Toc196247572)

[5.2 实验结果 65](#_Toc196247573)

[5.2.1 累计能量段打桩声降噪实验 65](#_Toc196247574)

[5.2.2 原始打桩声信号降噪实验 70](#_Toc196247575)

[5.2.3 非训练类型声信号降噪实验 75](#_Toc196247576)

[5.3 模型有效性与降噪方法对比分析 77](#_Toc196247577)

[5.3.1 海洋背景噪声降噪效果对比实验 78](#_Toc196247578)

[5.3.2 非训练信号降噪效果对比实验 79](#_Toc196247579)

[5.3.3 模块消融与多方法对比分析 81](#_Toc196247580)

[5.4 本章小结 83](#_Toc196247581)

[第六章 总结与展望 85](#_Toc196247582)

[6.1 论文的主要工作与创新点 85](#_Toc196247583)

[6.2 未来工作展望 86](#_Toc196247584)

[参考文献 87](#_Toc196247585)

[攻读硕士学位期间获得的成果 92](#_Toc196247586)

[致谢 93](#_Toc196247587)

[学位论文答辩委员会名单 94](#_Toc196247588)

第一章 引言

1.1 水下工程声信号降噪背景及意义

近年来，随着近海基础设施建设的快速推进[1][2]，水下打桩、桥梁施工、隧道掘进等工程活动在海洋环境中频繁开展。这些作业过程中产生的脉冲类水下声信号已成为海洋声学监测、水下施工状态评估与生态影响研究的重要对象[3]。由于该类信号能量密度高、传播距离远且结构特征明显，它们不仅反映了工程作业的动态过程，也对水下声场结构和海洋生态系统产生显著影响[4]。

在现代水声工程应用中，准确提取水下工程声信号的结构特征并实现精确的声源估计，是声场计算与声源定位的关键基础。这一能力对于提高远场声传播仿真精度、推进水下目标监测与生态风险评估的科学化管理具有重要意义[5]。

|  |
| --- |
| 图1-1水下工程声事件场景图 |

然而，在实际应用场景中，复杂的海洋环境为信号的获取与处理带来了诸多挑战。首先，水声传播过程具有强非平稳性，常伴随多径效应、边界反射和海底回波等现象，导致接收到的信号在时域上出现尾波拖尾、在频域内发生能量扩散，原始结构特征严重畸变[6][7]。尤其在远距离接收条件下，信号主体结构极易被延迟尾波所淹没，影响事件识别与声能评估的稳定性。其次，海洋背景噪声来源广泛，包括风浪、生物声源和船舶航行等多种自然与人为干扰，其频谱分布宽泛且叠加复杂，进一步加剧了信号与噪声分离的难度[8]。

尽管维纳滤波、小波变换等传统方法在理想条件下可实现一定程度的降噪，但面对工程脉冲信号时却往往导致信号结构压缩、频谱扭曲和尾波残留等问题。这些问题归根结底源于传统算法缺乏对水下结构性信号的特征感知，难以聚焦于脉冲主能量段并有效抑制非结构噪声干扰[9][10][11]。

近年来，深度学习技术在语音增强与音频建模领域取得了显著进展，为复杂噪声背景下的信号处理提供了新思路。然而，监督式深度学习方法高度依赖高质量纯净样本作为训练标签，而在实际海洋环境中获取无干扰的水下工程信号几乎不可能，导致此类模型难以直接应用于水下降噪任务[12]。此外，水下声信号在时频结构上显著区别于人声语音等常见训练样本，现有模型难以有效泛化至这类特殊信号[13]。

为突破数据依赖和结构建模的瓶颈，自监督学习策略被引入水下声信号降噪领域。该类方法通过设计伪造的训练任务，从无标签数据中自驱动学习有意义的特征表示，已经在图像复原和语音增强等任务中取得了初步成果[14] 。自监督方法的优势在于具备良好的迁移能力和工程适应性，特别适合处理真实场景中的大规模无标签数据。然而，现有自监督降噪方法主要针对结构相对平稳的信号[15]，对于具有瞬态脉冲特征和显著尾波混叠的水下工程声信号，建模能力仍显不足，在频谱连续性保持和主结构对齐方面存在明显局限[16]。

综上，面向无标签水下声学数据场景构建具备结构保持能力和频谱一致性约束的自监督降噪模型，不仅有助于提升水下工程声信号的结构还原精度，也为复杂海洋环境下的水声信息处理提供关键技术支撑和理论指导。例如，在跨海桥梁施工与运营过程中，结构振动产生的中低频水下工程噪声信号是开展声场计算与远程声源估计的重要依据，可用于辅助判断作业状态、分析能量传播路径并为生态影响评估提供参考。然而，该类信号常被复杂背景噪声和尾波效应所掩盖，传统方法难以恢复其完整的结构和能量特征。本研究提出的自监督降噪框架能够在无标签条件下对结构性脉冲信号进行高保真重构，显著增强信号的可识别性与可建模性，从而为跨海桥梁等典型水下工程场景下的精确声场建模、工程状态监测和生态影响评估提供关键支持[17]。

1.2 国内外研究现状

近年来，随着水下声学在海洋工程、海洋安全和生态保护等领域的持续拓展，水下声信号处理技术逐渐发展成为声学、人工智能与信息科学深度交叉融合的研究热点[18][19]。然而，复杂多变的海洋环境（如不确定的传播路径、水体运动引起的非平稳扰动以及多种环境噪声源的叠加）会显著干扰信号的稳定接收与结构提取，掩盖目标声源的关键特征[20][21]。

特别是在水下工程作业日益密集的情况下，背景噪声与尾波混叠问题愈发突出，严重限制了工程声信号的高保真提取与结构重建。如何在复杂声环境中保留信号的主体结构并提升远场建模的准确性，已成为声场计算与声源定位等关键环节亟待突破的技术瓶颈[5][8]。

在早期水声信号处理研究中，降噪主要依赖频域滤波和时频重构等传统技术。维纳滤波作为经典的去噪方法，通过构建信号和噪声的功率谱估计模型，在最小均方误差准则下实现最优滤波[22]。但该方法对噪声统计特性稳定性的依赖较强，在海洋中多源、非平稳背景干扰下效果会显著下降。当存在频谱重叠且目标信号能量微弱时，维纳滤波易造成信号主频成分削弱和细节丢失，难以保证结构性脉冲声信号的准确还原[23]。小波变换虽然通过多尺度分解能够提取局部瞬态特征，但其预设的固定频带划分难以自适应剧烈变化的海洋噪声[24][25]。因此，在尾波混响和非结构冗余干扰普遍存在的情况下，传统方法难以同时保留信号主结构并抑制无效噪声成分，面临明显的性能瓶颈[26][27]。

随着深度学习技术的发展，卷积神经网络（CNN）[28][29][30]、循环神经网络（RNN）[31][32]和Transformer[33][34]模型在语音增强与音频降噪任务中展现出卓越的特征提取和泛化能力。在水下声信号处理中，一些研究尝试引入这些深度模型用于信号分离、目标检测与波形重建，并采用监督学习策略通过带标签的数据训练模型，以学习噪声与信号的映射关系，从而取得了良好的降噪效果。例如，Zhou等人[35]的噪声感知深度学习模型利用噪声先验实现了复杂水下噪声的自适应抑制；Zhou等人[36]设计的DBSA-net利用双分支自注意力结构从不同尺度提取特征来有效提升水下声信号的降噪性能。此外，Song等人[37]基提出的双路径循环神经网络充分利用局部与全局信息，实现了对动态噪声的精准建模；同时，Zhou等人[38]将注意力机制与多尺度卷积相结合，在复杂水下环境中显著改善了信号的降噪效果。

然而，上述深度学习方法普遍依赖大量高质量的有标签样本，尤其需要目标信号与背景噪声一一对应的“纯净-混合”成对数据。在语音等领域，此类训练集通常可通过将纯净语音与标准噪声叠加合成的方式获得；但在水下环境中，由于干扰源多样、噪声不可控且传播机制复杂，获取不含背景干扰的水下工程信号几乎不可能，导致有标注的数据极为匮乏且获取成本高昂。

现有公开的水声数据资源十分有限且类型单一，多集中于舰船噪声、生物声或水下环境噪声等领域。例如，ShipsEar数据集[39]提供了多类型舰船的水下辐射噪声，用于船舶目标识别研究；DeepShip数据集[40]汇聚了大规模舰船噪声样本，用于舰船目标的分类与识别；Watkins海洋哺乳动物声库[41]收录了多种典型海洋哺乳动物的声信号，支持生物声学与生态监测；国内千岛Ear22数据集[42]则以多目标船舶噪声为主，支持噪声源检测与定位任务。主要包含多目标船舶噪声，用于噪声源检测与定位。然而，这些数据集中收录的信号普遍具有连续、周期或缓慢变化的特性，难以覆盖水下打桩声、桥梁振动声等工程脉冲信号突发、非周期、短时高能的特点。工程脉冲信号通常呈现尖峰—尾波的时域结构，频谱主要分布在中低频，并伴随多径反射、大量非结构冗余成分以及严重的尾波混叠现象[43][44][45]。在实际采集过程中，此类信号的结构特征易被回波、边界反射及远场干扰所扭曲或淹没，高质量标注样本几乎无法获取。因此，当前公开的数据集难以支撑针对水下工程脉冲信号的有效模型训练与降噪算法评估。

自监督学习作为摆脱人工标注依赖的一种有效策略，近年来在图像和语音处理领域展现出巨大潜力[46][47]，其核心思想是利用数据内部的关联构造伪监督任务，引导模型在无标签条件下学习具有语义或结构意义的特征表示[48]。对于能量集中且尾波混叠显著的水下脉冲信号来说，其独特的时频结构为自监督目标的构建提供了天然基础。已有的自监督音频降噪方法（如NerNT[49]、Noise2Noise(N2N)[50]、ONT[51]）虽然在无干净参考信号的条件下实现了一定的降噪效果，但这些方法主要针对语音信号设计，更注重语音语义清晰度或听觉感受的提升。将其应用于无语义、非周期的结构性工程脉冲信号时，降噪有效性和稳定性仍有待验证。同时，在应对多径尾波干扰、非结构噪声混叠以及频谱局部缺失等复杂水声现象时，上述方法的结构感知、频带恢复和主脉冲识别能力均难以满足工程实际需求。

综上，目前水下工程声信号降噪任务主要面临三大技术瓶颈：第一，复杂海洋环境中的背景噪声干扰强烈，传统方法在非平稳条件下的降噪性能严重下降，难以满足工程应用需求；第二，受多径传播、边界反射和海底回波等影响，目标信号往往伴随拖尾和能量分布不均，主结构特征在接收端难以完整保留，极大限制了降噪算法对工程信号的精确重构；第三，水下工程声信号的数据集建设仍处于起步阶段，高质量标注样本的缺乏严重制约了监督学习方法的泛化能力和工程部署效果。因此，亟需探索具备结构感知与自适应能力的新型建模途径，为水下工程声信号的高保真降噪与特征重建提供理论支撑和关键技术。

1.3 本文主要工作与技术路线

针对水下工程声信号在复杂背景噪声干扰下结构信息易被掩盖、主能量段难以准确恢复的问题，本文旨在提升信号的时频结构保持能力和复杂环境适应性，设计并实现了一种基于自监督学习的水下工程声信号降噪方法。研究工作从声信号特性分析入手，构建了涵盖理论分析、模型设计、系统实现与实验验证的完整技术路线（如图1-2所示）。本文的主要工作如下：

第一章阐述了本研究的背景与意义。随着海洋开发活动的频繁开展，例如海底打桩产生的大量短时强冲击声对环境保护和工程监测提出了挑战，水下脉冲声信号的精准获取成为支撑生态评估与结构健康监测的重要环节。在此基础上，梳理了国内外相关研究进展，明确当前方法在非平稳背景噪声干扰下存在结构还原不足、工程泛化性较弱等问题，进而提出以结构保持为核心的水下降噪思路。

第二章围绕了“指标体系—背景噪声—信号采集—特性分析”一体化闭环。首先建立峰值声压级、均方根声压级与声暴露级三维量化指标体系，梳理海洋自然噪声与人为噪声的谱域分布特点。随后，构建水下工程声信号采集系统，保障后续实验所需的高质量数据。在此基础上，对打桩声与桥梁振动声的时域‑频域结构特性进行了对比解析，为下一章评估经典降噪方法并引入自监督策略奠定了理论与数据基础。

第三章设计了自监督学习方案，以克服传统降噪方法与监督式深度学习方法的局限性。首先，以高斯白噪声与实测海洋背景噪声两类典型场景为基准，评估传统算法（维纳滤波、小波变化）在水下工程脉冲信号上的降噪效果。继而提出以自监督学习替代监督式训练的必要性，阐明其在无纯净标签条件下的可行路径。随后，提出含噪数据的伪目标构造方法，自动生成训练对，完成了自监督方案的设计。最后，剖析了子采样引发的频谱连续性缺失问题，并明确补偿频谱缺失对提升模型降噪性能的关键作用。上述工作为下一章深入阐述AMFNet的网络结构奠定了技术动机。

第四章完成了自监督深度降噪网络AMFNet与自监督水下降噪方案设计。在不依赖干净监督样本的前提下，通过伪标签生成机制训练AMFNet。该网络采用U型编码-解码主干结构，融合了频谱动态聚焦、复数域注意力、语义重校准和多尺度跳跃连接等模块，以增强对瞬态结构特征的建模能力。随后，设计了自监督损失函数，为AMFNet的稳定训练和泛化性能提供保障。上述工作为为后续实验验证与方法对比提供了必要的理论与实现基础。

第五章通过三类实验对所提方法进行了综合验证。首先，基于声暴露级理论构建了“累计能量段”打桩信号训练集，用于评估模型在结构约束条件下的信号重构能力。随后，将模型直接应用于未截断的原始打桩信号和未参与训练的桥梁振动信号样本，测试其面对非理想输入时的适应性与稳健性。实验设计了多组噪声叠加条件和信噪比设置，并结合时域波形、频谱图和帧级SNR曲线等可视化手段进行对比分析，同时以SNR、PSNR、RMSE三项指标为核心定量评估模型的降噪性能。进一步，引入传统方法（维纳滤波、小波变换）和经典深度学习模型对降噪效果进行对比，并通过模块消融实验验证了所提模型各结构单元的有效性与整体架构的合理性。

第六章对全文工作进行了总结，展望了未来的研究方向：一是面向多声源复杂干扰环境的复合信号分离研究；二是开发可在实际工程中部署的实时降噪系统，实现“边采集边处理”的智能化平台，推动本方法在海洋环境监测、近岸施工评估和生态保护等领域的落地应用。

|  |
| --- |
|  |
| 图1-2 技术路线图 |

第二章 水下工程声信号特性分析

2.1 水下工程声信号评估参数

水下工程声信号的声学特性通常通过一系列物理量加以表征，包括声压级（Sound Pressure Level, ）、峰值声压级（Peak Sound Pressure Level, ）、均方根声压级（Root Mean Square Sound Pressure Level, ）以及声暴露级（Sound Exposure Level, ）等，对于来自水下打桩、桥梁振动声信号等这类宽带脉冲型水下工程声信号，通常需要重点关注峰值声压级、均方根声压级与声暴露级这三个表征参数[3][4][5]。下面分别对这些关键评价参数进行定义和说明。

2.1.1 声压及相关参数

声波在介质中传播时会引起压强的微小变化，这种随时间波动的压力差称为声压（Pressure, ）。海洋环境中对声压的测量通常以帕斯卡（Pa）或微帕（μPa）为单位。声压的最大瞬时绝对值称为峰值声压（），它反映了声压波动的极值幅度，在强声源情况下通常可达数个数量级。为了便于量化和比较声信号的强度，常采用对数刻度定义的声压级（）进进行表征，其单位为分贝（dB），数学定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (2-1) |

其中， 为参考声压值，在水中通常取1 μPa，在空气中则取20 μPa。如果将峰值声压 入该公式，即可得到峰值声压级（），用于描述声信号的瞬时最大声压。

2.1.2 均方根声压级

在声信号分析中，均方根声压级 用于表征声信号随时间变化的均方根声压大小的对数值，单位为dB，其数学定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (2-2) |

其中， 为声压的均方根值，定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (2-3) |

其中， 和 分别表示所取波形的起始和终止时间点，单位为秒（s）， 为参考声压值，通常取1 μPa（水中）。 的单位为dB。

对于脉冲类声信号，在计算均方根声压级时常需要去除信号两端能量极低的部分，以更准确地反映信号的主要能量贡献。常用的方法是取信号总能量的90 % 进行统计，即剔除波形起始和末尾各约5 % 能量较低的部分。采用这种方式得到的均方根声压级能够更加准确地表征信号的有效声压水平，从而提高不同信号分析结果的稳定性和可比性[3][4][5][8]。

2.1.3 声暴露级

声暴露级（Sound Exposure Level, ）是用于衡量单次或离散水下声事件所包含声能量总和的重要指标。其数学定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (2-5) |

其中， 为参考时间（通常取1秒）， 和 为声暴露事件的起始和终止时刻， 为声压信号， 为参考声压（水下环境通常取1 μPa）。为了更准确地体现信号的实际能量分布，脉冲声的声暴露级计算常只考虑累计能量的主要持续时间段（例如90 % 累计能量所对应的持续时间）。通过这种方式，将不同时长的信号等效为相同累计能量的参考时间段信号，能够更客观地评估水下声事件对海洋环境可能产生的影响。在针对瞬态或间歇性高强度声源（如海上打桩作业等）的研究中，这一指标具有重要意义。

2.2 海洋背景噪声的来源与特性

2.2.1 背景噪声的分类

海洋背景噪声在水下探测与信号处理等应用中普遍存在，其复杂特性给目标信号的识别与提取带来了严峻挑战[52]。从噪声产生机制角度，可将海洋背景噪声分为自然背景噪声和人为背景噪声两大类。

自然背景噪声由海洋环境本身的物理现象和生物活动引起，其中风浪噪声尤为常见。风作用在海面产生波浪运动，并伴随大量气泡的形成和破裂；随着风速的增加，这类噪声能量显著提升，呈现连续且宽频带的分布特征[53]。此外，海洋生物的发声活动（如鲸类、鱼类、海豚及甲壳类动物等）也会在不同频段产生明显的噪声信号——这些生物声既是海洋生态系统内的重要交流方式，也使得水下声学环境更加复杂[54]。与此同时，海洋湍流、水体流动以及海底地质活动（如地震）等因素同样会产生背景声学干扰，进一步加剧了海洋噪声环境的多变性[52]。

人为背景噪声则主要来源于船舶航行产生的噪声，包括船体机械振动、发动机运转以及螺旋桨空化等，具有明显的周期性和结构化特征，对海洋声环境及水下探测系统造成持续且深远的影响[55]。总体而言，自然噪声和人为噪声在形成机制、空间分布及对环境因素的敏感度等方面存在显著差异：自然噪声广泛且相对均匀地充斥海洋，各种噪声源受气象、水文和季节等条件影响而呈现动态变化；而人为噪声则更具局域性，强度与人类活动的分布紧密相关，通常表现出更为明确的结构性和时空分布特征[56]。

2.2.2 背景噪声特性分析

如图2-1所示，海洋背景噪声在不同频率范围内表现出显著的频谱差异。这既与噪声来源的多样性密切相关，也深刻影响着水下声信号的传播机制和降噪策略的设计。通常根据频率范围可将海洋背景噪声划分为低频（<100 Hz）、中频（100 Hz至1 kHz）和高频（>1 kHz）三个典型频段，每个频段内的主导噪声来源及其与环境因素的耦合方式各不相同。

|  |
| --- |
| 图2-1 厦门某海域海洋背景噪声功率谱密度图 |

在低频段，船舶噪声和海洋水体运动是主要成因，尤以螺旋桨空化现象产生的噪声最为突出，典型主频约为50 Hz[57]。低频噪声具有传播距离远、衰减速率慢等特点，因此即使在远离人类活动区域的海洋环境中，仍能检测到明显的船舶航行噪声。Cui等[58]对曾母盆地的长期实测数据研究表明，在低于200 Hz 的频段内，船舶噪声始终占据主导地位；相比之下，风浪噪声在此频域内的能量贡献明显偏低。

中频段（100 Hz至1 kHz）中，风浪噪声与近距离的船舶噪声占据主导。于浅海环境中，海面风力形成的波浪和气泡运动会增强该频段内的风浪噪声水平。研究发现，随着频率的升高，船舶噪声的影响逐渐减弱，而风浪噪声显著提升，特别是在400 Hz 以上频率时，风浪噪声几乎成为该频段内能量的主要贡献源[58]。

高频段（>1 kHz）的背景噪声则主要来自降雨、生物活动以及水体湍流等因素。其中，降雨噪声既具备频谱上的集中性，也呈现出明显的瞬态波动特征。Meng等人[59]在浅海环境中对10至35 kHz 范围的高频背景噪声进行了长时间观测，发现其能量波动特征与环境状态高度相关——例如降雨强度和风速的变化都会引起高频噪声能量的显著波动。此外，即便是在植被覆盖情况各异的浅水海域，高频背景噪声的整体频谱变化趋势仍表现出相对一致的规律，这进一步表明高频噪声在各类浅水生态系统中具有较强的普适性[60]。

2.3 水下工程声信号采集系统

2.3.1 软硬件系统架构

研究构建的水下声信号采集系统主要由水听器、数据采集卡及笔记本电脑等组成，用于高精度获取和分析水下工程声信号。系统的核心部件包括：水听器用于声信号接收，数据采集卡用于模拟信号的模数转换，笔记本电脑用于数据记录和处理。

本系统选用丹麦Brüel & Kjær（B&K）公司生产的8105型水听器，内置球形压电陶瓷敏感元件，具有优良的全指向性。其频率响应范围为0.1 Hz 至160 kHz，涵盖了绝大多数水下工程声源的频段，且最大工作声压可达263 dB（9.8×10⁶ Pa），各项技术参数均满足采集工作需求，具体规格详见表2-1。

表2-1 B&K 8105型水听器主要参数

|  |  |
| --- | --- |
| 项目 | 参数 |
| 水听器型号 | 8105 |
| 标称电压灵敏度 | 56 μV/Pa |
| 电容 | 7250 pF |
| 频率响应 | 0.1 Hz - 160 kHz: +3.5/-10.0 dB |
| 垂直方向性 | 0° - 270° @ 80 kHz ±2.5 dB @ 100 kHz |
| 灵敏度随温度变化值 | 0 to -0.03 dB/°C |
| 灵敏度随静压力变化值 | 0 to -3×10⁻³ dB/Pa (0 to -0.03 dB/atm.) |
| 最大静水压力 | 9.8 MPa (1000 m 水深) |

为确保水下声信号的高精度采集与分析，本系统采用美国国家仪器（NI）公司生产的NI USB-4431数据采集卡，专为高精度声音与振动测量设计，采样率范围为1 S/s 至102.4 kS/s，动态范围达100 dB，确保高分辨率信号获取。同时，其模拟输出通道可与输入通道同步，更新速率达96 kS/s，动态范围为89 dB。该设备的高性能保障了水下声信号采集的准确性与稳定性，为后续数据分析提供可靠支持。采集程序使用NI公司推出的LabView进行编写，实现数据的自动采集并实时观察采集信号波形，如图2-2所示。

|  |
| --- |
| 图2-2 水下工程声信号采集系统LabView控制面板 |

2.3.2 采集环境与布放方案

如图2-3所示，为多维度分析水下工程声信号的传播特性，采集工作依托调查船（或海上作业平台）作为载体，在水体中垂直布设了多层（水深）水听器，并通过高精度数据采集系统对水下声信号进行实时记录。图2-3（a）给出了数据采集系统的组成示意，图2-3（b）则显示了现场水听器的布放方式。

|  |  |
| --- | --- |
| (a) | (b) |
| 图2-3 采集水听器布放示意图：（a）数据采集系统；（b）现场水听器布放 | |

在实际数据采集过程中，调查船（或平台）搭载采集设备，负责水听器的布放与回收。为避免船只自身噪声对数据的干扰，采集期间关闭了作业船的机械设备并抛锚定位，以尽可能消除船舶动力系统产生的噪声干扰，使测量环境接近“静音”状态。水听器采用分层布设的方案，即在不同水深位置分别安放表层水听器、中层水听器和底层水听器，以同步获取不同深度处的水下声信号。所有水听器均通过电缆连接至船上的数据采集系统，实现信号的实时传输与记录。另外，如图2-3（b）所示，为保证水听器阵列在水中的垂直稳定性，在阵列底端悬挂铅坠以抗衡水流扰动，确保水听器始终保持稳定姿态，避免因晃动导致的数据失真。在信号采集的同时，还详细记录了测量水域的水深、海况等环境参数，以辅助后续分析。

2.3.3 数据采集情况

图2-4展示了水下工程声信号现场采集的场景。在打桩声信号的采集过程中，通过采取一系列措施来保证数据质量和信噪比。首先，作业船在采集期间关闭动力装置并锚泊定位，从声源环境上减少了船舶噪声的干扰。在打桩平台上，水听器布放于距打桩锤约50 m处，以获取高信噪比、结构清晰的打桩声信号，同时有效避开机械振动干扰和信号传输畸变。对于桥梁振动声信号的采集，则选择在背景安静的时段进行，并确保在背景噪声测量期间无其他船舶经过，最大程度降低了外界人为噪声对测量的影响。

为保证数据的代表性和可靠性，两类声信号均进行了多次现场采集。例如，打桩声信号分别于2024年不同日期的5次实地采集，采集时的水深、潮汐状态和海况等环境参数基本相似（均为涨潮、1级海况），具体时间及环境信息见表2-2。与此同时，跨海桥梁振动声信号的采集选在厦门集美大桥的运营期内进行，记录了桥梁在交通荷载下产生的水下振动声，其相关采集时间和环境参数见表2-3。现场采集条件的一致性确保了所获取数据的可比性和有效性。

此外，系统本身的高性能为数据质量提供了保障。B&K 8105水听器的宽频带、高灵敏度和高耐受声压能力，确保了对强瞬态声信号的无失真捕获；NI USB-4431数据采集卡100 dB的动态范围，则保证了从微弱背景噪声到高强度脉冲信号的细节均被准确记录。综上，得益于严谨的实验设计和高规格的软硬件配置，本系统所采集的水下工程声信号具有高信噪比和良好的稳定性，为后续的特性分析奠定了坚实基础。

|  |  |
| --- | --- |
|  |  |
| f36909a7667eb41215e13e02c3dfe1a  (a) | 7feccc0d0e91b90d81b7dcd3abfbfec  (b) |
| (c) | 55f2e1911dd4ef7ebe19775141ceb28  (d) |
| 图2-4水下工程声信号采集场景：(a) 打桩船；(b) 打桩平台；(c) 跨海桥梁；  (d) 数据采集设备 | |

表2-2 打桩信号采集时间及环境参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 采集日期 | 水深 | 潮汐状态 | 海况 |
| 1 | 2024年5月10日 | 23.5 m | 涨潮 | 1级 |
| 2 | 2024年5月31日 | 23.2 m | 涨潮 | 1级 |
| 3 | 2024年9月2日 | 13.7 m | 涨潮 | 1级 |
| 4 | 2024年10月22日 | 14.4 m | 涨潮 | 1级 |
| 5 | 2024年12月3日 | 4.8 m | 涨潮 | 1级 |

表2-3 水下桥梁振动声信号采集时间及环境参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 采集日期 | 水深 | 潮汐状态 | 海况 |
| 1 | 2022年5月19日 | 9.2 m | 涨潮 | 1级 |
| 2 | 2022年10月26日 | 11.9 m | 涨潮 | 1级 |

2.4 打桩工程声信号特性分析

2.4.1 采集环境与布置

图2-5为厦门某海域的工程打桩现场。在该施工过程中采用直径为2.1 m 至3.0 m 的钢管复合桩。当钢管桩下贯遇到坚硬海底地层无法继续压入时，需借助大型液压冲击锤进行冲击打桩，以继续贯入桩基。冲击打桩过程中产生的强冲击噪声是施工区域主要的水下声信号来源[61]，也是水下打桩声信号的最突出特征声源。打桩作业的瞬时冲击性质使其水下声压级往往超过200 dB re 1 μPa，尤其在打桩初始阶段，峰值声压水平通常比后续打击阶段更为显著[62]。

|  |
| --- |
| 图2-5 厦门某海域打桩现场图 |

如图2-6所示，打桩工程声信号使用固定式与走船式方法进行采集并根据水深分别布设表层、中层和底层的水听器：固定式采集是在打桩平台上选取固定点位采集；走船式采集则通过调查船在打桩过程中移动，并抛锚固定至不同距离处的进行声信号采集。

|  |
| --- |
| 图2-6 打桩工程水下声信号采集布置 |

2.4.2 特性分析与讨论

为全面揭示水下打桩声信号的传播特性，本节选取距离打桩锤约50 m处固定测点采集的声信号，从时域波形结构、时频能量分布特征及频带能量分布规律三方面进行了深入分析。

图2-7所示为该打桩声信号的时域波形。可以看到，信号呈现出典型的瞬态脉冲特征：每次脉冲在极短时间内迅速达到峰值，随后快速衰减。主要脉冲之后伴随明显的尾波振荡现象，表明声波在传播过程中受到水面、海底等界面及周围结构的多次反射，出现了明显的多径传播效应，从而显著延长了信号的衰减时间。

|  |
| --- |
|  |
| 图2-7 打桩平台测得打桩声信号时域图 |

图2-8给出了该打桩声信号的频谱图。可以看出，打桩声在脉冲发生时刻的能量呈宽频带分布；在脉冲初始阶段，频谱中出现了明显的频率扩展现象，能量向高频方向延伸。随后，在较低频率范围内仍保持着时间较长的余振能量，对应于时域中观测到的尾波振荡和多径效应。

|  |
| --- |
|  |
| 图2-8 打桩平台测得打桩声信号频谱图 |

为进一步分析打桩噪声的能量分布特性，我们记录了同一测点在未施工时段的环境背景噪声，并将其作为对照。图2-9展示了打桩声信号与背景噪声的背景噪声的功率谱密度（PSD）与三分之一倍频程频带声压级图。

|  |  |
| --- | --- |
|  | |
| (a) | (b) |
| 图2-9 打桩平台测得打桩声信号能量分布情况：（a）功率谱密度图；（b）三分之一倍频程频带声压级图 | |

如图2-9（a）所示，打桩声信号在整个观测频率范围内（尤其是200 Hz 至1000 Hz 的中频段），打桩声的PSD显著高于背景噪声水平。其中在约400 Hz 处达到峰值，约为136 dB/Hz，体现出打桩噪声明显的中低频主导特性。随着频率的升高，打桩声与背景噪声的差异逐渐减小，说明打桩噪声的高频成分衰减较快。

为量化不同频带内的能量差异，图2-9（b）给出了打桩声与背景噪声的一组三分之一倍频程频带声压级比较结果。可以看出，打桩声的能量主要集中在4 kHz 以下频段，且在各个频带上均明显高于背景噪声。由于海洋背景噪声受水体运动的影响，其能量主要分布于中低频范围，因此在低频段打桩噪声与背景噪声的差异并不显著。而在约400 Hz 附近频带，二者差值达到最大值，约为40 dB；继续提高频率，二者差异趋于稳定，维持在约25 dB 左右。

综上所述，水下打桩声信号呈现出显著的瞬态脉冲特性和多径传播效应：时域上具有明显的尾波振荡，频域能量主要集中在中低频段，在中频范围（100 Hz～1000 Hz）与背景噪声水平存在显著差异。这些特征的深入分析为后续水下工程噪声的降噪方法设计和声信号特征提取提供了重要的理论基础。

2.5 跨海桥梁振动声信号特性分析

2.5.1 采集环境与布置

本研究选取集美大桥作为桥梁振动水下声信号采集的目标。如图2-10所示，桥梁全长约8.4千米，其为多桥墩多跨结构，桥面设置有钢制伸缩结构。当车辆行驶在跨海大桥上时，桥面和主梁产生的振动会通过桥塔传递至水下，从而辐射出低频声信号[68]。当桥面或梁体上的钢制构件（例如伸缩缝、支座等）在车辆荷载作用下发生振动时，这些振动能量沿着主缆、吊索或塔柱等路径向下传递，最终汇集至水下的基础构件，在界面处引起周围水介质的振荡并向外辐射声波[69]。

|  |
| --- |
| 图2-10 集美大桥 |

采用走船式布放的采集方案（图2-11（a）），在桥梁附近水下1 m、4 m以及接近河床处（约7～8 m）的深度分别布放水听器进行同步采集。如图2-11（b）所示，设置了A、B、C三个测点，分别距最近桥墩约32 m、34 m和32 m，其中B点位于桥梁钢结构件正下方。

|  |  |
| --- | --- |
| (a) | (b) |
| 图2-11采集布置与采集点位示意图：(a) 走船式采集布置；(b) 集美大桥桥墩及采集点A、B、C位置 | |

2.5.2 特性分析与讨论

如表3-4所示，各测点在不同水深下测得的桥梁振动水下声信号峰值声压级、均方根声压级和声暴露级如下：测点A的三项指标平均值分别为155.2 dB、120.4 dB和125.5 dB；测点B的平均值分别为158.5 dB、127.5 dB和133.0 dB；测点C的平均值分别为157.7 dB、122.9 dB和126.3 dB。

表2-4 各测点不同深度下三类声压级

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 测点 | 水深(m) | 峰值声压级(dB) | 均方根声压级(dB) | 声暴露级(dB) |
| A | 2 | 156.4 | 116.6 | 119.6 |
| 4 | 160.0 | **122.9** | 126.9 |
| 8 | 149.3 | 121.7 | **129.9** |
| B | 2 | 155.4 | 125.6 | 130.7 |
| 4 | 160.0 | **129.5** | **136.1** |
| 8 | 160.0 | 127.3 | 132.3 |
| C | 2 | 159.9 | 119.9 | 123.3 |
| 4 | 156.4 | 121.2 | 124.5 |
| 7 | 156.7 | **127.7** | **131.1** |

由上述结果可见，测点B在峰值声压级、均方根声压级和声暴露级三个指标上均为最高，测点A则为最低。不同测点之间的最大差值约为：峰值声压级3.3 dB，均方根声压级7.1 dB，声暴露级7.5 dB。此外，测点A和测点B的声压级最大值均出现在水下4 m处，而测点C的最大值出现在水下约7 m处。为深入揭示跨海桥梁振动噪声的传播特性，下文选取测点B（位于桥梁钢结构正下方，距桥墩约34 m）在4 m水深处记录的声信号进行时域和频域的详细分析。

如图2-12给出了桥梁振动产生的水下声信号时域波形。可以看出，该信号在时域上表现出显著的非平稳特性：总体背景噪声水平相对稳定，但在个别时刻出现了突出的短时脉冲峰值。在约58 s处捕捉到数次显著的脉冲事件（红色方框突出显示了1 s时间窗内声压的变化情况），其瞬时声压峰值达到约28 Pa（约149 dB re 1 µPa）。相比之下，非脉冲信号在时域波形中往往难以分辨，其峰值声压不足5 Pa（约134 dB）。特别是在大型车辆通过时，引起的桥墩和地基振动会向水下辐射声波，其中桥墩振动对水下声信号总声级的贡献最大[71]。

|  |
| --- |
| 图2-12 桥梁振动声信号时域图 |

如图2-13所示，该时段声信号的频谱进一步展示了其频域特征。图中红框部分对应于约58 s处多次强脉冲信号，引起了频谱的明显变化。这些脉冲信号在频域上主要集中于1 kHz以下的中低频范围，呈现出典型的短脉冲“柱状”频谱特征。整体来看，环境背景噪声主要分布在中低频段，表现较为平稳，但其中叠加了间歇出现的高能短时脉冲成分。结合表2-4的数据可知，位于桥梁钢结构下方的B测点，其峰值声压级较均方根声压级高出约31 dB，进一步说明该段声信号中包含强烈的瞬态脉冲分量。

|  |
| --- |
|  |
| 图2-13 桥梁振动声信号频谱图 |

为进一步揭示桥梁振动声信号的频谱组成，我们在远离桥梁的海域（距离集美大桥约680 m，离岸最近约784 m，采集过程中无其它船舶经过）采集了一段水下背景噪声信号，作为对比分析用的环境基线。图2-14展示了在4 m 水深处桥梁振动声信号与背景噪声的PSD曲线及三分之一倍频程声压级比较结果。

|  |  |
| --- | --- |
|  | |
| (a) | (b) |
| 图2-14 桥梁振动声信号能量分布情况：（a）功率谱密度图；（b）三分之一倍频程频带声压级图 | |

由图2-14（a）可见，桥梁振动声信号的能量主要集中在1 kHz以下的中低频段。在约100 Hz左右的频率范围内，PSD曲线上出现了显著峰值，对应的声压级超过110 dB。随着频率升高，信号能量迅速衰减：当频率高于100 Hz时，谱级已降至100 dB以下；超过500 Hz后进一步降至90 dB以下；1 kHz以上频率成分普遍低于85 dB，而4 kHz以上频段信号强度已接近背景噪声水平（低于80 dB）。

如图2-14（b）所示，为桥梁振动声信号与背景噪声的三分之一倍频程频带声压级对比结果。在1 kHz 以下的中低频段，两者声压级存在显著差异，尤其在约500 Hz 频带处达到最大差值，约25 dB。这一频谱特征表明，桥梁振动声信号的能量主要集中在中低频范围；随着频率进一步升高，其声压级逐渐趋近于周围海域的背景噪声水平。

桥梁振动声信号以中低频成分为主的特性与其传播机制密切相关。此类声源主要来自车辆荷载作用引发的桥梁结构振动，并通过钢制结构单元与水体的耦合辐射到水下，形成以低频为主的水下声场。尤其对于钢结构桥梁，由于其结构刚性强、传导性好，当车辆高速通过时，桥体结构中容易激发出强烈的低频结构波，并在与水体耦合的过程中转化为水下声波，形成持续或间歇的声辐射。

2.6 本章小结

本章从量化度量、环境噪声机理、数据采集体系以及典型工程声信号结构四个维度，系统阐述了水下工程声学研究的基础问题，旨在为后续降噪算法设计与实验评估奠定统一而可靠的技术基线。

首先，通过介绍了水下工程声信号的重要量化指标，包括峰值声压级 、均方根声压级 与声暴露级 。这些评估参数分别从瞬时能量峰值、时均能量和累积能量的角度对水下工程声信号进行了多维度衡量，为后续的信号分析与实验应用提供了重要的理论依据和工具支撑。

随后，结合海洋自然噪声与人为噪声的成因分析，揭示了低频段船舶噪声、中频段风浪噪声以及高频段降雨‑生物噪声交替占优的谱域规律，这不仅阐明了水下工程声信号易被中低频背景掩蔽的本质原因，也为后续网络模型的频带加权和噪声建模提供了理论依据。

在数据获取层面，本章给出了由B&K 8105宽带水听器与NI USB‑4431高分辨率采集卡构成的水下工程声信号采集系统，为后续的信号特性分析以及深度学习模型的训练提供了高质量数据源。

进一步的声信号特性分析表明，打桩声信号呈“尖峰‑尾波”强脉冲结构且集中于100 Hz–1 kHz 中频段，多径效应突出；桥梁振动声信号则以20 Hz – 1 kHz 中低频为主，表现为间歇脉冲与稳态混合模式。两类信号虽与背景噪声存在频谱分离窗口，但尾波拖长和低频冗余或许会导致传统静态滤波难以兼顾结构完整性与噪声抑制，在下一章进行进一步尝试验证。

综上，本章构建了“指标体系—背景噪声—信号采集—特性分析”的逻辑闭环，系统刻画了水下工程声信号在时‑频‑能量三维度的本质特征。下一章将在此基础上对经典降噪方法的局限性进行评估，并引入自监督深度学习思路，为面向频谱结构保持的水下降噪算法研究奠定理论与实验基础。

第三章 自监督学习方案设计

3.1 传统降噪方法能力分析

为评估传统降噪方法在水下工程声信号降噪任务中的实际表现，本节以实际采集的水下打桩声信号为研究对象（如图3-1所示），该信号主脉冲特征明显、能量集中且尾波结构清晰。在信噪比 SNR = 0 dB 的条件下，分别叠加高斯白噪声和实测海洋背景噪声，构建两种含噪场景，并采用维纳滤波和小波变换两种经典降噪方法对含噪信号进行处理。

|  |
| --- |
| 图3-1 水下打桩声信号（原始信号） |

3.1.1 高斯白噪声场景降噪

图3-2展示了在高斯白噪声干扰下的降噪结果。其中，图3-2（a）为叠加高斯白噪声后的波形，原始信号结构受到显著干扰；图3-2（b）为维纳滤波处理结果。作为基于统计特性的自适应滤波器，维纳滤波能够有效建模平稳噪声，主脉冲区域得到初步恢复，但尾波干扰未能充分抑制，仍存在结构模糊；图3-2（c）为小波变换的降噪结果。通过多尺度分解与阈值去噪策略，小波变换在抑制背景噪声方面表现出一定优势，波形整体趋于平滑，主脉冲结构得以保留，但尾波部分残留依旧明显。

综合上述分析，传统方法在面对高斯白噪声这一典型平稳噪声时，仍具备一定降噪能力，但在结构完整性和尾波恢复方面存在明显不足。

|  |
| --- |
|  |
| (a) |
| (b) |
| (c)  图3-2 混叠高斯白噪声（SNR=0）水下打桩声信号降噪情况：（a）含噪信号；（b）维纳滤波结果；（c）小波变化结果 |

3.1.2 海洋背景噪声场景降噪

图3-3进一步展示了在混叠真实海洋背景噪声条件下的降噪效果。图3-3（a）为信号与实测背景噪声混叠后的波形，噪声呈现出强非平稳性和高度随机性，主脉冲与尾波结构被显著掩盖，识别难度增加。图3-3（b）为维纳滤波处理结果，由于其建模依赖于频域统计特性，面对非平稳背景噪声难以准确匹配噪声模型，降噪效果不理想，信号轮廓模糊；图3-15（c）为小波变换处理结果。尽管小波变换在非平稳信号处理方面具有理论优势，但实际处理中受限于噪声的复杂性与非结构特性，降噪后信号中仍保留大量背景干扰，主结构恢复效果有限。

|  |
| --- |
| (a) |
| (b) |
| (c) |
| 图3-3 混叠真实海洋背景噪声（SNR=0）水下打桩声信号降噪情况：（a）含噪信号；（b）维纳滤波结果；（c）小波变化结果 |

通过对比图3-2和图3-3可以发现，传统降噪方法在不同噪声条件下展现出明显的性能差异。尽管维纳滤波和小波变换在应对高斯白噪声等平稳噪声方面具备一定优势，能够在一定程度上恢复信号主脉冲的能量，但在信号结构完整性和细节保真度方面仍存在明显不足。当面对频谱复杂、结构性弱且高度非平稳的海洋背景噪声时，这两种方法的结构保持能力均显著不足，尾波等细节难以有效重建。具体而言，维纳滤波由于高度依赖噪声频谱的精确估计，当噪声模型与实际环境不匹配时，容易过度抑制有效信号成分；小波变换虽然具有多尺度分析能力，但在复杂尾波结构和多源背景干扰叠加条件下，仍难以稳定提取信号的主结构特征。这些在信号结构重建、尾波抑制以及鲁棒性方面的不足，限制了传统方法在实际水下工程声信号降噪中的应用效果，难以满足高保真重构的任务需求。上述局限性进一步凸显了引入具备结构感知和非线性建模能力的深度学习方法的必要性。

3.2 自监督学习方案动因

近年来，深度学习在语音增强和音频建模领域展现出优异的性能，尤其在特征提取建模能力和非线性表达能力方面明显优于传统算法。然而，主流的监督式学习方法往往依赖于大量准确标注的纯净信号作为训练目标。在真实海洋环境下，获取完全不含背景干扰的目标声信号几乎不可行，纯净样本极度匮乏已成为限制监督学习应用推广的核心瓶颈。尤其是对于突发性强、结构复杂的水下工程脉冲信号而言，样本采集难度高、场景变化显著、数据分布不均等问题更加突出，导致现有监督式降噪模型在泛化能力和工程实用性方面存在明显限制。

基于上述挑战，本文引入无需依赖真实干净标签的自监督降噪策略，为水下工程声信号降噪提供新的解决思路。自监督方法通过构建伪标签或基于信号自身结构关系设计训练任务，使模型能够利用未标注的数据自主学习有效特征表示，具有显著的数据适应性和任务迁移能力。相较于传统方法对噪声统计分布等先验假设的依赖，自监督策略更贴近真实噪声环境下信号的结构规律，特别适用于具有明显主结构特征且尾波扰动复杂的脉冲类信号建模需求。该策略不仅有效缓解了缺乏纯净样本的数据困境，还可以在训练目标中融入对信号结构保持的约束，自适应增强信号的主能量部分并动态抑制冗余干扰成分，从而有望实现更高保真的水下声信号重构。

水下声信号降噪的传统方法通常依赖明确标注的干净信号作为监督信息，以训练深度神经网络实现信号增强或降噪。这类方法被称为“噪声到纯净”策略（Noise-to-Clean, N2C）[72]。在N2C策略中，网络模型通过最小化预测输出 与真实干净信号 之间的均方误差，来学习从含噪信号中提取纯净信号的映射关系：

|  |  |  |
| --- | --- | --- |
|  |  | (3-1) |

其中， 为神经网络对含噪音频输入 的估计输出， 为对应地纯净参考信号。含噪音频信号可表示为 ，而 表示叠加的噪声信号成分。然而，在真实的海洋环境中，大量干净信号数据往往难以获取，甚至在某些情景下无法获得，这导致N2C方法在实际应用中面临瓶颈。为了克服这种限制，近年来，自监督降噪方法逐渐引起研究人员关注。自监督降噪的优势在于仅依赖无标注的含噪音频数据进行训练，使模型在没有干净参考数据的条件下也能学习有效提取干净信号特征[15]。因此，自监督降噪方法在复杂海洋环境中具有广阔的应用前景和重要的现实价值。

针对上述问题，典型的自监督学习策略Noise2Noise（N2N）通过利用同一干净信号的两种不同含噪观测数据进行模型训练，以避免对干净信号数据的直接需求[50]。在该策略中，每对训练数据由同一干净信号与两个相互独立的噪声分量叠加生成，即训练数据对为 和 ，其中 和 表示互不相关的零均值随机噪声变量。

具体而言，N2N方法在训练过程中，以含噪数据 作为输入，另一个含噪数据 网络训练目标，通过最小化以下损失函数来优化网络：

|  |  |  |
| --- | --- | --- |
|  |  | (3-2) |

在N2N条件下，由于 是独立从零均值分布采样，可以得到 的期望值等于 的方差与其期望平方之和。利用该原理，可以对公式（4-2）中的第三项进行展开。样本分布的方差 等于总体方差除以样本大小，因此随着音频训练数据集规模的增大，公式（4-2）中的第二项和第三项趋近于零。可以得出，最终N2N的 损失值接近于N2C的 损失值，即：

|  |  |  |
| --- | --- | --- |
|  |  | (3-3) |

理论上，当训练样本趋于无限大时，N2N策略的性能能够逼近传统监督式N2C策略[50]。然而，在实际有限数据条件下，由于训练数据规模限制，N2N策略的性能仍略逊于N2C策略。此外，N2N策略的实际应用还面临更大的局限性——难以在真实环境下获得两个独立噪声叠加的同源信号对。尤其在复杂多变的水下场景中，环境噪声往往随时间快速变化，很难满足严格独立条件的数据对采集需求。

针对N2N方法在数据配对条件上的局限性，研究人员提出了Neighbor2Neighbor(Nbr2Nbr)[73]策略，通过子采样的方式从单个含噪数据中构造一对近似独立的伪噪声样本，以满足自监督训练的需求。这种方法无需严格的配对噪声信号，极大提升了训练数据构建的灵活性。在Nbr2Nbr基础上，Wu等人[51]将该策略进一步扩展到音频降噪领域，通过子采样构造伪训练目标对，实现了音频领域的高效自监督降噪训练。

具体来说，使用子采样器 从单个含噪音频数据 中生成一对音频训练数据 和 。与N2N方法不同的是，这里抽取得到的两段带噪音频 和 在期望意义上的真实值并不完全相同，因此产生了一个偏差项 ：

|  |  |  |
| --- | --- | --- |
|  |  | (3-4) |

通过引入偏差项的差值 ，其中 是噪声项，并假设 ，可以得到：

|  |  |  |
| --- | --- | --- |
|  |  | (3-5) |

考虑到理想的降噪网络 是用纯净的音频数据训练并采用MSE损失，它可以确保在给定含噪输入 ，对于任意的噪声项 的情况下，对于 ，满足 且 。因此，在理想降噪网络 条件下，以下结论成立：

|  |  |  |
| --- | --- | --- |
|  |  | (3-6) |

换言之，网络在训练过程中被迫忽略随机噪声成分，仅捕捉信号的结构性成分。自监督策略的优势正在于此：无需真实的无噪材料，仅通过单一含噪数据的子采样构造即可实现模型的有效训练。该策略利用了噪声信号中潜在的结构一致性，促使网络仅关注信号内在特征结构，自动忽略随机噪声部分，在实际应用中具有重要价值。

3.3 自监督子采样设计

本研究采用自监督学习方式训练模型，无法直接获取干净的目标信号，因此需要利用噪声信号自身构造伪“干净”目标，以提供有效的训练监督信号。借鉴了Nbr2Nbr策略[73]，在单通道水下噪声数据上应用随机子采样策略，以生成用于自监督训练的数据对。

设原始噪声信号为 ，信号采样点数为dim，可将其视为一维张量（单通道）。为了实现随机子采样，设定子采样步长 ，将信号按连续的两个采样点分组。设分组后信号的组数为dim1，计算如下：

|  |  |  |
| --- | --- | --- |
|  |  | (3-7) |

其中 为边界安全额度，以避免索引越界。为确保随机子采样的均匀性和数据覆盖率，在每一组数据中生成随机二值变量 ，其形状为 ，并定义两个布尔掩码：

|  |  |  |
| --- | --- | --- |
|  |  | (3-8) |

同时，设每组的起始索引序列为：

|  |  |  |
| --- | --- | --- |
|  |  | (3-9) |

则每一组的子采样信号可分别表示为：

|  |  |  |
| --- | --- | --- |
|  |  | (3-10) |
|  |  | (3-11) |

最后，将每组结果重塑为 的矩阵，得到两个子采样信号：

|  |  |  |
| --- | --- | --- |
|  |  | (3-12) |

|  |
| --- |
|  |
| 图3-4 子采样示意图 |

如图3-4所示，这种子采样方法确保在每一组中，每个采样点均被利用，并通过随机交叉分配，使得两次子采样在局部时域上的细节信息具有互补性。此外，由于子采样对整个时间段均匀覆盖，背景噪声的整体能量得以保持一致，从而避免了随机丢弃可能带来的能量偏差问题。该方法采用矢量化操作实现，可高效地在GPU上运行，提升模型训练效率。

接下来，将子采样信号 和 转换为频谱表示，分别得到 和 ，并构成训练对 。在每次训练迭代中， 作为模型输入，经过网络得到估计输出 ，然后计算 与参考目标 之间的损失。

这一训练机制的理论依据在于第3.2节介绍的噪声随机无关性假设。假设背景噪声是零均值的随机过程，则 和 之间的差异主要由噪声的随机波动部分构成，而它们的公共成分即为信号的稳定成分（即非随机水下工程脉冲信号）。因此，通过最小化 和 之间的损失，模型倾向于还原出 中 缺失的部分，这相当于在训练过程中主动补全子采样导致的频谱缺失信息。

3.4 子采样导致的频谱连续性缺失问题

引入子采样器的自监督策略虽然克服了N2N方法对成对训练数据的苛刻要求，但也带来了新的挑战：频谱连续性缺失问题。具体来说，子采样的预处理操作在提升训练数据获取灵活性的同时，破坏了信号频谱图时频结构的自然连续性，表现为时间分辨率和频率分辨率的降低。这种频谱结构的破损将导致降噪模型难以捕获原信号中的关键脉冲特征信息，影响最终降噪效果和音频重建质量。

为理解子采样对频谱的影响，首先回顾音频信号的短时傅里叶变换（Short-Time Fourier Transform, STFT）处理过程。STFT通过选取窗函数将长时域信号分割成多个短时帧，并对每个短帧进行傅里叶变换，从而获得信号在时间和频率两个维度上的联合表示。在STFT中，每个时间帧之间的间隔由连续窗口之间的采样点数 （即帧移长度）决定：

|  |  |  |
| --- | --- | --- |
|  |  | (3-13) |

其中， 为信号的采样频率。当使用子采样技术对信号进行预处理时，其等效采样率降低为原来的一半。此时，相邻STFT帧之间的时间间隔将被拉长为原来的两倍：

|  |  |  |
| --- | --- | --- |
|  |  | (3-14) |

这一变化对频谱图的影响在图3-5有所体现。图3-5a）展示了原始音频信号的STFT频谱图，其中红色方框突出显示了沿时间轴分布的一个柱状频谱结构（例如桥梁振动引起的一系列瞬时脉冲），该红框覆盖了约20个连续的时间帧。而在图3-5b）中，针对同一信号应用子采样处理后得到的频谱图可以看到，对应的柱状频谱结构被“压缩”到仅约10个时间帧内。原本在高时间分辨率下沿时间轴连续分布的能量特征，被挤压到更少的帧中呈现，造成频谱中细节信息与结构连续性的严重破坏。许多原本沿时间连续变化的谱纹理被割裂开来。这样的信息损失尤其影响对于具有短暂脉冲特征信号的识别和学习。例如，第2章分析的水下工程声信号的瞬态宽带脉冲部分，在高时间分辨率的频谱上表现为清晰的短时柱状能量分布；但经过子采样压缩后，这些脉冲在频谱上的表征变得模糊、拉伸，能量分布不再集中，干扰了网络对目标信号完整特征的准确学习，同时也降低了后续频谱重建的质量和准确性。

|  |
| --- |
|  |
| （a）    (b)  图3-5不同频谱场景下卷积核采样点分布情况：（a）原始信号频谱卷积核采样点分布情况；（b）子采样信号频谱卷积核采样点分布情况； |

除了时间分辨率下降，子采样对频率分辨率也有明显影响。在STFT频谱图中，每个垂直点代表某一时间窗内的频率成分，展示该时刻下各频率对应的能量或功率分布。频谱图中频率点之间的间距即为频率分辨率 ，由采样率 和FFT点数 共同决定，计算公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | (3-15) |

由公式（3-9）可知，当FFT点数 保持不变而采样率 减半时，频率分辨率 也降低一半，同时可显示的最高频率（奈奎斯特频率）变为原来的一半。这意味着频谱图的频率轴被压缩，频率范围变窄：原本分布在多个频率点上的谱能量将集中到更少的频点上，相邻频点间距增大，不同频率成分之间的区分度降低，削弱了频谱的细节解析能力。此外，那些原本高于新奈奎斯特频率上限的信号成分可能因欠采样产生的频谱混叠效应而折叠至低频段，进一步破坏频谱图的完整性与准确性。

综上所述，子采样操作虽然成就了自监督训练的数据构造，但伴随而来的频谱结构连续性受损不容忽视。第2章中我们了解到，水下工程引起的声信号在频域上通常表现为低频连续噪声叠加瞬时宽带脉冲的形式，要求模型既关注持续背景噪声又关注瞬态脉冲特征。然而子采样预处理使得频谱的局部纹理和整体结构遭到破坏，时间分辨率和频率分辨率的降低使网络难以有效捕捉瞬时宽带脉冲等关键结构信息。这种频谱连续性缺失削弱了网络的特征提取能力，也降低了降噪后信号的自然性和保真度。因此，如何在自监督框架下弥补因子采样引起的频谱信息损失，成为了提升模型降噪性能的关键问题。

3.5 本章小结

本章构建了“传统方法适用性验证—自监督需求提出—自监督策略设计—频谱连续性问题”的逻辑闭环，为下一章深入阐述 AMFNet 的网络结构与自监督训练细节奠定了坚实的理论基础与技术动机。

首先以两类典型噪声场景——高斯白噪声与实测海洋背景噪声，为实验对象，系统评估了维纳滤波与小波变换在水下工程脉冲信号降噪中的表现。实验结果显示：当噪声呈平稳特性时，两种传统方法虽能在主脉冲区域实现初步能量恢复，但在尾波结构与细节保真方面仍存在明显缺陷；当噪声高度非平稳且成分复杂时，两者对信号结构的保持能力急剧下降，已无法满足工程应用对高保真重构的要求。该结论充分暴露了传统线性与半线性算法在真实海域环境中的适用性边界。

针对上述局限，第3.2节阐述了以自监督学习取代监督式训练的必要性与可行性。自监督策略通过“用数据自身监督数据”的思想，避免了对纯净标签的依赖，可显著提升模型在标签稀缺场景下的泛化鲁棒性，尤其适用于瞬态脉冲与连续尾波混叠的水下工程声信号。

随后，第3.3节基于Neighbor2Neighbor子采样框架提出了伪目标构造方案，实现了在单通道含噪数据上构建训练对的全过程设计。

最后，第3.4节从短时傅里叶变换分辨率变化的角度，系统论证了子采样操作导致的时频纹理割裂与高频混叠等频谱连续性缺失问题，进一步明确了在自监督框架下弥补因子采样引起的频谱信息损失，成为了提升模型降噪性能的关键问题。

综上所述，本章完成了“传统方法适用性验证—自监督需求提出—自监督策略设计—频谱连续性问题”的逻辑闭环，为下一章深入阐述AMFNet的网络结构与自监督训练细节奠定了坚实的理论基础与技术动机。

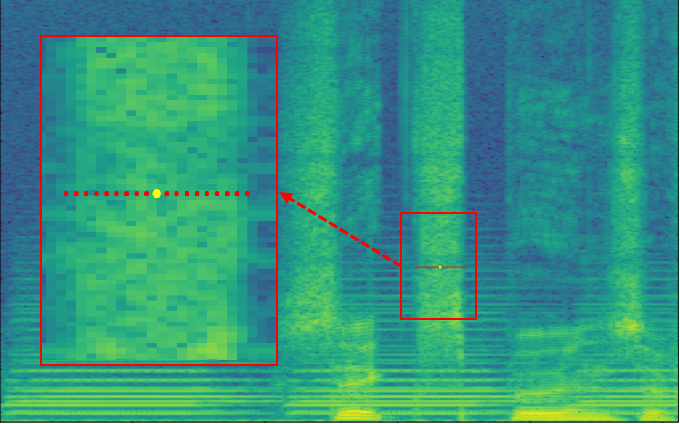
第四章 自监督降噪方案设计

本章的核心目标是提出一种基于自监督深度学习的水下工程声降噪方案，以提升水下工程脉冲信号（如打桩、桥梁振动信号）的降噪 能力。具体内容包括方案的总体思路、模型架构、伪“干净”目标构造策略、损失函数设计等。

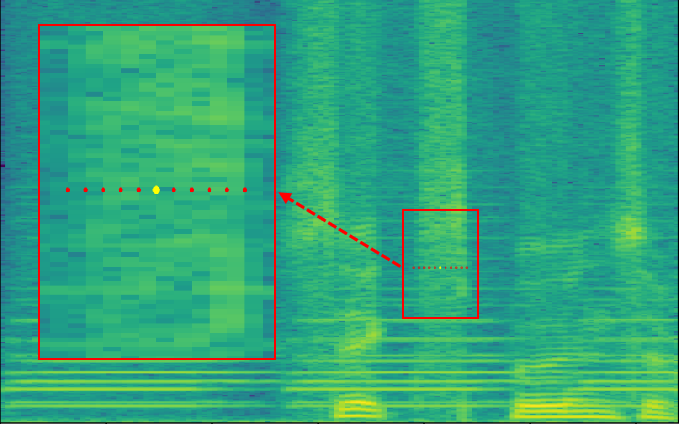
4.1 AMFNet网络与自监督方案

针对第3章总结的自监督策略中频谱信息丢失的不足，本研究提出了一种新型的自适应多重聚焦音频降噪网络（Adaptive Multi-Focus Audio Denoising Network， AMFNet），以弥补子采样造成的频谱连续性缺失，增强网络对短时宽带脉冲结构的识别能力。

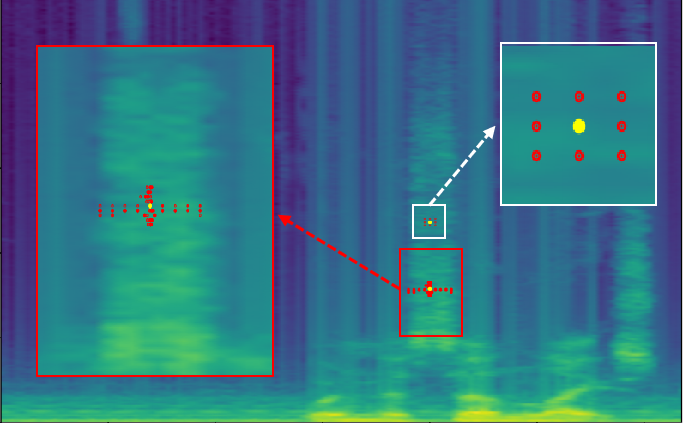
AMFNet从网络设计层面对卷积特征提取过程进行了改进，引入自适应卷积机制使模型能够更有效地聚焦目标信号的时频特征。如图4-1所示，传统卷积在处理子采样后的“压缩”频谱时，卷积核的采样点在频谱平面上仍是固定均匀分布的。图4-1(c)中白色虚线方框示意了这种传统卷积采样：无论频谱结构如何变化，卷积核在横向（时间轴）和纵向（频率轴）都按固定间隔取样，导致无法充分捕捉频谱随时间和频率变化的细节特征。而AMFNet通过卷积核采样位置的自适应调整，实现了对目标频谱特征的“多重聚焦”。图4-1(c)中的红色方框突出展示了AMFNet的作用效果：在横向（时间维），卷积核根据目标信号的频谱结构动态调整采样位置，使其能够准确对齐并抓取沿时间轴排列的脉冲谱柱等结构；在纵向（频率维），AMFNet显著提高卷积核在局部频带内的采样密度，以补偿子采样后频谱压缩所导致的信息空缺。通过这种自适应卷积机制，网络在提取特征时相当于同时具有“广角”关注整体趋势和“聚焦”关注细节的能力，对短时瞬态脉冲和细微谱变化更加敏感。此外，AMFNet在网络的各级卷积层中均配备了专门的多尺度特征提取模块（详见第4.2节），逐层融合不同尺度的谱信息，主动弥补子采样带来的信息损失。借助上述设计，AMFNet能够更有效地识别并提取水下工程声信号中的瞬态脉冲频谱特征，同时抑制海洋环境中的连续背景噪声，显著提升复杂噪声环境下的整体降噪性能。



（a）



(b)



(c)

图4-1 不同频谱场景下卷积核采样点分布情况：（a）原始信号频谱卷积核采样点分布情况；（b）子采样信号频谱卷积核采样点分布情况；（c）AMFNet网络处理频谱卷积核采样点分布情况

在建立了如上具有自适应特征提取能力的网络结构后，我们构建了相应的自监督训练方案流程，如图4-2所示。训练阶段， AMFNet完全基于含噪音频数据进行优化，无需任何干净信号介入。具体来说，首先将原始含噪音频信号 通过子采样器生成一对含噪音频训练数据 。其中，子采样后的信号 经过STFT得到频谱表示，作为训练网络的输入；网络输出降噪后的频谱经过逆短时傅里叶变换（Inverse STFT, ISTFT）得到 与 计算损失，以更新网络权重。

|  |
| --- |
| 图4-2 本研究提出的自监督策略概述 |

在训练过程，网络被迫学习两个子采样片段共同的信号结构，忽略所含噪声的差异部分。这意味着AMFNet掌握桥水下工程声信号典型的脉冲频谱结构，同时将随机的背景噪声视作可忽略的扰动加以抑制。随着训练的进行和损失的持续最小化，模型的参数逐渐收敛，网络对目标信号特征的提取能力不断增强。训练完成后的AMFNet在推理阶段只需将原始含噪音频经过STFT得到的频谱输入已经训练好的模型，网络便会直接输出对应的干净频谱估计 。最后，对 执行ISTFT重构回时域，即可得到降噪后的音频信号 。整个自监督降噪方案在训练和推理全流程中都不需要任何外部干净参考信号，完全依赖含噪数据自身的内部关联来驱动模型学习，充分体现了自监督学习在水下工程声降噪任务中的价值。

4.2 网络架构与关键模块

AMFNet采用编码器-解码器的U型网络整体结构，并融合多种特定设计的模块单元，有针对性地对水下工程声信号的频谱特性进行高效特征提取和重构降噪。

|  |
| --- |
| 图4-3 本研究提出的AMFNet网络架构 |

如图4-3所示，AMFNet的左半部分为编码器，用于逐级压缩并提取输入复数频谱的主要特征；右半部分为解码器，用于逐步重建并输出去噪后的细粒度频谱。在编码器和解码器对应层之间，引入了改进的跨层跳跃连接机制，以替代传统U-Net直接拼接的方式，实现不同尺度特征的深度融合；同时在网络的瓶颈层（编码器与解码器交汇处）设计了全局语义校准单元，以充分整合最高层次的特征信息。通过这些设计，AMFNet能够在无纯净标签监督的情况下，充分捕获水下脉冲信号的瞬态频谱脉冲特征并抑制宽带连续背景噪声，为后续高质量重构提供保障。下面将分别介绍AMFNet整体架构以及各关键模块的设计原理与优势。

4.2.1 编码器-解码器结构

AMFNet总体采用对称的编码器-解码器架构，以典型U-Net结构为基础骨干。整个网络由对称的编码端和解码端组成：编码端逐层提取并压缩输入信号的频谱特征，解码端逐步恢复出目标信号的细节频谱。编码器包括4个逐级串联的编码模块，对应4个不同的特征尺度；解码器同样包含4个逐级堆叠的解码模块，与编码器层次一一对应。通过这种由“粗到细”再“细到精”的分层结构，网络能够逐步从原始复杂频谱中抽取多尺度特征表示，并最终重建出净化后的目标信号频谱。

在编码端，每一级编码器首先接收来自上一层（或直接来自输入）的复数频谱特征图。为突出目标脉冲信号的显著特征并减少下采样导致的关键信息丢失，每个编码器层输入处都引入了一个创新模块——频谱动态聚焦单元（SDFU）。SDFU根据输入频谱的时频局部结构，动态调整卷积核的采样形状与位置，使卷积运算能够自适应地“聚焦”于频谱中能量显著的目标区域（详细机制见4.2.2节）。这一设计对于桥梁振动、打桩等脉冲信号尤为重要，因为此类水下工程声信号的频谱通常呈现不连续的高能量瞬时脉冲（例如在时频谱上表现为短暂出现的纵向高能量柱状结构）。借助SDFU对卷积感受野的动态调整，当频谱中出现这类瞬时突出的能量特征时，网络能够更有效地捕捉细节脉冲信息，避免其在后续下采样过程中被遗漏。

在SDFU预处理之后，每级编码器模块由复数注意力模块（CAM）和下采样操作组成。CAM模块旨在在复数域内高效提取频谱特征，同时建模幅度与相位信息的交互（详细原理见4.2.3节）。与只关注幅度的传统注意力机制不同，CAM对频谱的实部（幅度）和虚部（相位）特征同时赋予关注，以充分利用复数谱图中的信息。这使得网络在特征提取时能够兼顾频谱宏观能量分布和微观相位细节，更有效地区分目标信号与背景噪声。例如，当桥梁振动产生瞬时振动脉冲时，频谱上会同时出现幅度的突增和相位的剧烈变化；CAM可以敏锐捕捉并强调这些复数域特征，从而增强网络对目标脉冲的辨识能力，抑制环境噪声干扰。经过CAM处理后的特征再通过下采样模块进行空间尺寸压缩（例如使用步长为2的卷积或池化），以输入下一尺度的编码器。

编码器中的下采样模块主要实现频谱特征维度的压缩。本研究使用了专门的复数卷积层来完成下采样提取特征。复数卷积的计算可表示为对输入复数向量 与复数滤波器 进行卷积运算，即：

|  |  |  |
| --- | --- | --- |
|  |  | (4-1) |

复数卷积层具体由两个步长为2、大小为3×3的实值卷积核实现，分别处理输入频谱的实部和虚部。在下采样的同时，这种卷积操作高效提取频谱特征并压缩特征空间维度。此外，在复数卷积之后，我们配备了复数批量归一化层（Complex Batch Normalization, CBN）和复数ReLU激活函数（Complex ReLU, CReLU），这些复数域专用操作被证明能更稳定地处理复杂频谱特征[74]，保证网络在复杂水下噪声环境中的训练收敛性和特征表达能力。

与编码端对称，解码端同样由4个逐层对应的解码模块组成，用于逐步恢复编码压缩过程中丢失的频谱细节。每一级解码器接收来自上一层解码器的上采样特征，并融合来自编码端的跨层特征信息后，再经卷积生成更高分辨率的特征图。不同于经典U-Net直接将对应编码器特征与解码器特征在通道维度简单拼接，本模型在每层解码前引入复数多尺度跳跃连接单元（CMS-SCU）来替代直接跳跃连接。CMS-SCU通过对编码端不同尺度的特征进行对齐融合（详细见4.2.5节），实现编码特征与当前解码特征的深度整合。融合后，解码器模块内部还会再次使用CAM模块，以进一步细化频谱重建过程中的特征交互与优化。同时，解码器中的上采样模块采用复数转置卷积等操作将特征图逐步放大回原始尺寸。例如，我们使用两个3×3、步长为2的实值转置卷积核分别作用于复数特征的实部和虚部，实现空间上采样的同时重新引入高频细节；并配合CBN和CReLU对重建特征进行归一化和非线性变换，提升信号重建质量。

综上，AMFNet通过编码端的逐级特征压缩与解码端的逐级特征重建，形成一个从频谱粗略结构到精细细节的完整处理流程。尤其借助SDFU、CAM、SRU和CMS-SCU等创新模块的有机配合，网络能够有效捕获频谱中突出的脉冲结构和连续的背景噪声成分，实现多尺度下特征的自适应融合。在复杂水下噪声的自监督降噪任务中，该架构有效克服了子采样引起的频谱不连续性问题，在无需纯净参考信号的前提下突出并恢复出稳定的目标脉冲特征，显著提升了模型的降噪性能和鲁棒性。

4.2.2 频谱动态聚焦单元（SDFU）

频谱动态聚焦单元（Spectrum Dynamic Focusing Unit, SDFU）是AMFNet中为增强脉冲信号细节提取而设计的核心模块之一。其设计动机在于：编码器的下采样虽可提取抽象特征，但同时也会导致频谱图中部分细节和连续性丢失。当水下脉冲信号与背景噪声在频谱上高度叠加时，传统固定形状的卷积核难以同时兼顾全局结构和局部瞬时变化，可能遗漏短时突发的局部能量峰值。为了解决这一问题，SDFU引入了动态形变卷积（Deformable Convolution, DSConv）的思想[75]，通过可学习的卷积核偏移量来自适应调整卷积采样位置和区域，从而令卷积操作针对每个输入样本的频谱结构进行精准捕捉。

|  |
| --- |
| 图4-4 SDFU组成结构与效果示意图 |

如图4-4，SDFU在标准二维卷积的基础上增加了形变机制：卷积核的采样位置不再是固定规则网格，而是通过学习一系列位置偏移量 来动态调整。对于中心坐标为 的3×3卷积核 ，其采样位置定义为规则网格：

|  |  |  |
| --- | --- | --- |
|  |  | (4-2) |

受到可变形卷积[75]启发，而在SDFU的DSConv中，卷积核 沿着水平轴（时间轴）的采样位置定义变为：

|  |  |  |
| --- | --- | --- |
|  |  | (4-3) |

同样的，沿着垂直轴（频率轴）方向，卷积核 位置定义变为：

|  |  |  |
| --- | --- | --- |
|  |  | (4-4) |

上述公式中学习到的偏移量 和 通常为非整数，因此在实际实现中，引入了双线性插值方法计算非整数位置处的卷积核采样值。对于偏移后的位置 ，其卷积采样值通过周围四个整数栅格点：

|  |  |  |
| --- | --- | --- |
|  |  | (4-5) |

其中 表示由公式(4-12)和(4-13)计算出的非整数位置，而 表示其周围所有整数坐标位置, 表示的是卷积核 在整数坐标位置 处的取值。插值内核 由两个一维线性插值核 和 的张量积构成：

|  |  |  |
| --- | --- | --- |
|  |  | (4-6) |

在实现上，SDFU模块的最终输出由三个并行卷积分支的特征构成：沿时间轴（频谱横轴）的形变卷积特征 、沿频率轴（频谱纵轴）的形变卷积特征 和常规卷积特征 。这三个特征在通道维度拼接后，通过一个1×1卷积层进行初步融合，再经过组归一化（Group Normalization, GN）和挤压-激发注意力机制（Squeeze-and-Excitation Attention, SE）[76]进一步优化。其中，GN层对特征分布进行归一化处理，提升网络训练稳定性；SE机制则根据每个通道对输出结果的贡献，自适应增强与目标信号强相关的通道权重，降低与目标信号弱相关的冗余通道权重。对于输入复数频谱特征 而言，SDFU整个计算流程可表示为：

|  |  |  |
| --- | --- | --- |
|  |  | (4-7) |

其中符号 表示通道维度的拼接操作，函数 的具体定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (4-8) |

其中 表示1×1卷积，GN和SE分别表示组归一化和挤压-激发机制。SDFU引入的形变卷积机制显著提升了模型对频谱细节的捕捉能力。如图4-1（c）所示，对于缓变背景噪声中夹杂的瞬态脉冲信号，SDFU能够根据频谱图中能量分布的变化自动调整卷积感受野：在时间轴方向增加对瞬时突发事件的采样密度，在频率轴方向关注脉冲能量所在的频带范围。例如，对于桥梁振动产生的短促脉冲，其频谱常出现窄时间、宽频带的高能量带状结构，SDFU能够沿频率轴拉伸卷积核以覆盖该宽频带区域，确保脉冲能量被完整捕获；同时沿时间轴细化采样，刻画其短暂出现的瞬态特征。通过这种动态聚焦，SDFU有效避免了下采样过程中目标脉冲频谱细节的丢失。实验结果表明，引入SDFU后，编码器各层提取的特征中脉冲信号的显著性得到增强，为后续CAM模块的深入挖掘和区分提供了包含更多关键信息的输入。

4.2.3 复数注意力模块（CAM）

复数注意力模块（Complex Attention Module, CAM）是AMFNet网络中另一项重要创新设计，旨在提升网络对复数频谱特征的提取和表达能力。传统注意力机制大多仅针对幅度谱进行加权，忽略了相位信息；而在水下脉冲信号提取任务中，信号与噪声在幅度和相位上往往具有不同特征，仅关注幅度可能无法充分区分两者。为此，CAM在复数域同时对频谱实部（幅度分量）和虚部（相位分量）建模，通过特定结构设计实现幅度-相位信息的交互融合，提高了对复数谱信息的利用效率。如图4-5所示，CAM模块的整体架构可分为两个核心单元：复数特征交互单元和跨空间学习单元，共同作用以实现对频谱图局部细节与全局关联的自适应建模。

|  |
| --- |
| 图4-5 CAM组成结构示意图 |

在复数特征交互单元中，CAM首先针对输入的复数特征张量结构进行重新排布和并行处理。具体来说，对于输入的复数特征张量 （最后一个维度分别代表复数特征的实部和虚部）（最后一个维度分别代表复数特征的实部和虚部），CAM首先在最后一个维度将其拆分为实部和虚部两部分，随后将两部分沿通道维度重新连接，形成多个子特征组，以强化不同复数分量之间的交互：

|  |  |  |
| --- | --- | --- |
|  |  | (4-9) |

为了最大限度地加强实部和虚部之间的交互，CAM设计了三个并行子分支结构：其中两个分支以1×1的共享卷积为主，用于跨通道的信息融合；第三个分支以3×3卷积为主，用于提取局部空间细节特征。

在前两个共享的1×1卷积分支中，CAM借鉴坐标注意力（Coordinate Attention, CA）机制[77]的思想，分别针对实部特征和虚部特征在频谱图横向（时间轴）和纵向（频率轴）上的分布进行编码。具体过程是：通过在空间维度上分别进行水平方向和垂直方向的全局平均池化（Global Average Pooling, GAP），获取实部特征在各频率带的整体能量分布，以及虚部特征在各时间帧的整体相位分布。这相当于提取了实、虚部特征在两个正交方向上的全局描述。一方面，横向GAP捕获不同时间段上能量的变化趋势；另一方面，纵向GAP刻画不同频率分量上的总体激活水平。接着，将两个方向的全局特征向量分别通过共享的1×1卷积层进行投影和压缩，然后沿通道维度连接起来，形成综合的通道注意力向量。这一操作在降低特征维度的同时，实现了实部与虚部跨通道信息的融合交互，显著增强了两者之间的关联性。最终得到的通道注意力权重经过适当广播后，对原始特征图的各通道进行加权调制，自适应突出对目标信号有判别力的频带和特征通道。

与上述注重全局信息的两个分支并行，第三个3×3卷积分支则用于捕捉复数特征中实部与虚部的局部空间细节。该分支对输入的实、虚部特征在局部邻域内进行卷积操作，获取频谱图局部邻域的细微模式。例如，它可以关注到某一时刻相邻频点上共同出现的异常高幅度，或某一频带在相邻时刻的相位变化趋势。卷积输出经过sigmoid非线性激活生成一张空间注意力权重图，用于细粒度地调节原始特征图对应位置的值——相当于根据局部邻域模式对输入特征做像素级的增强或抑制。通过1×1和3×3卷积分支的互补，CAM在同一模块中同时建模了复数频谱特征的局部相关 （短程依赖）与全局相关 （长程依赖）。实践证明，这种多尺度相关性的融合策略在处理复杂多变的频谱结构时具有明显优势[78]。

在复数特征交互单元处理后，CAM利用跨空间学习单元进一步融合来自不同尺度的注意力信息，以产生最终的注意力映射。具体做法是：首先对前述得到的多尺度特征应用二维全局平均池化（GAP）操作，在空间维度上对特征图进行全局聚合：

|  |  |  |
| --- | --- | --- |
|  |  | (4-10) |

这一步获得的向量 表示特征图在整个时频平面的平均激活水平，捕获了长距离的空间依赖关系。然后，对该全局特征向量利用Softmax函数进行归一化缩放，确保其元素和为1且与原特征尺度一致。这相当于从全局视角提炼出一个空间注意力权重系数 。接着，将这一系数与之前得到的局部空间注意力权重 通过矩阵点乘的方式融合，生成最终的复数空间注意力图。该注意力图综合了短程局部细节和长程全局依赖两种空间相关信息，对频谱图上各位置的显著性进行评价。最后，利用Sigmoid激活将注意力图归一化至0-1范围，并作用于输入特征图，以精确突出目标频谱特征区域、有效抑制背景区域。通过以上一系列操作，CAM模块能够自适应调节频谱图中不同位置和通道的响应，使与目标脉冲信号相关的谱区得到强化，而噪声占优的谱区被淡化。

综上，CAM模块通过实部-虚部信息交互和多尺度空间注意力融合，大幅提升了网络对频谱细节特征的敏感度和表达能力。在复杂海洋噪声背景下，CAM可以更稳定地捕捉那些区分脉冲信号与背景噪声的关键模式——例如短时突发的能量峰值、异常的相位变动等——从而提高模型对目标信号的辨别能力和鲁棒性。值得强调的是，在自监督训练情景下，没有纯净标签直接指导模型关注哪些特征，CAM提供了一种有效机制引导模型自主关注关键的幅度和相位特征，这对于准确提取水下工程脉冲型声信号至关重要。

4.2.4 语义重校准单元（SRU）

在AMFNet中，编码器与解码器的衔接部分（即网络底部的瓶颈层）是信号由高维压缩表示过渡到重构输出的关键环节，也是全局特征信息汇聚与分流的桥梁。为充分融合编码端提取的高级语义特征并为解码端提供最优信息表征，本文在编码器输出与解码器输入的交汇处设计了语义重校准单元（Semantic Recalibration Unit, SRU）。SRU的作用在于重新编码并加权调整编码端的顶层特征，使之既保留原始的高层语义信息，又经过优化更适配解码端逐层细化重建的需求。换言之，SRU对编码器末层提取的特征进行“语义校准”，突出对后续重构有用的成分，削弱冗余或有害的信息，从而提高最终重建结果的质量。

如图4-6所示，为实现上述目标，SRU融合了全局特征建模和局部细节增强两方面策略，对应引入了两个子模块：复数全局多层感知器（CGMLP）和复数局部注意力（CLA）。其中，CGMLP模块侧重建模长程依赖和全局语义信息，通过通道维度的深度变换来整合频谱的宏观结构；CLA模块侧重捕捉局部细节和短程依赖，通过一种字典注意力机制来发掘频谱局部的重要模式。两个模块分别处理特征的不同方面，最后在SRU中协同工作，达到对编码顶层特征全方位“校准”的效果。

|  |
| --- |
| 图4-6 SRU组成结构示意图 |

SRU首先对来自编码器的顶层复数特征 进行预处理，以保证后续模块可以更高效地加以利用。具体而言， 的实部 与虚部 分别通过一个5×5卷积层（步长为1）、BN和ReLU激活函数，然后沿通道维度连接：

|  |  |  |
| --- | --- | --- |
|  |  | (4-11) |

其中函数 的具体定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (4-12) |

其中， 代表5×5卷积运算， 为批量归一化层， 为ReLU激活函数。该预处理稳定了特征分布，减小了不同通道之间的统计差异，为后续信息交互奠定基础。随后，将标准化后的特征 并行输入CGMLP和CLA两个子模块，分别进行全局特征融合和局部特征增强。

（1）CGMLP：复数全局多层感知器。CGMLP由多个子层叠加而成，每个子层各司其职，共同对输入特征的全局语义进行提炼和强化。首先，深度卷积残差模块采用1×1深度卷积（Depthwise Convolution）来处理 ，在保持各通道独立的前提下提取每个频谱通道的深层模式。由于使用深度卷积，此步骤几乎不引入跨通道计算，在提高效率的同时保留了原始特征结构。紧接着，引入通道缩放操作（Channel Scaling, CS）对经过深度卷积的特征进行线性变换，调整各通道的幅值范围，以增强特征的泛化能力：

|  |  |  |
| --- | --- | --- |
|  |  | (4-13) |

其中， 表示组归一化， 为1×1的深度卷积， 表示通道缩放操作。随后，在特征交互模块中，使用1×1卷积将实部特征与虚部特征在通道维度重新融合：

|  |  |  |
| --- | --- | --- |
|  |  | (4-14) |

如此一来，原本独立的实、虚部通道产生交互，从中学习复数域全局信息的关联模式。幅度信息和相位信息在全局尺度上得以整合，使网络可以更好地感知那些分散在不同通道的全局噪声模式或信号模式的整体趋势。接下来，经过一个通道MLP残差模块（即在通道维度上的多层感知机变换），进一步挖掘通道与通道之间的高阶非线性关系：

|  |  |  |
| --- | --- | --- |
|  |  | (4-15) |

其中， 代表通道维度的多层感知机（Multi-Layer Perceptron, MLP）[79]，通过跨通道的映射增强了复数特征之间的交互，能够捕获更复杂的全局依赖关系，提高模型对不同特征组合的适应性[80]。

此外（如图4-7所示），为抑制过拟合并强化全局特征的稳健性，我们对通道MLP输出的实部与虚部又施加了一次通道平均池化：

|  |  |  |
| --- | --- | --- |
|  |  | (4-16) |

其中， 为通道数，该操作相当于对每个通道的响应求平均，提取全局背景的平均能量水平，使网络对整体噪声强度的估计更加准确稳定。经过以上层层处理，CGMLP模块有效建模了频谱全局的语义特征和长程依赖：例如，它能识别出背景噪声在各频带的大致能量分布，提取出目标信号在整体上的微弱稳态特征。

|  |
| --- |
| 图4-7 通道平均池化 |

（2）CLA：复数局部注意力模块。CLA旨在对 中残留的细节信息进行强化，特别关注那些对重构有重要影响的局部模式。CLA引入了一种内置字典机制（Dictionary Mechanism）来实现这一目标。首先，将 通过1×1和3×3卷积层压缩通道维度，接着利用局部注意力计算相应特征的权重。设第 个编码子相关的所有特征信息 计算如下：

(4-17)

其中， 为局部编码向量， 表示缩放因子， 和 分别对应第 个输入向量的实部与虚部， 和 分别为第 个编码向量的实部与虚部。获得所有局部编码权重 后，CLA通过一个融合单元将它们整合，以突出那些对应关键模式的分量：

|  |  |  |
| --- | --- | --- |
|  |  | (4-18) |

其中， 由BN、ReLU、均值池化和全连接层构成。这一融合相当于统计所有局部模式出现的情况，找出最显著的几个模式类型并增强其权重。最后，CLA将得到的关注权重作用回原始顶层特征：具体地，将各局部注意力信号按照对应的位置映射回 ，并在实部和虚部通道间加入适当的交互，然后通过sigmoid归一化后与原特征逐元素相乘，再残差式地叠加回原特征：

(4-19)

|  |  |  |
| --- | --- | --- |
|  |  | (4-20) |

其中， 为sigmoid函数， 为通道维度点乘， 为通道维度加法。这一系列操作使与字典模式匹配的局部细节特征被显著强化，而无关细节被淡化。在水下脉冲信号场景中，这意味着CLA能够突出诸如脉冲峰值、瞬态尖锐变化等细节，为解码器提供更丰富准确的局部信息。

通过CGMLP和CLA的协同作用，SRU实现了对编码器顶层特征的全局-局部联合重校准：CGMLP从宏观上把握频谱整体结构和背景噪声水平，CLA从微观上雕琢关键脉冲细节特征，两者相互配合。经过SRU重校准后的特征既包含全局语义上下文（例如噪声分布、信号的大致形态），又保留局部判别细节（例如峰值位置、短时结构），为解码端的逐层信号重建奠定了坚实基础。在后续解码过程中，这些校准后的特征将指导每一层解码器更有针对性地恢复目标信号——既保证重构信号保留脉冲的瞬时特征，又有效压制背景噪声，最终实现高质量的去噪输出。

4.2.5 复数多尺度跳跃连接单元（CMS-SCU）

自监督学习场景下缺乏纯净参考，模型必须充分挖掘混合信号中蕴含的各种线索来区分信号与噪声。其中，频谱的多尺度信息利用至关重要：水下环境中的工程脉冲信号通常伴随缓慢变化的低频背景噪声，同时叠加短暂尖锐的高频脉冲。单一尺度的特征往往无法同时描述这两类截然不同的频谱成分。如果模型仅依赖于浅层（高分辨率）特征进行重建，可能缺乏对全局低频结构的把握，导致背景噪声残留；反之，若仅依赖深层（低分辨率）特征，则细微的高频脉冲纹理容易被忽略。为此，本文提出了复数多尺度跳跃连接单元（Complex Multi-Scale Skip Connection Unit, CMS-SCU），用于在解码阶段有效融合编码端不同尺度的特征表示，确保模型同时关注局部细节和全局背景，从而提升降噪性能。

|  |
| --- |
| 图4-8 CMS-SCU组成结构示意图 |

如图4-8所示，CMS-SCU的核心思想是在传统跳跃连接基础上，引入多尺度特征融合策略：每一层解码器在利用对应编码器输出进行特征补充时，不再仅使用单一层的特征，而是将其相邻的多个尺度的编码器特征一并融合进来。具体而言，对于解码端第  层，需要融合编码端第 层的输出特征以及其相邻的上、下尺度（例如第 和第 层）的特征。为了使这些尺度不同的特征能够直接融合，CMS-SCU首先对它们进行空间尺度对齐：对尺度较小（分辨率低）的特征进行上采样、对尺度较大（分辨率高）的特征进行下采样，最终将三者调整到相同尺寸。然后，将对齐后的特征在通道维度拼接，并通过一个1×1卷积进行初步融合。1×1卷积的作用一方面在于压缩通道维度，减弱直接拼接带来的冗余信息；另一方面是混合不同来源特征，有助于提炼跨尺度的综合表示。融合后的特征再送入一个通道注意力机制（Channel Attention, CA）进行加权筛选。通道注意力为融合特征的每一通道分配一个权重（范围0到1），表示该通道对当前重建任务的重要程度。这样，来自不同尺度的特征信息将通过注意力得到自适应的衡量：与目标信号关联度高的特征通道被赋予较大权重予以保留（例如代表脉冲结构的通道），而冗余或噪声相关的通道则被削弱。动态加权融合确保了网络在解码每层时，都能优先利用有益于脉冲信号重建的那部分编码器特征，同时抑制干扰信息，实现更纯净的重建。

借助CMS-SCU，多尺度特征在解码过程中得以充分交互和互补。特别地，在存在低频背景噪声与高频瞬态脉冲共存的情况下，CMS-SCU允许网络灵活调整感受野大小：对于描述缓变背景的全局模式，网络可以依赖融合进来的粗尺度特征进行刻画；而对于定位瞬时脉冲峰值这样的细节，网络则可以借助细尺度特征进行精确重构。这一融合策略保证了不同频率成分的信息在重建时被全面保留。从频谱角度看，低频部分的稳态噪声结构和高频部分的瞬态信号细节都不会在单一尺度的处理过程中被淹没或遗失，而是通过跨尺度的信息流得到统一考虑。这不仅增强了模型对频谱全局结构的理解力（确保背景噪声的整体形态被正确建模），也提高了其对局部细节的分辨力（确保脉冲信号的尖峰细节被精确重现）。因此，在自监督降噪任务中，CMS-SCU提供了一种稳定而高效的特征表示方式：即使缺乏明确的监督信号，网络也能凭借多尺度融合获得对目标信号更完整的表征，使最终输出的去噪频谱同时具备正确的整体噪声水平和细节信号特征，实现对水下目标脉冲信号更出色的提取与还原。

4.3 损失函数设计

在自监督学习框架下，损失函数的设计对于模型的训练效果至关重要。针对水下声信号的特点，本文构建的损失函数综合了时域和频域的误差项，以确保模型能够有效学习水下噪声环境中的目标信号特征。

设模型输出的时域信号为 ，其对应的目标信号为 ，两者的STFT分别记为 和 。在每一帧中，时域信号长度为 ，频域采样点数为 ，总帧数为 。综合考虑时域信号的重建精度、频域信号的频谱一致性以及模型的正则化约束，定义如下损失函数：

|  |  |  |
| --- | --- | --- |
|  |  | (4-21) |

其中， 主要用于保证时域信号的整体波形重建精度， 约束模型对频谱特征的精准重构，而 则用于在训练过程中引入额外的正则化约束， 和 为各项损失的平衡系数。

首先，在时域损失部分，采用MSE衡量模型输出与目标信号之间的差异：

|  |  |  |
| --- | --- | --- |
|  |  | (4-22) |

该项损失确保模型重建的时域波形与目标信号尽可能贴近，从整体上控制去噪信号的波形失真度。

其次，在频域损失部分，为了确保模型在不同频率段的重建精度，引入了一种加权误差策略。在计算STFT得到 和 后，定义频域损失为两者差异的加权MSE：

|  |  |  |
| --- | --- | --- |
|  |  | (4-23) |

此外，为进一步提升模型的泛化能力，并确保自监督学习过程中网络能更稳定地提取有用信息，我们在损失函数中加入了正则化约束项 ，其定义如下：

(4-24)

其中， 表示训练的降噪网络， 和 为Nbr2Nbr策略生成的两个子采样信号对。该正则项约束模型输出的自一致性，即要求网络在不同子采样条件下的降噪结果保持稳定，并减少过拟合风险。加入这一正则化约束能够提升模型对水下脉冲信号的敏感度，同时降低子采样带来的信息缺失影响，使模型更关注信号的稳定成分而非背景噪声的随机波动。

4.4 本章小结

本章面向水下工程脉冲信号频谱连续性缺失与尾波现象的核心痛点，提出了自监督深度降噪网络AMFNet及其配套训练策略，系统给出了模型框架、关模块与损失函数的完整实现。

首先，在4.1节梳理了自监督总体思路与AMFNet的功能定位，阐明通过子采样‑伪标签机制替代人工干净样本，迫使网络在频谱差分约束下学习信号公共结构、自动抑制随机噪声。

随后，4.2节由整体到局部解构了网络架构：在 U 型编码‑解码骨干中嵌入频谱动态聚焦单元（SDFU），通过可学习的二维采样权重实现对压缩频谱的自适应对齐；引入复数注意力模块（CAM）与语义重校准单元（SRU），分别增强复数域幅‑相信息耦合与多尺度语义一致性；结合复数多尺度跳跃式连接单元（CMS‑SCU），在跨层传播阶段弥补低‑高频信息缺口，共同构筑了针对瞬态脉冲与持续噪声的协同感知链路。

最后，4.3节综合了时域和频域的误差项、能量加权MSE与正则项约束，完成损失函数的设计，使网络能兼顾频谱局部细节与全局的重构。

综上所述，本章完成了自监督深度降噪网络AMFNet的构建，为后续章节的实验验证与方法对比分析奠定了理论基础和实现依据。

第五章 降噪实验与结果分析

5.1 实验设置

5.1.1 实验设计内容

为全面评估所提出自监督深度学习降噪方案在水下工程声场景中降噪的有效性与泛化能力，本研究围绕“标准训练—原始场景验证—跨类型迁移”的逻辑链构建了三个互补实验环节：累计能量段打桩声降噪、原始打桩声降噪及未见类型脉冲声（桥梁振动声信号）降噪。三项实验在目标信号类型、预处理方式与评估侧重点上逐级递进，既可验证模型在理想条件下的基准性能，又能检验其在复杂工程场景与任务迁移条件下的鲁棒性。

累计能量段打桩声降噪实验作为第一环节，主要用于构建训练数据并完成初步性能测试。实验选用距打桩作业点约50 m 处采集的高信噪比打桩声作为近似纯净的目标信号。第3章的分析表明，该距离范围内信号时频结构稳定、工程代表性强。鉴于水下传播的多径效应往往使主脉冲之后出现能量分散且信噪比较低的尾波，本研究基于声暴露级能量累积的工程需求，对原始波形进行积分处理，截取累计能量达到90 % 的有效波形段，用以剔除尾部多径扰动并提高样本一致性。随后，将所得主脉冲段与随机抽取的海洋背景噪声片段按设定信噪比线性叠加，构建训练集、验证集与基础测试集

原始打桩声降噪实验在不进行任何裁剪或预处理的情况下，直接将完整打桩波形与与前一环节相同的噪声片段混叠生成外部测试集，模拟采集系统实际接收到的原始含噪信号。通过对比可检验模型在面对尾波振铃、多径混叠等冗余干扰时的降噪能力，并进一步评估其自动聚焦主脉冲能量的能力，对模型在前端智能降噪与目标提取环节的实用性具有直接意义。

非训练类型工程声信号降噪实验，面向模型在非训练信号类型上的泛化能力评估。实验选取水下桥梁振动声信号作为目标，该类信号在时域结构和频谱特性上均与打桩声存在明显差异，未在模型训练阶段出现。该信号未参与模型训练，将其与同一噪声库随机抽取的背景噪声片段混叠后形成测试集，可验证模型面对“未见类型”信号时的适应性与泛化表现，为多源水下噪声抑制任务提供重要参考。

通过上述三级实验，本研究实现了从标准训练到原始复杂场景再到跨任务迁移的完整性能评估框架，为模型在不同水声应用环境中的部署提供了系统性支撑与理论依据，详见表5-1。

表5-1 实验环节设置

|  |  |  |
| --- | --- | --- |
| 实验环节名称 | 数据构成 | 数据用途 |
| 累计能量段打桩声降噪实验 | 90 % 能量段打桩声信号  + 海洋背景噪声 | 训练/验证/测试 |
| 原始打桩声信号降噪实验 | 原始打桩声信号  + 海洋背景噪声 | 测试 |
| 非训练类型声信号降噪实验 | 桥梁振动声信号  + 海洋背景噪声 | 测试 |

5.1.2 数据集构建与来源

本研究依据“目标信号采集—背景噪声采集—子集设计—信号混叠”四个步骤构建实验数据集，以系统评估所提出自监督降噪模型在复杂海洋噪声环境中的适应性与鲁棒性。

首先，目标工程声信号包括海上打桩声与跨海桥梁振动声两类，两者均采用 B&K 8105型水听器以65536 Hz / 32 bit 规格进行同步采集，相关布放方案与现场参数已在第3.1.3节阐述。为满足后续建模的片段化输入要求，打桩声依据单次脉冲进行切片（平均时长约0.8 s），桥梁振动声按 3 s 为单位切片，分别获得4887段与3544段目标信号，为模型训练与外部测试提供可靠基准。

其次，为获取时空覆盖度足够的噪声样本，在相邻海域非施工时段布设同型号水听器，采集周期横跨四季并覆盖不同时段，总采集时长约30 h。采样现场现场示意见图5-1，具体采集日期与海洋环境参数详见表5-2。采集完成后，原始噪声数据按5 s 为单位进行切片，得到海洋背景噪声数据库 。其中，样本数量 ，为后续数据集构建提供充足的环境样本。

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 230320杏林2  (a) | 230418莆田  (b) | 230529漳州  (c) |
| 269607fd6070b8023e50c44bc00b452  (d) | 230926厦门2  (e) | 240319厦门2  (f) |
| 图片2  (g) | 240902厦门2  (h) | 241022厦门  (i) |
| 241203厦门  (j) | 250121福清  (k) | 250318厦门  (l) |
| 图 5-1海洋背景噪声采集场景图 | | |

表5-2 海洋背景噪声采集时间及环境参数

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 采集日期 | 水深 | 潮汐状态 | 海况 | 场景图 |
| 1 | 2022年10月26日 | 4.8 m | 涨潮 | 1级 | — |
| 2 | 2023年3月20日 | 17.6 m | 涨潮 | 1级 | 图5-3(a) |
| 3 | 2023年4月18日 | 30.2 m | 涨潮 | 1级 | 图5-3(b) |
| 4 | 2023年5月29日 | 6.0 m | 涨潮 | 1级 | 图5-3(c) |
| 5 | 2023年9月14日 | 22.9 m | 涨潮 | 1级 | 图5-3(d) |
| 6 | 2023年9月26日 | 7.8 m | 涨潮 | 1级 | 图5-3(e) |
| 7 | 2024年3月19日 | 7.1 m | 涨潮 | 1级 | 图5-3(f) |
| 8 | 2024年5月10日 | 23.6 m | 涨潮 | 1级 | 图5-3(g) |
| 9 | 2024年5月31日 | 23.2 m | 涨潮 | 1级 | — |
| 10 | 2024年9月2日 | 13.7 m | 涨潮 | 1级 | 图5-3(h) |
| 11 | 2024年9月26日 | 5.8 m | 涨潮 | 1级 | — |
| 12 | 2024年10月22日 | 9.1 m | 涨潮 | 1级 | 图5-3(i) |
| 13 | 2024年12月3日 | 4.8 m | 涨潮 | 1级 | 图5-3(j) |
| 14 | 2024年12月24日 | 12.6 m | 涨潮 | 1级 | — |
| 15 | 2025年1月21日 | 16.3 m | 涨潮 | 1级 | 图5-3(k) |
| 16 | 2025年3月18日 | 7.2 m | 涨潮 | 1级 | 图5-3(l) |

基于上述信号资源，设计三类数据集以覆盖不同实验需求。第一类为“累计能量段打桩声子集”。为突出主脉冲能量，该子集从原始打桩声 ，中提取归一化累计能量5 % –95 % 区间，方法如下：设每采样点能量为 ，其归一化累计能量定义为：

|  |  |  |
| --- | --- | --- |
|  |  | (5-1) |

设起始点 和终止点 满足：

|  |  |  |
| --- | --- | --- |
|  |  | (5-2) |

则构建主能量波形段 ：

|  |  |  |
| --- | --- | --- |
|  |  | (5-3) |

如图5‑2所示：上图为原始波形，中图为累计能量曲线，下图为截取后的有效信号段。该数据集共4887段，并按8∶1∶1 划分训练、验证与测试集，是模型训练的主要数据来源。其二，“原始打桩声数据集”保留完整波形的同数量样本，仅用作外部测试，以检验模型在尾波振铃和多径混叠条件下的降噪性能。其三，“桥梁振动声数据集”直接采用3544段原始振动声信号，用于评估模型在未见类型脉冲声上的迁移能力，该数据集保持原始信号，不与背景噪声混叠。

|  |
| --- |
| 图 5-2：打桩信号有效能量段提取过程 上：原始信号；中：累计能量曲线；下：截取后的有效信号段 |

信号混叠仅针对前两类打桩声数据集执行。首先利用PyDub[81]将海洋背景噪声片段截断或循环拼接，使其时长与目标信号一致；随后根据设定信噪比 dB 计算比例系数：

|  |  |  |
| --- | --- | --- |
|  |  | (5-4) |

生成混合信号 。其中，海洋背景噪声片段从 中随机抽取；对于第一类与第二类数据集中来自同源的打桩声信号，混叠时始终复用同一噪声片段及信噪比，确保对比评估的一致性。至此，经“水下工程声信号采集—海洋背景噪声采集切片—结构化子集设计—可控信噪比混叠”四步，本研究构建了覆盖多噪声工况且结构多样的实验数据集，为后续算法训练与性能验证奠定了坚实基础，数据集构建与属性说明详见表5-3。

表5-3数据集构建与属性说明

|  |  |  |  |
| --- | --- | --- | --- |
| 数据集名称 | 构建方式 | 样本总数 | 数据用途 |
| 累计能量段打桩声数据集 | 混叠噪声 | 4887 | 训练/验证/测试 |
| 原始打桩声信号数据集 | 混叠噪声 | 4887 | 测试 |
| 桥梁振动声信号数据集 | 原始信号 | 3544 | 测试 |

5.1.3 实验环境与预处理流程

本实验在高性能计算服务器上进行，以确保模型训练和推理的计算效率。服务器配备 NVIDIA RTX 4090 GPU，采用Linux Ubuntu 20.04 操作系统，并运行 Python 3.8 作为开发环境。深度学习框架选用 PyTorch 1.10.0，并结合 CUDA 11.3 进行并行加速计算，以优化大规模数据处理性能。在模型训练完成后，最佳模型的参数将用于测试集推理，以评估其在不同数据集上的降噪性能。

数据预处理阶段，为确保所有音频样本具有一致的输入尺寸，所有音频数据统一截取或零填充为65280个采样点，并进行统一的归一化处理，采用STFT将时域波形转换为频域表示，便于网络学习频谱特征。该转换过程利用PyTorch中的torch.stft函数实现，具体变换公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | (5-5) |

其中， 表示在时间帧 和频率索引 处的复数频谱， 为输入音频信号， 为汉宁窗（Hanning Window），窗口长度设置为 1022，步长设置为 256，以保证时间分辨率和频率分辨率的平衡。STFT转换得到的复数频谱被拆分为实部和虚部，并分别作为网络输入的两个通道，以适应AMFNet复数卷积网络的输入格式。

在模型训练采用Adam优化器，初始学习率设置为0.001。若验证集损失在一个epoch内无显著下降，则将学习率按衰减因子0.1递减；连续5个epoch无改进时触发早停以防止过拟合。批大小设为16，在保证 GPU 显存利用率的同时兼顾模型的泛化性能。每个epoch结束后计算验证损失并更新最优权重，直至满足早停条件或达到预设训练轮次。

5.1.4 评价指标

为全面评估所提出自监督降噪方案在降噪任务中的表现，本文采用了多种主流且具有代表性的信号质量评估指标。考虑到水下声信号的非平稳特性及打桩声的瞬态冲击特征，本节选用的评价指标不仅关注整体信号增强效果，还兼顾信号保真度与重建误差的量化，具体包括信噪比（SNR）、峰值信噪比（PSNR）和均方根误差（RMSE）。

（1）信噪比（Signal-to-Noise Ratio, SNR）

SNR是衡量处理后信号中有用成分相对于噪声成分的比例，定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (5-6) |

其中， 为原始干净信号， 为去噪后信号。SNR越高，说明信号中的噪声成分越少，去噪效果越好。

（2）峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）

衡量最大信号值与重建误差之间的比例，尤其敏感于局部异常和信号峰值的还原质量，计算公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | (5-7) |

其中， 表示信号可能的最大幅值， 为均方误差。对于具有明显脉冲或峰值特征的打桩信号， 能有效评估模型对关键信息的保留能力。

（3）均方根误差（Root Mean Square Error, RMSE）

是衡量去噪信号与真实信号之间整体偏差的重要指标，能有效反映模型重建误差的绝对水平。其计算公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | (5-8) |

值越小，表示模型的重建误差越低，性能越优。

综上，实验通过上述四项指标从多角度对模型性能进行量化评估，以保证对自监督去噪模型在复杂海洋噪声背景下的有效性做出全面而可靠的判断。

5.2 实验结果

5.2.1 累计能量段打桩声降噪实验

本实验基于第5.1.2节中构建的累计能量段打桩声数据集，目标信号为通过SEL分析所提取的90 % 主能量波形段，噪声源为实际采集的海洋背景噪声。该实验旨在评估所提出模型在结构清晰、能量集中的典型脉冲信号条件下的降噪性能，尤其关注其对主能量区域的还原能力与干扰成分的抑制能力。

图5-3展示了信噪比为0 dB 条件下，目标信号、混叠信号与模型降噪输出的波形对比结果。如图5-3（a）所示，原始目标信号在0.5 s 附近具有典型的脉冲形态，结构紧凑、能量集中。如图5-3（b）所示，混叠信号中的海洋背景噪声显著干扰了脉冲结构识别，尤其是前后段被大量非结构噪声覆盖，信号起止模糊，波形细节几乎被完全掩盖。如图5-3（c）所示，模型输出结果呈现出显著改善：脉冲信号结构清晰恢复，两端的高频噪声与随机扰动被有效抑制，整体信号在时间轴上重新集中在目标段，体现出强大的波形重构能力与非结构成分抑制能力。

|  |
| --- |
| (a) |
| (b) |
| (c) |
| 图 5-3：SNR = 0 dB 条件下累计能量段打桩声信号波形图: (a) 目标信号；  (b) 混叠海洋背景噪声信号；(c) 降噪信号 |

进一步的频谱图展示见图5-4。如图5-4（a）所示，目标信号具有明显的脉冲特征，能量集中度高、带宽窄。如图5-4（b）所示，混叠信号频谱则被中低频随机噪声全面污染，几乎无法辨识出有效脉冲事件，主频段已与噪声混叠在一起。如图5-4（c）所示，模型输出展现出高度还原能力：主频段能量再次集中，原始脉冲特征基本重建，同时有效压制了非目标频段的能量泄露，表现出良好的频谱聚焦能力与中低频噪声抑制效果。

|  |
| --- |
| (a) |
| (b) |
| (c) |
| 图 5-4：SNR = 0 dB 条件下的频谱图: (a) 目标信号；  (b) 混叠海洋背景噪声信号；(c) 降噪信号 |

从工程声信号监测角度分析，水下打桩声信号属于典型的短时脉冲类信号，在时域上呈现出明确的突发能量集中结构。在实际应用中，声波在海水中传播过程中易受到多径效应、边界反射及海底回波等复杂干扰的影响，使得信号后半部分常出现幅度减弱但结构紊乱的“尾波”成分。此类成分虽具高能冗余，但对事件判别与声源分析价值有限，反而可能干扰声暴露评估与信号识别模型的稳定性。

基于SEL理论，声事件的有效波形段通常定义为累计能量达到90% 的时间区间，核心在于真实反映事件主体能量释放的过程，剔除环境响应引起的非结构成分。因此，该实验在数据构建上即采用累计能量截取方法提取目标信号，模型输出则应尽可能与此结构保持一致。

从图5-3的时域结果来看，模型显著提高了信号起止边界的清晰度，恢复了主段结构，同时在两端噪声主导区域实现了尾波成分的有效压制，这一特性对基于事件检测或能量累计的应用尤为关键；而从图5-4的频谱表现来看，模型输出不仅保留了目标信号在中低频段的能量主峰，还抑制了混叠信号中由背景噪声引入的高频干扰，增强了信号的频域可分性与结构一致性。时域与频域双重维度的降噪效果共同验证了模型对水下工程声信号的强适配能力，为打桩事件的自动识别、暴露级计算与传播路径分析等下游任务提供了坚实基础。

此外，为更细致地评估降噪模型在不同时刻的表现，本文采用帧级信噪比（frame-level SNR）可视化方式，对目标信号与背景噪声在时序上的分离效果进行深入分析。将音频信号以30毫秒的帧长进行分帧，并设置75%的帧重叠率，可在时间轴上获得连续的帧级信号段。随后分别计算原始含噪信号与模型降噪输出在每一帧上的SNR值，并绘制成曲线图，从而实现对模型时域性能的可视化追踪。

图5-5所示为打桩声信号在0–1秒范围内的帧级SNR对比结果，其中蓝色曲线表示原始含噪信号的帧级SNR，绿色曲线表示模型降噪后的帧级SNR，红色阴影区域则标注了打桩信号帧段（即5.1.2节处理帧段），时间约0.5 s 附近。

从图像整体趋势可以观察到，在非打桩帧段中，原始含噪信号的帧级SNR长期维持在较低水平，约为 -60 dB，而模型输出信号的帧级SNR稳定提升至约 -40 dB，说明模型能够在背景区域有效降低噪声干扰。在打桩帧段开始时，模型输出信号的帧级SNR出现明显提升，表明模型不仅成功保留了冲击类打桩信号的能量特征，而且显著抑制了其混杂背景。随着打桩段结束，帧级SNR值快速回落至稳定水平，未观察到明显的残留伪影，显示模型对瞬态事件具有较好的响应能力和时间分辨性。

|  |
| --- |
| 图 5-5帧级SNR可视化对比图（蓝线：噪声信号；绿线：降噪信号；红色区域为打桩段） |

为系统评估模型性能，本文在三种信噪比（0 dB / 5 dB / 10 dB）混叠条件下，分别计算噪声信号与降噪输出在SNR、PSNR与RMSE三项指标下的表现。各项指标的定义与计算方法遵循音频增强与音频重建领域的通用规范。结果汇总见表5-4。

表5-4 模型降噪性能指标对比（累计能量段信号）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 混叠信噪比(dB) | 信号类型 | SNR (dB) | PSNR (dB) | RMSE |
| 0 | 噪声信号 | -1.20 | 23.90 | 0.0064 |
|  | 降噪信号 | **10.25** | **36.42** | **0.0014** |
| 5 | 噪声信号 | 3.75 | 28.65 | 0.0037 |
|  | 降噪信号 | **13.48** | **39.65** | **0.0010** |
| 10 | 噪声信号 | 8.45 | 33.12 | 0.0020 |
|  | 降噪信号 | **16.78** | **42.65** | **0.0007** |

5.2.2 原始打桩声信号降噪实验

为进一步验证模型在非预处理条件下的降噪能力，本节采用未经累计能量截断处理的原始打桩声信号作为目标信号，与实际采集的海洋背景噪声进行混叠，构建新的测试数据集，并开展降噪实验。该实验旨在评估模型在面对更复杂、未规整的原始信号输入时，是否仍具备对关键能量区域的自动提取能力，并能有效抑制由于水声多径传播导致的尾部混响干扰，为水下声信号采集监测等工程应用提供更具通用性的降噪策略支持。

相比累计能量段信号，原始打桩声信号包含完整的打桩事件响应过程，前期为能量突发阶段，中后段则常出现因海底或水面反射引发的尾波与混响。该类尾部信号虽具一定能量，但其结构往往紊乱、分布离散，且难以用于事件强度评估或源定位分析，属于典型的多径冗余信号。因此，降噪模型若能自动识别主能量段并弱化尾波结构，将在实用性与鲁棒性上表现更佳。

图5-6展示了信噪比为0 dB 条件下，原始打桩声信号的混叠波形与模型降噪结果。从图中可见，原始信号前段呈现出强烈脉冲结构，而尾段则被背景噪声与多径混响共同覆盖，导致时域轮廓模糊、边界不清。模型输出信号则实现了前段结构的清晰恢复，并在中后段显著抑制了多余尾波成分，有效聚焦于主事件区域，体现出对打桩事件的自动识别与结构提取能力。

|  |
| --- |
| (a) |
| (b) |
| (c) |
| 图 5-6：SNR = 0 dB 条件下原始打桩声信号的波形图: (a) 目标信号；  (b) 混叠海洋背景噪声信号；(c) 降噪信号 |

图5-7展示了该信号在频谱域的可视化结果。从图5-7（a）可见，原始打桩信号在约0.5 s 处的主频段能量集中，但中后段受多径传播效应影响，频谱表现出明显的模糊扩展特征，尾部混响能量扩散明显，反映出典型的声波在水中遇边界反射或海底折返造成的混响效应，但结构离散，对频谱贡献有限。如图5-7（c）所示，模型输出信号频谱结构更为集中，不仅保留了主要冲击事件频率成分，还实现了对尾部非结构能量的有效压制。特别是在中高频区域，模型显著削弱了多径混响造成的尾段杂散频率分布，使频谱恢复紧凑，轮廓清晰，进一步提升了后续能量评估、事件识别与频率特征提取的可靠性。

|  |
| --- |
| (a) |
| (b) |
| (c) |
| 图 5-7：SNR = 0 dB 条件下条件下原始打桩声信号的频谱图: (a) 目标信号；(b) 混叠海洋背景噪声信号；(c) 降噪信号 |

进一步地，图5-8展示了原始信号条件下的帧级信噪比可视化结果。图中，蓝色曲线表示原始含噪信号帧级SNR，绿色曲线为降噪后信号，红色阴影区域则标注了打桩信号帧段（即5.1.2节处理帧段），时间约0.5 s 附近。可观察到：在背景噪声段，模型能稳定将帧级SNR提升约20 dB；在打桩主频段附近，模型帧级SNR峰值提升至 + 30 dB 左右，而原始信号峰值仍不足 15 dB。在尾段，降噪信号帧级SNR快速回落至稳定水平，尾部混响未形成伪影拖尾。以上表明模型在保留信号主结构特征的同时，显著降低了背景噪声与多径混响效应的干扰。



图 5-8帧级SNR可视化对比图（蓝线：噪声信号；绿线：降噪信号；红色区域为打桩段）

表5-5汇总了在三种信噪比下，原始信号条件下模型降噪前后的定量指标（SNR、PSNR、SDR、RMSE）对比情况。对比第5.2.1节中表5-4可见，累计能量段信号在0 dB 混叠条件下降噪后的SNR达10.25 dB，略高于原始信号条件下的9.12 dB，表明模型在结构规整输入下能获得更佳性能。PSNR呈现相同趋势，主要归因于原始信号包含完整尾部波形，存在更多由水声多径传播引起的混响干扰，结构更复杂，降噪难度提升。但在信噪比为5 dB 和10 dB 情况下，模型在两类信号上的性能差距显著缩小，表现出较强的一致性和适应性，说明所提出模型具备良好的泛化能力，能够在结构理想与实际复杂场景中均实现有效降噪。

表5-5 模型降噪性能指标对比（累计能量段信号）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 混叠信噪比(dB) | 信号类型 | SNR (dB) | PSNR (dB) | RMSE |
| 0 | 噪声信号 | -1.20 | 24.10 | 0.0060 |
|  | 降噪信号 | **9.12** | **34.28** | **0.0019** |
| 5 | 噪声信号 | 3.16 | 27.56 | 0.0041 |
|  | 降噪信号 | **12.84** | **38.17** | **0.0012** |
| 10 | 噪声信号 | 8.03 | 31.95 | 0.0025 |
|  | 降噪信号 | **16.32** | **41.68** | **0.0009** |

5.2.3 非训练类型声信号降噪实验

为进一步检验所提出模型在非训练场景中的泛化能力与类型迁移适应性，本节选取实际工程中另一类典型的水下工程冲击类声源——水下桥梁振动声信号，作为测试样本，输入模型进行推理并分析降噪效果。该类信号具有突发性、瞬时性与多频成分特征，但其时域形态、频谱结构及背景环境均未参与模型训练，故可用于评估模型在“未知类型事件”场景下的适应表现。

由于该实验样本为实际采集的噪声混合信号，缺乏对应的纯净标签信号，因此无法采用SNR、PSNR、RMSE等传统定量指标进行性能评估。为了确保分析的科学性与工程有效性，实验采用可视化分析法对模型的结构保持能力与干扰抑制效果进行综合评估，结合桥梁结构声传播特点与脉冲事件识别需求，重点考察模型是否能突出恢复主能量段，并有效剔除背景扰动与结构尾波。

图5-9展示了某一桥梁振动信号片段的时域波形降噪效果对比。从图中可见，原始信号中存在强背景底噪干扰，脉冲主段被大量非结构能量覆盖，信号边界模糊。降噪结果中，模型显著提升了信号的对比度，脉冲主段结构更加清晰，两侧底噪得到明显削弱，具备良好的冲击事件聚焦能力。

|  |
| --- |
| (a) |
| (b) |
| 图 5-9：水下桥梁振动声信号降噪前后波形对比图: (a) 原始采集含噪信号；  (b) 降噪信号 |

图5-10为对应的频谱分析结果，显示降噪前信号频谱能量在全频段扩散，低频部分尤其受到环境噪声和混响的显著污染；而模型输出的频谱结构表现为能量向中低频主段回归、杂散频率成分明显压缩，整体频谱轮廓更为集中，反映出模型对结构主频段的识别与保持能力。

|  |
| --- |
| (a) |
| (b) |
| 图 5-10：水下桥梁振动声信号降噪频谱对比图: (a) 原始采集含噪信号；  (b) 降噪信号 |

从工程声学监测角度出发，有效提取此类信号中的主能量段对于声场计算、事件识别及异常检测具有关键意义。模型在未训练该类型信号的条件下，仍能准确识别其脉冲核心段并压制背景干扰，说明所提出架构具备良好的结构通用性与时频自适应能力。尤其在模型未显式学习该类事件的语义特征、传播路径或信号形式的情况下，仍可实现工程有效的信号增强，进一步验证了模型的类型迁移能力与弱监督鲁棒性。

5.3 模型有效性与降噪方法对比分析

本章将对提出的自监督水下降噪模型进行进一步的有效性验证，通过设计不同实验场景，分别针对纯海洋背景噪声信号与典型水下工程信号（桥梁振动信号）进行降噪效果分析，并与传统降噪方法的表现进行对比，深入评估本文模型在工程实际应用场景中的优势。此外，通过消融实验验证模型内部模块的实际贡献，凸显所提出方法在架构设计上的科学性与有效性。

5.3.1 海洋背景噪声降噪效果对比实验

为验证本文提出的自监督深度学习模型在实际复杂海洋环境中的降噪性能，本节选取典型的实际海洋背景噪声信号，分别应用传统降噪方法（维纳滤波、小波变换）与本文所提出模型进行降噪实验，通过波形和频谱的详细对比分析，深入评估各方法的降噪能力与实际工程适用性。

图5-11 展示了不同方法降噪后的时域波形对比。原始的海洋背景噪声波形（图5-11(a)）呈现明显的随机波动与尖峰扰动，整体表现为典型的非平稳随机噪声特性。维纳滤波方法（图5-11(b)）对背景噪声的随机波动仅有轻微的抑制效果，随机扰动成分仍明显存在，未能有效改善信号的整体稳定性。小波变换方法（图5-11(c)）表现出与维纳滤波类似的问题，随机波动未能显著减轻，信号整体仍然表现为较为杂乱的状态，降噪效果有限，无法有效提高后续水声信号识别分析的可靠性。相比之下，本文提出的自监督模型（图5-11(d)）的降噪效果极为显著，波形变得非常平稳，随机噪声成分被大幅压制，波动幅值明显减小，展现出极佳的噪声抑制能力和工程适用价值。

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | (d) |
| 图5-11海洋背景噪声降噪时域波形图对比：(a) 原始信号；(b) 维纳滤波降噪；(c) 小波变换降噪；(d) 本文模型降噪 | |

进一步通过频谱分析（图5-12）验证各方法的降噪性能差异。从原始噪声频谱（图5-12(a)）可以明显观察到能量广泛分布于整个频带内，随机扰动强烈且频谱结构复杂。维纳滤波降噪后的频谱（图5-12(b)）尽管对部分高频噪声稍有抑制，但频谱整体杂乱程度并未显著改善，随机扰动特征依然突出，显示了维纳滤波在非平稳随机背景中的局限性。小波变换降噪后的频谱（图5-12(c)）也表现出类似的特征，随机噪声仍较明显，频谱整体未出现明显的结构改善或扰动压制迹象，难以有效改善信号品质。而本文所提出模型的降噪频谱结果（图5-12(d)）表现出极为显著的改善，频谱能量高度集中在低频区域，随机扰动和高频噪声成分几乎被完全抑制，频谱清晰、稳定且干净，显著提升了频谱结构的辨识度与可靠性，展现出该模型在非平稳随机噪声降噪方面的卓越性能。

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | (d) |
| 图5-12海洋背景噪声降噪频谱图对比：(a) 原始信号；(b) 维纳滤波降噪；(c) 小波变换降噪；(d) 本文模型降噪 | |

综合以上详细的时域与频谱域分析可得，传统方法（维纳滤波、小波变换）在面对非平稳随机海洋背景噪声时，存在明显的适用性问题，表现为随机噪声压制效果弱、频谱结构无法有效改善，降噪性能受限。与之对比，本文提出的自监督深度学习模型则充分体现了其结构稳定性与频谱高效聚焦优势，不仅有效抑制了随机扰动，还极大提升了频谱结构清晰度与可靠性，具备明显的工程应用价值与优势。

5.3.2 非训练信号降噪效果对比实验

为进一步验证本文提出的自监督深度学习模型在非训练信号上的结构适应能力与工程实用性，本节选取实际采集的水下桥梁振动声信号作为输入样本，分别采用维纳滤波、小波变换和本文模型进行降噪处理。该类信号在实际采集中常伴随海洋背景噪声与结构传播干扰，信号特征表现为非周期突发冲击与背景拖尾并存，结构不稳定、干扰复杂，具备显著的非平稳性与脉冲性特征。

图5-13展示了各方法处理下的时域波形图。原始信号（图5-13(a)）包含多个幅度不一的冲击波动，并夹杂有持续性的背景噪声干扰，信号结构边界模糊。维纳滤波（图5-13(b)）压制了部分高频噪声成分，但对于突发冲击结构仍存在模糊趋势，主能量段辨识度不高；小波变换结果（图5-13(c)）虽然保留了部分结构边缘，但对背景扰动抑制有限，且在多个位置引入高幅值伪信号，削弱了原始结构清晰度。相比之下，本文模型处理结果（图5-13(d)）有效突出了多个冲击事件的结构波动，信号主体成分突出，起止界限明确，同时对背景波动实现了更强的压制，波形清晰紧凑。

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | (d) |
| 图5-13海洋背景噪声降噪时域波形图对比：(a) 原始信号；(b) 维纳滤波降噪；(c) 小波变换降噪；(d) 本文模型降噪 | |

频谱分析结果如图5-14所示，进一步验证了各方法的频域降噪能力。原始信号（图5-14(a)）呈现多个短时冲击频率带叠加现象，但同时存在大量高频杂散分量，频谱能量分布松散。维纳滤波（图5-14(b)）削弱了部分高频能量，但未能聚焦真实频率结构，频谱依旧混杂；小波变换（图5-14(c)）频谱结构略有收敛，但信号主体频带仍被背景所淹没，难以提供有效事件支撑。本文模型频谱结果（图5-14(d)）则呈现出良好的频谱聚焦性，冲击成分在频域中被完整还原，背景频带能量下降显著，主频段结构突出，特别在低频段表现出更强的能量集中度，说明模型在频谱层面具有更优异的分离与还原能力。

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | (d) |
| 图5-14海洋桥梁振动声信号降噪频谱图对比：(a) 原始信号；(b) 维纳滤波降噪；(c) 小波变换降噪；(d) 本文模型降噪 | |

从图5-13和图5-14的联合观察中可进一步发现，在时域约0.85 s 处存在明显的非平稳背景噪声扰动，频谱图中该处信号不具备明确主频结构，属于典型的海洋背景噪声波动。在原始波形图中，该波动以突增扰动的形式出现，对信号整体判断构成干扰。传统方法如维纳滤波与小波变换均无法有效区分该背景扰动与真实冲击信号，导致残余能量保留甚至误识为主信号成分。而本文模型在该区域展现出较强的抑制能力，成功压制该无效扰动，使主能量段的结构边界更加清晰，频谱中也未出现对应杂散频带，进一步体现出模型在时频域上的协同降噪优势。

综上所述，本文模型在面对非训练脉冲信号时，仍能准确保留其主结构波动并有效压制背景扰动，表现出良好的时域结构保持能力与频域聚焦能力。传统方法在该类非平稳信号处理上存在明显的残留与结构模糊问题，难以满足工程信号处理中的实际需求。本文方法不仅具备高适应性与强鲁棒性，同时展现出在复杂水下背景中的广泛适用潜力。

5.3.3 模块消融与多方法对比分析

为进一步验证本文模型结构设计的合理性与性能表现的先进性，本节在累计能量段打桩信号数据集（混叠信噪比为0 dB）上开展模块消融实验与多方法对比实验，并在比较中引入原始含噪信号的性能指标作为参考基准，以更直观展示各模型对信噪质量的提升能力。

在模块消融部分，保持模型整体结构与训练配置一致，依次去除SDFU、CAM、CMS-SCU，构建三种对照变体，并与完整模型在相同数据集上进行性能评估。表5-6汇总了各模型在SNR、PSNR和RMSE三项指标下的表现情况。结果显示，完整模型在三项指标上均表现最优，SNR达10.25 dB，PSNR为 36.42 dB，RMSE降至 0.0014，相较于原始噪声信号（SNR = -1.20 dB，PSNR = 23.90 dB，RMSE = 0.0064）实现了大幅提升。去除SDFU后模型的结构对齐能力减弱，影响主能量聚焦能力，SNR 降为8.31 dB，RMSE升至0.0021。实验验证了多模块协同在复杂结构信号恢复中的关键作用。

表5-6 模型结构消融实验性能指标对比（累计能量段打桩信号，SNR = 0 dB）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型编号 | 模型结构 | SNR (dB) | PSNR (dB) | RMSE |
| 原始信号 | 噪声信号 | -1.20 | 23.90 | 0.0064 |
| 模型A | **完整模型** | **10.25** | **36.42** | **0.0014** |
| 模型B | 去除SDFU | 8.31 | 33.66 | 0.0021 |
| 模型C | 去除CAM | 7.54 | 32.83 | 0.0023 |
| 模型D | 去除CMS-SCU | 6.74 | 31.85 | 0.0027 |

为进一步论证模型的工程实用性，如表5-7所示，本节引入近年自监督领域音频降噪模型、维纳滤波与小波变换，构建完整性能对比体系，结果如表5-6所示。维纳滤波在平稳背景中具有一定抑噪能力，但面对非结构冲击成分时出现明显频谱模糊现象，SNR仅为1.17 dB，RMSE为0.0048；小波变换虽在某些脉冲信号上保留边缘，但抑制能力不稳定，SNR为0.02 dB，RMSE为0.0055。U-Net及其他自监督降模型具备良好的整体建模能力，但对非稳态尾波扰动及信号边界定位仍显不足。相比之下，本文模型在三项指标上均实现性能最优，结构还原与能量重构能力显著优于对照方法，充分说明模型结构设计的针对性与有效性。

表5-7 模型与其他方法降噪性能对比（累计能量段打桩信号，SNR = 0 dB）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法类型 | 模型方法 | SNR (dB) | PSNR (dB) | RMSE |
| — | 原始含噪信号 | -1.20 | 23.90 | 0.0064 |
| 传统方法 | 维纳滤波 | 1.17 | 26.35 | 0.0048 |
| 小波变换 | 0.02 | 25.2 | 0.0055 |
| 自监督策略模型 | U-Net[28] | 5.47 | 30.65 | 0.0030 |
| NerNT[49] | 7.18 | 32.42 | 0.0025 |
| N2N[50] | 7.82 | 33.07 | 0.0022 |
| ONT[51] | 8.63 | 33.81 | 0.0020 |
| **本文模型** | **10.25** | **36.42** | **0.0014** |

5.4 本章小结

本章围绕所提出的自监督水下结构脉冲声降噪模型，系统设计并实施了一系列具有工程针对性与科学严谨性的实验，全面评估模型在多种复杂场景下的适应能力与降噪效果。具体工作涵盖数据集构建、多段实验、模型对比与消融验证三个方面，具体内容如下：

首先，在数据集构建方面，基于实测的水下打桩声与海洋背景噪声信号，设计了三类数据集与对应的降噪实验任务：累计能量段打桩声降噪实验、原始打桩声信号降噪实验以及非训练类型工程声信号降噪实验。其中，累计能量段实验聚焦于SEL理论下的主结构能量恢复，通过提取信号中累计达到90%能量的有效波形段，体现模型在结构性特征提取方面的适应性；原始信号实验则考察模型在未进行预处理情况下对完整信号的结构自适应能力及其抗尾波干扰性能；非训练类型实验引入水下桥梁振动声作为泛化验证样本，用于分析模型在跨类型冲击声处理中的拓展能力。

其次，在降噪性能评估方面，针对不同实验任务，从波形对齐、频谱保持及帧级SNR可视化等维度开展分析。结果显示，在累计能量段实验中，模型能够准确增强主能量段、有效抑制尾波扰动，输出波形具备良好可视解释性，频谱主带重建清晰，帧级SNR在打桩声段显著提升，体现出较强的瞬态识别能力。在原始信号实验中，模型表现出良好的结构起始定位能力、频谱抖动修复效果及尾部多径干扰抑制能力，展现出高度稳健性。在非训练样本实验中，尽管声源结构与训练数据不同，模型仍能有效识别脉冲峰值、压制背景非稳态噪声，验证了其良好的泛化能力与工程适应性。

此外，为进一步验证模型结构设计的有效性与先进性，本章基于累计能量段数据集开展了模块消融实验与多方法对比实验。消融实验结果表明，完整模型在SNR、PSNR和RMSE等指标上均优于去除任一模块的简化结构，凸显多模块协同设计的重要性。对比实验结果也显示，本文所提出模型在整体性能上显著优于传统方法（如维纳滤波、小波变换）以及现有自监督深度模型，体现出其在结构保持与噪声抑制方面的综合优势。

综上所述，本章通过多维度、系统性的实验验证，充分论证了所提模型在复杂海洋环境下对结构性脉冲声信号的识别与降噪能力，尤其在低信噪比条件下表现出更高的稳定性与工程应用价值。为后续在实际水下工程场景中的部署与应用奠定了坚实的技术基础。

第六章 总结与展望

6.1 论文的主要工作与创新点

本论文围绕水下工程脉冲声信号的降噪问题，针对传统方法在复杂背景干扰条件下结构保持能力不足、现有深度模型对监督样本依赖强以及模型工程适应性差等问题，提出了一种融合结构感知机制的自监督水下降噪方法，并围绕数据构建、模型设计与性能验证展开系统研究，构建了完整的工程任务闭环方案。本文的主要研究工作与创新点概括如下：

（1）从实际采集的打桩与背景噪声数据出发，结合声压级、声暴露级等典型声学参数，系统分析了水下脉冲类信号的能量分布规律与非结构噪声的干扰特性。在此基础上，设计基于累计能量段提取的结构性参考信号构建方案，采用实际采集的海洋背景噪声与工程脉冲声信号进行混叠，取代传统高斯噪声模拟方式，构建贴近真实场景的自监督训练集，解决了无纯净标签条件下模型训练目标难以定义的问题。

（2）针对脉冲类信号的主能量段识别与尾波干扰压制需求，构建了一种融合频域建模与通道注意力机制的深度神经网络结构。模型引入频谱动态聚焦模块、复数域注意力机制与多尺度结构融合策略，实现对非平稳噪声背景下信号主结构的多维感知与显著增强。各子模块具备明确的功能分工与协同作用，显著提升了模型在复杂背景中的结构保持能力与误差控制能力。

（3）论文设计了多层级、多场景的实验验证体系，涵盖基于累计能量段的模型训练、原始未截断信号的降噪效果验证，以及非训练类型工程声信号（如桥梁振动声）的泛化能力测试，构建了从结构恢复到工程应用的完整性能评估框架。同时通过与传统、神经网络等方法进行对比，并辅以模块消融实验，全面验证了模型的结构设计合理性与性能优势。

综上，本文在模型结构设计、训练目标构建与实验验证框架三方面形成系统创新，具体体现在以下三点：

（1）提出基于声暴露级理论的结构性参考信号构建机制，采用累计能量段提取目标信号有效结构，解决了无纯净标签条件下自监督训练目标缺失的问题。

（2）设计融合频谱建模与注意力机制的结构保持型降噪网络，面向工程脉冲信号特性，实现对主能量段的精确恢复与尾波干扰的有效压制。

（3）构建覆盖训练内与训练外样本的多场景验证体系，实现模型在结构恢复精度、泛化能力与工程适应性等方面的系统性评估。

这些创新为水下工程声信号的智能降噪处理提供了切实可行的技术路径，为复杂水声环境中的结构事件监测、能量评估与声源感知奠定了方法基础与工程支撑。

6.2 未来工作展望

本文提出了一种面向水下工程脉冲声信号的自监督结构保持型降噪方法，围绕信号结构感知、背景噪声抑制及无标签训练机制进行了系统设计与验证，在多个任务场景下取得了良好的实验效果。然而，考虑到实际应用需求的复杂性与模型推广的可行性，当前研究仍存在若干值得进一步拓展的方向，未来的研究工作可在以下几个方面展开：

（1）多声源干扰下的复杂信号建模与分离仍是亟待解决的重要问题。实际海洋施工环境中常同时存在船舶运行、设备轰鸣、环境扰动等多类背景噪声，这些声源之间存在频谱重叠、时序交错与非结构混合等现象，传统基于单一目标的降噪方法在此类环境中易产生误削与信号混淆。未来研究可结合语义建模与特征判别机制，增强模型在复杂声源场景中的目标聚焦能力与干扰识别能力。

（2）当前研究仍以离线处理为主，模型部署与工程系统的结合程度有限。随着智能传感技术与边缘计算设备的持续发展，面向实时降噪与系统集成的模型轻量化研究将成为关键方向。未来可探索嵌入于水下观测平台或声学监测终端，实现打桩声等事件的在线检测、噪声屏蔽与结构感知闭环处理。

综上，未来工作将在多源干扰感知、工程系统集成等方面持续推进，进一步提升降噪模型在复杂水声环境中的实用性与工程应用价值，为智能海洋监测、生态评估与近岸作业声控系统提供更加稳健的技术支撑。

参考文献

1. 金永明.新时代海洋强国战略的演进历程与重要成就[J].国家治理,2024,(02):40-45.
2. 王春娟,辛庞晨雨,刘大海.中国海洋工程装备国产化进程及其高质量发展趋势[J].中国软科学,2024,(S1):379-387+413.
3. 张然.海上风电场水下打桩噪声研究[D].厦门大学,2019.
4. 章蔚.海上风电场水下噪声传播研究[D].上海海洋大学,2016.
5. 王晨卉,陶毅,许肖梅.海上风电场运营期水下噪声的仿真研究[C]//中国声学学会水声学分会,山东声学学会,中国造船工程学会船舶仪器仪表学术委员会.中国声学学会水声学分会2021～2022年学术会议论文集.厦门大学海洋与地球学院;水声通信与海洋信息技术教育部重点实验室;,2022:429-431.DOI:10.26914/c.cnkihy.2022.062110.
6. Lurton X. An introduction to underwater acoustics: principles and applications[M]. Springer Science & Business Media, 2002.
7. Gussen C M G, Diniz P S R, Campos M L R, et al. A survey of underwater wireless communication technologies[J]. J. Commun. Inf. Sys, 2016, 31(1): 242-255.
8. 王晨卉.海上风电场运营期水下噪声研究[D].厦门大学,2022.DOI:10.27424/d.cnki.gxmdu.2022.002902.
9. Ou H, Allen J S, Syrmos V L. Frame-based time-scale filters for underwater acoustic noise reduction[J]. IEEE Journal of oceanic engineering, 2011, 36(2): 285-297.
10. Raj K M, Murugan S S, Natarajan V, et al. Denoising algorithm using wavelet for underwater signal affected by wind driven ambient noise[C]//2011 International Conference on Recent Trends in Information Technology (ICRTIT). IEEE, 2011: 943-946.
11. Liu F, Li G, Yang H. Application of multi-algorithm mixed feature extraction model in underwater acoustic signal[J]. Ocean Engineering, 2024, 296: 116959.
12. Aslam M A, Zhang L, Liu X, et al. Underwater sound classification using learning based methods: A review[J]. Expert Systems with Applications, 2024, 255: 124498.
13. Zhou A, Li X, Zhang W, et al. A novel cross-attention fusion-based joint training framework for robust underwater acoustic signal recognition[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-16.
14. Jaiswal A, Babu A R, Zadeh M Z, et al. A survey on contrastive self-supervised learning[J]. Technologies, 2020, 9(1): 2.
15. Liu S, Mallol-Ragolta A, Parada-Cabaleiro E, et al. Audio self-supervised learning: A survey[J]. Patterns, 2022, 3(12).
16. Zhang W, Yang X, Leng C, et al. Modulation recognition of underwater acoustic signals using deep hybrid neural networks[J]. IEEE Transactions on Wireless Communications, 2022, 21(8): 5977-5988.
17. Gao R, Liang M, Dong H, et al. Underwater acoustic signal denoising algorithms: A survey of the state-of-the-art[J]. IEEE Transactions on Instrumentation and Measurement, 2025.
18. Wu Y, Ta X, Xiao R, et al. Survey of underwater robot positioning navigation[J]. Applied Ocean Research, 2019, 90: 101845.
19. Zhang W, Wu W, Teng Y, et al. An underwater docking system based on UUV and recovery mother ship: design and experiment[J]. Ocean Engineering, 2023, 281: 114767.
20. Xie Y, Ren J, Xu J. Unraveling complex data diversity in underwater acoustic target recognition through convolution-based mixture of experts[J]. Expert Systems with Applications, 2024, 249: 123431.
21. Jiang J, Wu Z, Huang M, et al. Detection of underwater acoustic target using beamforming and neural network in shallow water[J]. Applied Acoustics, 2022, 189: 108626.
22. Chen J, Benesty J, Huang Y, et al. New insights into the noise reduction Wiener filter[J]. IEEE Transactions on audio, speech, and language processing, 2006, 14(4): 1218-1234.
23. Hong J, Bae I, Seok J. Wiener filtering-based ambient noise reduction technique for improved acoustic target detection of directional frequency analysis and recording sonobuoy[J]. The Journal of the Acoustical Society of Korea, 2022, 41(2): 192-198.
24. Raj K M, Murugan S S, Natarajan V, et al. Denoising algorithm using wavelet for underwater signal affected by wind driven ambient noise[C]//2011 International Conference on Recent Trends in Information Technology (ICRTIT). IEEE, 2011: 943-946.
25. Gur B M, Niezrecki C. Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform[J]. The Journal of the Acoustical Society of America, 2007, 122(1): 188-199.
26. Ganapathi S, Kumar S M, Deivasigamani M. Noise reduction in underwater acoustic signals for tropical and subtropical coastal waters[C]//2016 IEEE/OES China Ocean Acoustics (COA). IEEE, 2016: 1-6.
27. Liu F, Li G, Yang H. Application of multi-algorithm mixed feature extraction model in underwater acoustic signal[J]. Ocean Engineering, 2024, 296: 116959.
28. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer international publishing, 2015: 234-241.
29. Xu Y, Du J, Dai L R, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE/ACM transactions on audio, speech, and language processing, 2014, 23(1): 7-19.
30. Rethage D, Pons J, Serra X. A wavenet for speech denoising[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 5069-5073.
31. Botinhao C V, Wang X, Takaki S, et al. Investigating RNN-based speech enhancement methods for noise-robust text-to-speech[C]//9th ISCA speech synthesis workshop. 2016: 159-165.
32. Weninger F, Erdogan H, Watanabe S, et al. Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR[C]//Latent Variable Analysis and Signal Separation: 12th International Conference, LVA/ICA 2015, Liberec, Czech Republic, August 25-28, 2015, Proceedings 12. Springer International Publishing, 2015: 91-99.
33. Kim J, El-Khamy M, Lee J. T-gsa: Transformer with gaussian-weighted self-attention for speech enhancement[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 6649-6653.
34. Wang K, He B, Zhu W P. TSTNN: Two-stage transformer based neural network for speech enhancement in the time domain[C]//ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2021: 7098-7102.
35. Zhou A, Zhang W, Li X, et al. A novel noise-aware deep learning model for underwater acoustic denoising[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.
36. Zhou A, Zhang W, Xu G, et al. DBSA-net: Dual branch self-attention network for underwater acoustic signal denoising[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 1851-1865.
37. Song Y, Liu F, Shen T. A novel noise reduction technique for underwater acoustic signals based on dual‐path recurrent neural network[J]. IET Communications, 2023, 17(2): 135-144.
38. Zhou A, Li X, Zhang W, et al. An attention-based multi-scale convolution network for intelligent underwater acoustic signal recognition[J]. Ocean Engineering, 2023, 287: 115784.
39. Santos-Domínguez D, Torres-Guijarro S, Cardenal-López A, et al. ShipsEar: An underwater vessel noise database[J]. Applied Acoustics, 2016, 113: 64-69.
40. Irfan M, Jiangbin Z, Ali S, et al. DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification[J]. Expert Systems with Applications, 2021, 183: 115270.
41. Sayigh L, Daher M A, Allen J, et al. The Watkins marine mammal sound database: an online, freely accessible resource[C]//Proceedings of Meetings on Acoustics. AIP Publishing, 2016, 27(1).
42. Du X, Hong F. QiandaoEar22: a high-quality noise dataset for identifying specific ship from multiple underwater acoustic targets using ship-radiated noise[J]. EURASIP Journal on Advances in Signal Processing, 2024, 2024(1): 96.
43. Wilkes D R, Gourlay T P, Gavrilov A N. Numerical modeling of radiated sound for impact pile driving in offshore environments[J]. IEEE Journal of Oceanic Engineering, 2016, 41(4): 1072-1078.
44. Tsouvalas A. Underwater noise emission due to offshore pile installation: A review[J]. Energies, 2020, 13(12): 3037.
45. Peng Y, Tsouvalas A, Stampoultzoglou T, et al. A fast computational model for near-and far-field noise prediction due to offshore pile driving[J]. The Journal of the Acoustical Society of America, 2021, 149(3): 1772-1790.
46. Caron M, Touvron H, Misra I, et al. Emerging properties in self-supervised vision transformers[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 9650-9660.
47. Liu S, Mallol-Ragolta A, Parada-Cabaleiro E, et al. Audio self-supervised learning: A survey[J]. Patterns, 2022, 3(12).
48. Gui J, Chen T, Zhang J, et al. A survey on self-supervised learning: Algorithms, applications, and future trends[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
49. Fujimura T, Koizumi Y, Yatabe K, et al. Noisy-target training: A training strategy for DNN-based speech enhancement without clean speech[C]//2021 29th european signal processing conference (EUSIPCO). IEEE, 2021: 436-440.
50. Kashyap M M, Tambwekar A, Manohara K, et al. Speech denoising without clean training data: A noise2noise approach[J]. arXiv preprint arXiv:2104.03838, 2021.
51. Wu J, Li Q, Yang G, et al. Self-supervised speech denoising using only noisy audio signals[J]. Speech Communication, 2023, 149: 63-73.
52. Hildebrand J A. Anthropogenic and natural sources of ambient noise in the ocean[J]. Marine Ecology Progress Series, 2009, 395: 5-20.
53. Dahl P H, Miller J H, Cato D H, et al. Underwater ambient noise[J]. Acoustics Today, 2007, 3(1): 23-33.
54. Cato D H. Ambient noise and its significance to aquatic life[J]. Bioacoustics, 2008, 17(1-3): 21-23.
55. Berdnikova J, Klauson A, Mustonen M, et al. Underwater ship noise pattern detection and identification[J]. The Journal of the Acoustical Society of America, 2017, 142(4\_Supplement): 2686-2686.
56. Rodrigo F J, Ramis J, Carbajo J, et al. Underwater anthropogenic noise pollution assessment in shallow waters on the south-eastern coast of Spain[J]. Journal of Marine Science and Engineering, 2022, 10(9): 1311.
57. Wittekind D, Schuster M. Propeller cavitation noise and background noise in the sea[J]. Ocean Engineering, 2016, 120: 116-121.
58. Cui X, Yang H, Hu Q, et al. Noise Sources and Statistical Characteristics of the Marine Ambient Noise in the Zengmu Basin[C]//2024 OES China Ocean Acoustics (COA). IEEE, 2024: 1-5.
59. Chunxia M, Dan S, Feng C, et al. Statistical characteristic of spectrum for ambient noise at high frequencies in shallow water[C]//2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2017: 1-4.
60. Felisberto P, Silva J P, Silva J, et al. Background noise in areas covered by marine plants in the Ria Formosa lagoon during the summer[C]//2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). IEEE, 2018: 1-5.
61. Bailey H, Senior B, Simmons D, et al. Assessing underwater noise levels during pile-driving at an offshore windfarm and its potential effects on marine mammals[J]. Marine pollution bulletin, 2010, 60(6): 888-897.
62. Yang L, Xu X, Huang Z, et al. Recording and analyzing underwater noise during pile driving for bridge construction[J]. Acoustics Australia, 2015, 43: 159-167.
63. Dahl P H, de Jong C A F, Popper A N. The underwater sound field from impact pile driving and its potential effects on marine life[J]. Acoustics Today, 2015, 11(2): 18-25.
64. Reinhall P G, Dahl P H. Underwater Mach wave radiation from impact pile driving: Theory and observation[J]. The Journal of the Acoustical Society of America, 2011, 130(3): 1209-1216.
65. Dahl P H, Dall'Osto D R, Farrell D M. The underwater sound field from vibratory pile driving[J]. The Journal of the Acoustical Society of America, 2015, 137(6): 3544-3554.
66. 时文静,王志陶,方亮,等.打桩水下噪声对长江江豚影响初探[J].水生生物学报,2015,39(02):399-407.
67. 于晋源,许肖梅,张鑫海,等.水下打桩噪声场的有限元分析[C]//中国声学学会水声学分会.中国声学学会水声学分会2019年学术会议论文集.厦门大学海洋与地球学院;水声通信与海洋信息技术教育部重点实验室;,2019:406-408.
68. 陆炜.跨江桥梁水下振动噪声特性研究[D].东南大学,2022.DOI:10.27014/d.cnki.gdnau.2022.001807.
69. Song X, Yin L, Xiong W, et al. Underwater noise prediction and control of a cross-river subway tunnel: An experimental and numerical study[J]. International Journal of Environmental Science and Technology, 2024, 21(4): 4045-4062.
70. Song X, Zhang X, Xiong W, et al. Experimental and numerical study on underwater noise radiation from an underwater tunnel[J]. Environmental Pollution, 2020, 267: 115536.
71. Song X, Lu W, Xiong W, et al. Sound contribution of the low frequency underwater noise radiated from a suspension bridge[J]. Journal of Low Frequency Noise, Vibration and Active Control, 2024, 43(1): 3-19.
72. Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2Noise: Learning image restoration without clean data[J]. arXiv preprint arXiv:1803.04189, 2018.
73. Huang T, Li S, Jia X, et al. Neighbor2neighbor: Self-supervised denoising from single noisy images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14781-14790.
74. Trabelsi C, Bilaniuk O, Zhang Y, et al. Deep complex networks[J]. arXiv preprint arXiv:1705.09792, 2017.
75. Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.
76. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
77. Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 286-301.
78. Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.
79. Tolstikhin I O, Houlsby N, Kolesnikov A, et al. Mlp-mixer: An all-mlp architecture for vision[J]. Advances in neural information processing systems, 2021, 34: 24261-24272.
80. Nosrati L, Fazel M S, Ghavami M. Improving indoor localization using mobile UWB sensor and deep neural networks[J]. IEEE Access, 2022, 10: 20420-20431.
81. Robert J, Webbie M. Pydub (2018)[J]. URL http://pydub. com.

攻读硕士学位期间获得的成果

1. **X**

致谢

123

**学位论文答辩委员会名单**

|  |  |  |  |
| --- | --- | --- | --- |
| 主 席： | 许肖梅 | 教授 | 厦门大学海洋与地球学院 |
| 委 员： | 万磊 | 副教授 | 厦门大学信息学院 |
|  | 陶毅 | 助理教授 | 厦门大学海洋与地球学院 |