

分 类 号: TN911.7
研究生学号: 2021551026

单位代码: 10183
密 级: 公 开



吉 林 大 学

博士学位论文

基于深度学习的地震勘探数据去噪及速度模型反演技术研究

Research on Deep Learning-Based Seismic Exploration Data
Denoising and Velocity Model Inversion

作者姓名: 王洪洲

专 业: 电气工程

研究方向: 智能化地震数据处理

指导教师: 林君 教授

培养单位: 仪器科学与电气工程学院

2024 年 06 月

基于深度学习的地震勘探数据去噪及速度模型反演技术
研究

Research on Deep Learning-Based Seismic Exploration
Data Denoising and Velocity Model Inversion

作者姓名：王洪洲

专业名称：电气工程

指导教师：林君 教授

学位类别：工程博士

论文答辩日期： 2024 年 5 月 31 日

授予学位日期： 年 月 日

答辩委员会组成：

	姓名	职称	工作单位
主席	王赞	教授	中国地质大学（北京）
委员	嵇艳鞠	教授	吉林大学
	孙锋	教授	吉林大学
	李刚	教授	吉林大学
	裴明波	教授	大庆油田有限责任公司
	刘立超	教授	中国地质大学（武汉）

摘要

基于深度学习的地震勘探数据去噪及速度模型反演技术研究

地震勘探是探查油气的重要方法,通过由人工激发向地下传播后反传的地震波,在地面采集形成地震记录作为地层构造成像与解释的基础。然而,地震数据经常受到来自仪器、环境及非地震动力学过程的噪声干扰。因此去噪成为地震数据处理中一个重要的步骤,通过这一环节移除噪声干扰并保留地震信号,为后续从地震波场数据中推断更重要的信息提供基础。在压制噪声后,进一步分析地震数据以揭示地下结构成为必要的后续步骤,此阶段的核心任务是速度模型反演,该过程通过地震数据来估算地下介质的波速分布,从而解析地层构造及其物质成分。传统的反演方法通常基于迭代优化过程,不仅计算量大,而且容易陷入局部最优解。深度学习模型能够在大量数据中学习到复杂的模式和特征,这些模型不仅可以被训练来去除噪声并尽可能保留有用的地震信号,还能够从地震数据中直接学习到速度模型。相对于传统基于迭代优化的方法,深度学习技术能够显著减少计算时间并提高反演的准确性。

然而,将深度学习应用于地震数据去噪和速度模型反演也面临着挑战和问题,包括标记数据的缺乏、对模型的远程特征捕获能力要求高以及模型在实际数据中的泛化能力差等。尽管如此,基于深度学习技术执行这两项处理任务仍然是一个前沿且充满潜力的研究领域。本文聚焦于开发有效的训练算法和模型结构,目标是在噪声压制和速度模型反演两项任务上,将深度学习技术有效应用至实际地震数据中。本文的主要创新点如下:

1. 地震资料中噪声和信号往往分布不均,不同区域的噪声水平可能存在显著差异,且信号强度变化范围大。常规的卷积神经网络(Convolutional Neural Networks, CNN)在处理这类记录时往往性能不佳。为了解决这一问题,本研究基于密集连接卷积神经网络(Dense Connected Convolutional Networks, DenseNet)提出一种密集连接去噪架构,通过在卷积神经网络中引入宏观和微观的密集跳跃连接,在不引入过大计算量的情况下,解决线性堆叠结构的CNN不同层之间缺乏直接的信息流动路径的问题。在网络中,每层的特征图都被传入后

续的层中，从而使得网络能够组合低级特征和高级特征来重建地震有效信号，具有更好的重建效果。同时，密集复杂连接提供了输出层与中间各层之间的短路径，从而减缓了梯度消失问题，使网络更易训练。在构建的高质量的训练集上进行训练后，该方法在模拟地震记录上能够实现 24.0578 dB 的信噪比提升，且在实际地震记录的处理中，能有效压制噪声，恢复同相轴信号。

2. 在极低信噪比条件下执行地震数据去噪任务时，CNN 架构在获取信号的远程相关性和全局一致性方面存在局限性，使其难以完整恢复信号。本研究提出一种基于 Transformer 架构的模型用于地震信号去噪。利用自注意力机制，模型克服了 CNN 架构在特征提取过程中的局限性，能够有效捕捉地震信号中的远程特征。为充分保证 Transformer 架构在实际地震记录去噪中的泛化性，本文设计了一种结合掩码自监督预训练和微调的两阶段训练策略，使用无标注数据的大规模地震数据集来对模型进行预训练，之后使用配对的“含噪-纯净数据”对模型进行微调。该方法在极低信噪比的模拟地震记录上能够实现平均 21.53 dB 的信噪比提升，且在多种实际记录的处理中能够有效衰减复杂地震噪声且恢复微弱信号。

3. 除了应用于去噪任务外，深度学习方法还可以用于从原始地震数据中直接估计推断速度模型，但是多数基于 CNN 的反演网络架构设计使得此类方法倾向于只提取局部特征，而忽略了反映同一界面信息的不同特征之间的空间相关性。远程特征对于反演任务更为重要。此外，地震数据的时变特性也不利于权重共享的 CNN 架构提取特征。针对这些问题，本研究基于 Transformer 架构开发了一种新的反演网络模型。模型在解编码结构中使用自注意力机制来捕捉地震数据中信号的远程相关性，为构建速度模型提供更多远程特征。使用合成模型的数值实验验证了所提出方法的有效性和可靠性。与传统的全波形反演（Full Waveform Inversion, FWI）方法和现有的基于 CNN 的速度反演方法相比，该方法在速度值、地下结构和地质界面方面与目标具有更大的一致性。

4. 基于迭代优化的常规地震反演方法如 FWI 往往高度依赖初始速度模型。而基于深度学习的反演算法，已可以生成相对有效的速度模型。在此基础上，本研究提出一种结合深度学习反演模型（数据驱动）和常规 FWI 方法（物理驱动）的混合反演策略。通过使用改进的特征对齐 Transformer 反演网络从原始地震数

据中推理估计初始速度模型，并利用 FWI 方法迭代优化获得最终反演结果，保证反演结果的正确性。混合反演策略可以充分发挥数据驱动和物理驱动两类方法的优势。在合成和实际地震数据上进行测试。实验结果表明混合反演策略可以显著提高反演结果的准确性，并能有效推广至实际地震数据的速度反演中。

关键词：

地震数据去噪，速度模型反演，深度学习，自监督预训练，多头自注意力

Abstract

Research on Deep Learning-Based Seismic Exploration Data Denoising and Velocity Model Inversion

Seismic exploration is a crucial method for oil and gas exploration, where artificially induced seismic waves are transmitted underground and then reflected back, forming seismic records on the surface. These records serve as the foundation for imaging and interpreting subsurface geological structures. However, seismic data are often plagued by noise from instruments, environmental factors, and non-seismic dynamic processes. Therefore, denoising becomes an essential step in seismic data processing, aiming to remove noise interference while preserving the seismic signal. This process lays the groundwork for inferring more critical information from seismic wavefield data. After suppressing noise, further analysis of seismic data to reveal subsurface structures is a necessary subsequent step. The core task at this stage is velocity model inversion, which estimates the distribution of wave speeds in the subsurface medium through seismic data, thereby deciphering geological structures and their material components. Traditional inversion methods typically rely on iterative optimization processes, which are not only computationally intensive but also prone to becoming trapped in local optima. Deep learning models, capable of learning complex patterns and features from large datasets, can be trained to remove noise while retaining useful seismic signals and directly learn velocity models from seismic data. Compared to traditional iterative optimization-based methods, deep learning techniques can significantly reduce computation time and improve inversion accuracy.

However, applying deep learning to seismic data denoising and velocity model inversion faces challenges, including a lack of labeled data, high requirements for capturing distant features by the model, and maintaining model generalizability. Despite these issues, leveraging deep learning for these tasks remains a cutting-edge and promising research field. This paper focuses on developing effective training algorithms and model architectures to apply deep learning techniques to real seismic

IV

data for noise suppression and velocity model inversion. The main innovations of this study include:

1. Seismic records often exhibit uneven distribution of noise and signal, with significant variations in noise levels across different areas and wide-ranging signal strengths. Conventional Convolutional Neural Networks (CNNs) typically underperform in processing such records. To address this issue, we propose a dense denoising architecture based on Densely Connected Convolutional Networks (DenseNet), introducing macro and micro dense skip connections within the CNN to overcome the lack of direct information flow paths between layers in linearly stacked structures. This network design allows for the integration of low-level and high-level features to reconstruct effective seismic signals, achieving superior reconstruction outcomes. The dense and complex connections provide short paths between the output and intermediate layers, mitigating the vanishing gradient problem and facilitating easier network training. Trained on a high-quality dataset, this method achieved a 24.0578 dB signal-to-noise ratio (SNR) improvement on synthetic seismic records and effectively suppressed noise while recovering coherent signals in real seismic data.

2. When performing seismic data denoising tasks under extremely low SNR conditions, CNN architectures face limitations in capturing long-range correlations and global consistency of signals, making complete signal recovery challenging. This study introduces a model based on the Transformer architecture for seismic signal denoising. Utilizing self-attention mechanisms, the model overcomes the limitations of CNNs in feature extraction, effectively capturing long-range features within seismic signals. To ensure the Transformer architecture's generalizability in real seismic record denoising, we designed a two-phase training strategy combining masked self-supervised pre-training and fine-tuning with large-scale unlabeled seismic datasets, followed by fine-tuning with paired "noisy-clean data". This approach achieved an average of 21.53 dB SNR improvement on synthetic seismic records under extremely low SNR conditions and effectively attenuated complex seismic noise while recovering faint signals in

various real records.

3. Deep learning methods can also be used to directly estimate velocity models from the seismic data, in addition to their application in denoising tasks. However, most CNN-based inversion network designs tend to extract only local features, ignoring the spatial correlations among different features reflecting the same interface information. Additionally, the time-varying nature of seismic data poses challenges for feature extraction in CNN architectures with shared weights. Addressing these issues, our research developed a new inversion network model based on the Transformer architecture. By employing self-attention mechanisms in an encoder-decoder structure, the model captures long-range correlations in seismic data signals, providing more distant features for constructing velocity models. Numerical experiments with synthetic models validated the effectiveness and reliability of the proposed method. Compared to traditional Full Waveform Inversion (FWI) and existing CNN-based velocity inversion methods, our approach showed greater consistency with the target in terms of velocity values, subsurface structures, and geological interfaces.

4. Traditional seismic inversion methods, like FWI, heavily rely on initial velocity models and are based on iterative optimization. In contrast, deep learning-based inversion algorithms can generate relatively effective velocity models. Building on this, we propose a hybrid inversion strategy that combines data-driven deep learning inversion models with physics-driven FWI methods. By inferring initial velocity models from raw seismic data using an improved feature-aligned Transformer inversion network and optimizing with FWI methods iteratively, the hybrid strategy ensures the correctness of inversion results. This mixed approach maximizes the advantages of both data-driven and physics-driven methods. Tests on synthetic and real seismic data demonstrated that the hybrid inversion strategy significantly improves the accuracy of inversion results and can be effectively extended to real seismic data velocity inversion.

Keywords:

Seismic exploration data denoising, velocity model inversion, deep learning, self-supervised pretraining, multi-head self-attention (MSA)

目 录

第 1 章 绪论.....	1
1.1 课题的研究背景及意义.....	1
1.2 国内外研究现状.....	4
1.2.1 地震背景噪声分类及特征.....	4
1.2.2 地震数据去噪技术研究现状.....	5
1.2.3 地震数据速度反演技术研究现状.....	8
1.3 研究内容.....	11
1.4 论文的组织结构.....	14
第 2 章 基本原理与方法概述	17
2.1 引言.....	17
2.2 地震勘探方法介绍.....	18
2.2.1 地震勘探的方法原理及基本流程.....	18
2.2.2 地震波传播规律.....	20
2.3 深度学习基本原理.....	23
2.4 全波形反演基本理论.....	29
2.4.1 有限差分正演模拟.....	30
2.4.2 时间域梯度计算.....	35
2.4.3 速度模型迭代更新.....	39
2.5 本章小结.....	41
第 3 章 基于密集连接卷积神经网络的地震数据去噪算法	43
3.1 引言.....	43
3.2 DnDenseNet 地震数据去噪算法.....	46
3.2.1 宏观密集网络结构.....	47
3.2.2 微观密集连接块结构.....	47
3.2.3 数据集设置.....	48
3.2.4 去噪原理.....	51
3.3 算法性能验证.....	53
3.3.1 模拟记录处理结果.....	54
3.3.2 实际记录处理结果.....	60
3.4 本章小结.....	64
第 4 章 基于自监督学习预训练 Transformer 的地震数据去噪算法	67
4.1 引言.....	67
4.2 SDT 基本理论	69
4.2.1 SDT 网络结构	70

4.2.2 自监督预训练及微调算法.....	74
4.2.3 数据集构建.....	78
4.3 算法性能验证.....	81
4.3.1 实验数据集及实验设置.....	81
4.3.2 模拟实验结果与分析.....	83
4.3.3 实际实验结果与分析.....	89
4.4 算法分析讨论.....	92
4.4.1 网络的通用性验证.....	92
4.4.2 去噪网络的归因分析.....	96
4.4.3 去噪网络的消融实验分析.....	98
4.4.4 预训练对结果的影响.....	100
4.5 本章小结.....	100
第 5 章 基于 Transformer 的地震速度模型反演算法	103
5.1 引言.....	103
5.2 地震速度模型反演 Transformer 算法.....	106
5.2.1 深度学习速度模型反演问题定义.....	106
5.2.2 深度学习速度反演任务数据集构建.....	107
5.2.3 速度模型反演网络 SVIT.....	111
5.2.4 速度模型反演网络 SVIT 方法原理.....	117
5.3 算法性能验证.....	119
5.3.1 实验设置.....	119
5.3.2 实验结果与分析.....	120
5.3.3 算法分析讨论.....	134
5.4 本章小结.....	136
第 6 章 基于特征对齐 Transformer 地震速度混合反演策略	139
6.1 引言.....	139
6.2 特征对齐 Transformer 混合反演策略.....	142
6.2.1 深度学习速度反演任务数据集构建.....	142
6.2.2 特征对齐 Transformer 网络结构.....	144
6.2.3 混合反演策略研究.....	147
6.3 算法性能验证.....	150
6.4 本章小结.....	158
第 7 章 总结与展望	159
参考文献.....	163

重要英文缩写解释

英文缩写	英文全称	中文解释
BN	Batch Normalization	批归一化层
CAB	Channel Attention Block	通道注意力块
CNN	Convolutional Neural Network	卷积神经网络
Curvelet	Curvelet Transform	曲波变换
DAS	Distributed Acoustic Sensing	分布式光纤传感技术
DenseNet	Densely Connected Convolutional Networks	密集卷积网络
DnCNN	Feedforward Denoising Convolutional Neural Network	前馈去噪卷积神经网络
EMD	Empirical Mode Decomposition	经验模态分解
FCNVMB	Fully Convolutional Neural Network for Velocity Model Building	全卷积速度反演网络
FFN	FeedForward Network	前馈网络
FWI	Full Waveform Inversion	全波形反演
GAN	Generative Adversarial Network	生成对抗网络
GPU	Graphics Processing Unit	图形处理单元
IALM	Inexact Augmented Lagrangian Multiplier	不精确增广拉格朗日乘子
LN	Layer Norm	层归一化层
MAE	Mean Absolute Error	平均绝对误差
MSE	Mean-Square Error	均方误差
MSSIM	Mean Structural Similarity	平均结构相似性
PML	Perfectly Matched Layer	完全匹配层边界条件
RED-Net	Residual Encoder-Decoder Networks	残差编码器-解码器网络
ReLU	Rectified Linear Unit	修正线性单元
ResNet	Residual Network	残差网络
RPCA	Robust Principal Component Analysis	鲁棒主成分分析
Shearlet	Shearlet Transform	剪切波变换
SNR	Signal-to-Noise Ratio	信噪比
SSA	Singular Spectrum Analysis	奇异谱分析

续前表：

英文缩写	英文全称	中文解释
SSIM	Structural Similarity	结构相似性
SSL	Self-Supervised Learning	自监督学习
SW-MSA	Shifted Windows Multi-Head Self-Attention	基于移位窗口进行 自注意力计算
VSP	Vertical Seismic Profile	垂直地震剖面测量
VMD	Variational Mode Decomposition	变分模态分解
W-MSA	Windows Multi-Head Self-Attention	基于窗口的多头自 注意力
WNNM	Weighted Nuclear Norm Minimization	加权核范数最小化
WT	Wavelet Transform	小波变换

第1章 绪论

1.1 课题的研究背景及意义

我国经济自改革开放以来,经历了翻天覆地的变化。据世界银行数据,我国的年均GDP增长率自1978年以来一直保持在两位数,这一成就在世界经济史上是前所未有的。我国的经济增长不仅改变了全球经济格局,也极大地推动了国内各行各业的发展。然而,这种高速增长背后,伴随着对能源特别是油气资源的巨大需求。国际能源署的报告指出,中国已成为全球最大的能源消费国之一,其能源消耗量在全球占据了显著的份额。油气资源在国家发展中扮演着重要角色。这不仅因为它们工业生产的基础能源,更因为它们与国家的政治安全、经济独立以及社会稳定息息相关。近年来,我国政府深知外部依赖所带来的风险,大力推动国内油气资源的勘探与开发。根据中国国土资源部的数据,自21世纪初以来,我国在油气勘探和开发上的投资增长了数倍,特别是在西部大开发战略的推动下,西北地区的油气勘探取得了突破性进展。地震勘探作为探查油气资源的关键工具,在我国的能源勘探与开发中扮演着重要角色^[1],如图1.1所示,以陆地地震勘探为例,其过程分为三个主要部分:

(1) 数据的采集阶段:这一阶段主要通过激发弹性波(如爆炸或撞击产生的波)并记录这些波在地层中的反射情况。地震波在地层中的传播过程中,会遇到不同的地质结构而产生反射,这些变化被地表的检波器捕捉并记录下来,形成地震记录。这些记录中包含了丰富的地下信息,使研究者能够直接或间接了解地下构造,为油气开采活动提供指导^{[2]-[3]}。

(2) 数据的处理与分析阶段:在数据采集中不可避免地会混入噪声,这些噪声可能会覆盖有效信号,从而降低地震资料的品质。因此,噪声的消减成为处理阶段的关键环节。除了噪声消减,这一阶段还包括数据的校正、过滤、提取有效信号等多个步骤。通过这些处理,可以优化地震资料的质量,为地震解释提供更准确的基础^{[1]-[3],[5]}。

(3) 地震构造的解释阶段:这一阶段主要使用一系列反演成像技术来获取

地下的分布情况，并借助一定的经验知识对地下构造及储层进行解释。地震勘探的每个阶段都非常关键，它们共同构成了一套复杂且精密的系统，用于探索和评估地下油气资源^{[1]-[3]}。

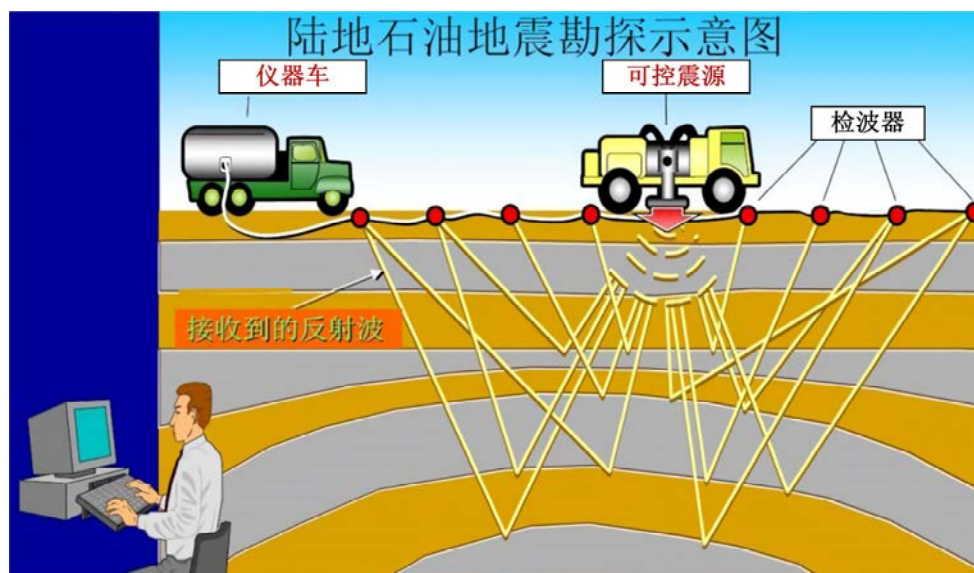


图 1.1 陆地地震勘探示意图

在数据采集过程中，高精度的地震测量设备可以更准确地探测到地下的油气藏。同时，随着信息技术的发展，数据处理和分析的精度也有了显著提升^{[4]-[6]}。例如，通过使用人工智能和大数据技术，可以更有效地从大量地震数据中提取有用信息，从而提高勘探的成功率。中国地质科学院的研究显示，这些技术的应用显著提高了油气资源探测的准确性和效率。为了进一步提高地震数据的质量，我国的研究者们提出了“高信噪比、高分辨率、高保真度”的技术目标。这些目标不仅推动了地震资料处理技术的发展，还引领了整个地震勘探领域的创新。为了实现这些目标，研究者们开发了一系列高效的噪声消减方法，比如基于机器学习的噪声识别和消除技术，以及基于物理模型的信号重构方法。这些技术不仅提升了数据质量，还极大地提高了地震资料的解读准确性，为油气资源的勘探和开发提供了更为可靠的数据支持。

除了数据的采集和处理外，地震资料的速度模型反演也是一个至关重要的环节。速度模型反演是通过解析地震数据来估计地下岩石的速度分布，这对于准确定位油气藏具有重要意义。通过对地震波速度的精确估计，能够更好地理解地下结构，从而为油气勘探提供重要的地质信息^[7]。随着计算技术的发展，地震资料

的速度模型反演技术也得到了显著提升。现代的速度模型反演方法往往结合了高性能计算和先进的算法,如全波形反演(Full Waveform Inversion, FWI)技术^{[7],[8]}。这种技术能够利用地震波的全波形信息,提供比传统方法更为精确和详细的地下速度模型。同时,人工智能和机器学习在速度模型反演中的应用也日益增多。通过训练算法识别地震数据中的模式和特征,可以更快速、准确地完成速度模型的建立和更新。这种方法不仅提高了反演过程的效率,而且有助于提升模型的准确度。

但是,在全球能源需求日益增长的大背景下,我国地震勘探的范围正日益扩大,勘探的深度和复杂性也在不断增加。原本集中于易开采的浅层地区的勘探,现在正逐渐转向更深层、难以开采的区域。同时,勘探活动也由平原地带拓展至地形更加复杂和恶劣的山地、沙漠等区域^[4]。这些地质环境的变化,使得在处理收集到的地震勘探数据时面临着前所未有的挑战。这些数据往往质量较差,包含了各种复杂且随机的噪声,这些噪声极大地干扰了有效信号的提取和识别,严重影响了数据的后续处理工作^{[9],[10]}。此外,不同地质环境下速度模型的差异性要求反演算法具备较高的适应性和灵活性,以应对各种复杂情况^[9]。尤其是对于全波形反演技术来说,选择合适的初始速度模型至关重要。再者,速度模型反演的结果受限于所采用的地震波传播理论和模型。不同的理论和方法可能会得到不同的反演结果,这要求研究人员在选择方法时必须谨慎,以确保结果的准确性和可靠性^[10]。最后,地震数据中的非线性特性使得速度模型反演变得更加复杂。非线性问题通常难以用传统的线性方法解决,需要开发新的算法和技术来处理这些复杂情况^[9]。因此,随着地震勘探工作的不断深入,地震数据的前期噪声处理及速度模型反演面临的问题和挑战也同时在不断增加。这也要求研究者不断创新和改进技术方法,以提高数据处理的精确度和效率。

本文立足于实际,为了解决上述问题,以“基于深度学习的地震勘探数据去噪及速度模型反演技术研究”为选题,旨在设计提高地震勘探中数据处理的效率与准确性的算法,克服实际应用中的瓶颈问题,通过对深度学习技术的研究与应用,为地震勘探中数据噪声压制和速度反演两个环节提供高效、精准的解决方案。

1.2 国内外研究现状

1.2.1 地震背景噪声分类及特征

在地震勘查领域,使用的检波器能够捕捉到观测点周围的大量震动数据^{[1]-[3]}。这些数据中,对于油气探测有价值的被称为有效波或信号;而那些干扰地震探测、妨碍有效信号识别、以及污染信号的部分,一般称之为噪声^[2]。当然,在地球物理探测领域,尤其是在反射地震法的应用过程中,由于技术的多元化和地质问题的复杂性,某些波在特定的探测场景中可能成为关键的信息来源,而在其他场合则可能被认定为噪声,如面波,在本文中不对这些特定的技术进行讨论。

地震数据中的噪声可以从多个维度进行分类,其中包括:

基于出现模式,噪声可分为规律性较强的相干噪声和分布广泛、无明显规律的随机噪声。相干噪声因其较为一致的特性(如稳定的视速度和频率)而容易识别,同时也因其高能量特性而容易降低数据质量。而随机噪声的不确定性和广泛分布使其难以控制^[11],对数据也有着较大的干扰。

通过传播方式,不同类型的波,根据其传播特性也可被分为不同类别^{[12]-[13]},如面波、多次波、声波、折射波以及套管波。

依据频谱属性,可以将噪声分为低频、高频和特定频率的干扰,这依赖于频率分布的特点。

根据来源分类,可区分为环境噪声和震源噪声,前者由如风、人类活动等外界因素产生^{[10]-[13]},后者则直接来源于震源活动^[14]。

下面探讨几种常见的地震噪声类型:

声波噪声:该类噪声在使用炸药或重锤等震源时表现显著,其速度接近空气中的声速,特点为高频率和长持续时间^{[1],[2]}。

面波:这种波沿地面传播,其特征为传播速度相对一致且持续时间较长,能量在垂直方向上迅速减少而在水平方向上缓慢衰减,形成的振幅大的波形可能掩盖有价值的信号^[14]。

多次波:由于地下强反射界面的反复反射形成,这类波与初次反射波在性质上相似,但具有更低的视速度和更长的旅行时间^{[1],[2]}。

伴随波：人造震源产生的地震波在地面的反射导致的波，这些波的传播和反射增加了数据的复杂度，使其与常规反射波在多个特性上极为相似，难以区分[10],[11]。

频率干扰：在不同介质中，如沼泽或坚硬岩石中地震波的传播会引起特定频率范围内的干扰，这包括低频和高频干扰，其分布广泛，给信号识别带来挑战[10],[11]。

微震：自然界和人类活动等因素引起的微小震动，这些活动导致的力度变化大，频率范围宽，主要受环境影响，频率最高可达 150Hz^{[10],[11]}。

综合而言，地震噪声的多样性和复杂性要求采用精细化的分析和处理技术，以提升地震数据的可用性和解析力。通过深入研究和应用先进的噪声抑制技术，将那些原本可能被忽略的数据转化为可靠的地质信息，对地震勘探的成功至关重要。

1.2.2 地震数据去噪技术研究现状

地震勘探作为一种有效的地下结构探测手段，被广泛应用于油气资源的勘查。在这一过程中，科研工作者们通过地震波的传播特性，能够解析出地下地层的构造，从而为油气资源的发现提供科学的指引和依据。然而，地震勘探过程中所遇到的一个主要挑战便是随机噪声的干扰。这些噪声通常来源于各种不可预测的环境因素，如自然地震、人为活动造成的震动等，它们会对地震信号产生影响，导致采集到的地震资料在质量上出现下降^{[10],[11]}。为了解决这一问题，噪声消减技术应运而生，成为地震数据处理中不可或缺的一环。在噪声消减技术的发展历程中，研究人员不断探索和创新，不断追求更有效地剥离有用信号和噪声。这些算法大致可以分为以下几类：

基于时频的去噪方法利用有效信号与噪声的时频差异性进行信噪分离，典型的方法包括带通滤波^[14]、时频峰值滤波^{[15]-[17]}等，其基本原理是将含噪的时间序列通过时频变换转入时频域，不断修改所得到的时频系数，以衰减和噪声相关的系数，并找到信号系数的估计值。将修正后的系数进行逆变换，重构有效信号。从这个过程中可以看出，时频方法的去噪性能依赖于精确的滤波参数，通过人为

手动调整参数往往不能保证精准度，且费时费力。

模态分解是时频分析中常用的一种信号处理方式，常应用于复杂噪声抑制。这类方法包括经验模态分解（Empirical Mode Decomposition, EMD）^{[18]-[20]}、变分模态分解（Variational Mode Decomposition, VMD）^{[21],[22]}等，其基本原理是将复杂信号分解为有限个包含原始信息在不同时间尺度上的局部特征的固有模态函数，选取反映信号特征的模态进行去噪记录的重构，从而实现信噪分离。在这个过程中，各个模态之间不可避免地存在混合的问题。当地震噪声和有效信号在多个频率范围重叠时，模态混叠将更加严重，使得这些方法去噪效果不佳。

与基于时频的方法类似，稀疏变换方法被认为是一种更准确有效的方法，如小波变换（Wavelet Transform, WT）^{[23],[24]}、剪切波变换（Shearlet Transform）^{[25]-[27]}、曲波变换（Curvelet Transform）^{[28]-[30]}和字典学习^{[31],[157],[158]}方法，已成功应用于地震勘探数据处理。它们可以将原始地震数据转换为不同的稀疏域，利用信号和噪声之间转换后的稀疏性差异来实现信号识别，然后通过使用少量参数精确地表示所需的信号。尽管稀疏变换方法具有优异的噪声衰减性能，但它们对阈值常数比较敏感，影响去噪精度，带来了一些额外的时间成本。

基于低秩的方法假设理论上干净的地震数据是低阶结构，噪声污染会增加其秩，因此可以通过降秩来衰减噪声，并恢复信号。这些方法包括鲁棒主成分分析（Robust Principal Component Analysis, RPCA）^[32]，Cadzow 滤波^[33]和奇异谱分析（Singular Spectrum Analysis, SSA）^[34]。它们因为在处理复杂地震勘探数据中具有优异的性能而引起了人们的关注，但是，去噪后的结果中仍然有一些可见的残余噪声，性能可以进一步提高。另外，其他一部分经典方法还包括 F-X 反褶积^{[35]-[39]}、Random 域滤波^{[40]-[44]}、 $\tau-p$ 变换^[45]、形态学滤波^{[46]-[48]}等。

近年来，随着图形处理单元（Graphics Processing Unit, GPU）的发展及并行计算的推广，深度学习得以快速发展并在不同领域的数据处理中获得了广泛应用。在图形领域的广泛应用包括图像识别、图像超分辨率，图像去噪，图像分割等，并发展出 VGG^[49]、U-Net^[50]、ResNet^[51]、DenseNet^[52]等一系列结构和生成对抗网络（Generative Adversarial Network, GAN）^[53]、自监督学习（Self-Supervised Learning）^{[54],[55]}等一系列框架。在图像领域普遍的成功也激励了深度学习在地震

数据处理领域的应用。在去噪领域,已经有了很多丰富的方法被应用如前馈去噪卷积神经网络(Feedforward Denoising Convolutional Neural Network, DnCNN)^{[56],[57]}、U-Net^[58]和生成对抗网络(Generative Adversarial Networks, GAN)^[59],被应用于复杂地震噪声衰减。基于卷积神经网络(Convolutional Neural Network, CNN)的去噪模型可以在训练过程中对参数进行优化,以提取地震数据和期望信息之间的隐藏模式和多维映射关系,它们取得了令人印象深刻的结果。

深度学习的方法不仅适应了现代地震勘探的自动化和智能化需求,而且解决了传统方法在处理大量多维数据时的局限性。随着技术的不断进步,基于深度学习的去噪技术在地震数据处理中的作用将越来越显著。Yu 等人利用 CNN 构建了一般地震噪声衰减的统一框架,包括随机噪声、地滚波和多次波,尽管这种统一框架在去噪质量和自动化程度方面都优于传统方法,但也存在缺乏训练数据集、超参数不合适以及训练和测试数据集之间存在明显差异导致模型性能差等问题^[80]。Yang 等人提出了一种结合多个双层卷积块的改进残差 CNN,该网络训练效率高,对地震随机噪声具有良好的抑制性能^[169]。为了克服分布式光纤声学传感地震数据低信噪比的挑战,Dong 和 Li 基于去噪器和鉴别器之间的对抗训练提出了卷积对抗去噪网络,该网络可以有效抑制 DAS 噪声并恢复微弱的上行反射信号^[170]。为了加强对弱信号的保留,Zhong 等人提出了一种结合分层结构的新型多尺度 CNN 来衰减地震随机噪声,实验表明这种提出的多尺度 CNN 在弱信号恢复方面比现有的三种基于 CNN 的方法性能更好,但采用多尺度结构会导致计算成本的增加^[117]。Liu 等人提出了一种基于相似性自学习的方法来解决缺少无噪声地震数据做标签时的去噪问题,并开发了一个专门的工作流程以充分利用含噪地震数据的自相似性来构建训练对,然后训练去噪模型,实验表明了所提出的方法用于地震数据去噪的有效性和稳定性^[171]。同时,Liu 等人提出一种噪声样本到噪声样本的无监督学习方法,使用单个含噪地震数据训练去噪网络,实验表明网络可以学习有效信号的特征并抑制不可预测的随机噪声,这些方法通过训练模型学习地震信号和噪声的特征,实现更为智能和高效的噪声消减^[172]。这些研究成果都表明了神经网络等机器学习技术的应用,不仅提高了噪声消减的准确度,也为地震资料处理带来了新的研究方向和突破。但是,这类研究中使用的结构多是

CNN 结构,在捕获长程依赖性方面存在局限性。具体来说,CNN 主要依靠滑动卷积核提取不同位置的特征,并将这些特征在空间上进行局部的加权和融合,从而得到具有位置不变性和层次结构的特征表示。卷积算子是局部算子,为了提取远程特征,CNN 需要堆叠卷积核形成更大的感受野。这种运算模式限制了其从复杂地震数据中捕捉远程特征的能力,而地震记录中信号往往是连续的、较长范围的。另外,为了有效地重建信号,对不同噪声模式的准确感知起着关键作用。由于卷积运算的固有性质,基于 CNN 的一个潜在限制是仅依靠在输入数据的局部视图中学习的特征来识别噪声模式(如几类噪声混合的模式)。然而,在执行整个地震记录的恢复时,获得整个数据的全局感知是至关重要的。在全局视图中,噪声与信号的关系往往更容易学习。为了解决上述问题,研究者们开始探索深度学习中的新方法,以弥补 CNN 在处理长程依赖性方面的不足。这其中一个显著的方法是引入了基于自注意力机制(Self-Attention)的模型,如 Transformer^{[60],[61]}。Transformer 模型通过自注意力机制能够捕捉到数据中长距离的依赖关系。在自注意力机制中,模型会计算输入数据中各个元素间的关系,从而能够关注到整个数据集的全局信息。这种方式在处理地震数据时尤为有效,因为它可以更好地捕捉地震信号中远程和复杂的模式,提升去噪和信号重建的质量。此外,为了进一步提高地震数据处理的效率和精度,研究者们还尝试将深度学习与其他先进技术相结合。例如,集成学习和迁移学习在地震数据处理中的应用逐渐增多。集成学习^[62]通过组合多个模型来提高预测的准确性和稳健性。而迁移学习^[63]可以利用在其他任务上训练的模型,快速适应新的地震数据处理任务,节省了大量的训练时间和资源。随着深度学习技术的不断发展和完善,结合自注意力机制、集成学习、迁移学习等先进技术,地震数据处理的效率和准确性大幅提升,未来这些技术的结合能够不断提升地震数据去噪和重建的质量。

1.2.3 地震数据速度反演技术研究现状

地震勘探通过人工激发、观测和分析地层中传播的地震波来查明地下地层形态和性质。其中,地震波在地层中的传播速度是地震勘探中最重要的参数之一,其对于高精度的地震成像及预测储层岩性具有重要意义^{[64],[65]}。地震信号处理后

续的一系列步骤高度依赖于通过反演获取精确的速度模型,比如,叠前深度偏移成像也极大地依赖于地层速度模型的精度^[64]。随着地震勘探环境与条件的复杂化,对速度估计的精度和效率都提出了越来越高的要求^[65]。如何快速而精确地获得速度模型成为一个重要的问题。鉴于速度模型的重要性,研究者们开发了很多求解它的反演方法。传统反演方法中,偏移速度分析方法以及走时层析反演方法只能得到宏观速度场,也就是低频部分,高分辨率的速度模型则无法获得,这严重限制了其实际应用^{[66],[67]}。地震全波形反演^{[68]-[70]}是近年来备受关注的方法,其利用地震波场的运动学和动力学信息建立地层速度模型,能够精确刻画模型细节,是近来研究的热点。具体来说,全波形反演方法通过迭代来最小化地震数据和合成数据间的差异来重建速度模型。但是,其严重依赖于初始模型来避免局部最小值^[71]。同时,由于噪声污染等因素,全波形反演方法有着极大的非唯一性,且存在局部极小问题。对于全波形反演方法的局部极小问题,诸多的改进策略被提出以解决它,例如转换到其他域的处理^{[72]-[76]}。另外,全波形反演方法也在多个方面得到拓展,如应用于粘弹性介质^[77]和联合反演中^{[78],[79]}。但是,依赖初始模型的问题依旧需要新的方法去解决。同时,迭代优化带来巨大的计算量,使得耗时较多,难以广泛应用到实际数据中,这也成为全波形反演方法一个值得注意的问题。

近年来,随着硬件计算能力的快速增长,在大规模数据集上学习的深度学习模型无论在地震记录的去噪^[80]、反演^[81]、解释^[82]等任务上都展现出优于传统方法的有效性。随着模型的表示能力的不断提高,数据驱动方法为速度模型反演提供了新思路,使得这部分的研究更受瞩目。Roth 等人首次使用神经网络方法从地震数据中反演了一维模型^[83],Araya 等人通过神经网络从叠前数据计算的速度谱立方体中进行速度模型重建^[84],Wu 和 Lin 通过使用解编码器 CNN 实现速度模型的构建^[85]。Yang 和 Ma 提出的基于 U-Net 的 FCNVMB 方法(Fully Convolutional Neural Network for Velocity-Model Building, FCNVMB)直接从从叠前数据重建速度模型,在盐丘速度模型上取得了极好的效果^[86],迁移学习方法也被应用到速度模型反演任务以进一步增加其普适性场景应用^[87]。Li 等人提出了 SeisInvNet 模型进一步提高了使用深度学习进行速度模型反演时的效果^[88],并提供了详细的

机制研究, Wang 和 Ma 提出了一种由全连通层和卷积层组成的速度模型反演网络并命名为 VMB-Net, 将多炮数据映射到速度模型上^[89], 并取得了很好的效果。此外, Zhang 等人提出了一种基于生成对抗性网络的反演网络并命名为 VelocityGAN, 它被端到端地训练以学习从原始地震数据到速度模型的映射函数^[90], 这些都展现了深度学习方法在地震速度模型反演中的光明前景。

然而, 现有的深度学习反演网络结构多是基于 CNN, 尽管有着非常好的效果, 但 CNN 应用于反演时依旧存在不少潜在问题, 多数的 CNN 会在提取地震记录里信号特征的过程中提供适当的归纳偏差, 具体来说, CNN 中的卷积运算使用了两个重要的空间约束以提高特征的学习和提取能力: 一是权重共享机制, 使得卷积层所提取的特征具有平移不变性, 模型更在乎对目标任务决定性的特征是否存在, 这样的特性使得其在去噪这种局部回归任务以及目标识别等任务时可以产生极好的作用。二是卷积算子, 卷积的特征图具有局部敏感性, 这使得其有着强大的特征提取能力, 但随之而来的弊端是每次卷积操作只提取到局部信息。CNN 的归纳偏差使得它缺乏对输入数据的整体把握, 难以提取全局数据之间的长距离特征。以上特点成为 CNN 应用于速度反演任务时的潜在的问题。因此, 开发更适用于地震数据反演的神经网络结构是一个重要的研究热点。

此外, 基于深度神经网络实现的速度模型估计方法主要分为学习和推断两个步骤, 具有一次学习 (建立特征数据和速度模型之间的映射关系), 即可在不同数据上进行推断 (反演) 的能力, 相较于传统基于迭代实现的反演成像方法, 其在计算效率上具有一定优势。但是, 深度学习方法的泛化性不强、可解释性较弱等缺点, 是制约其应用于实际数据反演中的重要原因。目前深度学习技术大多数的工业应用, 还是作用在降低人力、经济成本以及提高效率方面。但是, 在深度学习框架下进行地震信号处理、成像与参数建模是勘探地震学发展的必然趋势。未来比较有前景的方向可能是将数据驱动的深度学习方法与物理驱动的常规方法相结合。对于正过程还不清晰的问题, 应用数据驱动方法。对于正过程已经清晰的问题, 物理驱动的方法可以更加有效的把握数据规律。因此, 将数据驱动与物理驱动相结合极大可能是未来研究的方向。

1.3 研究内容

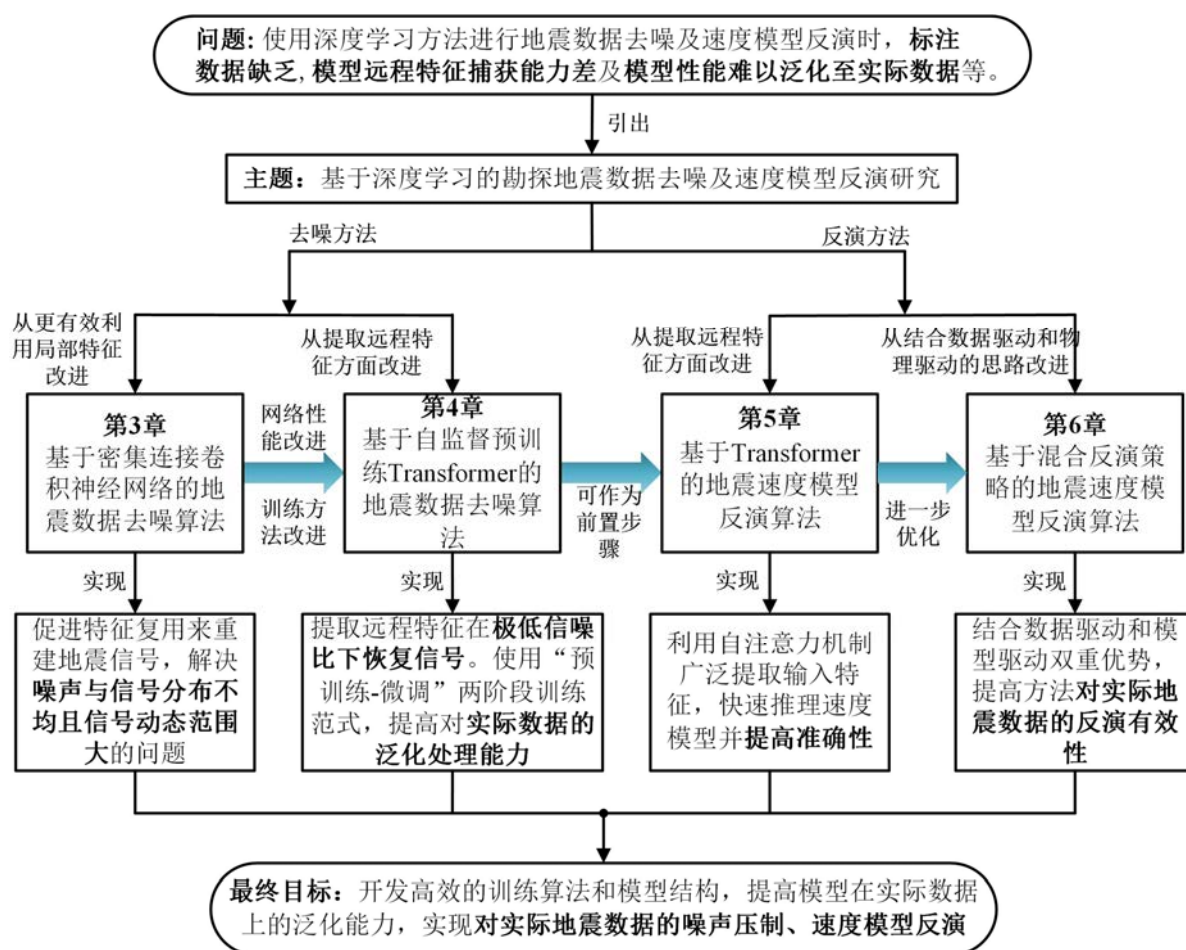


图 1.2 本文的研究思路 and 主要研究内容

本文面向在实际数据上利用深度学习技术进行地震数据去噪及速度模型反演任务时遇到的问题, 通过一系列网络架构和方法策略改进, 充分发挥深度学习数据驱动方法在这两项任务中的优势, 以完成“基于深度学习的勘探地震数据去噪及速度模型反演研究”为目标, 层层开展对具有低信噪比地震数据的去噪算法及能够有效应用至实际案例中的速度模型反演算法的研究。本文的研究思路与主要研究内容如图 1.2 所示。其中, 第 2 章介绍了基本原理与方法, 包括地震勘探方法、深度学习基本原理和全波形反演基本理论。第 3 章研究基于密集连接卷积神经网络 (Densely Connected Convolutional Networks, DenseNet) 的去噪架构, 利用具有密集的跳跃连接的 CNN 结构来提高对地震记录中不同信噪比区域的处

理能力,同时解决网络加深时的优化困难,提高 CNN 应用于地震数据去噪任务时的训练效率和去噪效果。第 4 章针对 CNN 架构的局限性,提出了基于 Transformer 架构的地震数据去噪模型,通过自注意力机制,该模型有效捕捉地震信号的长程特征,实现了在复杂噪声环境中对地震数据的有效去噪。通过设计自监督预训练流程,网络能够有效在预训练阶段捕获更多信号特征,保证模型在实际数据去噪应用中的泛化性。第 5 章针对速度模型反演任务的挑战,基于 Transformer 架构开发了一种新的反演网络模型,捕捉信号的远程相关性,利用更多的远程特征,提高了速度模型的反演准确性。第 6 章针对实际应用中速度模型反演任务的挑战,提出了一种结合深度学习和常规全波形反演方法的混合反演策略,提出了特征对齐的 Transformer 网络估计初始速度模型,再通过 FWI 进行优化,显著提高了反演结果的准确性,并在合成和现场地震数据上的应用均取得了良好效果。具体内容如下:

(1) 基于密集跳跃连接卷积神经网络的地震数据去噪算法研究

基于深度学习的方法在地震数据去噪问题上已经能够有着不错的性能,尤其是去噪卷积神经网络的使用可以有效的提高地震记录的信噪比。实际中的地震资料中往往存在信号和噪声分布不均的问题,这对线性堆叠的 CNN 结构造成巨大挑战,常规 CNN 在逐层传递信息过程中,可能发生信号损失,这会限制卷积神经网络在实际地震数据去噪的应用。为了解决这一技术难题,本研究使用一种基于 DenseNet 的去噪架构,通过在卷积神经网络中引入密集复杂的跳跃连接,从而在增加网络层数的同时充分促进特征复用。同时在网络中每层的特征图都被传入后续的层,使得网络能够组合低级特征和高级特征,针对不同区域的信噪情况,灵活地重用和组合特征,来重建地震有效信号。同时,密集复杂连接提供了输出层与中间各层之间的短路径,从而减缓了梯度消失问题,解决了网络加深时的优化困难,使网络更容易训练。通过使用模拟信号和采集到的实际噪声制作了高质量的训练集,用来训练网络,在模拟和实际地震记录的处理中,训练好的网络拥有比传统方法和常规卷积神经网络更好的处理效果。

(2) 基于自监督预训练 Transformer 的地震数据去噪算法研究

CNN 在分析含有长距离信号特征的复杂地震数据时具有局限性,难以应对

低信噪比场景。针对卷积算子在捕获地震信号远程特征方面存在局限性的问题,本文提出了一种替代方案——用于地震信号处理的 Transformer 模型。在网络结构设计中,本文结合通道注意力和自注意力机制,使得网络能够提取更多的远程特征来用于地震信号的重建。由于 Transformer 结构相比于 CNN 结构归纳偏差更弱,因此需要在一定数量的数据上进行预训练,为此本文设计了一种自监督预训练的策略,以充分利用现场地震数据,通过预训练进一步提升 Transformer 性能。在此基础上,对基于 Transformer 的地震数据去噪网络形成一种“预训练-微调”两阶段的训练范式。与传统方法和最近提出的先进的去噪网络相比,结果表明所提出方法具有复杂的地震噪声衰减和弱信号幅度保持的优点。该方法在两类实际地震记录的去噪任务中取得了较好的效果。

(3) 基于 Transformer 的地震数据速度模型反演算法研究

除了地震数据的噪声问题以外,获得精确的速度模型也是地震勘探中重要的任务。传统反演方法中 FWI 存在高度依赖初始模型,耗时且计算量巨大等问题,近年来逐渐发展的深度学习方法基于数据训练而非先验知识假设从共炮点数据中直接推理速度模型,有望成为另一种解决方案。但是,多数深度学习反演网络基于卷积神经网络构建,由于卷积运算的固有性质,这些方法的一个潜在限制是速度模型只能依靠在地震记录的局部视图中学习的特征来预测,而忽略了整个视图的全局感知,全局特征对于反演更为重要。为此,本文基于 Transformer 结构提出一种名为 Seismic Vector Inversion Transformer (SVIT) 的网络结构来利用地震数据重建速度模型。该模型是一种基于自注意力机制的 Transformer 编码器-解码器架构,是一种自相关强表征模型,通过多尺度下利用自注意力机制来广泛提取输入道中的特征,并在此基础上推理速度模型。SVIT 的优势就在于利用自注意力的方式来捕获全局的上下文信息,对各种反射事件建立起远距离关系,这些远距离的依赖对于重建速度的不同区域更为重要,从而提取出更为稳健的特征用于反演任务,使得模型泛化能力大大加强。自注意力机制中所生成的特征图不会像卷积计算一样具有空间限制。相反,能够根据反演目标和网络中该层的位置来学习最合适的归纳偏差,能够极大地规避 CNN 在反演任务中的缺陷。在多种不同类型地质结构的模拟数值实验中,验证了本文方法的结果在速度值、地下结

构、界面等方面性能更加出色。

(4) 基于混合反演策略的地震速度模型反演算法研究

基于迭代优化的常规地震反演方法如 FWI 往往高度依赖初始速度模型。而基于深度学习的反演算法,虽然可以生成相对有效的速度模型,但是其面对更加复杂的实际场景时往往存在泛化性能差,且反演结果往往缺乏可解释性。为了充分发挥两种方法的优势,解决实际反演过程中遇到的困难,本文提出一种结合深度学习反演模型(数据驱动)和常规 FWI 方法(物理驱动)的混合反演策略。通过使用提出的特征对齐 Transformer 反演网络从原始地震数据中推理估计初始速度模型,并利用 FWI 方法迭代优化获得最终反演结果,保证反演结果的正确性。该策略可以充分发挥数据驱动和物理驱动两类方法的优势,有效互补。本方法在合成数据和现场地震数据上均进行了测试。结果表明,混合反演策略可以显著提高反演结果的准确性,且该流程更加适配实际地震数据反演的任务场景。

1.4 论文的组织结构

本文总共包括六章,组织结构如下:

第 1 章 阐述了研究地震勘探数据处理中去噪算法和速度模型反演算法的背景和意义、国内外研究进展,分析了两类任务中常规算法及近几年发展的深度学习算法存在的问题,并针对这些问题引出本文的研究内容。

第 2 章 介绍了一般地震数据处理的工作流程、深度学习的基础概念和相关算法理论、全波形反演的基础概念和相关算法理论,这些都是本文研究内容涉及到的基础知识。

第 3 章 针对去噪卷积神经网络在处理噪声和信号分布不均且动态范围较大的地震记录时性能不佳的问题,提出一种基于密集连接卷积神经网络的去噪架构,通过在卷积神经网络中引入密集的跳跃连接,充分促进网络中的特征复用,针对不同区域的信噪情况,灵活地重用和组合特征,有效提升去噪性能。

第 4 章 针对在处理低信噪比数据时,卷积神经网络结构在捕获地震信号远程特征方面存在局限性,不利于信号重建的问题,提出了一种用于地震信号处理的 Transformer 模型。在网络结构设计中本文结合通道注意力和自注意力机制,

使得网络能够提取更多的远程特征来用于地震信号的重建。同时设计了一种自监督预训练策略，利用现场地震数据对 Transformer 结构进行充分预训练，从而形成一种“预训练-微调训练”二阶段的训练范式。模拟及实际数据的测试结果表明该方法具有复杂的地震噪声衰减和弱信号幅度保持的优点。

第5章 针对基于 CNN 的反演网络架构远程特征提取范围有限，且地震数据的时变特性也不利于权重共享的 CNN 架构提取特征的问题，基于 Transformer 架构提出一种新的速度模型反演网络。通过使用自注意力机制来捕捉地震数据中信号的远程相关性，从而为构建速度模型提供更多长程特征。与传统的全波形反演方法和现有的基于卷积神经网络的速度反演方法相比，所提出的方法的结果在速度值、地下结构和地质界面方面与目标具有更大的一致性。

第6章 针对迭代优化的常规地震反演方法如 FWI 往往高度依赖初始速度模型、基于深度学习的反演算法面对复杂的实际场景泛化性能差的问题，本文提出一种结合深度学习反演模型（数据驱动）和常规 FWI 方法（物理驱动）的混合反演策略。通过使用特征对齐 Transformer 反演网络从原始地震数据中推理估计初始速度模型，并利用 FWI 方法迭代优化获得最终反演结果，以充分发挥数据驱动和物理驱动两类方法的优势。实际数据中的反演结果表明，混合反演策略可以显著提高反演结果的准确性。

第7章 总结与展望。系统总结了全文的主要工作与贡献，客观分析了本文所提方法的优缺点，指出了需要改进的地方，同时展望了接下来的研究方向。

第2章 基本原理与方法概述

2.1 引言

随着地球科学领域的不断发展,地震勘探作为一种重要的地球物理探测手段在资源勘探、地质研究等领域发挥着关键作用。然而,地震勘探数据常常受到复杂的地下介质影响,导致获取的地震信号混杂着各种噪声,给后续的数据解释和模型反演带来了挑战。为了提高地震数据的质量并准确获取地下结构信息,研究人员越来越关注利用深度学习技术来进行地震勘探数据的去噪和速度模型反演。借助深度学习的强大数据处理和特征学习能力,可以大大缓解地震勘探数据中存在的复杂噪声问题。与传统方法相比,深度学习技术能够更好地识别和分离地震信号与噪声,为速度模型的反演提供了更为可靠的数据基础。同时,也有望高效实时地实现速度模型的快速推理。

但是,对于反演任务来说,全波形反演技术作为一种物理驱动的反演方法,其利用地震波的完整波形信息来精确反演地下介质的速度模型,从而获得高分辨率的地下结构信息,是一种更易被信任的反演方法,也一直是反演领域研究的热点。然而,尽管全波形反演在理论上具有极高的反演精度,它在实际应用中却常常受到数据质量和计算复杂性的限制。深度学习技术的引入,为克服这些挑战提供了新的可能。通过学习大量地震数据中的特征,深度学习模型能够有效地去噪和识别地震信号,从而为全波形反演提供更清晰、更高质量的输入数据。更进一步,深度学习可以直接参与到反演过程中,通过数据驱动的方式为全波形反演建立一个合适的初始模型。这种结合了数据驱动的深度学习和物理驱动的全波形反演的方法,不仅能够提高反演的准确性和效率,也能够扩大全波形反演在复杂地质条件下的适用范围。因此,全波形反演与深度学习技术的结合代表了地震勘探领域的一种未来趋势,两者相辅相成,互相补充。全面掌握这些技术的基本原理和应用方法,是极其必要的。

本章中,将首先介绍地震勘探的基本原理及常见方法,随后介绍深度学习的基本原理,最后详细介绍全波形反演的基本原理。这几种研究的结合对于提高地

震勘探数据处理的效率和准确性，推动地下结构解释和资源勘探具有重要意义。

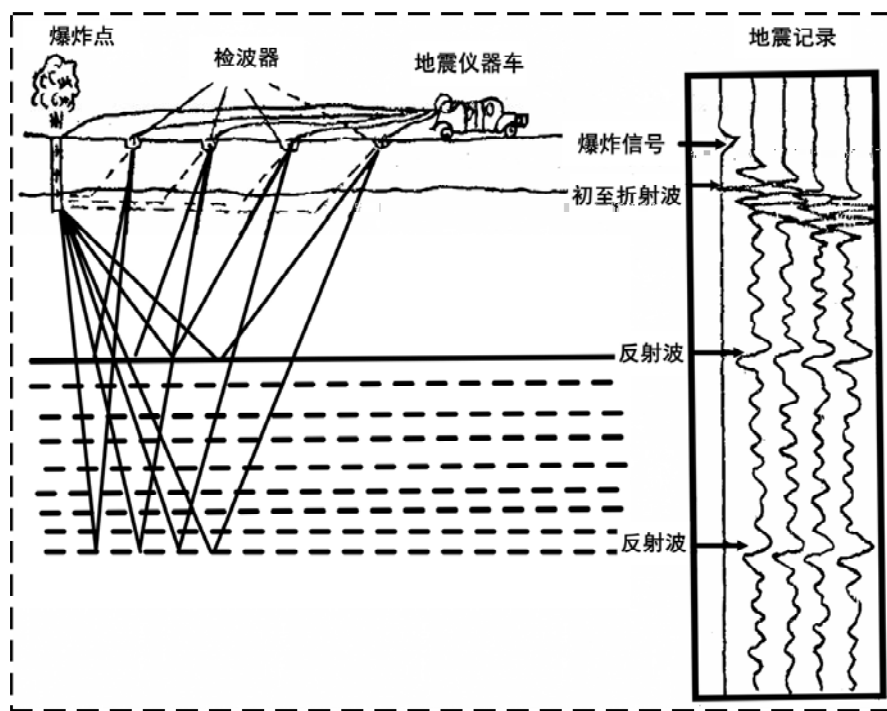
2.2 地震勘探方法介绍

在当前油气勘探领域，地震探测技术因其广泛的应用和在解决勘探难题方面的有效性而备受推崇。本小节中首先对地震探测的基本原理和关键流程进行了简要的介绍。

2.2.1 地震勘探的方法原理及基本流程

天然地震是地球内部发生运动而引起的地壳振动。而地震勘探是通过人为方式（例如利用炸药引发的爆炸或者可控的震源设备）产生地表振动，然后通过特定的观测装置捕捉这些振动信号。通过分析这些振动信号，可以揭示地下结构的特征^[1]。地震波通过地壳层传递，携带着关于地壳结构及其物理性质的信息。其中，地震波的速度取决于介质的密度和弹性模量，而这些物理特性与地壳的组成及状态有密切关联。因此，在地震波传播的研究中，通常通过分析地震波的速度来推断地壳的物理状态。如图 2.1 所示，例如，在一定的直线测量区域内，如果在某一点引爆炸药，所产生的地震波会向下传播。这些波在遇到不同地层的界面时会产生反射。与此同时，地面上的检测设备会记录从各个界面反射回来的波及其引发的振动。通过计算地震波往返于地面与地层界面之间的时间，结合已知的波速，可以计算出地层界面的深度^[2]。

对于二维地震探测来说，通过在地面沿一定的线路布置观测点并分析数据，能够得到反映地下岩层界面深度变化的地震剖面图。这些图上的反射波振幅的振幅极大值的连接线（在地震探测中被称为同相轴）直观地显示了地层界面的轮廓。这些数据显示，由于地层界面的不同，反射波在不同位置的传播时间会有所不同，这导致同相轴出现弯曲。从图 2.1 中可以观察到地层界面为水平形态，检波器在靠近震源的地方，传播时间短，而远离震源的地方传播时间长。由此，相应反射波的同相轴就是弯曲的。通过一定的处理方法，可以通过地震波在不同介质中的传播时间来探测地壳的结构差异。这就是地震勘探的基本原理。

图 2.1 地震勘探原理示意图^[1]

地震勘探利用地震波的传播原理,对地下结构进行探测,以揭示地壳的构造。基于地震波在不同介质中的速度差异,通过对地震波传播速度的测量,可以推断出地下结构的特性。其实施通常包括以下几个流程:

- 1) 通过使用炸药、可控震源或气枪等手段,人为激发地震波;
- 2) 采用动圈型、压电型或数字型接收器及其对应的记录装置来捕捉这些地震波;
- 3) 借助地震数据的处理分析以及综合解释,重塑地震波的传播路径,从而精确映射地下结构。

实践中,运用这套理论探查地下地质结构面临许多挑战。以沙漠环境下的地震勘探为例,地震波的激发不仅会引发来自地层界面的反射波,还会产生多种波。同时,环境噪声也可能被仪器捕获,这些噪声会干扰反射波的准确接收,并可能产生误导。此外,要通过反射波的行进时间来判断地层分界面的深度,关键在于准确测量地震波在各层的速度,这是一个颇具挑战的任务。这两项内容,也是本研究聚焦的重点内容。

2.2.2 地震波传播规律

(1) 波阻抗定义

地震波在地下传播时，其行为受到地下介质物理性质的强烈影响。当地震波从一个介质传播到另一个具有不同物理性质的介质时，其传播路径和速度会发生变化，这一现象是地震勘探中用于探测地下结构的基础。

当地震波遭遇两种不同介质的交界时，如介质 1 与介质 2，它会经历反射和透射的过程，如图 2.2 所示。这意味着，一部分地震波会被反射回原来的介质（介质 1），而另一部分则穿透界面，进入另一种介质（介质 2），这种穿透的波通常被称作透射波。仅当两种介质具有不同的波阻抗时，才会观察到反射和透射现象。波阻抗是介质密度与波速乘积的度量，对于介质 1 和介质 2，其波阻抗分别用 $\rho_1 v_1$ 和 $\rho_2 v_2$ 表示。这里， ρ_1 和 v_1 分别代表介质 1 的密度和波速，介质 2 的相关参数则由 ρ_2 和 v_2 表示。表 2.1 展示了一些典型介质的波阻抗范围值。

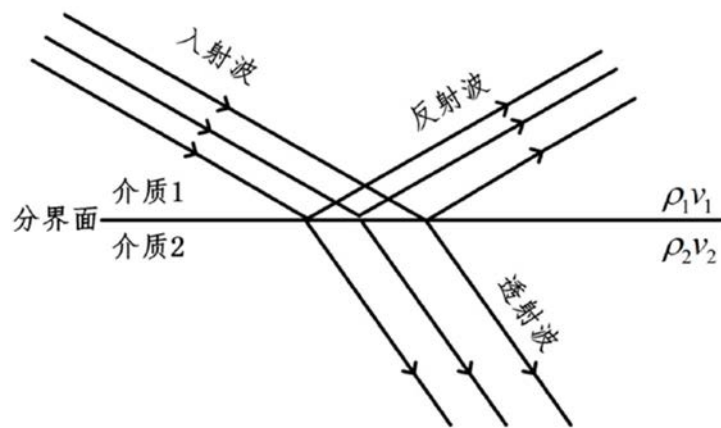


图 2.2 地震波的反射与透射

进一步地，地震波的反射和透射效应可以通过图 2.3 来加以说明，其中地震波的入射、反射和透射角度在不同的情况下有所不同。地震波在达到分界面 P 点时，按入射角 θ_1 沿着 OP 方向传播，其中一部分能量以反射角 θ_1' 被反射形成反射波，而另一部分则以透射角 θ_2 透射，穿过分界面形成透射波。斯奈尔定律描述了

分界面上方介质波速 v_1 、下方介质波速 v_2 、 θ_1 、 θ_1' 、 θ_2 的关系为

$$\frac{\sin \theta_1}{v_1} = \frac{\sin \theta_1'}{v_1} = \frac{\sin \theta_2}{v_2} = k \dots\dots\dots (2.1)$$

其中参数 k 是射线参数。

表 2.1 不同地层介质波速、密度及波阻抗数值范围

介质类型	波速 (m/s)	密度 (g/cm ³)	波阻抗 (g/s×cm ² ×10 ⁴)
土壤	200-800	1.1-2.0	2.2-16
砂层	300-1300	1.4-2.0	4.2-26
黏土	1800-2400	1.5-2.2	27-52.8
砂岩	2000-4000	2.1-2.8	42-112
石灰岩	3200-5500	2.3-3.0	73.6-165
结晶岩	4500-6000	2.4-3.4	108-204

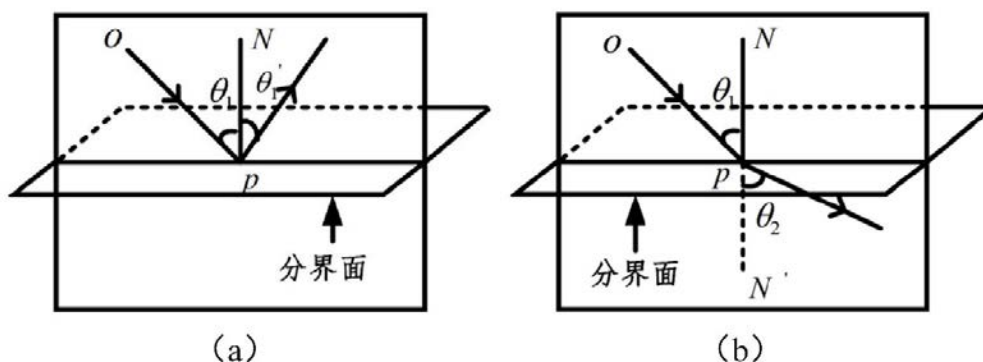


图 2.3 地震波传播示意图 (a) 地震波反射, (b) 地震波透射

(2) 一次反射地震波时距方程推导

在探讨地震波的反射时距方程时, 以地层的理想水平界面为案例, 如图 2.4 所示。设想震源位于 O 点, 而检波器分别位于 D_1 、 D_2 、 D_3 点, 与震源的距离分别为 x_1 、 x_2 、 x_3 。介质的分界面被记作 W_1 和 W_2 , 地震波在这两种介质中的传播速度分别为 V_1 和 V_2 。

时距曲线描述了地震波到达观测点的时间与震源到观测点距离的关系。通过构造震源 O 的镜像点 O' ，并以分界面为对称轴，可以简化反射波路径的计算。以 D_1 点和反射点 A_1 为例，反射波的路径长度 S 可视为 O 到 A_1 再到 D_1 的距离之和。基于几何关系和勾股定理，可知

$$S^2 = (2h)^2 + x_1^2 \cdots \cdots \cdots (2.2)$$

可以推导出观测点 D_1 接收到反射波的时间 t_1 之间的函数关系为

$$t_1 = \frac{S}{V_1} = \frac{\sqrt{(2h)^2 + x_1^2}}{V_1} \cdots \cdots \cdots (2.3)$$

化简后可得 t_1 与 x_1 之间的函数关系如下：

$$t_1^2 = \frac{x_1^2}{V_1^2} + \left(\frac{2h}{V_1}\right)^2 \cdots \cdots \cdots (2.4)$$

其中 $\frac{2h}{V_1}$ 为自激自收时间。

同样的方法可用于计算 D_2 和 D_3 点接收到反射波的时间 t_2 和 t_3 与它们对应的距离 x_2 和 x_3 之间的关系为

$$t_2^2 = \frac{x_2^2}{V_1^2} + \left(\frac{2h}{V_1}\right)^2 \cdots \cdots \cdots (2.5)$$

$$t_3^2 = \frac{x_3^2}{V_1^2} + \left(\frac{2h}{V_1}\right)^2 \cdots \cdots \cdots (2.6)$$

最终，得到地震波一次反射的时距方程如下：

$$t^2 = \frac{x^2}{V_1^2} + \left(\frac{2h}{V_1}\right)^2 \cdots \cdots \cdots (2.7)$$

其中 t 和 x 分别表示地震波传播时间及炮检距。该方程揭示了传播时间 t 与炮检距

x 之间呈双曲线关系。这意味着在地震数据的共炮点记录中,同相轴的轮廓通常与双曲线相似。

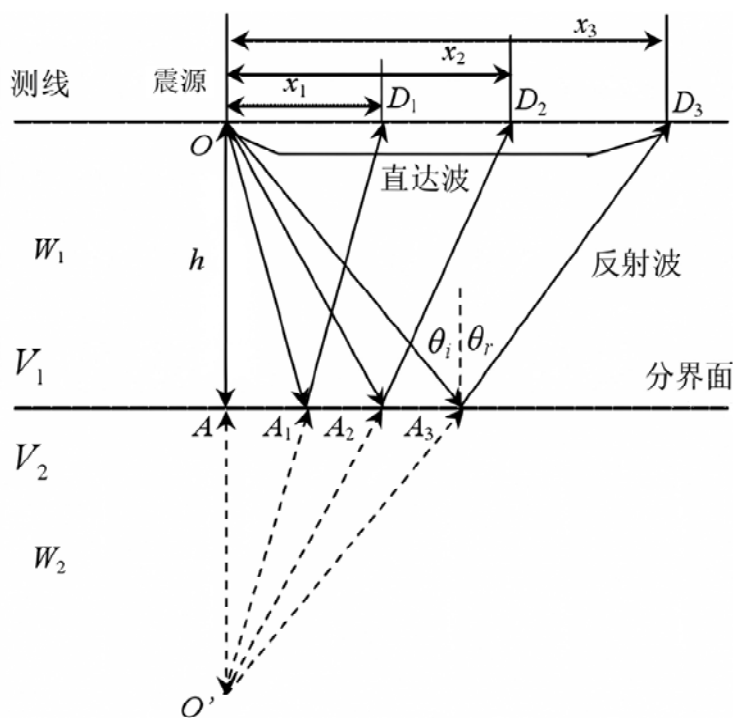


图 2.4 反射地震波传播路径示意图

2.3 深度学习基本原理

在当今快速发展的人工智能领域,深度学习已经成为一种不可或缺的技术,它在图像识别、自然语言处理、自动驾驶等众多领域展现出了巨大的潜力和实用价值。深度学习的核心是模仿人类大脑结构的神经网络,通过复杂的数据处理和学习能力,实现对复杂模式的识别和决策。本节将简要介绍深度学习的基本原理,包括神经网络的结构、前向传播和反向传播机制、损失函数的作用以及优化算法的应用。首先,神经网络作为深度学习的基础结构,它的多层次设计和非线性处理能力是实现复杂功能的关键。随后,介绍了前向传播过程,它是数据在神经网络中传递和转换的基本机制。接着,本节将介绍反向传播过程,这是一种高效的学习算法,用于根据输出结果的误差来优化网络参数。最后,本节将重点介绍讨论优化算法,其对于改善网络性能、加快学习过程至关重要。本节中通过对这些

基本原理进行详细介绍,作为本研究中深度学习技术在地球物理领域应用的理论基础。

(1) 神经网络

神经网络是深度学习的核心组件,它模仿人脑神经元的工作方式来处理复杂的数据问题。一个基本的神经网络通常包括以下几个部分:

输入层 (Input Layer): 输入层是网络的第一层,它负责接收输入数据。每个输入节点代表数据集中的一个特征。

隐藏层 (Hidden Layers): 隐藏层位于输入层和输出层之间,可以有一个或多个。隐藏层的神经元将输入数据进行转换和组合,提取更高级的特征。每个隐藏层的神经元都通过权重和偏置连接到上一层。

输出层 (Output Layer): 输出层是神经网络的最后一层,它将隐藏层的处理结果转换为最终的输出,如分类标签或回归预测。

尽管这两年神经网络的架构不断发展,出现了诸多“明星架构”比如卷积神经网络、循环神经网络、Transformer 等,但是究其结构依旧能够将其按照这些基本组成部分划分。在这些层中,神经元是神经网络的基本单位,它接收来自前一层的输入,通过激活函数进行处理后传递给下一层。其中,每个神经元中往往有着各自的权重和偏置,其中,权重控制着输入信号的重要性,而偏置则是独立于输入信号的调整参数。激活函数决定了神经元是否应该被激活,它给神经网络引入非线性,使网络能够学习和执行更复杂的任务。常用的激活函数包括修正线性单元 (rectified linear unit, ReLU)、Sigmoid 函数和 Tanh 函数。

因此,一个简单的单层神经网络可以用下面的数学公式表示为

$$y = f(W \cdot x + b) \dots\dots\dots (2.8)$$

其中, x 是输入向量, W 是权重矩阵, b 是偏置向量, f 是激活函数, y 是输出。

以多层感知机 (Multilayer Perceptron, MLP) 为例,多层感知机是一种基本的前馈神经网络,它包括一个输入层、若干隐藏层和一个输出层。在 MLP 中,每一层的输出是下一层的输入,通过连续的层级结构,网络能够学习数据中的复杂模式。

神经网络通过这些基本的结构组件和数学公式,能够处理各种复杂的数据问题,从而在许多领域内发挥重要作用,如图像识别、自然语言处理、地震信号处理等。

(2) 前向传播

前向传播是神经网络中的一个基本过程,它指的是数据在网络中从输入层向输出层传播的过程。在这个过程中,每一层的输出将成为下一层的输入。前向传播的目的是根据给定的输入数据计算出网络的预测输出。首先,输入数据被送入神经网络的输入层。每个神经元接收到来自前一层的输入后,会根据自己的权重和偏置计算加权和。之后,加权和通过激活函数进行处理,以引入非线性,使得网络可以学习复杂的模式。激活后的值传递给下一层,直到最后的输出层。

设 $x^{(i)}$ 是第 i 层的输入, $W^{(i)}$ 和 $b^{(i)}$ 分别是第 i 层的权重和偏置, $f^{(i)}$ 是第 i 层的激活函数,那么第 i 层的输出可以表示为

$$y^{(i)} = f^{(i)}(W^{(i)} \cdot x^{(i)} + b^{(i)}) \dots\dots\dots (2.9)$$

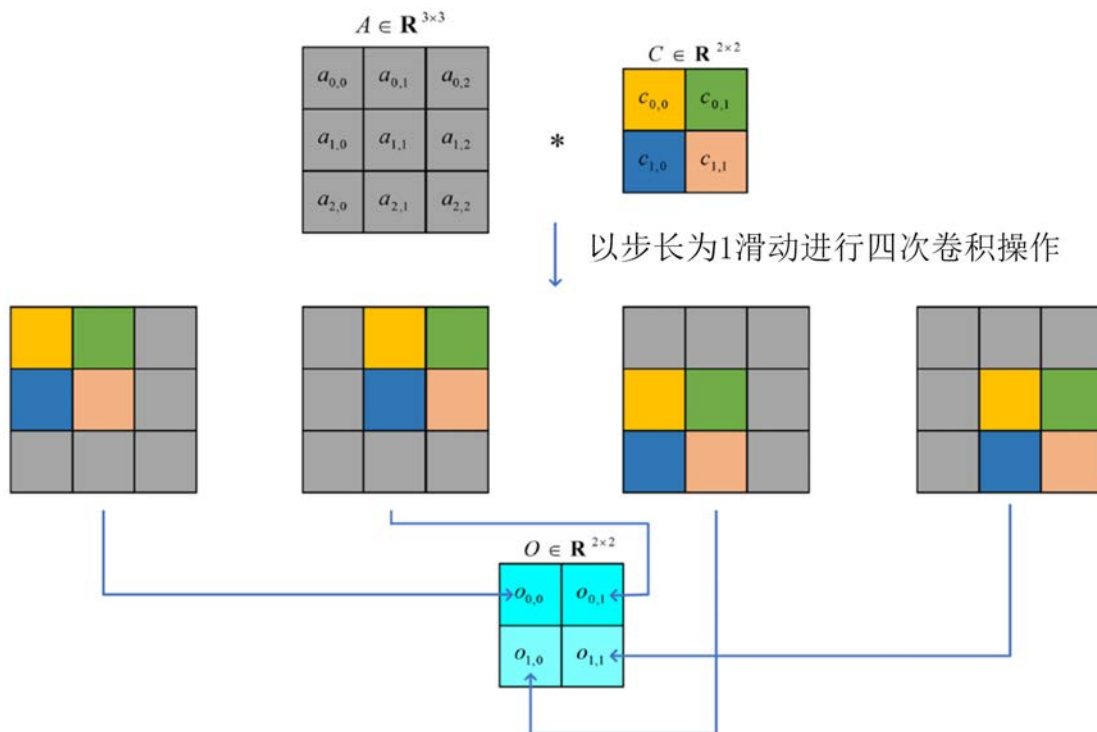


图 2.5 卷积操作示意图

以卷积神经网络中的卷积单元为例,如图 2.5 所示, $A \in R^{3 \times 3}$ 为输入矩阵,其子元素为 $a_{k,l}$, $k, l \in [0, 2]$; $C \in R^{2 \times 2}$ 为卷积核,其子元素为 $c_{i,j}$, $i, j \in [0, 1]$ 。卷积核运算结果为 $O \in R^{3 \times 3}$, 其子元素为 $o_{m,n}$, $m, n \in [0, 1]$ 。则三者中元素的关系为

$$\begin{cases} o_{0,0} = a_{0,0} \cdot c_{0,0} + a_{0,1} \cdot c_{0,1} + a_{1,0} \cdot c_{1,0} + a_{1,1} \cdot c_{1,1} \\ o_{0,1} = a_{0,1} \cdot c_{0,0} + a_{0,2} \cdot c_{0,1} + a_{1,1} \cdot c_{1,0} + a_{1,2} \cdot c_{1,1} \\ o_{1,0} = a_{1,1} \cdot c_{0,0} + a_{1,2} \cdot c_{0,1} + a_{2,1} \cdot c_{1,0} + a_{2,2} \cdot c_{1,1} \\ o_{1,1} = a_{1,0} \cdot c_{0,0} + a_{1,1} \cdot c_{0,1} + a_{2,0} \cdot c_{1,0} + a_{2,1} \cdot c_{1,1} \end{cases} \dots\dots\dots (2.10)$$

对于多层网络,前向传播的过程可以看作是这些操作的连续应用。例如,在一个有两个隐藏层的网络中,前向传播过程可以表示为

$$y^{(output)} = f^{(3)}(W^{(3)} \cdot f^{(i)}(W^{(2)} \cdot f^{(1)}(W^{(1)} \cdot x^{(input)} + b^{(1)}) + b^{(2)}) + b^{(3)}) \dots\dots (2.11)$$

其中, $x^{(input)}$ 是输入层的输入, $y^{(output)}$ 是输出层的输出。

前向传播是神经网络进行预测的基础,它通过一系列计算将输入数据转换为输出。在多数实际场景中如实际地震数据去噪任务,优化后的神经网络会固定参数不再进行调整,这类场景下对神经网络的应用仅涉及使用前向传播。

(3) 反向传播

反向传播是用于训练神经网络的一种算法,主要用于根据损失函数的梯度来更新网络的权重和偏置。它通过计算损失函数相对于网络参数的梯度,并根据这些梯度调整参数以最小化损失函数。其步骤可概括为:

- 1) 前向传播计算输出:通过前向传播得到网络预测输出并计算出损失函数。
- 2) 计算输出层的误差:计算输出层的预测值与实际值之间的差异(误差)。
- 3) 反向传递误差:将误差反向传播回网络,即从输出层向输入层逐层传递。
- 4) 计算梯度:在每一层计算误差相对于该层权重和偏置的梯度。
- 5) 更新权重和偏置:根据计算出的梯度和学习率更新网络的权重和偏置。

设损失函数为 L , 第 i 层的权重为 $W^{(i)}$, 偏置为 $b^{(i)}$, 则权重和偏置的更新公式如下为

$$W_{new}^{(i)} = W^{(i)} - \eta \frac{\partial L}{\partial W^{(i)}} \quad \dots\dots\dots (2.12)$$

$$b_{new}^{(i)} = b^{(i)} - \eta \frac{\partial L}{\partial b^{(i)}} \quad \dots\dots\dots (2.13)$$

其中, η 是学习率, $\eta \frac{\partial L}{\partial W^{(i)}}$ 和 $\eta \frac{\partial L}{\partial b^{(i)}}$ 分别是损失函数相对于权重和偏置的偏导数。

反向传播的核心是链式法则, 它用于计算损失函数相对于各层权重和偏置的偏导数。对于第 i 层, 链式法则可以表示为

$$\frac{\partial L}{\partial W^{(i)}} = \frac{\partial L}{\partial y^{(i)}} \cdot \frac{\partial y^{(i)}}{\partial W^{(i)}} \quad \dots\dots\dots (2.14)$$

$$\frac{\partial L}{\partial b^{(i)}} = \frac{\partial L}{\partial y^{(i)}} \cdot \frac{\partial y^{(i)}}{\partial b^{(i)}} \quad \dots\dots\dots (2.15)$$

其中, $\frac{\partial L}{\partial y^{(i)}}$ 是损失函数相对于第 i 层输出的偏导数, $\frac{\partial y^{(i)}}{\partial W^{(i)}}$ 和 $\frac{\partial y^{(i)}}{\partial b^{(i)}}$ 分别是第 i 层输出相对于权重和偏置的偏导数。

反向传播是深度学习中最重要算法之一, 它使得神经网络能够通过学习减少预测误差, 其对于神经网络的训练过程至关重要。

(4) 优化算法

优化算法在深度学习中扮演着至关重要的角色, 它们负责调整神经网络中的参数 (权重和偏置), 以最小化或最大化一个目标函数 (通常是最大化或最小化损失函数)。这个过程涉及到计算目标函数相对于网络参数的梯度, 并利用这些梯度来更新参数。

在深度学习中, 优化算法的选择对模型的训练效果和收敛速度有着重要影响。随机梯度下降 (Stochastic Gradient Descent, SGD) ^[95] 和 Adam (Adaptive Moment Estimation) ^[98] 是两种广泛使用的优化方法。它们虽然基于不同的原理, 但共同目标是通过调整网络参数来最小化损失函数。接下来, 以这两种方法为例介绍深度学习优化算法的基本原理。

首先, 随机梯度下降是最基础的优化算法之一。它的核心思想是在每一步中使用数据集的一个随机子集 (通常称为一个批次) 来计算梯度, 然后用这个梯度

来更新模型的参数。这样做的主要优点是计算效率高,特别是对于大规模数据集。SGD 的更新公式为

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; X^{(i)}, y^{(i)}) \dots\dots\dots (2.16)$$

其中, θ 表示模型参数, η 是学习率, $\nabla_{\theta} J(\theta; X^{(i)}, y^{(i)})$ 是在第 i 个批次数据上计算得到的损失函数 J 相对于 θ 的梯度。

对于大规模数据集, SGD 可以快速进行参数更新。但是其缺点是每次更新只使用一个批次的数据,更新过程可能会比较震荡,特别是当梯度的估计非常嘈杂时。

Adam^[98]是一种自适应学习率的优化算法,它结合了动量(Momentum)和均方根传播(Root Mean Square Propagation, RMSprop)两种优化方法的思想。Adam 通过计算梯度的一阶矩估计(即平均值)和二阶矩估计(即未中心化的方差)来调整每个参数的学习率。Adam 的参数更新公式如下: 表示为

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \dots\dots\dots (2.17)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \dots\dots\dots (2.18)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \dots\dots\dots (2.19)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \dots\dots\dots (2.20)$$

$$\theta = \theta - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \dots\dots\dots (2.21)$$

其中, g_t 是在时间步 t 的梯度, m_t 和 v_t 分别是梯度的一阶和二阶矩的估计, β_1 和 β_2 是衰减率,通常取值接近 1。Adam 优化方法有着自适应学习率,Adam 会为每个参数计算独立的学习率,这通常可以带来更好的性能。相比于基本的 SGD, Adam 通常能够更快地收敛,并且对超参数的选择更加鲁棒。

总结来说,SGD 和 Adam 都是优化神经网络的有效方法。SGD 简单且有效,特别适合于大规模数据和资源有限的情况。而 Adam 通过更复杂的计算,提供了更快的收敛速度和对超参数选择的鲁棒性,适用于需要快速获得结果的场景。

2.4 全波形反演基本理论

全波形反演技术是一种通过综合利用地震波场的全面信息(包括振幅、相位和走时等)来精确反演地下介质的物性参数的地震数据处理方法。该技术通过将问题近似为线性并在目标函数的指导下,使用局部优化技术反复迭代更新初始地下模型,以期使模拟产生的数据与实际观测数据之间的差异最小化,从而实现对地下介质结构的精确描述^[168]。本研究面向的是在各向同性声学介质中反演纵波速度,其流程如图 2.6 所示,可以简要总结为:

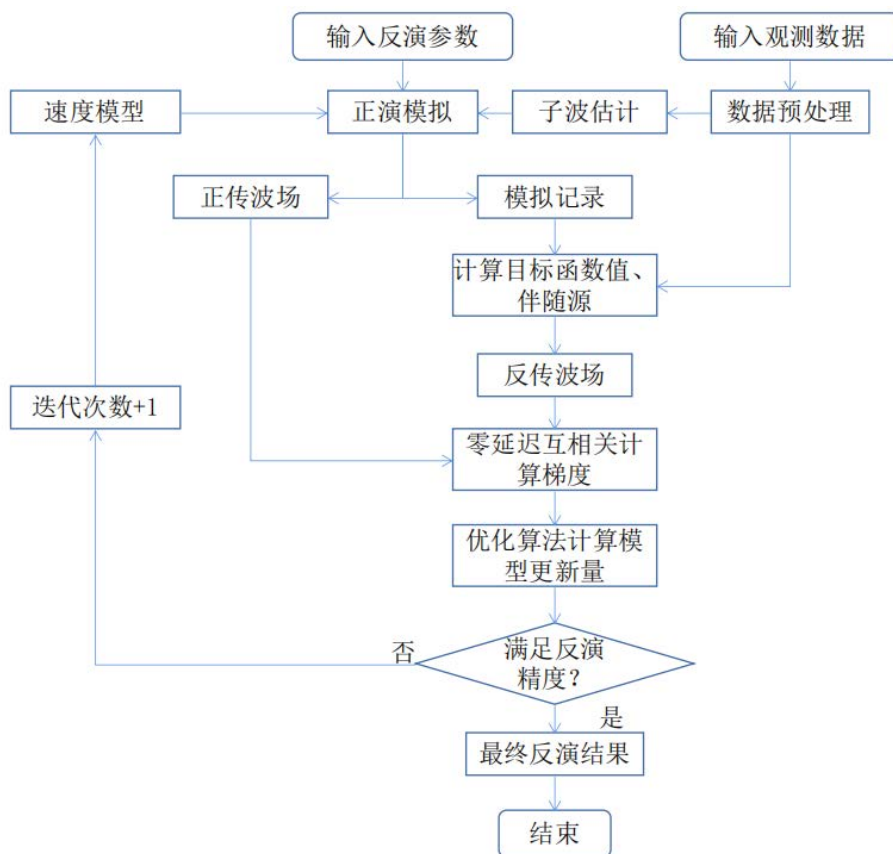


图 2.6 时间域常规 FWI 基本流程图

(1) 初始化阶段及数据预处理: 设置观测系统参数, 提供起始速度模型,

并输入采集到的实际地震数据。需要设定观测系统的关键参数，并输入一个初始的速度模型，以及实际收集的地震数据。接着，对这些观测数据进行必要的预处理工作，同时对地震的震源子波进行估算。

(2) 正演模拟：依托于前期定义的初始模型，并利用数值模拟技术正演，从而获得模拟产生的数据。

(3) 差异分析：计算模拟数据与实际观测数据之间的差异，以生成伴随源，进而进行逆时反传计算。

(4) 梯度计算：应用伴随状态法，通过零延迟互相关分析计算梯度。

(5) 模型更新：将计算的梯度应用于局部优化算法，更新模型速度参数。

(6) 迭代判断：如果达到预设的最大迭代次数，或者目标函数的数值已经不再继续下降，则结束迭代过程，否则继续迭代。

2.4.1 有限差分正演模拟

在二维声学介质中，速度与位移的波动方程可以具体表述为

$$\frac{\partial^2 u}{\partial t^2} = \frac{1}{\rho} \left[\frac{\partial K}{\partial x} + \frac{\partial K}{\partial z} \frac{\partial u}{\partial z} + K \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} \right) \right] \dots\dots\dots (2.22)$$

这些参数中， u 为波场位移， t 是时间， ρ 代表密度， K 为体积模量。在常密度各向同性介质中，上述方程可以简化为

$$\frac{\partial^2 u}{\partial t^2} = v^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} \right) \dots\dots\dots (2.23)$$

上述公式中， v 代表声波速度。可借助有限差分法对此偏微分方程进行近似处理，这一方法通过迭代计算波场，可按照时间序列迭代求解波场。

(1) 空间有限差分

记 Δx 和 Δz 分别为水平和深度方向上的网格间距，分别在水平和垂直两个方向等距划分网格，即 $\Delta x = \Delta z$ 。使用差分方法近似求解波动方程。以水平 x 方向上差分公式的推导为例，若 $u(x)$ 有 $2L+1$ 阶导数，那么 $u(x)$ 在 $x = x_0 + \Delta x$ 的 $2L+1$

阶泰勒展开为

$$u(x+\Delta x)=u(x)+\frac{\partial u}{\partial x}\Delta x+\frac{1}{2!}\frac{\partial^2 u}{\partial x^2}(\Delta x)^2+\frac{1}{3!}\frac{\partial^3 u}{\partial x^3}(\Delta x)^3+\cdots \\ +\frac{1}{(2L)!}\frac{\partial^{2L} u}{\partial x^{2L}}(\Delta x)^{2L}+O(\Delta x^{2L+1}) \quad \cdots \cdots (2.24)$$

$u(x)$ 在 $x=x_0-\Delta x$ 处的 $2L+1$ 阶泰勒展开为

$$u(x-\Delta x)=u(x)-\frac{\partial u}{\partial x}\Delta x+\frac{1}{2!}\frac{\partial^2 u}{\partial x^2}(\Delta x)^2-\frac{1}{3!}\frac{\partial^3 u}{\partial x^3}(\Delta x)^3+\cdots \\ +\frac{1}{(2L)!}\frac{\partial^{2L} u}{\partial x^{2L}}(\Delta x)^{2L}+O(\Delta x^{2L+1}) \quad \cdots \cdots (2.25)$$

将以上两个公式相加, 可得二阶中心差分表达式为

$$\frac{u(x+\Delta x)-2u(x)+u(x-\Delta x)}{2}=\frac{1}{2!}\frac{\partial^2 u}{\partial x^2}(\Delta x)^2+\frac{1}{4!}\frac{\partial^4 u}{\partial x^4}(\Delta x)^4+\cdots \\ +\frac{1}{(2L)!}\frac{\partial^{2L} u}{\partial x^{2L}}(\Delta x)^{2L}+O(\Delta x^{2L+1}) \quad \cdot (2.26-1)$$

以此类推, 则波动方程的 $2L$ 阶差分格式表达式如下:

$$\frac{u(x+L\Delta x)-2u(x)+u(x-L\Delta x)}{2} \\ =\frac{1}{2!}\frac{\partial^2 u}{\partial x^2}(L\Delta x)^2+\frac{1}{4!}\frac{\partial^4 u}{\partial x^4}(L\Delta x)^4+\frac{1}{(2L)!}\frac{\partial^{2L} u}{\partial x^{2L}}(L\Delta x)^{2L}+\cdots \cdots (2.26-L) \\ +O(\Delta x^{2L+1})$$

将以上 L 个公式的等号左侧差分表示为如下格式:

$$f_1=\frac{u(x+\Delta x)-2u(x)+u(x-\Delta x)}{2} \\ f_2=\frac{u(x+2\Delta x)-2u(x)+u(x-2\Delta x)}{2} \quad \cdots \cdots (2.27) \\ \vdots \\ f_L=\frac{u(x+L\Delta x)-2u(x)+u(x-L\Delta x)}{2}$$

等号右侧的泰勒展开式不同阶导数项为

$$a_1 = \frac{\partial^2 u}{\partial x^2} (\Delta x)^2, a_2 = \frac{\partial^4 u}{\partial x^4} (\Delta x)^4, \dots, a_L = \frac{\partial^{2L} u}{\partial x^{2L}} (\Delta x)^{2L} \dots\dots\dots (2.28)$$

联立公式 (2.26) 中所有 L 个方程, 可进一步得到波场对空间的二阶偏导数为

$$\begin{aligned} & \frac{\partial^2 u}{\partial x^2} \\ &= \frac{1}{(\Delta x)^2} \sum_{m=1}^L D_m f_m \\ &= \frac{1}{(\Delta x)^2} \sum_{m=1}^L D_m \left[\frac{u(x+m\Delta x) - 2u(x) + u(x-m\Delta x)}{2} \right] \dots\dots (2.29) \\ &= -\frac{1}{(\Delta x)^2} \sum_{m=1}^L D_m u(x) + \frac{1}{(\Delta x)^2} \sum_{m=1}^L \frac{D_m}{2} [u(x+m\Delta x) + u(x-m\Delta x)] \\ &= \frac{1}{(\Delta x)^2} \left\{ C_0 u(x) + \sum_{m=1}^L C_m [u(x+m\Delta x) + u(x-m\Delta x)] \right\} \end{aligned}$$

该式中 C_m 是最终差分系数, D_m 是待定差分系数。规则网格中, 不同阶导数的有限差分系数值可见表 2.2。

表 2.2 规则网格二阶导数高阶有限差分系数值

2L	C_0	C_1	C_2	C_3	C_4	C_5
2	-1	0.5				
4	-1.25	0.6667	-8.3333E-2			
6	-1.3611	0.75	-7.5E-2	5.5556E-3		
8	-1.4236	0.8	-0.1	1.2698E-2	-8.9285E-4	
10	-1.4836	0.8333	-1.1905E-1	1.9841E-2	-2.4802E-3	1.5873E-4

(2) 时间有限差分

为了在时间轴上实现迭代计算, 除了对空间进行有限差分近似, 同样需要对时间进行差分近似。通过在 $u(t)$ 时间点 $t = t + \Delta t$ 和 $t = t - \Delta t$ 处进行泰勒级数展开, 可以得到关于时间轴中心差分的二阶精确度表达式为

$$\begin{aligned}\frac{u(t+\Delta t)-2u(t)+u(t-\Delta t))}{2} &= \frac{1}{2!} \frac{\partial^2 u}{\partial t^2} (\Delta t)^2 + \frac{1}{4!} \frac{\partial^4 u}{\partial t^4} (\Delta t)^4 + \dots \\ &= \frac{1}{2!} \frac{\partial^2 u}{\partial t^2} (\Delta t)^2 + O((\Delta t)^2)\end{aligned}\quad \dots (2.30)$$

通过这种方法,得到的波场位移关于时间二阶导的有限差分公式具有二阶精确度,其中 Δt 表示时间间隔, $O((\Delta t)^2)$ 表示 Δt 从二阶开始的更高阶项。那么位移对时间二阶导数的二阶精度有限差分为

$$\frac{\partial^2 u}{\partial t^2} = \frac{1}{(\Delta t)^2} [u(t+\Delta t) - 2u(t) + u(t-\Delta t)] \quad \dots (2.31)$$

(3) 二维声波方程规则网格高阶有限差分格式

将空间方向的高阶差分格式与时间方向的二阶差分格式结合,代替位移相对于时间和空间二阶偏导数,则二维声学各向同性介质中的高阶差分近似公式可表示为

$$\begin{aligned}u_{i,j}^{k+1} &= 2u_{i,j}^k - u_{i,j}^{k-1} + \left(\frac{v\Delta t}{\Delta x}\right)^2 \left[C_0 u_{i,j}^k + \sum_{n=1}^N C_n (u_{i+n,j}^k + u_{i-n,j}^k) \right] \\ &\quad + \left(\frac{v\Delta t}{\Delta z}\right)^2 \left[C_0 u_{i,j}^k + \sum_{n=1}^N C_n (u_{i,j+n}^k + u_{i,j-n}^k) \right]\end{aligned}\quad \dots (2.32)$$

这些参数中, i 和 j 分别代表 x 轴和 z 轴的坐标, k 表示时间离散点, C_n 为差分系数,其取值如表 2.2 所示。

(4) 稳定性条件

在对波动方程进行有限差分模拟时,随着时间推进通过迭代过程进行波场模拟不断推进其进行。这一过程依赖于差分方程,该方程通过省略泰勒级数的高阶项来简化计算。然而,这种简化会导致随时间积累的截断误差,不仅影响波场模拟的准确性,引发数值频散现象,还可能使模拟过程无法稳定进行。根据 Lines 等人于 1999 年的研究^[163],波动方程的时间域有限差分模拟的稳定性可以通过特定条件进行判断:

$$\frac{v_{\max} \Delta t}{h} \leq \sqrt{\frac{a_1}{a_2}} \dots\dots\dots (2.33)$$

此处, $v_{\max}$ 代表速度模型中的最高速度, a_1 为时间偏导数差分权重的绝对值之和, a_2 代表空间偏导数差分权重的绝对值之和, h 代表了网格间距。

(5) 边界条件

为了将地震波在无限空间内的数值模拟限制在有限的计算区域内, 本项研究采用了特殊的模拟边界处理, 即完美匹配层边界条件 (Perfectly Matched Layer, PML) [146]。这种方法通过在模拟区域的外部添加一个吸收层来实现, 从而有效避免计算过程中的反射问题, 确保地震波在计算域中的传播更加接近真实情况。通过在计算域外增加衰减层, 并在衰减层中设置特定的衰减函数来减少边界反射。这些衰减函数能够根据波的入射方向, 有效衰减波能量, 从而模拟一个更加接近真实情况的波场传播。完全匹配层的配置如图 2.7 所示。

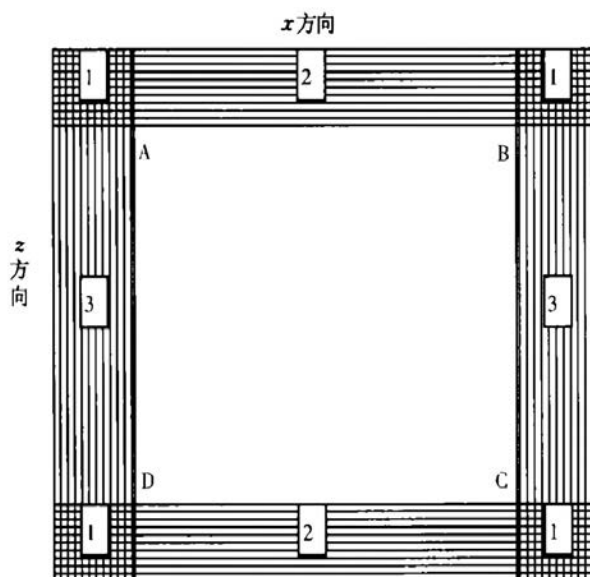


图 2.7 PML 技术示意图

假如 x 方向上和 z 方向上的衰减函数分别为 $d(x)$ 和 $d(z)$, 那么区域 1 中, $d(x) \neq 0, d(z) \neq 0$; 区域 2 中, $d(x)=0, d(z) \neq 0$; 区域 3 中, $d(x) \neq 0, d(z)=0$ 。通过精心设计 PML 边界条件中的衰减层和衰减函数, 可以显著减少因边界反射

带来的干扰,使得模拟的波场更加贴近实际地震波在无限介质中传播的情况,为复杂地质结构下的地震波传播提供一个准确可靠的数值模拟环境。

2.4.2 时间域梯度计算

(1) 伴随状态法

在 FWI 中,核心任务是通过计算梯度来确定模型参数的最优更新路径。FWI 通常采用基于 L2 范数的目标函数来衡量模拟数据与实际地震数据之间的差异,目标函数可表示为

$$E(m) = \frac{1}{2} [d_{syn}(m) - d_{obs}]^2 \dots\dots\dots (2.34)$$

此处, E 是目标函数, m 代表模型参数,而模拟与实测地震数据分别由 d_{syn} 和 d_{obs} 表示。寻找最优收敛方向即求解目标函数的梯度,此过程涉及到对地下介质的物理参数进行导数计算,公式展示为

$$\frac{\partial E}{\partial m} = \left\langle \frac{\partial d_{syn}}{\partial m}, d_{syn} - d_{obs} \right\rangle \dots\dots\dots (2.35)$$

在数学上,这里用到的导数 $\frac{\partial d_{syn}}{\partial m}$ 被称为 Fréchet 导数,用以描述模拟数据对介质物性参数变化的敏感度。直接对每个网格点进行 Fréchet 导数的计算不仅计算量巨大,而且不易于实际操作。因此,引入伴随状态法,它允许仅通过两次正演计算即可获得所需梯度,从而显著降低了计算成本。这种方法在 FWI 的计算优化中发挥了重要作用。

定义函数 $J(m)$ 来计算使用伴随状态法得到的梯度,其中 $J(m)$ 的值取决于 $u(m)$ 。这一过程通过映射 h 实现:

$$J(m) = h(u(m), m) \dots\dots\dots (2.36)$$

这里, u 代表状态变量,存在于状态空间 $\mathbf{U}$ 中,可能是实数或复数; m 为模型参数,存在于模型空间 $\mathbf{M}$ 中,属于实数。 h 是从 $\mathbf{U} \times \mathbf{M}$ 空间到实数 $\mathbf{R}$ 空间的映射函

数，而 J 是从 $\mathbf{M}$ 空间到 $\mathbf{R}$ 空间的函数。状态变量 u 服从的映射关系为

$$F(u(m), m) = 0 \dots\dots\dots (2.37)$$

其中， F 表示正演问题，它是从 $\mathbf{U} \times \mathbf{M}$ 空间到 $\mathbf{U}$ 空间的映射函数。当 u 满足 $F(u, m) = 0$ 时， u 被视为 F 的一个实际解。 δm 、 δu 和 δJ 分别代表 m 、 u 和 J 的变化量，当 m 变化 δm 时， $u + \delta u$ 成为 $m + \delta m$ 的一个可能解。在考虑泰勒展开的基础上，仅关注一阶变化量

$$\begin{aligned} & F(u + \delta u, m + \delta m) \\ &= F(u, m) + \frac{\partial F(u, m)}{\partial u} \delta u + \frac{\partial F(u, m)}{\partial m} \delta m = 0 \dots\dots\dots (2.38) \end{aligned}$$

同时，可以得

$$\frac{\partial F(u, m)}{\partial \tilde{u}} \delta u = - \frac{\partial F(u, m)}{\partial m} \delta m \dots\dots\dots (2.39)$$

对公式 (2.36) 进行泰勒展开并仅考虑一阶项

$$\delta J = \left\langle \frac{\partial h(u, m)}{\partial \tilde{u}}, \delta u \right\rangle_u + \frac{\delta h(u, m)}{\delta m} \delta m \dots\dots\dots (2.40)$$

假设 $\mathbf{M}$ 空间中任意 m 都对应 $\mathbf{U}$ 空间中的唯一解，那么 $u + \delta u$ 是 $F(u + \delta u, m + \delta m) = 0$ 的唯一解。可发现 δu 是公式 (2.39) 的唯一解，表达式如下：

$$\delta u = - \left[\frac{\partial F(u, m)}{\partial \tilde{u}} \right]^{-1} \frac{\partial F(u, m)}{\partial m} \delta m \dots\dots\dots (2.41)$$

将公式 (2.41) 代入 (2.40)，可得到如下形式：

$$\delta J = \frac{\delta h(u, m)}{\delta m} \delta m - \left\langle \left[\frac{\partial F(u, m)}{\partial \tilde{u}} \right]^{-1} \frac{\partial h(u, m)}{\partial \tilde{u}}, \frac{\partial F(u, m)}{\partial m} \delta m \right\rangle_u \dots\dots (2.42)$$

此处的*代表伴随算子,而伴随算子操作的变量位于 $\mathbf{U}$ 的对偶空间 $\mathbf{U}^*$ 中。在公式 (2.42) 的内积表示中,与 δm 无关的部分被统一处理,目的是为了免去直接求解 Fréchet 导数的需求。引入变量 λ 后可以得到

$$\left[\frac{\partial F(u, m)}{\partial \tilde{u}} \right]^* \lambda = \frac{\partial h(u, m)}{\partial \tilde{u}} \dots\dots\dots (2.43)$$

则上式可表示为

$$\delta J = \left[- \left\langle \lambda, \frac{\partial F(u, m)}{\partial m} \right\rangle_{\mathbf{U}} + \frac{\delta h(u, m)}{\delta m} \right] \delta m \dots\dots\dots (2.44)$$

其中 λ 为伴随状态变量,属于空间 $\mathbf{U}^*$ 。通过使用以上的两个公式,来直接计算 $\frac{\partial J}{\partial m}$, 伴随状态法可以简化计算,减少计算消耗。

(2) 时间域伴随状态法梯度算子

提取常密度声波方程波动算子 $L = \sigma^2 \frac{\partial^2}{\partial t^2} - \Delta$, 其中模型参数 m 的慢度由 σ 表示。震源 f_s 引发的应力场 u_s 需要遵循的初始和边界条件为

$$\begin{cases} u_s(0) = 0 \\ \frac{\partial u_s(0)}{\partial t} = 0 \\ Lu_s = f_s \end{cases} \dots\dots\dots (2.45)$$

在地震勘探中,全波形反演用于从地震数据中重建地下介质的速度模型,其核心是最小化观测数据与模拟数据之间的差异。常规的时间域 FWI 的目标函数一般可设置为

$$J(m) = \frac{1}{2} \sum_{s,r} \int_0^T [S_{s,r} u_s(t) - d_{s,r}(t)]^2 dt \dots\dots\dots (2.46)$$

在此式子内, T 代表了整个采样周期; $S_{s,r}$ 代表着波场约束算子,该算子从检波器记录的波场数据中提取出模拟的地震数据。通过结合伴随状态 $\tilde{\mu}_s^0$ 和 $\tilde{\mu}_s^1$ 与初始边界条件,并将 $\tilde{\lambda}_s$ 与波动方程相联,可以形成增广的泛函表达式为

$$\begin{aligned}
 \Gamma[(\tilde{u}_s), (\tilde{\lambda}_s), (\tilde{\mu}_s^0), (\tilde{\mu}_s^1), m] &= \frac{1}{2} \sum_{s,r} \int_0^T [S_{s,r} u_s(t) - d_{s,r}(t)]^2 dt \\
 &- \sum_s \int_0^T \left\langle \tilde{\lambda}_s(t), m \frac{\partial^2 \tilde{u}_s(t)}{\partial t^2} - \Delta \tilde{u}_s(t) - f_s(t) \right\rangle_x dt \quad \cdots (2.47) \\
 &- \sum_s \langle \tilde{\mu}_s^0, \tilde{u}_s(0) \rangle_x - \sum_s \left\langle \tilde{\mu}_s^1, \frac{\partial \tilde{u}_s(0)}{\partial t} \right\rangle_x
 \end{aligned}$$

对上式中 $\tilde{u}_s$ 对 t 的二阶导数作两次分部积分，能够得到

$$\begin{aligned}
 \int_0^T \left\langle \tilde{\lambda}_s, m \frac{\partial^2 \tilde{u}_s}{\partial t^2} \right\rangle_x dt &= \int_0^T \left\langle m \frac{\partial^2 \tilde{\lambda}_s}{\partial t^2}, \tilde{u}_s \right\rangle_x dt \\
 &+ \left\langle \tilde{\lambda}_s(T), m \frac{\partial \tilde{u}_s(T)}{\partial t} \right\rangle_x - \left\langle \tilde{\lambda}_s(0), m \frac{\partial \tilde{u}_s(0)}{\partial t} \right\rangle_x \quad \cdots (2.48) \\
 &- \left\langle m \frac{\partial \tilde{\lambda}_s(T)}{\partial t}, \tilde{u}_s(T) \right\rangle_x + \left\langle m \frac{\partial \tilde{\lambda}_s(0)}{\partial t}, \tilde{u}_s(0) \right\rangle_x
 \end{aligned}$$

依据上述两个公式，进而可以导出伴随状态方程为

$$\begin{cases}
 \lambda_s(T) = 0 \\
 \frac{\partial \lambda_s(T)}{\partial t} = 0 \\
 m \frac{\partial^2 \lambda_s}{\partial t^2} - \Delta \lambda_s = \sum_r S_{s,r}^T (S_{s,r} u_s - d_{s,r}) \quad \cdots (2.49) \\
 \mu_s^0 = m \frac{\partial \lambda_s(0)}{\partial t} \\
 \mu_s^1 = m \lambda_s(0)
 \end{cases}$$

其中的 T 代表转置操作。因此，目标函数 J 对 $\mathbf{x}$ 位置的模型参数梯度表达式如下：

$$\frac{\partial J}{\partial m}(\mathbf{x}) = - \sum_s \int_0^T \lambda_s(\mathbf{x}, t) \frac{\partial^2 u_s(\mathbf{x}, t)}{\partial t^2} dt \quad \cdots (2.50)$$

这一公式是时间域 FWI 的梯度计算的核心，基于伴随状态法的框架进行推导。通过计算正向波场的时间二阶偏导，并使用伴随源作为震源进行零延迟的互相关计算，从而实现所需梯度信息的获取。值得注意的是，在每一次计算目标函数梯度的过程中，只需执行一次正向波场和一次反向波场的正演模拟。该方法不仅能

够有效地提取地下介质的信息，而且在计算过程中也能够节省时间和计算资源。

2.4.3 速度模型迭代更新

在速度模型的逐步优化方面，全波形反演过程涉及复杂的非线性问题求解，通过将问题线性化并运用迭代优化算法逐步逼近理想解。由于地下介质的物性参数估计涉及庞大数据量，直接去使用全局优化策略会带来高昂的计算成本及缓慢的收敛速度。因而，通常采用局部优化技术以提高模型参数更新的速度。在此研究中，所采纳的局部优化策略是有限内存 BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno, L-BFGS) 算法。接下来，对其进行详细介绍。

在一次模型参数的更新过程中，

$$m_{k+1} = m_k + \alpha \Delta m \quad \dots\dots\dots (2.51)$$

m 代表速度模型， Δm 为速度模型的更新量， α 为更新的步长。为了使目标函数趋向于其全局最低点，需要满足一定的条件，即通过确保目标函数对地下介质的物理参数导数趋向于零。因此，目标函数的导数与其通过泰勒级数进行展开的关系可表示为

$$\frac{\partial J(m)}{\partial m} = \frac{\partial J(m_0)}{\partial m} + \frac{\partial^2 J(m_0)}{\partial m^2} \Delta m + O(m^3) + \dots = 0 \quad \dots\dots\dots (2.52)$$

在这里， m_0 代表初始模型参数，而 $O(m^3)$ 则指泰勒级数中的更高阶项。当忽略这些高阶项后，简化的表达式为

$$\frac{\partial J(m)}{\partial m} = \frac{\partial J(m_0)}{\partial m} + \frac{\partial^2 J(m_0)}{\partial m^2} \Delta m = 0 \quad \dots\dots\dots (2.53)$$

根据上述简化，可以推导出模型参数的更新量为

$$\Delta m = \left[\frac{\partial^2 J(m_0)}{\partial m^2} \right]^{-1} \frac{\partial J(m_0)}{\partial m} \quad \dots\dots\dots (2.54)$$

据此，全波形反演过程中的迭代公式可以表达为

$$m_{k+1} = m_k + \alpha \mathbf{H}_k^{-1} \mathbf{g}_k \quad \dots\dots\dots (2.55)$$

在这一系列参数中， $\mathbf{H}$ 代表 Hessian 矩阵，而 g 是反演过程中计算得到的梯度。通过反复应用公式 (2.55) 进行迭代，可以逐步优化初始模型参数，进而达到全波形反演的高精度结果。然而，直接计算及存储 Hessian 矩阵的逆所需的计算资源非常庞大。为了既保证收敛的精度，又提升计算效率和节约存储空间，L-BFGS 算法通过采用近似 Hessian 矩阵的逆进行迭代求解。下面将详细解释 L-BFGS 算法，目标函数的更新量的泰勒级数展开式表示为

$$J(m_k + s_k) = J(m_k) + \nabla J(m_k)^T s_k + \frac{1}{2} s_k^T \mathbf{H}_{k+1} s_k + o(m_k)^3 \cdots \cdots (2.56)$$

其中， s 表示 m 的扰动量， $s_k = m_{k+1} - m_k$ 。舍弃上式中的高阶项并对 s_k 进行求导，可以得到：

$$J(m_k + s_k) = J(m_k) + \mathbf{H}_{k+1} s_k \cdots \cdots (2.57)$$

令 $y_k = \nabla J(x_k + s_k) - \nabla J(x_k)$ ，则上式表达式如下：

$$s_k = \mathbf{H}_{k+1}^{-1} y_k \cdots \cdots (2.58)$$

根据 Sherman-Morrison-Woodbury 公式，可以得到近似 Hessian 矩阵逆的表达式，如下：

$$\mathbf{B}_{k+1}^{-1} = (I_k - \frac{s_k v_k^T}{v_k^T s_k}) \mathbf{B}_k^{-1} (I_k - \frac{s_k v_k^T}{v_k^T s_k}) + \frac{s_k v_k^T}{v_k^T s_k} \cdots \cdots (2.59)$$

在这里， $\mathbf{B}$ 是近似 Hessian 矩阵。如果 $\rho_k = 1 / y_k^T s_k$ ，则上式可整理为

$$\mathbf{B}_{k+1}^{-1} = (I_k - \rho_k s_k v_k^T) \mathbf{B}_k^{-1} (I_k - \rho_k s_k v_k^T) + s_k \rho_k s_k^T \cdots \cdots (2.60)$$

如果 $v_k = I_k - \rho_k y_k s_k^T$ ，则上式可整理为

$$\mathbf{B}_{k+1}^{-1} = v_k^T \mathbf{B}_k^{-1} v_k + s_k \rho_k s_k^T \cdots \cdots (2.61)$$

在分析上述方程的基础上，可以观察到，逆向近似 Hessian 矩阵的求解依托于一

个逐步迭代和递归的流程,其中每个迭代步骤的结果均紧密依赖于先前所有迭代步骤所累积的逆向近似 Hessian 矩阵的结果。特别是在执行全波形反演过程中对模型参数进行更新的环节里,这个逆向近似 Hessian 矩阵会随着迭代次数的增长而逐步靠拢于其真实的逆向 Hessian 矩阵值。将方程(2.61)中得到的逆向近似 Hessian 矩阵值应用到方程(2.55)中,便可进行基于 L-BFGS 策略的 FWI,且能够大幅减少内存占用。

2.5 本章小结

本研究主要聚焦地震数据处理流程中的数据噪声处理、速度模型反演两个阶段的任务。本文的主要工作是基于深度学习算法研究更为有效的地震数据去噪及速度模型反演方法。因此,为了方便地引出后续核心研究内容,本章对地震勘探的基本原理、深度学习的阐述理论以及反演常用的全波形反演方法进行了详细介绍。

在本章中,首先对地震勘探的基本原理和常见方法进行了阐述。之后,深度学习基本原理一节系统性地介绍了深度学习技术的核心概念和基本原理。通过深入讨论神经网络结构、前向传播、反向传播算法和优化算法,展示了深度学习在处理复杂数据中的流程,为解决地震勘探数据处理中的挑战提供工具。最后一节概述了时间域全波形反演的基本流程,介绍了这些步骤的细节。以上是本文的预备知识,为后续深入讨论提供了必要的基础和理论支持。

第3章 基于密集连接卷积神经网络的地震数据去噪算法

3.1 引言

高质量的地震记录是后续处理解释的基础,对噪声的压制往往是地震信号处理的基础步骤。因此,处理噪声是地震记录处理过程中需要考虑的优先任务。目前,已经开发有一系列方法应用于地震信号的去噪中,如反褶积方法^[91],曲波变换^{[92],[93]},经验模态分解(Empirical Mode Decomposition, EMD)^[94]等方法,这些方法都取得了非常好的效果。但是传统方法都有着特定的应用场景,或者依赖某种先验条件,难以推广至广泛的应用场景中,比如:反褶积方法在处理不规则噪声时表现不佳;使用曲波变换滤波时对阈值的选择会影响最终去噪效果;使用EMD方法时存在频谱混叠问题。地震勘探记录中往往存在着复杂的噪声干扰,其中包括面波、随机噪声。这些噪声具有非高斯非平稳的特征,有的噪声与地震子波序列存在严重的频谱混叠,而且其能量较强,地震记录中有效信号被噪声严重淹没,难以辨别。传统的方法往往在此类问题上表现不佳,这驱使研究者不断寻找新的解决方案。

近年来,基于卷积神经网络的方法在对特定地区地震勘探记录的噪声的压制中表现突出,引起研究者广泛关注。深度学习的概念由Hinton^[95]在2006年提出,其中,卷积神经网络是目前较为流行的深度学习方法,先后有LeNet、VGGNet^[49]、ResNet^[51]、DenseNet^[52]等新结构被提出,这些结构被广泛应用于图像识别^[100],图像分割^[50]和图像分类^[51]中,并在图像去噪问题中有了不俗的表现。卷积神经网络具有权值共享和局部连接特性,其在应用于去噪任务时通过将低维特征逐步汇聚成为高维特征,并在最终的重建层依据高维特征重构原始信号,这种特性非常适合对地震记录中噪声的处理压制。

但是,对于同一个地震记录来说,不同区域的信号噪声性质往往存在差异,具体体现在两个方面:第一,噪声和信号分布不均,地震记录中的噪声水平在不同区域可能存在显著差异,某些区域可能噪声较大,而其他区域信号相对明显且噪声较小。这种不均匀的分布要求去噪模型具有高度的适应性和灵活性。第二,

信号往往具有高动态范围，地震记录中信号的强度变化范围广泛，这要求去噪模型能够处理从极弱到极强的信号，确保重要信息的保留而不丢失细节。

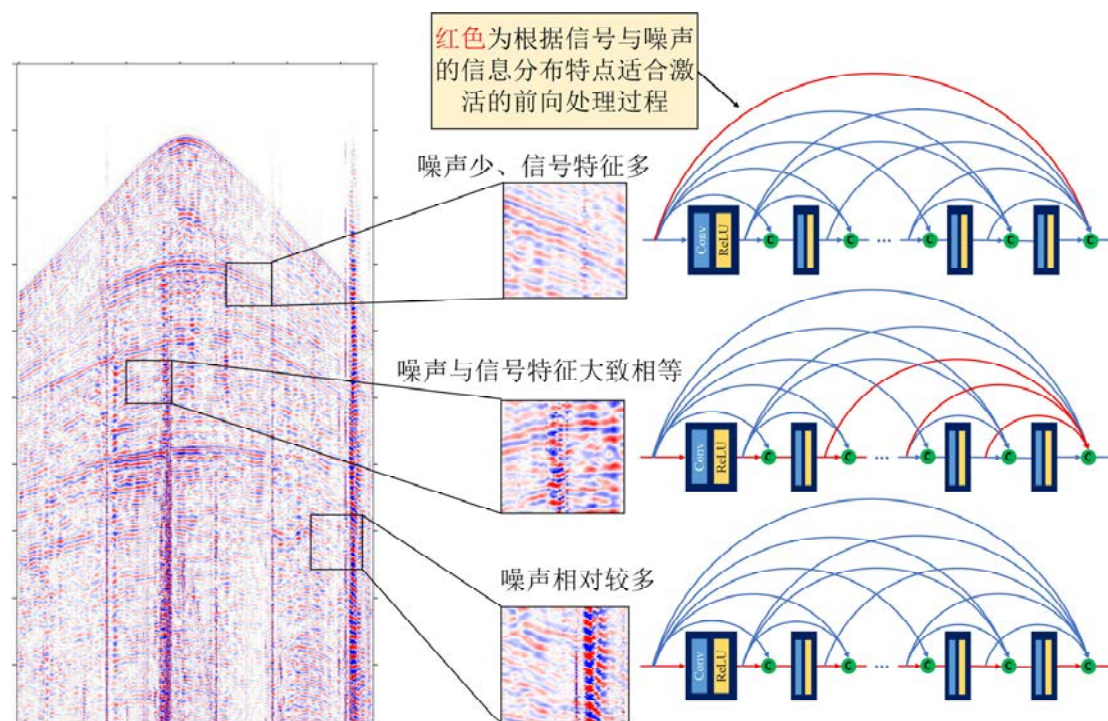


图 3.1 处理不同信噪比区域时网络最适合的前向过程

一般来说，神经网络层数的加深能够扩大感受野的范围，提取更多特征。使用大的感受野，能够使 CNN 在含噪信号中获取更多的上下文信息作为信号重建时的参考，提升去噪性能。但是伴随着网络层数的加深，会出现梯度消失和模型退化的问题。一系列正则化技术被提出并应用，例如批归一化层（Batch Normalization, BN）^[96]的出现可以在一定程度上解决梯度消失的问题，前馈去噪卷积神经网络（Feedforward Denoising Convolutional Neural Network, DnCNN）^[56]就应用了这种设置在图像去噪中实现了极好的去噪效果。通过跳层连接改善网络的信息流动也能让模型有更好的性能，如残差网络（Residual Network, ResNet）^[51]通过设置旁路连接，也能减轻卷积神经网络梯度消失和模型退化问题的产生。残差编码器-解码器网络（Residual Encoder-Decoder Networks, RED-Net）^[97]通过在卷积层和反卷积层之间设置镜像对称的跳跃连接，使得低维特征能够连接到高维特征中，实现了极好的去噪效果。但是，这些网络也难以根据地

震记录中不同区域的信噪比灵活地组合特征,往往无法实现对噪声强和噪声弱的区域的同时去噪,这种局限性使得模型在处理具有复杂噪声特性的地震数据时,难以达到既保留重要信号细节,又有效去除背景噪声的理想效果。2017年,Huang提出密集卷积网络(Densely connected convolutional networks, DenseNet)结构[52],[159],通过极致密集的跳跃连接,让信息在网络中充分流动,为网络性能提升带来一种高效方案。这也启示着密集连接可以作为改进用于地震记录去噪的卷积神经网络的一种方案,如图3.1所示,这种网络结构能够在处理具有不同信噪比的地震记录区域时,灵活地选择最适合的前向传播路径,并有效地组合不同层级的特征。这种密集连接的方式不仅能够增强模型对信息的捕捉能力,也使得网络在处理不同信噪比的区域时具有更好的适应性和灵活性,对于实现精准和有效的去噪有着至关重要的作用。

本章节中开发了一种基于 DenseNet 的去噪架构,与在卷积神经网络中采用彻底的密集连接不同,架构中采用宏观和微观的密集连接相结合的方式,在不增加太多的计算消耗的条件下实现网络中的密集连接。密集连接能够改善网络中的特征流动,缓解层数加深带来的梯度消失问题,使得网络更容易训练。同时,采用密集复杂的连接使得底层的特征重用,避免网络学习冗余特征。最重要的,密集连接的设计使得网络能够有效地组合低维和高维的特征,以实现地震有效信号的高质量重建。这极大地提升了模型在处理不同信噪比区域的信号恢复能力。通过这种方式,去噪架构不仅提高了去噪效果,而且增强了模型对复杂地震数据的适应性。通过采用声波正演获得纯净的地震数据,并与实际记录中采集的一部分地震噪声组成训练集对网络加以训练后,网络可以在特定测区的数据上发挥去噪作用。本研究在模拟地震数据上进行了实验验证,通过定量分析可以看出所提出方法的去噪效果要明显优于无跳跃连接的 DnCNN,表明密集连接的引入对网络去噪性能的明显增益。在实际记录的处理上,选择了海底电缆采集(Ocean Bottom Cable, OBC)的地震记录^{[150]-[156]}作为海洋地震记录的典型例子进行处理,训练好的网络获得了比传统方法更优的处理效果,且能够比无跳跃连接的卷积神经网络 DnCNN 拥有更好的处理效果,尤其是对于不同噪声水平区域的信号,本章提出的方法都能实现相对完整的恢复。

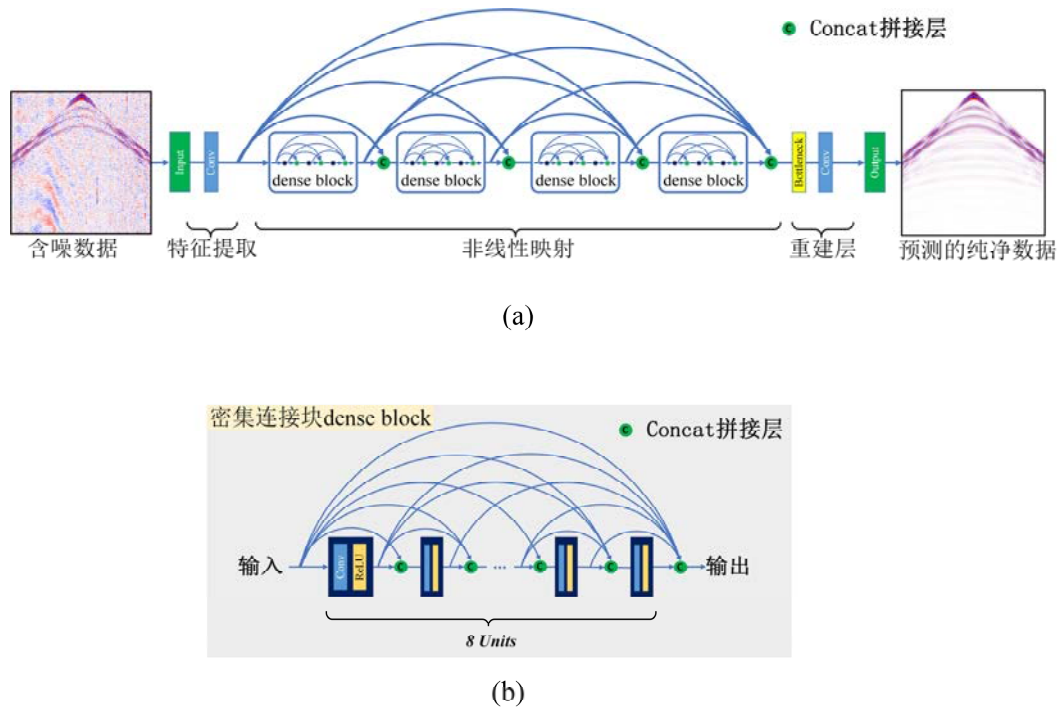


图 3.2 提出的 DnDenseNet 架构 (a) DnDenseNet 网络总体结构, (b) 微观密集连接块的结构

3.2 DnDenseNet 地震数据去噪算法

在去噪卷积神经网络中引入密集连接能够有效提升神经网络的去噪能力,在本节中,将从宏观和微观两个角度去描述网络的结构。地震数据中对代表地层反射的同相轴的恢复是重要的,然而,噪声干扰往往使同相轴难以识别。神经网络算法可以消除噪声,在信号重构过程中发挥着重要作用。为了使地震信号恢复更准确,本节提出的方法中向卷积神经网络中引入密集连接,有效地结合低维度和高维度特征,在准确恢复信号细节的同时实现噪声抑制。密集连接的引入可以促进梯度的反向传播,从而更好地优化网络。为了叙述方便,把这种网络称为 DnDenseNet (Denoising Densely connected convolutional networks)。此外,为了有效地训练网络,构建了一个完整的训练集来指导网络训练。接下来,从以下四个方面进行介绍:宏观的密集网络结构、微观密集连接块结构(网络的基本组成块)、训练网络的数据集构建和方法的去噪原理。

3.2.1 宏观密集网络结构

所使用的网络结构如图 3.2 所示, 其中的具体层的定义可见表 3.1, 整个网络结构包含有三个部分, 第一层卷积层用于提取信号特征, 最后一层的卷积层用于重建信号, 为了防止重建层的输入通道过多, 在输入层前添加瓶颈层作为特征降维的方法来降低通道数量, 以提高计算效率。网络中间的层部分用于实现非线性映射, 其中每一个组成块中都具有密集的连接。为了防止数据在前向传播过程中特征过度增长, 采用宏观密集连接和微观密集连接相结合的方式。微观密集连接块是网络用于实现非线性映射的中间层部分的基本组成块, 其由几个密集连接的卷积层组成, 用于实现局部的密集连接。微观密集连接块彼此之间也有着宏观的密集连接, 以确保各个层次的特征都能在重建层中被使用, 而且误差信号可以更直接地传播到前边的层中, 形成一种隐含的深度监督, 误差可以直接从最终层传递到前边的层, 前边的层可以从最终的重建层直接获得监督。

表 3.1 网络中各层的具体说明

层	描述	功能
Conv	卷积层	$y = W * x + b$
ReLU	线性修正单元	$y = \max(0, x)$
Bottleneck	瓶颈层	减小特征维度
Concat	特征连接层	合并特征图

3.2.2 微观密集连接块结构

微观密集连接块结构是网络的基本组成部分, 其中使用了 DenseNet 结构, 每个块的结构如图 3.1 (b) 所示, 具体层的定义见见表 3.1。微观密集连接块中, 每层都会接受前边所有层输出的特征作为本层的输入, 并进行处理, 这一层输出的特征也可以在后边的每一层中重用。因此, 第 i 层的输出为

$$X_i = H(X_0, X_1, \dots, X_{i-1}) \dots\dots\dots (3.1)$$

其中, $X_0, X_1, \dots, X_{l-1}$ 代表前边的网络层产生的特征映射的串联, H 表示该网络层的运算, 具体来说, 在本章中, 微观密集连接块中每层的处理都是一组卷积层和非线性激活层的组合, 因此第 l 层的输出可具体表示为

$$X_l = \max(0, w * (X_0, X_1, \dots, X_{l-1})) \dots \dots \dots (3.2)$$

其中 w 表示卷积层中的权重参数, $*$ 表示卷积符号, 可以看出在该层和前方每一层之间已经建立了短路径, 从而加强了网络中的特征传播, 也更加有利于训练时梯度的回传, 缓解了梯度消失的问题。在本研究中, 定义每个构建块由 8 个处理层组成, 每个处理层由一个卷积层和一个非线性激活层组成。设每个卷积层产生 g 个特征映射作为输出, 则一个微观密集连接块产生的特征映射数量为 g , g 被称为增长率, 其决定每个处理层产生多少信息用于微观密集连接块最后的输出。本研究中, 设定 $g=8$, 因此每个微观密集连接块可以产生 64 个特征图作为输出的计算中。此外, 由于瓶颈层的存在, 网络的输出不会过度膨胀。在瓶颈层中, 通过采用 1×1 的卷积操作, 减少特征图的通道, 同时保持特征的主要信息, 可以有效降低后续卷积层的输入特征图数量。这种压缩机制使得网络能够在保持较深层次结构的同时, 避免了由于参数过多导致的过拟合问题, 并且使得网络在训练过程中更加稳定和高效。

3.2.3 数据集设置

深度学习通过利用特征学习来解决以往需要人工设计特征提取算子的重要难题。数据集作为深度学习算法的重要基础。训练去噪算法的模型的数据集往往需要两部分: 含噪信号与对应匹配的纯净信号。在地震勘探中, 往往只能获得含噪信号而难以获得与之匹配的纯净的地震信号, 这会使得有监督学习算法无法运行。DnDenseNet 是有监督学习模型, 这意味着需要高质量的信号集和噪声集用于指导网络的训练。在地震勘探中, 纯净的信号难以获得, 对已有数据去噪后的数据往往质量不佳。因此采用正演的方式生成理论纯净信号, 并与无源情况下现场采集的噪声混合构建含噪信号。

为了获得模拟的纯净地震数据, 本研究放置了一个人工震源来激发地震波, 并沿水平方向放置了 280 个接收器来记录地震波场。本研究中构建了 100 个不同

地层构造的速度模型。图 3.3 (a) 展示了其中一个速度模型。在图中, 三角形表示震源, 黑线表示地震检波器。关于正演的更多细节如表 3.2 所示。在这项研究中, 模拟纯净数据是基于时域中的声波方程进行建模的, 方程如公式(3.3)所示。

$$\frac{\partial^2 u(x, y, t)}{\partial t^2} = v(x, y)^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + s(x, y, t) \dots\dots\dots (3.3)$$

其中 v 表示速度, u 表示声波场, s 表示源函数。 (x, y) 表示空间位置, t 表示时间。将源类型设置为雷克子波。公式如下:

$$f(t) = A \left[1 - 2 \times (\pi f_0 (t - t_0))^2 \right] \times e^{(-\pi f_0 (t - t_0))^2} \dots\dots\dots (3.4)$$

其中 A 表示信号幅度, t_0 表示起始时间, f_0 表示主频, 通过数值模拟生成与每个速度模型相对应的地震数据。

表 3.2 正演参数设置

参数	含义	具体设置
震源子波类型	地震信号的小波函数的种类	雷克子波
子波频率	地震信号的频率范围	30-50 Hz
波速范围	地震波在介质中传播的速度范围	1500-3700 m/s
检波器间隔	检波器接收数据时的空间间隔	25 m
采样时间间隔	地震数据记录中连续采样数据的时间间隔	2×10^{-3} s
地层深度	用于指定模拟地震波传播的地层深度	3000 m
接收器数量	地震数据采集中使用的接收器的数量	280
最大传播时间	地震波从震源到接收器的最长传播时间	4 s

图 3.3 (b) 显示了与图 3.3 (a) 中的速度模型相对应的地震记录。通过求解声波方程, 总共生成了 100 个类似的模拟干净地震记录。由于尺寸过大的数据送入网络会带来巨大的计算量, 因此本研究中通过裁剪数据, 使用切片的方式训练网络。由于实际记录中潜在的纯净信号与模拟的纯净信号在局部特征更加一致, 因此能够保证网络在处理实际信号时候依旧保持泛化性能。采用步长为 10, 尺寸为 100×100 的滑动窗裁剪数据, 将振幅归一化的合成地震数据划分为大小为

100×100 的贴片，以获得干净的信号集。振幅归一化是必要的，以确保输入数据在不同样本中具有相似的幅度，这是许多深度学习算法的常用设置。具体来说，将不同的比例因子应用于数据中每个样本，使每个数据的最大绝对值变为 1，即这个比例因子是计算每个样本的最大绝对值得到的。这一步骤对于地震数据去噪任务尤其重要，由于检波器和震源之间的距离变化，地震数据可能具有宽范围的振幅。通过执行振幅归一化，能够提高去噪算法对地震数据进行信号恢复和噪声抑制的性能和可靠性。

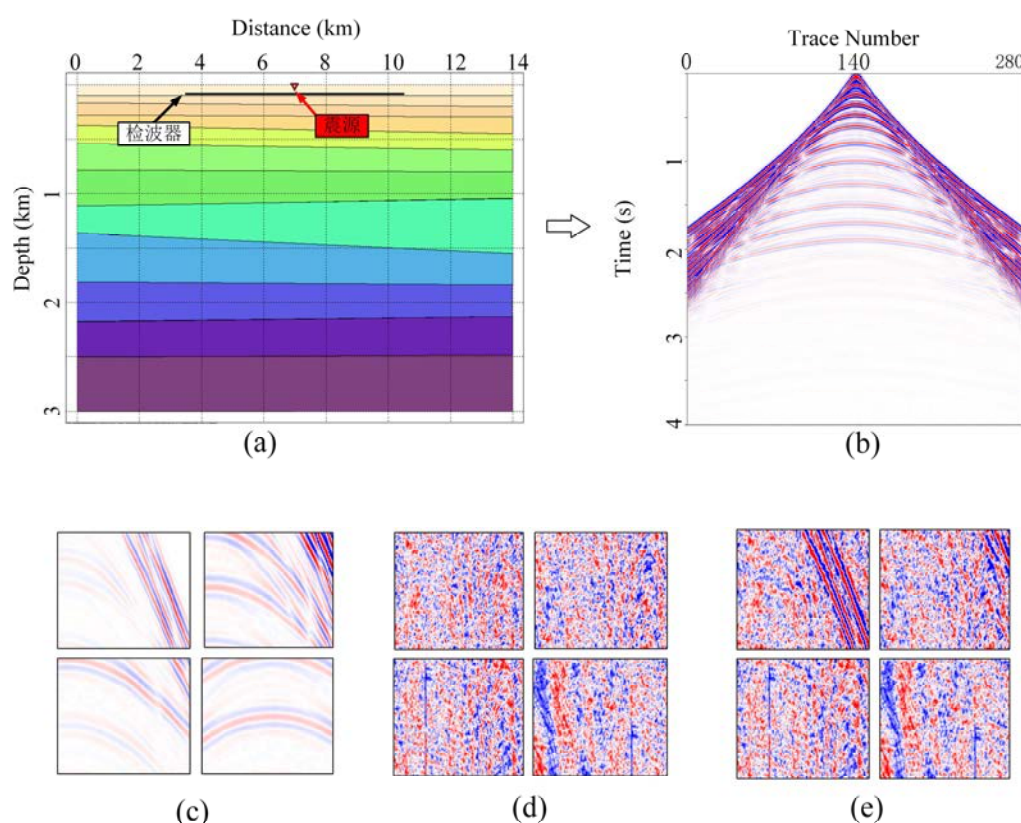


图 3.3 数据集的构建过程 (a) 速度模型, (b) 地震记录, (c) 部分信号数据集切片, (d) 部分噪声数据集切片, (e) 部分含噪数据集切片

此外，噪声集的构建也是极为重要的，决定了网络去除噪声的类型。本文的策略是针对要处理的具体的实际记录，选择在无源条件下收集的噪声和实际地震记录中初至波之前的噪声来形成噪声集。使用步长为 10、大小为 100×100 的滑动窗口来裁剪所收集的噪声数据，以生成噪声训练集。

接下来，无重复随机抽选信号集中的一组数据和噪声集中的一组数据，随后，通过将它们相加来获得噪声信号。构造一个有噪信号集和相应的标签集在有监督

学习范式下训练去噪模型。对应的纯净信号是网络在输入为含噪信号时的标签，对两者进行相同幅度的归一化。在图 3.3 中，还展示了一些干净的信号切片、噪声切片和含噪信号切片的形态。

3.2.4 去噪原理

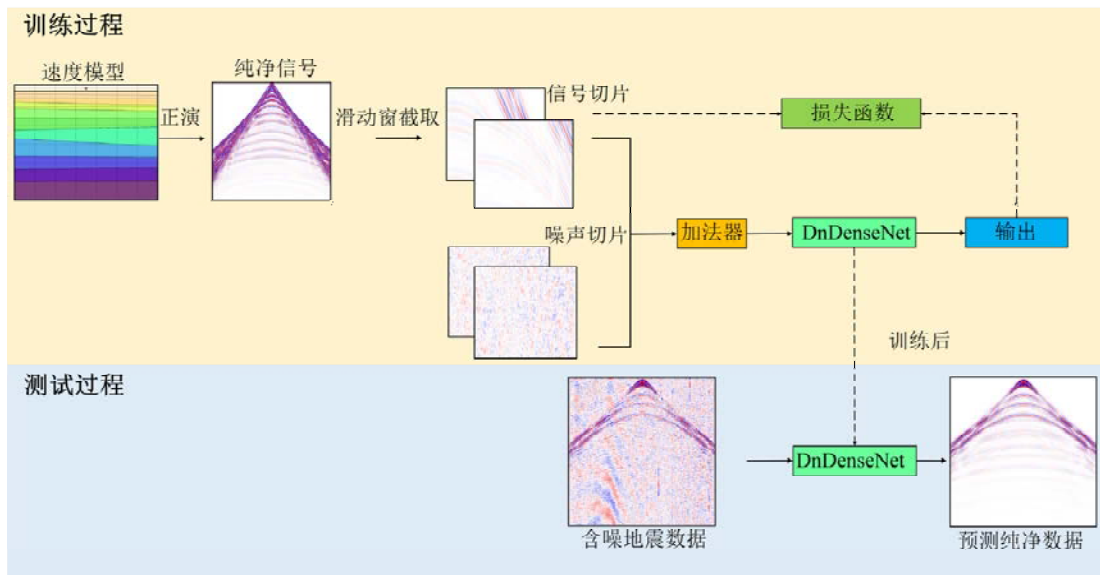


图 3.4 所提出方法的去噪流程图

地震记录在采集中往往会受到大量随机噪声干扰，含噪信号可以表示为

$$y = s + n \quad \dots\dots\dots (3.5)$$

其中， y 表示含噪信号， s 表示有效信号， n 表示噪声信号。使用的网络旨在学习从含噪信号到去噪信号的端到端映射函数 D ，进而通过它得到纯净信号的估计：

$$\tilde{s} = D(y; \Theta) \quad \dots\dots\dots (3.6)$$

其中， Θ 是网络中的参数，整个网络本质上是一个拥有巨大参数量的复杂函数，它可以拟合高度复杂的映射关系。训练的重要一环是设定损失函数，反馈给网络当前的性能表现。设定的损失函数为

$$L = \frac{1}{2N} \sum_{i=1}^N \|D(y_i; \Theta) - s_i\|_F^2 \quad \dots\dots\dots (3.7)$$

算法 3.1 密集连接卷积神经网络去噪系统

输入:

批处理大小 N ; 迭代周期数 E ; 一个迭代周期内的迭代次数 I ; 幅度归一化后的噪声集 n ; 幅度归一化后的信号集 x ; 网络模型想要学习到的映射关系 D ; 测试样本 $\hat{y}$ 。

训练阶段:

- 1: **for** $m=1,2,3,\dots,E$ **do**
- 2: **for** $n=1,2,3,\dots,I$ **do**
- 3: 从归一化后的信号集中取第 i 组信号数据 $\{x^{(i)}\}_{i=1}^N \in x$
- 4: 从归一化后的噪声集中取第 i 组噪声数据 $\{n^{(i)}\}_{i=1}^N \in n$
- 5: 叠加获得不同信噪比的含噪数据 $\{noisy^{(i)}\}_{i=1}^N = \{x^{(i)}\}_{i=1}^N + \{n^{(i)}\}_{i=1}^N$
- 6: Input $\{noisy_{nr}^{(i)}\}_{i=1}^N = \frac{\{noisy^{(i)}\}_{i=1}^N}{\{max(noisy^{(i)})\}_{i=1}^N}$
- 7: Label $\{x_{nr}^{(i)}\}_{i=1}^N = \frac{\{x^{(i)}\}_{i=1}^N}{\{max(noisy^{(i)})\}_{i=1}^N}$
- 8: $\theta \leftarrow \nabla_{\theta} \left[\frac{1}{2N} \sum_{i=1}^N \|D(input) - label\|^2 \right]$
- 9: **end for**
- 10: **end for**

测试阶段:

- 1: **for** 测试样本 y **do**
- 2: 用训练阶段生成的随机权重和偏置获得参数矩阵 Θ
- 3: 根据式 (3.8) 计算输出结果 $\tilde{s} = D(\hat{y}; \Theta)$
- 4: **end for**

输出: $\tilde{s}$

其中 (y_i, s_i) 代表 N 对训练样本, 其中, s_i 是网络输入为 y_i 时网络的标签也即期望的网络输出, $\|\cdot\|_F^2$ 代表 Frobenious 范数。神经网络的训练过程是通过不断地调整网络参数, 以减小网络输出与期望输出之间的差异, 即减小损失函数的值。在训练初期, 网络参数是随机初始化的, 这时候网络的输出与期望输出之间的差距通常较大。通过计算损失函数, 可以得知当前网络的性能表现。损失函数是衡量网络性能的关键, 它反映了网络输出与实际标签之间的差异大小。为了优化网络, 采用梯度下降算法来优化网络, 这种方法通过计算损失函数对网络参数的梯度, 能够指示损失函数下降最快的方向。神经网络的训练是一个迭代过程, 每一次迭

代包括前向传播和反向传播两个步骤。前向传播是指输入数据在网络中逐层传递直至输出层,得到网络的输出;反向传播则是根据损失函数对每一层参数进行梯度计算,并更新参数。这个过程会不断重复,直到网络性能满足预定的条件或达到最大迭代次数。图 3.4 展示了所提出算法的流程图,具体算法如算法 3.1 所示。由于梯度下降算法可能会导致训练过程中的步长过大或过小,使用其变种即 Adam 算法^[98]来优化网络。训练过程的更多细节如表 3.3 所示。经过训练后,使用训练后的网络来处理有噪声的地震数据。测试阶段的目的是评估网络在未见过的数据上的性能,即泛化能力。本研究中使用独立于训练集的测试集来进行这一评估。通过对测试集中的数据进行处理,并将网络的输出与实际标签进行比较,且通过计算出诸如信噪比等性能指标,帮助评估网络是否已经足够好地学习到了数据的内在规律,以及是否能够有效地应用于实际的地震数据去噪问题中。对于输入 $\hat{y}$, 去噪后的数据 $\tilde{s}$ 可以表示为

$$\tilde{s} = D(\hat{y}; \Theta) \cdots \cdots \cdots (3.8)$$

表 3.3 网络超参数设置

超参数	含义	具体设置
批次大小	在训练神经网络时一次输入的样本数量	32
轮次	整个训练数据集在神经网络上反复迭代的次数	100
学习率范围	控制神经网络权重更新幅度的参数	$[1 \times 10^{-4}, 1 \times 10^{-3}]$
训练贴片大小	用于训练时的数据集的尺寸大小	100×100
优化器	用于调整模型权重以最小化损失函数的算法	Adam 算法

3.3 算法性能验证

为了验证方法对于地震数据中噪声的去噪能力,本研究在模拟记录 and 实际记录中展开实验,并与不同的对比方法进行比较。模拟实验中使用量化指标对去噪

后的记录进行量化评估,并在时域和频域中对去噪结果进行分析,以量化各个方法去噪性能。在实际实验中,使用不同的去噪方法在从海洋地震勘探采集的 OBC 实际地震记录上进行去噪实验。本研究在搭载有 Intel Xeon E5-2620 CPU 和两个 NVIDIA GeForce 1080Ti GPU 的工作站上进行实验。实验中使用 Pytorch 深度学习框架^[141]构建网络,实验在 Python 环境下进行。实验中对比方法包括带通滤波器、F-X 反褶积^[35]、小波变换滤波^[142]、鲁棒主成分分析 (Robust Principal Component Analysis, RPCA)^[165]、加权核范数最小化 (Weighted Nuclear Norm Minimization, WNNM)^[143]和前馈去噪卷积神经网络 (Denoising Convolutional Neural Network, DnCNN)^[56]。其中,带通滤波器是一种线性滤波器,可以通过设置截止频率,抑制其他频率。使用一个低截止频率为 10 Hz、高截止频率为 50 Hz 的四阶巴特沃斯滤波器来去除低频和低频噪声,并将有用的信号信息保留在特定的频率范围内。F-X 反褶积是一种可以补偿地下速度变化引起的小波拉伸和压缩的方法。使用正则化参数为 0.1 的整形正则化方法来执行 F-X 反褶积。小波变换滤波器是一种时频分析工具,可以将数据分解为不同的尺度和频率,并抑制高频分量中的噪声。使用了级别为 3 的 db4 小波来提高信噪比。RPCA 是一种可以分离数据中的信号和噪声成分的方法。使用秩为 10、正则化参数为 1 的不精确增广拉格朗日乘子 (Inexact Augmented Lagrangian Multiplier, IALM) 算法来执行 RPCA,最大迭代次数设置为 1000。WNNM 是一种利用加权核范数最小化方法来抑制数据中噪声的去噪技术。具体来说,WNNM 通过解决使加权核范数和数据保真度项之和最小化的凸优化问题来达到去噪的目的。加权核范数鼓励信号的低秩,而数据保真度项能确保去噪信号与原始信号尽可能接近地匹配。本研究中,核范数的权重系数设置为 0.54,数据保真度项的正则化参数设置为 2.8。为确保方法训练条件一致,对 DnCNN 的设置也如表 3.3 中所示。此外,为了便于描述,将本章提出的方法一律简称为 DnDenseNet。

3.3.1 模拟记录处理结果

首先,通过正演生成模拟地震数据,图 3.5 (a) 为一个七层的二维地质模型,横坐标为水平距离,纵坐标为深度,设置 P 波从顶层到底层的速度分别为 1500, 1900, 2100, 2500, 2700, 2900 和 3100 m/s。其中,第一层代表海水,其中速度设置为 1500 m/s。因此,此正演实验模拟的是海洋勘探的一个场景。图中,

倒三角代表震源，水平线代表构成的测线。源与海面的距离为 20 m，水听器与海面的距离设置为 50 m。时间采样间隔为 2 ms，检波器的空间采样间隔为 25 m。测试过程中关于正演的设置如训练时即如表 3.2 中的设置一致。通过求解声波方程，可以得到理论上的无噪声地震数据，如图 3.5 (b) 所示。向其中添加如图 3.5 (c) 所示的实际采集过程中检波器记录的噪声，这些噪声来自于海洋地震勘探 OBC 记录中的初至前噪声。图 3.5 (d) 显示了模拟的含噪的地震数据，本研究中合成了 30 个不同信噪比的含噪记录作为测试集。用于组成测试集的数据与训练集数据相互独立。

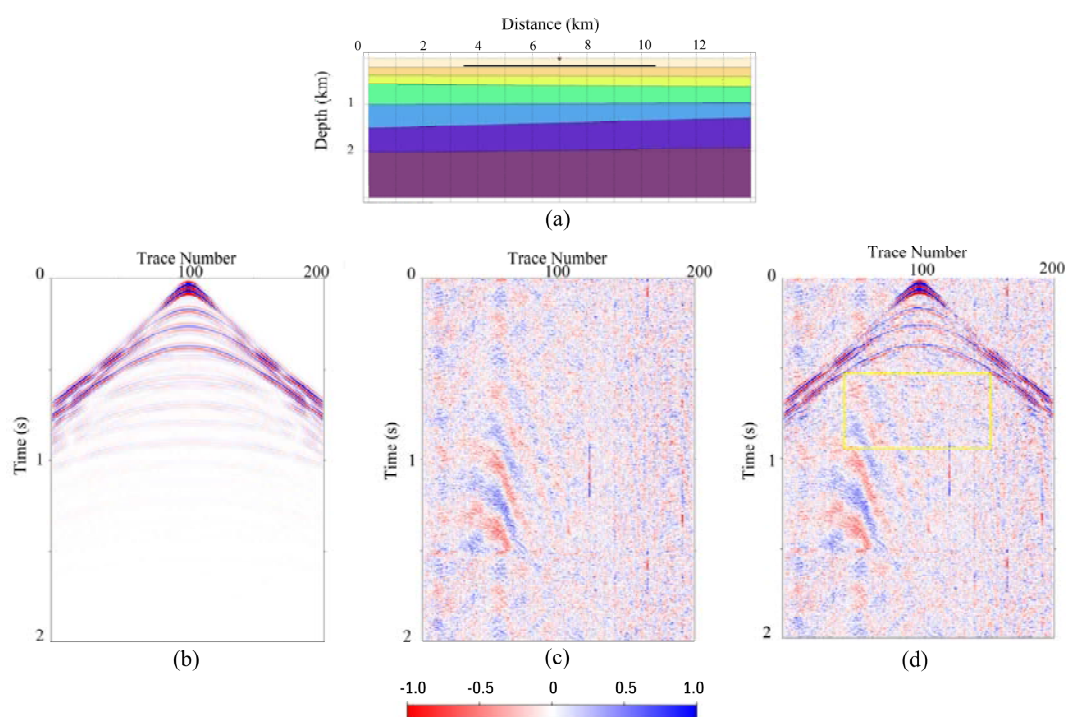


图 3.5 模拟含噪地震数据的构建过程示意图 (a) 速度模型, (b) 纯净地震数据, (c) 加入的噪声, (d) 合成含噪地震数据

图 3.6 展示本章提出的 DnDenseNet 方法的结果和其他对比方法的结果。图 3.7 则展示了不同方法去噪后的差值。本章方法处理后，结果和差值分别如图 3.6 (b) 和 3.7 (b) 所示，其中图 3.7 (b) 所示的差值是指在图 3.5 (d) 和 3.6 (b) 之间的差异，也可以认为它是方法所预测的噪声，从中可以间接观察评估去噪方法对于信号的保幅能力。可以看出，本章提出的方法可以完全抑制噪声。从信号恢复的角度来看，波场中的各种信号都被恢复，原本被噪声污染的直达波、反射

波可以清晰地观察到,这说明方法可以实现期望的效果。本研究中也在频域中对各种方法的信号和差值进行了对比分析,在频域中,可以观察到各种方法对信号分量和噪声分量的分离情况。不同方法结果对应的频谱图如图 3.8 所示,不同方法对应的差值图的频谱图如图 3.9 所示。整体来看,频域的分析也支持相同的结论,即 DnDenseNet 方法在信号的保留上相比其他几个对比方法有着更好的效果。

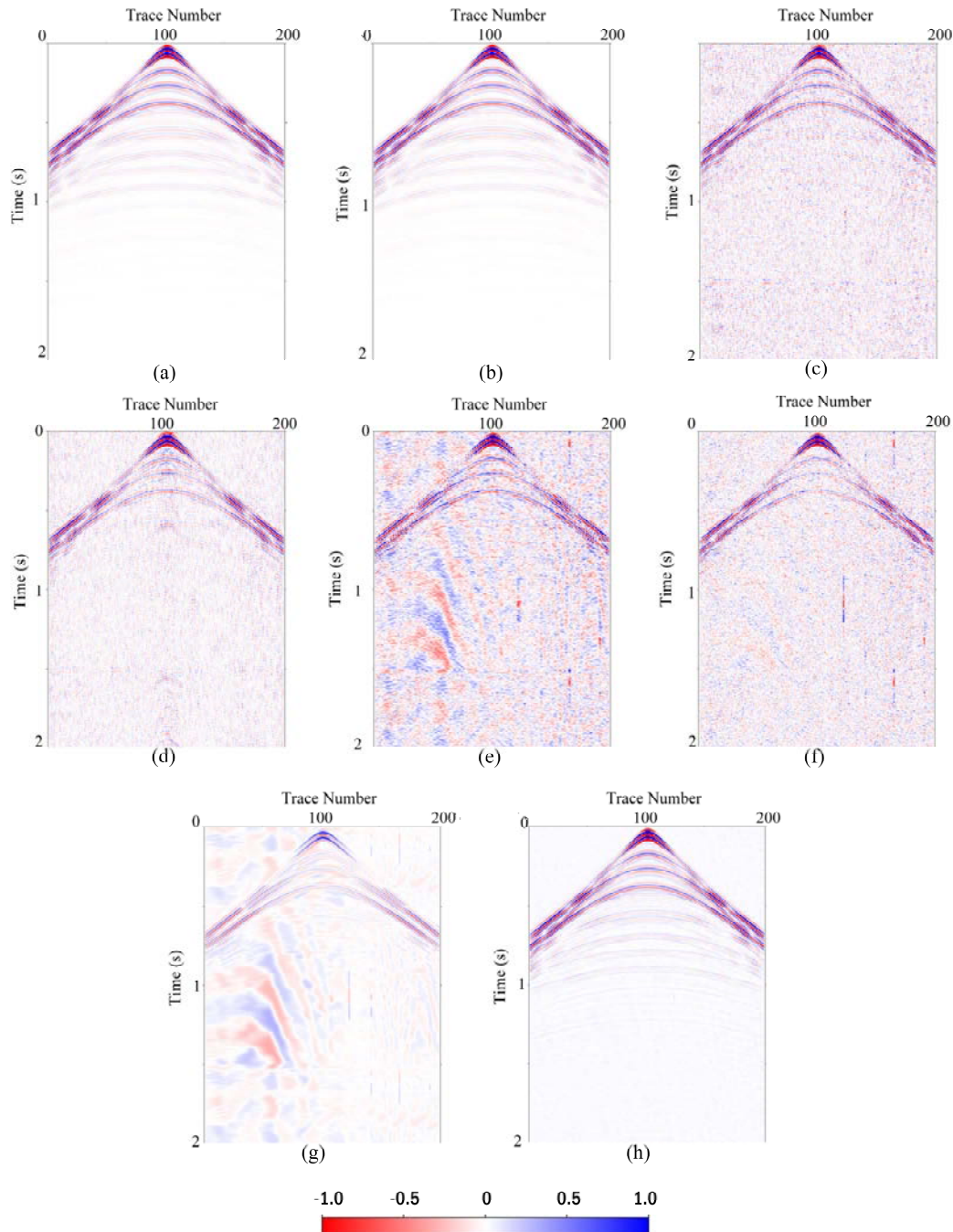


图 3.6 不同方法的去噪结果对比 (a) 纯净信号, (b) DnDenseNet 结果, (c) 带通滤波器结果, (d) F-X 反褶积结果, (e) 小波变换滤波结果, (f) RPCA 结果, (g) WNNM 结果, (h) DnCNN 结果

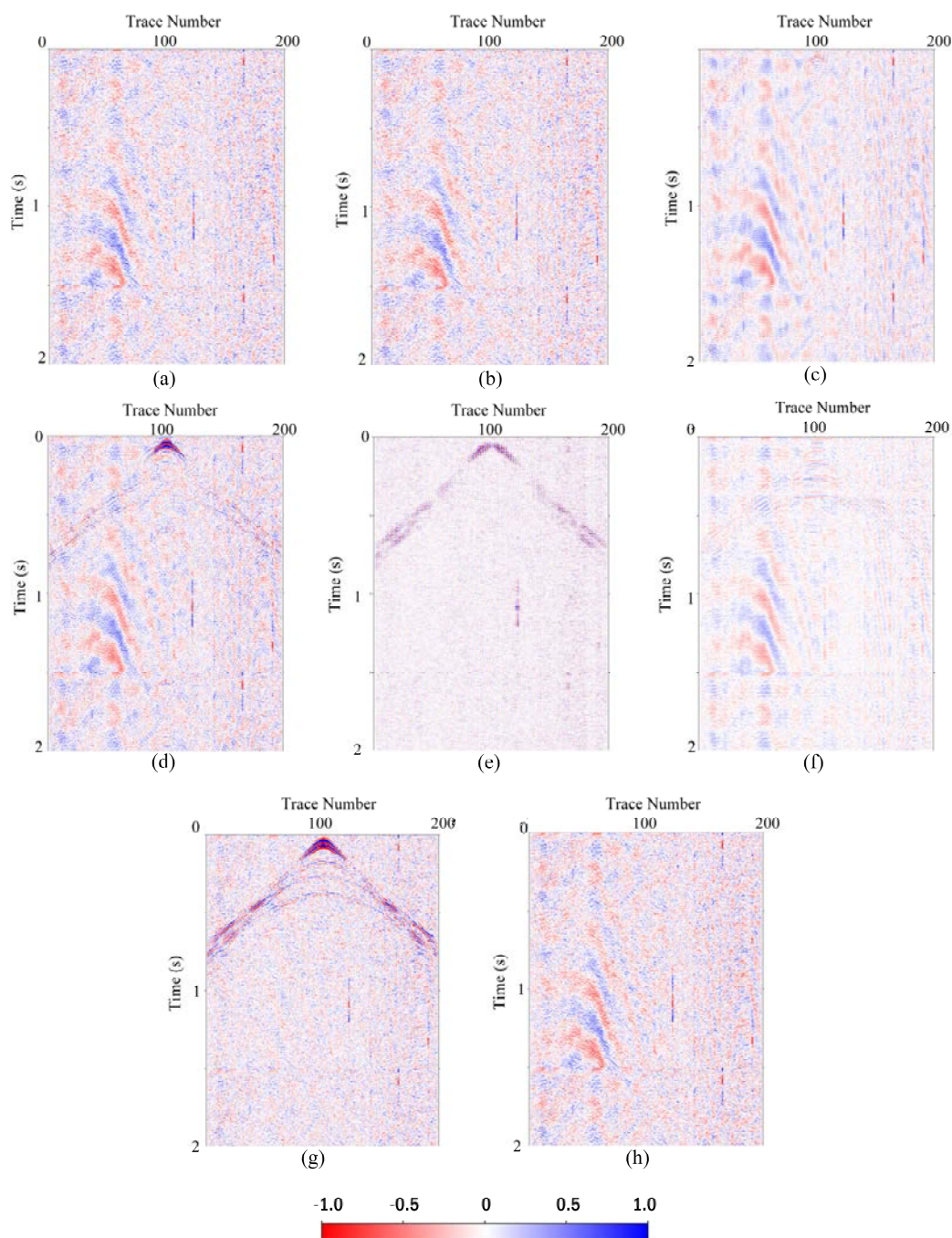


图 3.7 不同方法去噪后的差值对比。(a) 纯噪声数据, (b) DnDenseNet, (c) 带通滤波器, (d) F-X 反褶积, (e) 小波变换滤波, (f) RPCA, (g) WNNM, (h) DnCNN 差值

尽管如此,在将本方法与同为深度学习类去噪网络的 DnCNN 对比时,依旧难以直接从时域和频域的结果中分辨出具体哪种方法有着更好的去噪效果。因此,需要借助后续的量化指标来对比两者。

为了反映地震数据的整体质量,衡量整体去噪效果,选取了几项指标用于客观评估。利用信噪比 (Signal-to-noise ratio, SNR)、平均绝对误差 (Mean absolute

error, MAE)、均方误差 (Mean-square error, MSE), 以及结构相似性 (Structural similarity, SSIM) [99]四个指标对几种方法的结果进行定量评估。信噪比、平均绝对误差、均方误差以及结构相似性的公式如下所示:

$$SNR = 10 \log \frac{\sum_x \sum_t |s(x, t)|^2}{\sum_x \sum_t |d(x, t) - s(x, t)|^2} \dots\dots\dots (3.9)$$

$$MAE = \frac{1}{MN} \sum_x \sum_t |d(x, t) - s(x, t)| \dots\dots\dots (3.10)$$

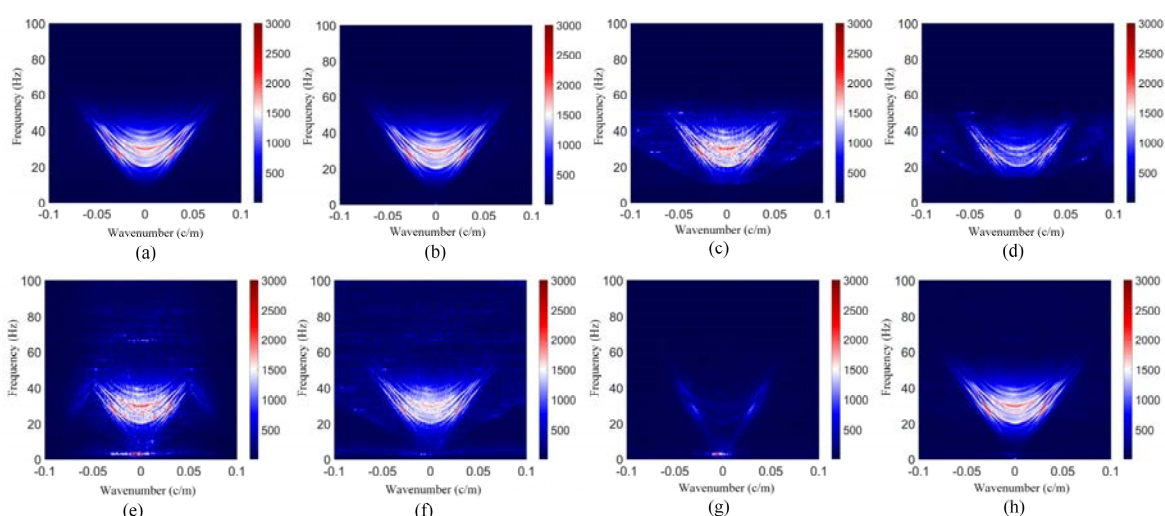


图 3.8 不同方法去噪结果的 F-K 谱分析 (a) 纯信号数据, (b) DnDenseNet, (c) 带通滤波, (d) F-X 反褶积, (e) 小波变换滤波, (f) RPCA, (g) WNNM, (h) DnCNN

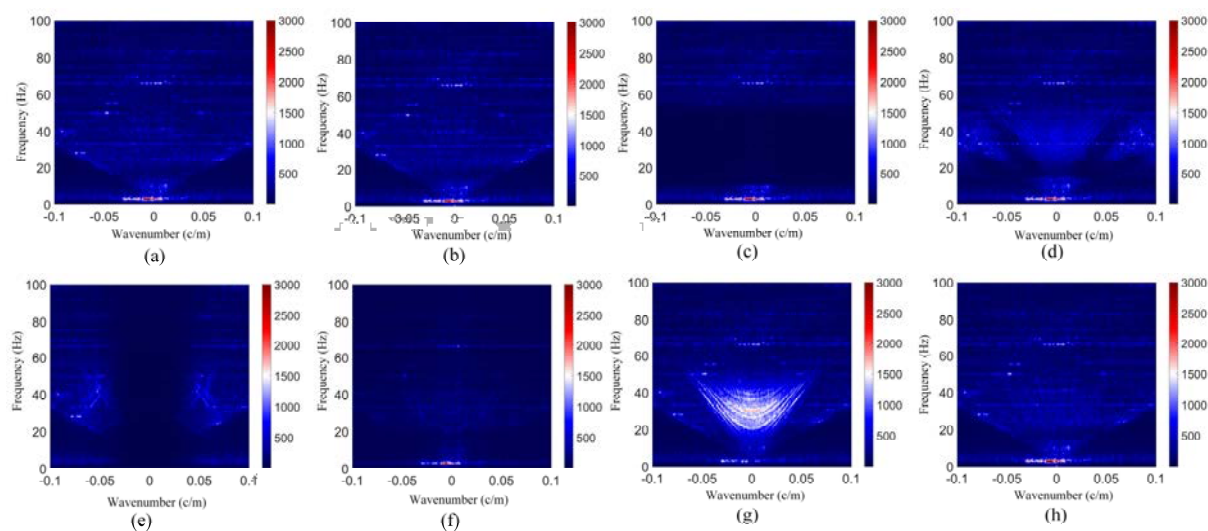


图 3.9 不同方法差值的 F-K 谱分析 (a) 纯噪声数据, (b) DnDenseNet, (c) 带通滤波, (d) F-X 反褶积, (e) 小波变换滤波, (f) RPCA, (g) WNNM, (h) DnCNN

$$MSE = \frac{1}{MN} \sum_x \sum_t |d(x, t) - s(x, t)|^2 \dots\dots\dots (3.11)$$

$$SSIM = \frac{(2\mu_s\mu_d + c_1)(2\sigma_{sd} + c_2)}{(\mu_s^2 + \mu_d^2 + c_1)(\sigma_s^2 + \sigma_d^2 + c_2)} \dots\dots\dots (3.12)$$

一般来说,更好的去噪结果往往意味着更高的信噪比、结构相似性和更小的平均绝对误差和均方误差。表格 3.4 显示了各个方法去噪前后的记录的各项指标,从中可以看出,从四种指标的评估上,本章提出的 DnDenseNet 方法在不同的信噪比场景下都有着最高的信噪比提升,其他指标也均验证了其提升不仅优于传统方法,更要优于深度学习方法 DnCNN,尤其是与 DnCNN 的对比说明了密集连接的引入对于改进传统卷积神经网络结构的去噪性能有着重要的作用。

表 3.4 不同方法的去噪性能的量化指标对比

不同方法	MAE	MSE	SNR	SSIM
原始记录	1.613×10^{-2}	4.21×10^{-4}	3.1731	0.52963
带通滤波	9.946×10^{-3}	1.58×10^{-4}	7.4293	0.75907
F-X 反褶积滤波	8.03×10^{-3}	2.044×10^{-4}	6.3169	0.85928
小波变换滤波	1.323×10^{-2}	3.099×10^{-4}	4.5093	0.59541
RPCA	1.097×10^{-2}	2.302×10^{-4}	5.8008	0.7113
WNNM	6.975×10^{-3}	6.862×10^{-4}	1.0575	0.82899
DnCNN	1.043×10^{-3}	1.88×10^{-6}	26.6802	0.99017
DnDenseNet	1.027×10^{-3}	1.6561×10^{-6}	27.2309	0.99046

为了对各个区域的去噪情况进行衡量,使用局部信噪比来描述局部地震数据的质量,以定量评估所提出方法的去噪性能。具体来说,使用一个大小为 5×5 、步长为 1 的移动窗口来分割去噪前后的地震数据,并进一步在移动窗口中计算数据的信噪比作为局部信噪比。窗口中的局部信噪比可以表示为

$$L-SNR(x, t) = 10 \log_{10} \frac{\sum_{i=t-(w-1)/2}^{t+(w-1)/2} \sum_{j=x-(w-1)/2}^{x+(w-1)/2} (s(i, j) - \bar{s}(i, j))^2}{\sum_{i=t-(w-1)/2}^{t+(w-1)/2} \sum_{j=x-(w-1)/2}^{x+(w-1)/2} (d(i, j) - \bar{d}(i, j))^2} \dots\dots (3.13)$$

其中, w 是窗口长度, s 是无噪声数据, $\bar{s}$ 是 s 的均值, d 是去噪数据。图 3.10 显示了不同方法去噪之后的局部信噪比结果图, 其中横坐标为地震道数, 纵坐标为采样点数, 右侧颜色指示条表示信噪比, 单位是分贝 (dB)。其中, 合成的含噪地震数据的局部信噪比如图 3.10 (a) 所示。受噪声影响, 不同的区域的局部信噪比呈现不同程度的下降。图 3.10 (b) 显示了 DnDenseNet 方法处理的结果的局部信噪比, 可以看出, 有效信号区域的局部信噪比明显得以提升。同时, 相比于其他几种方法, 本文方法相对来说在局部信噪比上的提升更加明显。尤其值得强调的是, 在与 DnCNN 方法的结果的对比中可以看出, DnDenseNet 方法处理后记录的局部信噪比提升更加明显, 这说明密集连接的引入能够大大增加卷积神经网络对局部信号的恢复能力和水平。

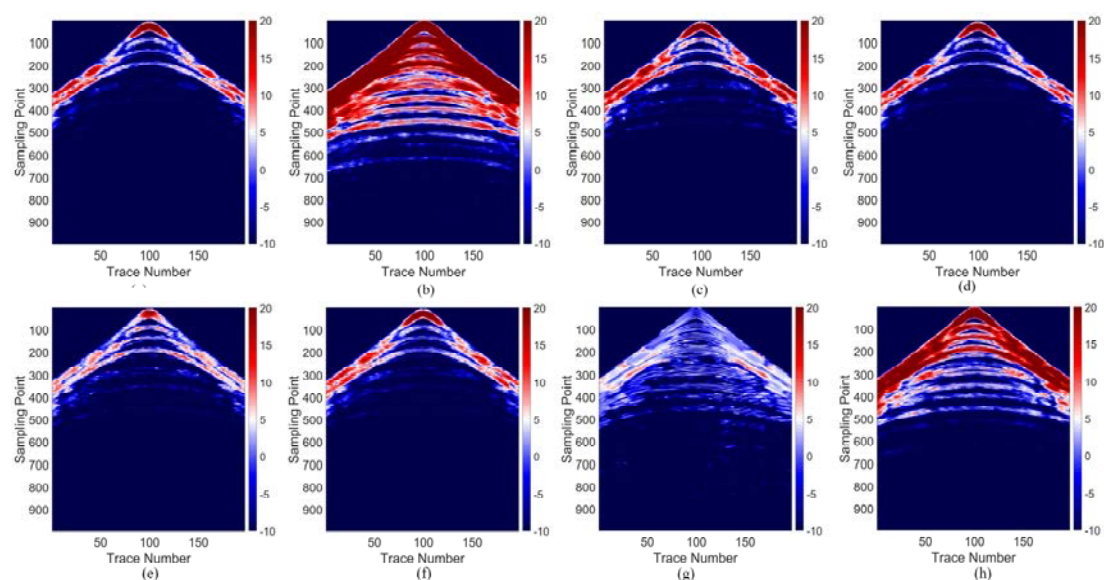


图 3.10 不同去噪方法结果的局部信噪比分析 (a) 合成含噪数据, (b) DnDenseNet, (c) 带通滤波, (d) F-X 反褶积, (e) 小波变换滤波, (f) RPCA, (g) WNNM, (h) DnCNN

3.3.2 实际记录处理结果

在深入研究 DnDenseNet 方法在地震数据去噪中的应用之前, 本研究已经在模拟数据上进行了广泛的测试。这些模拟数据能够在信噪比受控的场景下模拟地震数据的各种特性, 包括噪声强度以及信噪比等。通过在模拟数据上的测试, 已经验证了 DnDenseNet 方法在噪声压制中的有效性。

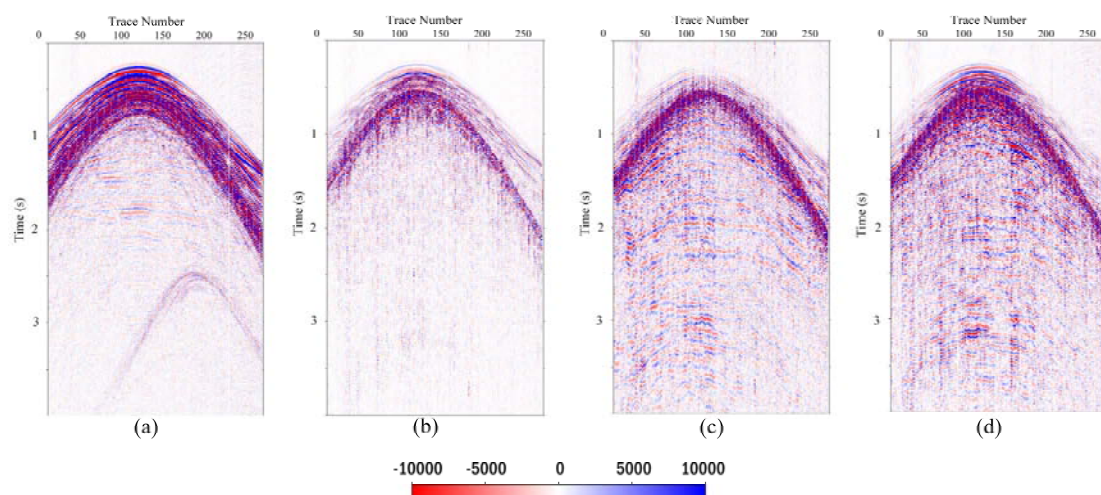


图 3.11 OBC 实际数据 (a) P 分量, (b) X 分量, (c) Y 分量, (d) Z 分量

在模型验证的基础上,接下来将研究重点转向实际地震数据的处理,以评估 DnDenseNet 方法在实际应用场景中的去噪性能。实际地震数据相比模拟数据更为复杂,因为它包含了多种未知的噪声来源和信号特性,这对去噪方法提出了更高的要求。在实际地震数据的处理中,本研究选择了南海涠西南试验区采集的 OBC 地震数据中的 P 分量作为研究对象。OBC 地震数据是一种在海洋地震勘探中使用的数据,主要通过海底铺设的电缆来收集。这种技术允许研究者通过电缆上的地震传感器捕捉到由地震震源(如气枪)产生的声波反射信号。这项技术特别适用于复杂海底地形的地区,可以反映更加精确的地下结构信息。选择 OBC 的 P 分量作为去噪对象有着特殊的考虑。首先,OBC 数据由于其采集环境特殊,常受到水流、设备运动等因素的影响,导致数据中存在大量噪声,特别是在中深层位置,这对数据的进一步处理和解释提出了挑战。其次,P 分量(即水听器数据)是当前地震数据中用于后续成像的主要成分,直接关系到地震信号的识别和分析。因此,对 P 分量数据的有效去噪不仅能够提高数据质量,还能够为后续的反演成像、地震解释提供更准确的信息。在 OBC 的 P 分量数据中,常见的噪声类型包括:(1)流水噪声,由于海水流动造成的噪声,这种噪声随着水流速度的变化而变化,可能会掩盖或模糊地震信号,往往呈现高频;(2)设备噪声,由于海底地震仪或电缆移动、振动等造成的噪声,往往呈现低频;(3)面波噪声:在海底地震数据中,面波可能会产生较强的噪声,干扰信号的识别。此外,除了噪声外,多次波处理是 OBC 地震数据处理中的一个重要方面,但它需要专门的方法和技术。本研究中选择专注于噪声处理,暂时不对多次波压制进行研究,因为

在获得清晰的信号之前，首先需要解决的基本问题是如何有效去除数据中的噪声。

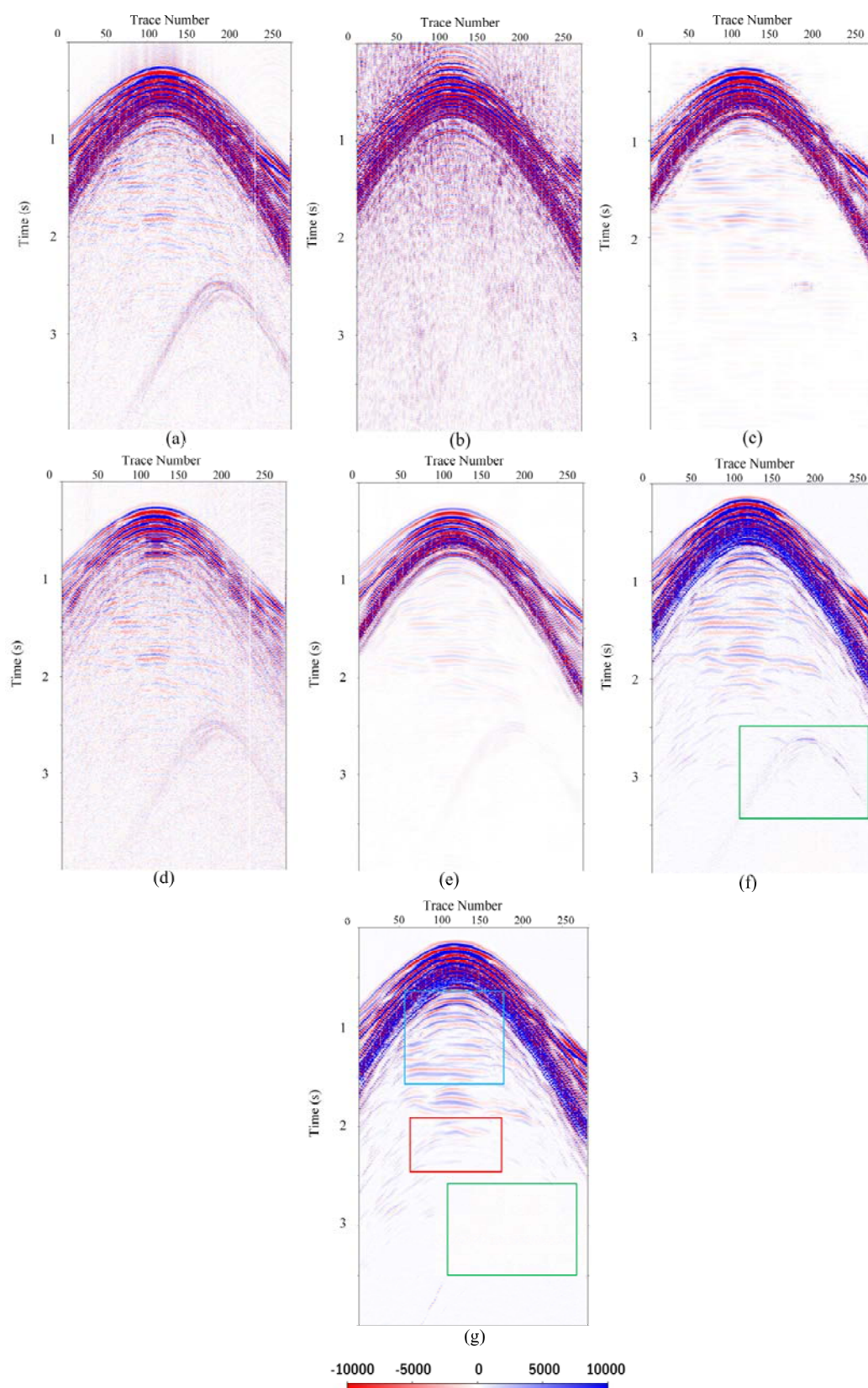


图 3.12 不同方法对实际 OBC 数据 P 分量处理后的去噪结果对比 (a) 带通滤波, (b) F-X 反褶积, (c) 小波变换滤波, (d) RPCA, (e) WNNM, (f) DnCNN, (g) DnDenseNet

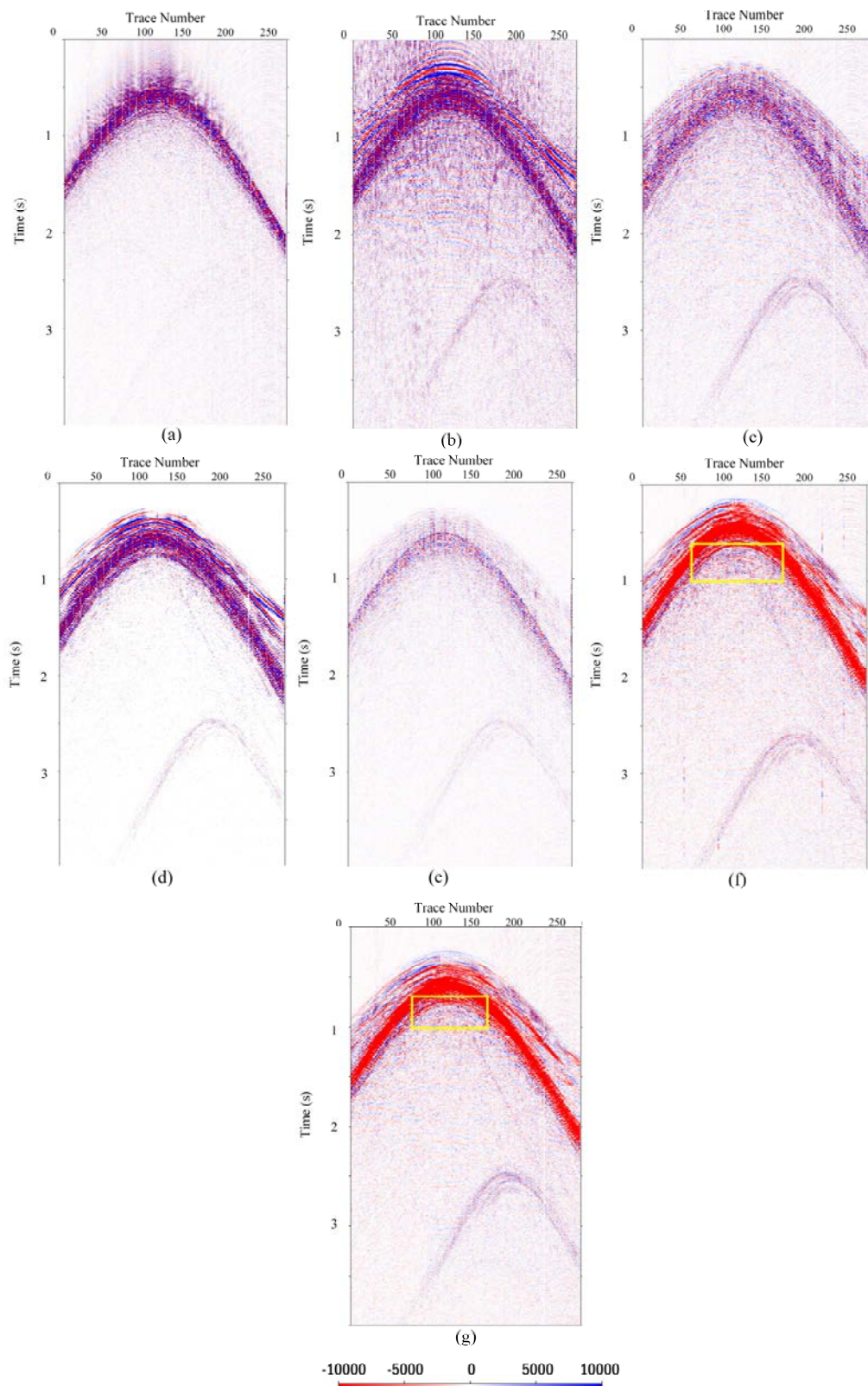


图 3.13 不同方法对实际 OBC 数据 P 分量去噪后差值的对比 (a) 带通滤波, (b) F-X 反褶积, (c) 小波变换滤波, (d) RPCA, (e) WNNM, (f) DnCNN, (g) DnDenseNet

图 3.11 显示了带噪声的 OBC 数据, 包括 P 分量数据 (即水听器数据)、X 分

量、Y 分量和 Z 分量。本研究中将图 3.11 (a) 中显示的现场有噪声的现场 OBC 记录的 P 分量数据作为要处理的数据。数据采集过程中检波器的空间采样间隔为 25 m, 采样频率为 5000 Hz, 数据的时间长度为 4 s。横坐标为地震数据记录道, 纵坐标为时间。需要说明的是, 在图 3.11 中, X、Y 和 Z 分量仅用于提供 OBC 四分量数据的完整图。这三个分量没有用于数据处理, 本研究中重点关注 P 分量数据的去噪和信号恢复, 以测试不同方法的去噪能力。可以观察到, 图 3.11 (a) 所示的 OBC 数据 P 分量数据在中层和深层位置上质量较差。影响数据质量的主要因素包括噪声、面波和外源干扰。在中深层, 有效信号几乎被噪声覆盖, 无法有效识别, 因此需要对噪声进行衰减, 处理上主要是去除噪声和面波影响。

应用不同的方法对数据进行处理后, 图 3.12 展示了不同方法的去噪效果。可以看出, 相比于传统的去噪方法, 本章提出的 DnDenseNet 方法对实际记录中信号的恢复和对噪声的压制更加有效。在浅层区域如图 3.12 (g) 中蓝色框标注的区域所示, 原本受噪声严重影响的同时相轴在去噪后变得更清晰、更连续。在中深层区域如图 3.12 (g) 中红色区域所示, 能量微弱的反射信号, 在噪声和面波的干扰下几乎无法被观察。在经过本方法处理后, 其明显得到恢复。相比之下, 传统方法在这两个区域的去噪效果都不尽如人意。同时, 对于外源干扰的处理, 从图 3.12 (f) 和图 3.12 (g) 中绿框标记的区域来看, 本章提出的 DnDenseNet 方法在抑制外源干扰噪声方面优于 DnCNN。DnCNN 不能很好地抑制外源干扰, 去噪后的数据中留下了大量的外源干扰的残留。

同时, 图 3.13 显示了各种方法的差值图, 从中可以看出各种方法去除的噪声。从差值图中, 可以看到, DnDenseNet 的差值图中基本没有信号泄露, 而传统方法中明显有着较多的信号泄露。与 DnCNN 相比, DnDenseNet 方法的信号泄露明显更少, 这表明 DnDenseNet 方法有着更好的保留有效信号的能力, 这可以在图 3.13 (f) 和 3.13 (g) 中黄色方框标记的区域中观察到, 这验证了所提出的方法具有更好的信号保留能力, 也进一步证实了所提出的方法在处理实际地震数据时具有更好的性能。

3.4 本章小结

地震记录中的噪声水平在不同区域可能存在显著差异, 某些区域可能噪声较

大,而其他区域信号相对明显且噪声较小。这种不均匀的分布要求去噪模型具有高度的适应性和灵活性。此外,地震记录信号的强度变化范围广泛,这要求去噪模型能够处理从极弱到极强的信号,确保不丢失细节。常规 CNN 是线性堆叠结构,网络的不同层之间缺乏直接的信息流动路径,限制了模型在深层网络中对于浅层特征的复用能力。针对这一问题,本章基于 DenseNet 架构提出一种密集连接的去噪模型 DnDenseNet。在网络中,密集连接块被用作网络的基本构建块,用于实现微观的密集连接,此外,密集连接块之间采用宏观的密集连接,从而在微观和宏观两个层面上促进特征的充分利用。密集的连接加强特征传播,鼓励特征复用,通过融合低维特征和高维特征,使得不同级别的特征融合后为被地震信号的重建提供丰富的信息,从而实现更优的信号重建质量。同时,密集连接缓解了梯度消失问题,使得网络更易训练。在实验阶段,本研究构建了不同信噪比的测试数据集用于模型去噪性能定量测试,使用多个评价指标来测试和评估模型性能。结果表明,DnDenseNet 模型在模拟数据去噪测试中获得了最好的信噪比提升,同时在多个指标上要优于其它传统对比方法和常规无跳跃连接的卷积神经网络。在实际地震数据去噪测试中,选择 OBC 记录中的 P 分量数据进行处理测试,DnDenseNet 模型能够在该类实际记录处理中对噪声进行有效压制,从中恢复有效信号,其去噪性能高于其他对比方法,验证了本文方法在实际数据地震去噪任务中的出色性能。

第4章 基于自监督学习预训练 Transformer 的地震数据去噪算法

4.1 引言

第3章基于卷积神经网络中的 DenseNet 架构提出了去噪密集连接卷积神经网络去噪模型,即 DnDenseNet,在处理模拟和实际地震数据的去噪问题上表现出色,但该模型在捕捉长期依赖关系的能力上还存在不足。一般来说,卷积神经网络(Convolutional Neural Network, CNN)模型通过使用滑动卷积核捕捉不同位置的信息,将这些信息在空间层面上结合,从而构建出具有平移不变性和分层特性的特征表达。然而,卷积操作的这种局部性质意味着,为了提取更远距离的信息,需要通过增加卷积层来扩展其感知范围。DnDenseNet 引入的密集连接虽然能够使 CNN 模型加深层数,但是对于更远范围的地震同相轴信号,这种方法在分析含有长距离信号特征的复杂地震数据时显示出其局限性。此外,模型训练中使用的合成信号和噪声相结合的方式,在训练集和实际数据之间产生了分布偏差,这影响了模型在实际应用中的去噪效果。本章节主要聚焦解决这个问题。

近年来,随着图形处理单元(Graphics Processing Unit, GPU)的发展及并行计算的推广,深度学习得以快速发展并在不同领域的数据处理中获得了广泛应用。在图形领域的广泛应用包括图像识别^[100]、图像超分辨率^[101],图像去噪^[56],图像分割等^{[100]-[102]}。发展出 VGG^[49]、U-Net^[50]、ResNet^[51]、DenseNet^[52]等一系列结构和生成对抗网络(Generative Adversarial Network, GAN)^[53]、无监督^[127]、自监督^[128]等一系列训练框架。这些结构和框架的主干网络基本都是选择的 CNN 结构。因为 CNN 由于其卓越的特性,如权重共享和局部感知,可以提供良好的处理结果。在图像领域普遍的成功应用也激励了深度学习在地震数据处理领域的应用^{[103]-[117]}。在去噪领域,已经有了很多丰富的方法被应用,如前馈去噪卷积神经网络(Feedforward Denoising Convolutional Neural Network, DnCNN)^[57]、U-Net^[58]和生成对抗性网络(Generative Adversarial Network, GAN)^[59],被应用于复杂地

震噪声衰减。基于 CNN 的去噪网络可以在训练过程中对参数进行优化,以提取含噪地震数据和期望信息(即纯净信号)之间的隐藏模式和多维映射关系,这种研究思路取得了较好的成功实践和应用结果,但在捕获长程依赖性方面存在局限性。具体来说, CNN 主要依靠滑动卷积核提取不同位置的特征,并将这些特征在空间上进行局部的加权和融合,从而得到具有位置不变性和层次结构的特征表示。卷积算子是局部算子,为了提取远程特征, CNN 需要堆叠卷积核形成更大的感受野。这种运算模式限制了其从复杂地震数据中捕捉远程特征的能力,而地震记录中信号尤其是同相轴信号往往是连续的、远程的^{[118]-[125]}。另外,为了有效地重建信号,对不同噪声模式的准确感知起着关键作用。由于卷积运算的固有性质,基于 CNN 的去噪方法的一个潜在限制是仅依靠在输入数据的局部视图中学到的特征来识别噪声模式(如几类噪声混合的模式)。然而,在执行整个地震记录的信号恢复和噪声压制时,获得整个数据的全局感知是至关重要的。在全局视图中,空间相干性和合成模式(例如,噪声与信号的相互关系)往往更容易学习。

作为 CNN 的替代方案,近年来提出的 Transformer 结构设计了一种自注意力机制来捕捉上下文内容之间的全局交互,这种结构被认为在一定程度上能够克服 CNN 的一系列缺点,被视作新一代的神经网络的架构。目前,Transformer 结构已经发展出了诸如 Swin Transformer^[126]等多种变体,并在计算机视觉领域引发了较多的关注。本研究中提出了一种用于地震信号处理的 Transformer 模型,以进一步发挥深度学习技术在地震数据去噪任务中的潜力。为叙述方便,提出的模型称为 SDT (Seimisc Denoising Transformer)。SDT 由三个模块组成:浅层特征提取、远程特征提取和数据重建模块。浅层特征提取模块利用卷积层提取浅层特征。远程特征提取模块主要由改进的 Swin-Transformer 块^[126]组成,其中结合了通道注意力和自注意力机制,同时跨窗口交互提取特征。受益于这几项设计,使得远程特征提取模块能够提取更多的远程特征来用于地震信号的重建。最后,在重建模块中实现高质量的地震数据重建。由于 Transformer 相比于 CNN 不具有归纳偏置,因此大规模的数据预训练非常重要,本章节设计了一种数据的自监督预训练方案^[128],仅使用大量的实际数据而无需配对数据就可以来对模型进行预训练。接下来,通过正演模拟纯净信号,并与现场地震噪声数据相结合,生成一

个高质量的含噪信号-纯净信号训练数据集，对模型进行了有监督微调训练，在微调过程中，网络具备更强泛化性的去噪能力。实验结果表明 SDT 具有衰减复杂地震噪声和保持弱信号幅度的优点，与传统方法和最近提出的先进的去噪网络相比具有更好的去噪效果。实验中，本章提出的方法在处理采集自沙漠地带的地震记录时取得了较好的效果。另外，在通用性研究时，网络在处理分布式光纤传感技术垂直地震剖面测量（Distributed Acoustic Sensing - Vertical Seismic Profile, DAS-VSP）记录^[120]时也保持了较好的去噪性能。这两者分别可作为复杂地形环境采集的地面记录和新型仪器记录的代表，验证了该算法的有效性。

4.2 SDT 基本理论

Transformer 已经在计算机视觉界引起了极大关注，并被开发用于视觉任务中，包括图像分类^[49]、目标检测^[100]、图像分割^{[129]-[131]}等任务。然而，Transformer 结构在用于图像恢复任务时通常将输入图像划分为具有固定大小（例如， 48×48 ）的切片，并独立处理每个切片，这制约了远程特征的提取，且全局的注意力运算带来巨大的计算负担。Swin Transformer 结构融合了 CNN 和 Transformer 的优势。一方面，通过局部注意力机制，节省计算量，使得其具有与 CNN 结构一样处理大尺寸图像的优势。另一方面，使用移位窗口方案可以对远程特征进行建模。Transformer 在建模远程特征依赖性方面显示出其优势，但也有研究工作表明，卷积可以帮助 Transformer 实现更好的视觉表示^{[132],[133]}。此外，由于地震记录的庞大，因此如何在大的地震记录上充分提取远程特征也是一个重要问题。因此，现阶段 Transformer 结构在地震记录处理任务上的研究应主要围绕以下两点进行：

（1）如何将 Transformer 结构与 CNN 结构进行更好的聚合；（2）如何在降低自注意力机制复杂度的同时更好地计算远程特征依赖。本章中介绍的 SDT 方法就是一种依据以上思路，并基于 Transformer 结构提出的地震数据去噪架构。

本节中对 SDT 方法进行了全面描述，涵盖了几小节，包括模型的网络结构、结合自监督预训练和有监督训练微调的训练方法，以及数据集构建，从而对 SDT 方法的实施细节进行全面的介绍。

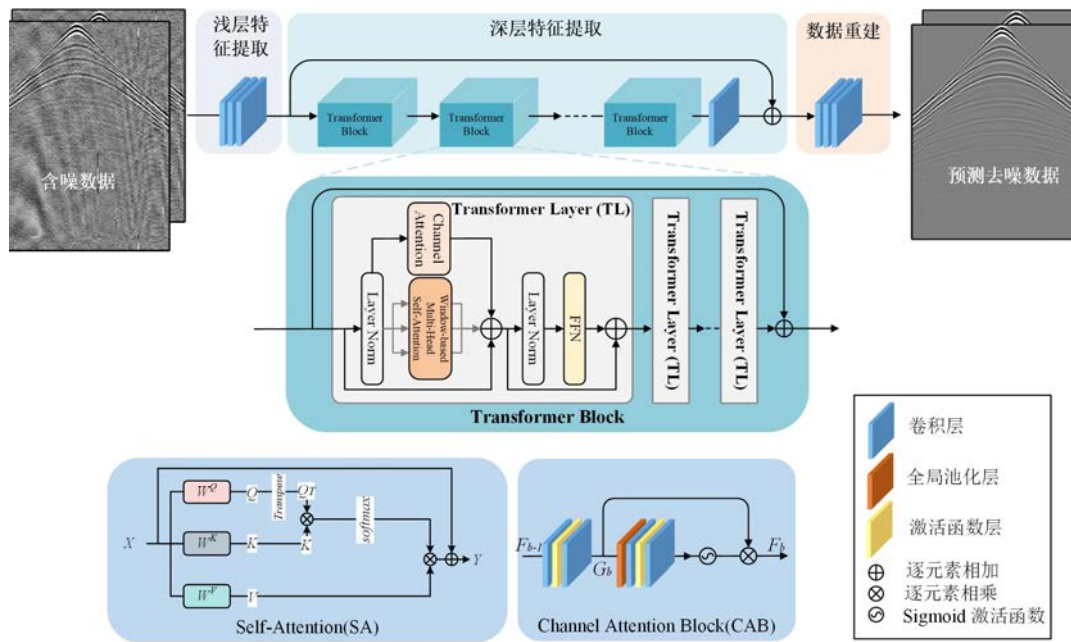


图 4.1 地震数据去噪 Transformer 方法（SDT）的网络结构图

4.2.1 SDT 网络结构

对于去噪神经网络，能在小图块上训练，并且直接在大图上推理的属性被称之为“全卷积”属性。全卷积属性对于地震去噪任务来说至关重要，因为去噪任务需要对输入数据进行采样点级的操作，输出地震数据需和输入地震数据尺寸一样，且尺寸往往无固定大小限制。一般来说，Transformer 结构因为缺乏“全卷积”性质，而难以直接应用于地震记录去噪中。Swin-Transformer 结构^[126]的提出改进了这一点，通过引入层次化的结构，譬如提出了局部自注意力机制来解决计算复杂度问题。这使得 Swin-Transformer 具有“全卷积”属性，其根本原因是自注意力机制被限定作用在一个小的窗口内，而在整个空间内的不同窗口都是共享权重的。如果在更大的尺寸下推理，只需要更多的窗口即可。作为其代价，Swin-Transformer 结构也损失了一部分捕获远程特征的能力。本研究中提出的 SDT 方法就是基于 Swin-Transformer 结构进行设计的，并增加了一些设计着力于增强其捕获远程特征的能力。如图所示，整个网络由三部分组成，如图 4.1 所示，包括浅层特征提取模块、Transformer 远程特征提取模块和地震信号重建模块。对于给定的含噪的输入，首先利用一个卷积层来提取浅层特征。然后，使用一系列

Transformer 远程特征提取模块进行远程特征提取, 具体来说, 在每个 Transformer 远程特征提取模块中通过多头自注意力及通道自注意力来实现这一过程。最后通过重建模块来重建去噪后的结果, 重建模块由卷积层构成。网络中各层的定义如表 4.1 所示。

表 4.1 SDT 网络中各层的具体说明

层	描述	功能
Conv	卷积层	特征提取
Global pooling	全局池化层	一种降维操作, 用于将卷积层或其他层的输出降低为单个值
GELU ^[135]	GELU 激活函数	将缩放双曲正切 (tanh) 函数应用于输入值, 并对其进行缩放和移位, 以确保输入数据以 0 为中心, 且靠近 0。
Sigmoid	Sigmoid 激活函数	将输入值映射到 0 到 1 之间的范围, 它常用于二分类问题中, 用于输出模型的概率估计
LN / Layer Norm ^{[136],[137]}	层归一化	通过标准化每个层的输入, 使其均值接近 0 并具有单位方差, 有助于加速训练过程和提高模型的泛化能力。
Softmax	Softmax 激活函数	将输入值转化为表示每个类别概率的输出。Softmax 函数会使所有类别的概率总和为 1, 以便进行分类决策

这一小节中主要详细介绍各个 Transformer 模块中的结构。如图 4.1 所示, 通道注意力块 (Channel Attention Block, CAB)^[134]与基于窗口的多头自注意力 (Windows Multi-Head Self-Attention, W-MSA) 模块并行地插入到第一个层归一化层 (LayerNorm, LN) 层^{[136],[137]}之后的标准 Swin Transformer 块中。在上述的结构中, W-MSA 模块是所提出的特征提取结构的核心部分。对于自注意力的计

算, 给定大小为 $H \times W \times C$ 的输入特征, 首先将其划分为 $\frac{H \times W}{M^2}$ 个大小为 $M \times M \times C$ 的特征块, 这些特征块被称为“键”, 然后在每个键之间计算自注意力。这里, H 、 W 和 C 分别表示特征的高度、特征的宽度和通道的数量, M 表示单个键的大小。在 SDT 中, 将数据中的每个采样点视为一个编码的键, 即 M 设置为 1。对于给定的输入特征, W-MSA 的计算过程如下:

$$X = \{X^1, X^2, \dots, X^N\}, N = \frac{H \times W}{M^2} \dots\dots\dots (4.1)$$

$$Y_h^i = \text{Attention}(X^i W_h^Q, X^i W_h^K, X^i W_h^V), i=1, \dots, N \dots\dots\dots (4.2)$$

$$\hat{Y}_h = \{Y_h^1, Y_h^2, \dots, Y_h^M\} \dots\dots\dots (4.3)$$

在 Transformer 结构中, 往往执行多头注意力计算, 其中, 自注意力头是执行每个自注意力计算的具体结构, 因此上述公式中 $\hat{Y}_h$ 是第 h 个头的输出, W_h^Q, W_h^K, W_h^V 分别表示第 h 个头的“查询向量”、“关键向量”和“数值向量”的映射矩阵。自注意力机制的输出可以被视为各个键之间的相关系数。具体而言, 每个采样点都被视为嵌入的键 X^i 。对于特征 $X^N \in \mathbb{R}^{M \times M \times C}$, 利用“查询向量”、“关键向量”和“数值向量”三个映射矩阵, 将特征线性映射为对应的三个向量 Q 、 K 和 V 。注意力计算可以用以下形式表示:

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d_h}} + B\right)V \dots\dots\dots (4.4)$$

其中 d 表示键的维度, B 表示相对位置编码^[138]。在 Transformer 结构中, Q 、 K 和 V 是自注意力机制中使用的三个输入表示向量。它们是通过线性变换从原始输入向量 (通常是编码表示) 得到的。

Q 、 K 、 V 矩阵在自注意力机制中的作用如下: Q 矩阵作为查询向量, 用于在自注意力机制中查询特定位置的信息, 可以看作是要关注的信息, 即需要从输入数据中提取的特征。 K 矩阵作为关键向量, 表示各个位置的特征重要性, 可以看作是输入数据中的特征的权重, 用于确定每个位置对于输出的影响程度。 V 矩阵

作为数值向量,代表与查询向量和关键向量匹配的结果,可以看作是每个位置的数值信息,用于计算最终的输出结果。这三个矩阵通过自注意力机制的计算,可以生成一个权重分布,用于加权求和得到关联的数值向量。该过程可被视作一种特征选择的过程^[60]。

为了最大化所使用的采样点的范围,本研究中利用大窗口进行自注意力计算。在本研究中窗口大小设置为 16。此外,为了在相邻的非重叠窗口之间建立连接,并尽可能多地捕获远程依赖关系,采用了移位窗口方法^[126]。在连续的 Transformer 块中,以固定的间隔长度使用基于移位窗口进行自注意力计算(Shifted Windows Multi-Head Self-Attention, SW-MSA),将移位大小配置为窗口大小的一半,从而确保更宽范围的特征提取。

此外,通道注意力部分也是网络的重要组成部分。通过将基于通道注意力的卷积块合并进入标准的 Transformer 块中,网络的表示能力得以增强。当采用通道注意力时,可以利用更多的全局信息来计算通道注意力权重。使用通道注意力模块^[134]即 CAB 模块来自适应地重新缩放通道特征。CAB 模块被设计为捕获通道间的依赖性,并突出由前面的卷积层生成的特征图中信息量最大的通道。通过这样,网络能够关注重要特征并抑制不太相关的特征。CAB 模块如图 4.1 中右下角所示,CAB 模块将来自前面的卷积层的输出特征图作为输入,这些特征图具有多个通道,其中每个通道都包含特定的特征信息。为了捕获全局上下文内容并计算基于通道的注意力,CAB 模块在每个通道的空间维度上执行全局平均池化,此操作对每个通道上的特征值取平均值,得到长度等于通道数量的一维张量(向量)。然后,将 Sigmoid 函数应用于输出的张量上,以确保每个通道的注意力权重落在合理的范围内(通常在 0 和 1 之间),然后使用从 Sigmoid 激活函数获得的注意力权重来重新缩放原始特征,具有较高注意力权重的通道将被增强,而具有较低注意力权重的通道将被抑制。此外,将一个平衡系数乘以 CAB 的输出,用于平衡 CAB 的输出所占据的权重。

在上述过程之后,将自注意力和通道注意力计算后的特征输出发送到前馈网络(Feed Forward Neural Network, FFN)进行继续处理,特征在 FFN 中的计算可以表示为

$$FFN(X) = MLP(LN(X)) \dots\dots\dots (4.5)$$

其中 MLP 表示多层感知器。

综上所有过程，对于给定的输入特征 X ，整个过程计算可表示如下：

$$X' = LN(X) \dots\dots\dots (4.6)$$

$$X'' = X' + SWMSA(X') + \alpha CA(X') \dots\dots\dots (4.7)$$

$$Y = FFN(X'') \dots\dots\dots (4.8)$$

其中， X' 和 X'' 表示中间特征， Y 表示FFN的输出，也是整个Transformer特征提取模块的输出。在本研究中，平衡系数 α 具体被设置为0.15。

4.2.2 自监督预训练及微调算法

SDT 的输入是有噪声的地震数据，表示为

$$y = x + n \dots\dots\dots (4.9)$$

其中， x 表示纯净信号， n 表示地震噪声， y 表示含噪信号。SDT通过训练将纯净信号 x 与地震噪声 n 分离，以在含噪信号 y 和纯净信号 x 之间建立非线性映射关系。在此基础上，通过将含噪信号 y 送入网络，通过网络进行推理预测，可以获得估计的信号为

$$\tilde{x} = SDT_{Trained}(y, \Theta) \dots\dots\dots (4.10)$$

其中， Θ 表示训练参数，这些参数由权重矩阵和偏置组成。由于Transformer模型不像是CNN模型那样具有归纳偏置，因此大规模的数据预训练对于释放此类模型的潜力至关重要。因此本研究中设计了一种可以充分使用实际含噪记录的自监督预训练方法。预训练过程中，借助自定义的前置任务，网络可以对输入的数据进行掩码破坏后再尝试修复数据，在这一过程中，网络参数被初步地优化。同时，这一阶段不需要带标签的数据，掩码后的数据与原始数据自然而然地构成一对训练网络的数据组合。预训练阶段完成后，模型已经通过掩码自修复过程学习到了地震数据的基本特征和结构，这为模型提供了一个良好的初始状态。在此之后，再通过有监督训练对其参数进行微调。具体将分为以下两个步骤：

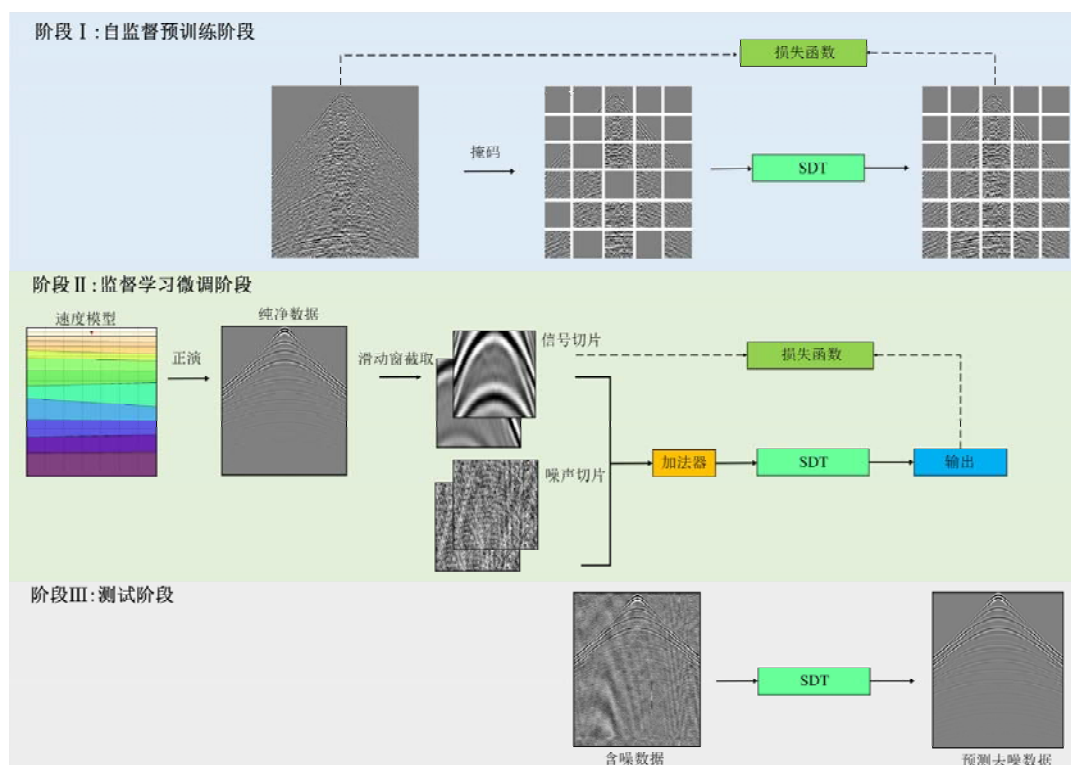


图 4.2 自监督预训练 Transformer 模型训练及测试流程

(1) 自监督预训练

首先需对 SDT 进行预训练。自监督表示学习的目标是对地震记录的特征进行编码。本研究中为地震数据表示学习开发了一个代理任务，如图 4.2 的第一阶段所示，在预训练过程中，输入的训练数据由现场收集的真实地震记录中随机裁剪的区域组成。通过使用一种称为掩码增强的技术，在局部窗口内随机隐藏特定的区域。掩码增强的过程是在子体积内随机选择特定的区域，并将其掩码屏蔽。具体来说，使用正态分布生成的值填充选定区域，从而有效地隐藏或遮挡选定区域，同时引入一定程度的随机性。使用掩码增强技术是为了鼓励模型从部分可见的数据中学习，并建立基于不完整信息进行预测或重建的能力。体积比率决定了将被隐藏的子体积的比例或范围，在本研究中将体积比率设置为 0.2，也就是说，大约 20% 的子体积将被随机屏蔽。这种掩码策略可以将掩码数据视为缺失值，从而将自监督预训练过程近似定义为神经网络用于缺失内容重建的训练过程。通过采用这种策略，模型能够在预训练过程中学习一般的地震数据特征。此外，模型的预训练不依赖于干净的数据，允许在预训练过程中使用大量的含噪实际数据。

对于任何输入, 部分区域被屏蔽后, 所得到的具有隐藏区域的子体积被用作 SDT 网络的输入。重建的目标是最小化原始数据和重建数据之间的差异。这是通过测量两组数据之间的 L1 损失来实现的。重建目标定义为

$$L_1 = \frac{1}{2N} \sum_{i=1}^N |SDT_{Pretrained}(y_i^M; \Theta) - y_i| \quad \dots\dots\dots (4.11)$$

其中 y 表示原始数据, y^M 表示屏蔽数据。 $SDT_{Pretrained}$ 表示预训练阶段的模型。在地震数据预训练任务中选择 L1 损失主要是因为它对异常值的鲁棒性, 而异常值是地震数据中的一个常见特征。与 L2 损失不同, L2 损失会过分强调大误差并导致结果过度平滑, L1 损失通过关注数据分布的中值来保留地震信号中的基本特征。这一方面对于保持地震数据中同相轴信号的突变和不同模式至关重要。形式上, 通过用掩码训练 SDT 来最小化损失函数。由于这种预训练是自监督的, 只需要包含有信号的真实含噪地震记录即可用来驱动预训练过程。这允许在没有配对的含噪信号-干净信号的情况下利用大量含噪的数据来预训练网络。

(2) 有监督学习的微调训练

在随后的微调阶段, 如图 4.2 中的第二阶段所示, 使用公式 (4.12) 所示的损失函数对网络进行微调。

$$L = \frac{1}{2N} \sum_{i=1}^N \|SDT_{Finetuning}(y_i; \Theta) - x_i\|_F^2 \quad \dots\dots\dots (4.12)$$

其中 x_i 和 y_i 分别表示训练数据集中配对的干净数据和噪声数据块。 $SDT_{Finetuning}$ 表示微调阶段的模型。此外, $\|\cdot\|_F$ 表示 Frobenius 范数, 该范数因其能够全面测量所有矩阵元素的误差而备受青睐, 使其成为深度学习算法中正则化的常见选择。在此基础上, 通过最小化损失函数来获得网络的最优参数^{[139],[140]}, 这一过程中, 不仅能够精确地衡量干净数据块和去噪数据块之间的差异, 而且还能够确保模型在整个训练过程中向着减少重建误差的方向优化。使用梯度下降法在反向传播的过程中逐层调整权重, 直到损失函数下降到一个极小的值或迭代次数到达规定值。此时, 模型可被认为已经优化至期望的状态。

算法 4.1 自监督预训练 Transformer 网络去噪系统**输入:**

批处理大小 N ; 迭代周期数 E ; 一个迭代周期内的迭代次数 I ; 幅度归一化后的预训练含噪数据声集 y ; 幅度归一化后的噪声集 n ; 幅度归一化后的信号集 x ; 网络模型想要学习到的映射关系 SDT ; 测试样本 $\hat{y}$ 。

预训练阶段:

```

1: for m=1,2,3,..., E do
2:   for n=1,2,3, ..., I do
3:     从归一化后的含噪数据集中取第  $i$  组数据  $\{y^{(i)}\}_{i=1}^N \in y$ 
4:      $\theta \leftarrow \nabla_{\theta} \left[ \frac{1}{2N} \sum_{i=1}^N \|SDT_{Pretrained}(y_i^M; \Theta) - y_i\|_1 \right]$ 
5:   end for
end for

```

训练微调阶段:

```

1: for m=1,2,3,..., E do
2:   for n=1,2,3, ..., I do
3:     从归一化后的信号集中取第  $i$  组信号数据  $\{x^{(i)}\}_{i=1}^N \in x$ 
4:     从归一化后的噪声集中取第  $i$  组噪声数据  $\{n^{(i)}\}_{i=1}^N \in n$ 
5:     叠加获得不同信噪比的含噪数据  $\{noisy^{(i)}\}_{i=1}^N = \{x^{(i)}\}_{i=1}^N + \{n^{(i)}\}_{i=1}^N$ 
6:     Input  $\{noisy_{nr}^{(i)}\}_{i=1}^N = \frac{\{noisy^{(i)}\}_{i=1}^N}{\{max(noisy^{(i)})\}_{i=1}^N}$ 
7:     Label  $\{x_{nr}^{(i)}\}_{i=1}^N = \frac{\{x^{(i)}\}_{i=1}^N}{\{max(noisy^{(i)})\}_{i=1}^N}$ 
8:      $\theta \leftarrow \nabla_{\theta} \left[ \frac{1}{2N} \sum_{i=1}^N \|SDT_{Finetuning}(input; \Theta) - label\|_F^2 \right]$ 
9:   end for
10: end for

```

测试阶段:

```

1: for 测试样本  $\hat{y}$  do
2:   用训练微调阶段生成的权重和偏置获得参数矩阵  $\Theta$ 
3:   计算输出结果  $\tilde{s} = SDT(\hat{y}; \Theta)$ 
4: end for

```

输出: $\tilde{s}$

接下来, 通过训练的 SDT 来预测最终恢复的信号, 整个预训练、微调训练

和测试过程如图 4.2 所示。对于 SDT 模型的超参数设置，采用了表 4.2 所示的配置，包括预训练阶段和微调训练阶段都在此配置下进行。

表 4.2 网络超参数设置

超参数	含义	具体设置
批次大小	在训练神经网络时一次输入的样本数量	32
轮次	神经网络在整个训练数据集上反复迭代的次数	500
学习率范围	控制神经网络权重更新幅度的参数	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
训练切片大小	用于训练时的数据集的尺寸大小	128×128
优化器	用于调整模型权重以最小化损失函数的算法	Adam 算法
自注意力头数	多头自注意力机制中的注意力头数	16
窗口大小	执行窗口自注意力的窗口大小	16×16
“键”大小	用于形成进行自注意力计算的每个“键”覆盖的范围	1×1

4.2.3 数据集构建

深度学习算法利用特征学习来克服传统方法需要手动设计特征提取算子的问题。一般来说，基于深度学习的去噪方法的去噪质量在很大程度上取决于训练数据集。训练去噪模型的数据集往往由两部分组成：有噪信号及其对应的纯净信号。不幸的是，由于信号和噪声在时频域中的混叠，干净的信号无法从实际地震记录中分离出来，这使得在地震勘探中，含有噪声的数据很容易获得，但匹配的纯净地震信号则很难获得。一般来说，有监督学习算法在这类场景下必然会面临跨数据分布的泛化性能挑战。在本研究中，采用自监督预训练加有监督微调的训

练模式, 以让模型尽可能多地具备在实际数据中保持性能的泛化能力。本章提出的 SDT 是一种基于自注意力机制的模型, 它对数据中的每个位置都通过自注意力机制进行远程的信息交互, 更加关注远程的上下文关系。这种大范围的关注和信息交互使得 Transformer 对于数据中的细节和结构都更为敏感, 需要更多的数据来进行训练以捕捉到这些细微的模式和关系。Transformer 往往需要比 CNN 更多的数据驱动训练。CNN 在设计上具有一些天然的先验假设, 例如平移不变性和局部关联性。这些假设使得 CNN 在训练过程中可以更有效地利用有限的数。而 Transformer 在图像和信号处理领域中的应用相对较新, 对于其设计和训练过程的最佳实践还在不断探索和发展中。因此, 为了充分发挥 Transformer 的优势并取得良好的性能, 通常需要更大规模的数据集来提供足够的样本和多样性, 从而让模型能够更好地学习到数据中的复杂模式和结构。因此, 无论在自监督预训练阶段, 还是有监督微调阶段, 都需要构造尽可能多的训练数据。

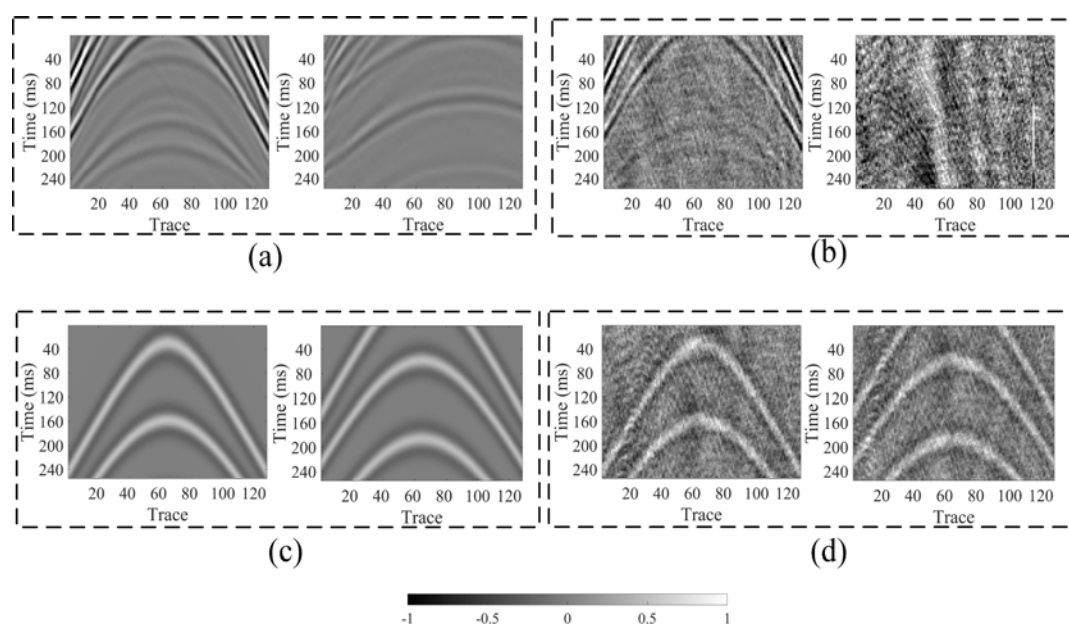


图 4.3 用于训练的数据集样本 (a) 正向建模信号切片, (b) 正向建模含噪信号切片, (c) 模拟生成信号切片, (d) 模拟生成含噪信号切片

在预训练阶段, 使用现场地震记录进行预训练, 这些记录是从我国不同的调查区域获得的, 采样率为 500 Hz。它们被裁剪成 128×128 大小的切片进行训练。在微调阶段, 通过正演建模生成理论上干净的信号, 并将它们与在无源条件下收集的噪声混合, 以构建噪声记录, 从而形成了用于微调网络的噪声干净训练对。

为了获得模拟的干净数据，本研究中构建了 100 个不同地质结构的速度模型。放置人工震源以生成地震波，并沿水平方向放置 300 个接收器以记录地震波场，从而得到观测地震记录。在这项研究中，合成数据是基于时域中的声波方程建模的，如下方程所示：

$$\frac{\partial^2 u(x, y, t)}{\partial t^2} = v(x, y)^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + s(x, y, t) \dots\dots\dots (4.13)$$

其中 v 表示波速， u 表示声波场。 $s(x, y, t)$ 表示震源， (x, y) 表示空间位置， t 表示时间。通过数值模拟生成对应于每个速度模型的纯净地震数据。通过求解声波方程，总共生成了 100 个模拟的纯净地震数据。表 4.3 提供了关于正演建模的更多细节。

表 4.3 正演参数设置

参数	含义	具体设置
震源子波类型	地震信号的小波函数种类	雷克子波
子波频率	地震信号的频率范围	15-30 Hz
波速范围	地震波在介质中传播速度范围	1200-4000 m/s
密度范围	介质中密度设置的范围	1300-4000 kg/m ³
采样时间间隔	地震数据记录中连续采样数据的时间间隔	2×10 ⁻³ s
接收器数量	地震数据采集使用的接收器的数量	300
最大传播时间	地震波从震源到接收器的最长传播时间	3 s

此外，本研究中使用雷克子波模拟了一组相似的波形信号，以增强类似同相轴信号的特征，这些信号如图 4.3(c) 所示。这些信号也被合并到信号数据集中。通过裁剪数据，并以切片的形式训练网络。本研究中将振幅归一化地震记录划分为大小为 128×128 的块，以获得总共 20576 个干净的信号块。该数据集虽然大小适中，但由于其多样性，足以用于监督学习去噪任务。

选择在无源条件下收集的噪声以形成噪声数据集。使用大小为 128×128 、步长为 16 的滑动窗口对收集的噪声数据进行裁剪，以生成噪声训练集。通过随机选择一组信号切片和噪声切片并相加从而构建一个有噪声的数据。在监督学习下，构建了一个有噪声的输入数据集及其相应的纯净数据集即标签集，以训练去噪模型。在图 4.3 中，展示了一些干净信号切片和含噪声信号切片的例子。其中图 4.3 (a) 和 (b) 分别为通过正演获得的纯净记录和含噪记录，图 4.3 (c) 和 (d) 分别为利用雷克子波构建得到的纯净记录和含噪记录。

4.3 算法性能验证

本节通过广泛实验验证所提 SDT 算法在地震数据去噪任务上的有效性和优越性，具体内容包括构建的测试数据集，相关实验设置，实验结果和对结果的分析，与本领域常规去噪方法及先进去噪方法的性能对比。其中，本章方法在具有如下参数的服务器运行：CPU Intel_6330_Xeon, 2.0GHz, RAM 64GB, 8 个 Tesla A10 GPU。该算法在服务器的 Python 环境下基于 Pytorch^[141]深度学习框架部署实施。

4.3.1 实验数据集及实验设置

为了验证 SDT 模型的去噪能力，本研究使用了图 4.4(a)所示的 Marmousi 地层模型，该地层模型包含具有许多不同速度和形态的地层。图 4.4 (a) 中的红色倒三角形表示人工震源，黑线表示接收器的位置。利用基频为 20 Hz 的雷克子波来作为震源进行正演，并获得图 4.4 (b) 中干净的合成记录。

在此基础上，通过叠加采集自我国塔里木盆地沙漠地区的背景噪声，并通过放大噪声的幅值来获得不同信噪比的记录。如图 4.4 (d) 所示为一个信噪比为 -5dB 的噪声记录。从中可以看出，有效信号受到干扰的严重影响，尤其是一些深层的信号在标记区域难以被观察到。需要说明的是，该模型及其对应的正演数据未包含在训练数据集中。本研究中，使用 SDT 与其他传统方法，包括带通滤波 (Band-Pass Filter, BPF)，小波变换 (Wavelet Transform Filter, WTF)^[142]，变分模态分解 (Variational Mode Decomposition, VMD)^[21]，加权核范数最小化

(Weighted Nuclear Norm Minimization, WNNM)^[143]和基于深度学习的网络,如 DnCNN^[57], U-Net^[58]和 MIRNet^[144]等网络模型进行比较,以分析去噪的有效性和准确性。

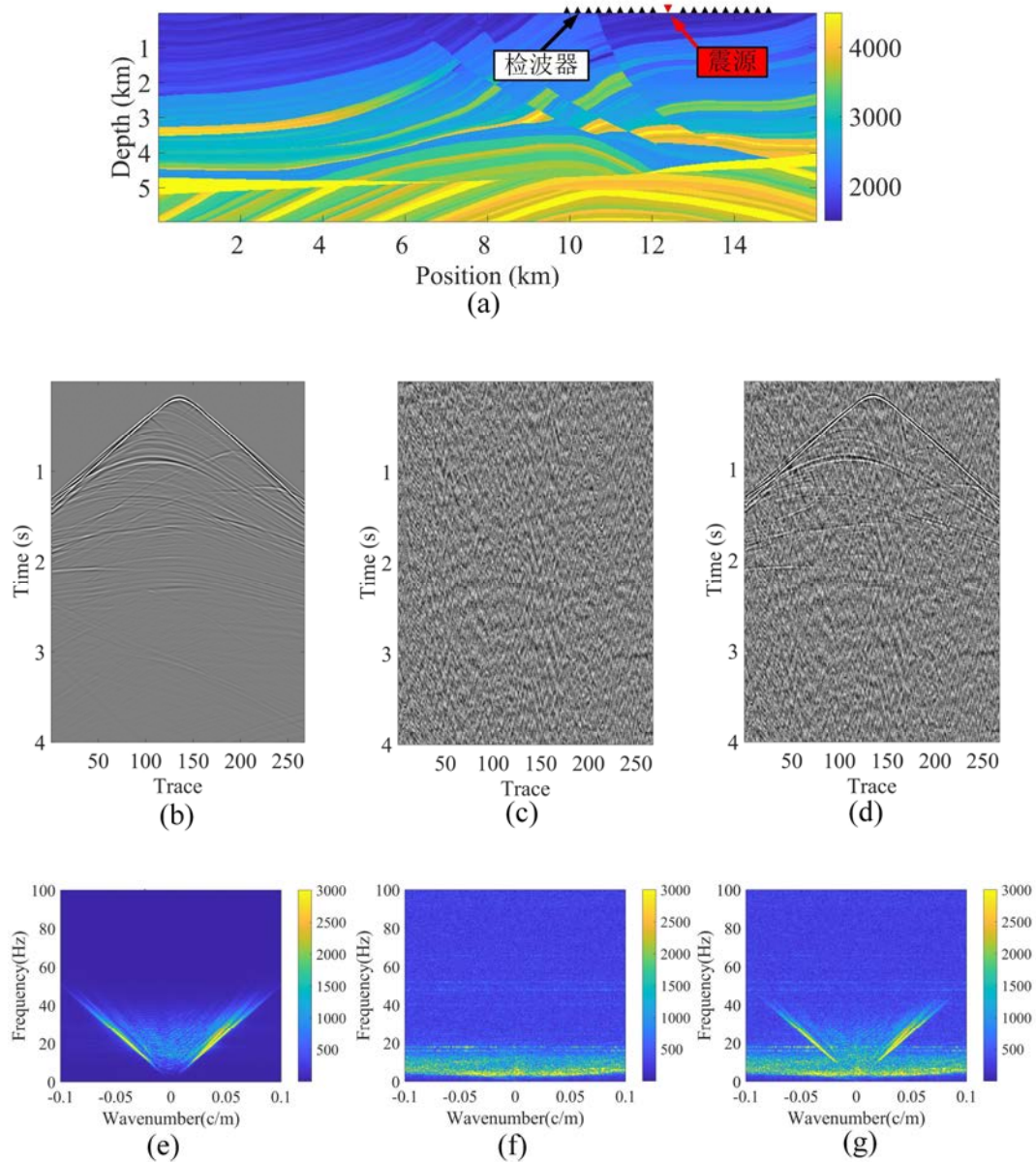


图 4.4 模拟含噪地震数据的构建过程 (a) Marmousi 速度模型, (b) 纯净信号, (c) 噪声, (d) 含噪地震信号, (e) 纯净信号的 F-K 谱, (f) 噪声的 F-K 谱, (g) 含噪信号的 F-K 谱

BPF 作为一种基本的频率域滤波方法,其通带范围设置被设定为 30-70 Hz,目的是过滤掉低频和高频的噪声分量,保留中频带的有效信号。WTF 利用软阈值策略提取系数,并基于 db5 小波函数将分析的信号分解为 15 层,进而提取和重构信号的有效成分。VMD 方法通过自适应分解地震信号到多个模态,然后选

取特定模态组合作为去噪结果,从而捕捉信号的本质特征,设置第一模态和第二模态相加作为去噪结果。WNNM 则是通过加权核范数的最小化来达到对数据降噪的目的,其参数设置遵循原始论文^[143]中的推荐配置。对于深度学习的方法,所有方法的网络超参数设置都一致。为了进行公平的比较,使用相同的数据集训练了 CNN、U-Net 和 MIRNet。不同方法的去噪结果和对应频谱图分别如图 4.5 和图 4.6 所示。不同方法差值及对应频谱图分别如图 4.7 和图 4.8 所示。

4.3.2 模拟实验结果与分析

(1) 时域结果与分析

从 4.4 (d) 中可以看出,强烈的噪声会严重影响反射事件的识别,特别是对于深层的弱信号。在已知信号的频带范围的前提下,BPF 能够滤出大部分噪声,并使信号被观测到。但是,一部分和信号同频带的噪声会被滞留到去噪结果中。同时,从预测噪声中也可以看出一部分信号也被过滤到了记录里。WTF 未能准确分离有效信号和噪声,仅滤除了高频噪声分量。WNNM 和 VMD 虽然也能够提取出信号,但是其对于信号的保幅度性较差。与传统方法相比,基于深度学习的方法基本都能对信噪比进行大幅度的提升改善,这些网络几乎完全消除了复杂背景噪声造成的负面影响,并重建了同相轴信号。但是,在 DnCNN 的去噪结果中仍然能观察到残余的噪声,且在预测的噪声中会发现细微的信号泄露,如图 4.7 (e) 中红色和黄色箭头标注所示。U-Net 在模拟记录中的表现要优于 DnCNN,这可能归因于解编码结构赋予的更大范围的感受野,但是差值图中仍有信号泄露,如图 4.7 (f) 中红色和黄色箭头标注所示。MIRNet 的结果也类似,只能根据一部分差值结果中的区域判断出该方法存在微量的信号泄露,如图 4.7 (g) 中红色和黄色箭头标注所示。整体看来,SDT 在信号恢复和噪声衰减方面对比其他方法更具有优势。例如,弱的同相轴信号可以通过充分提取远程特征来被更准确地恢复,如图 4.5 (h) 中红色、黄色和绿色箭头标注所示,同时在差值图中,也几乎不存在有效信号的泄露。总结来看,对弱信号的恢复方面和对信号的保幅方面,SDT 优于 DnCNN, U-Net 和 MIRNet,同时去噪效果也要远超传统去噪方法。

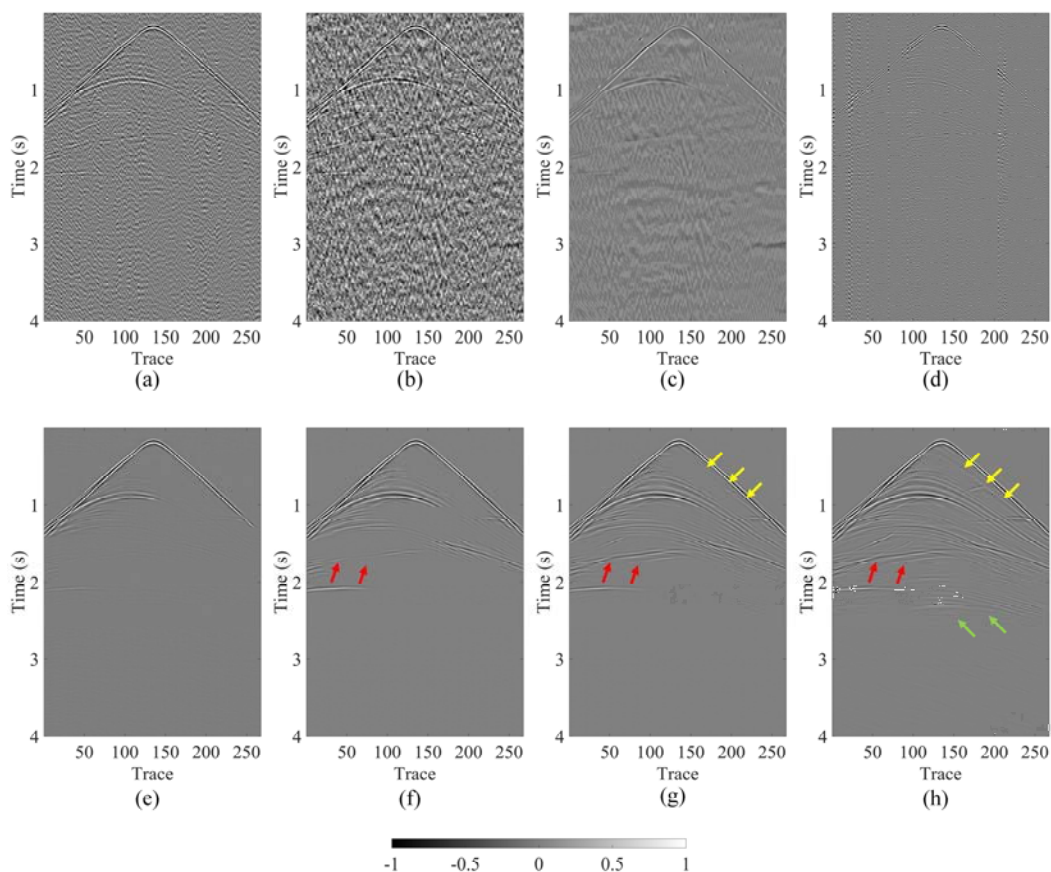


图 4.5 不同方法的去噪结果 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

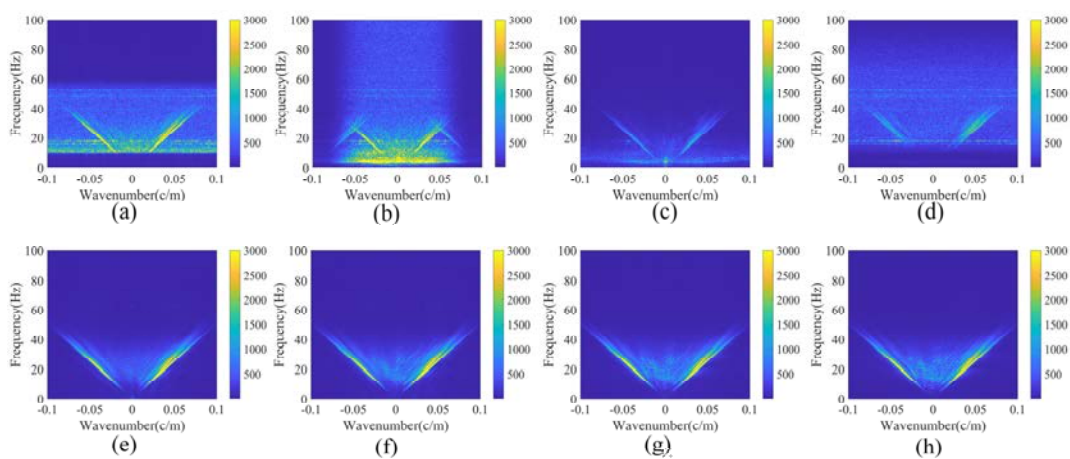


图 4.6 不同方法的去噪结果的 F-K 谱分析 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

(2) 频域结果与分析

频谱分析是反映信号恢复的有效方法,这在时域分析中是无法清楚观察到的。

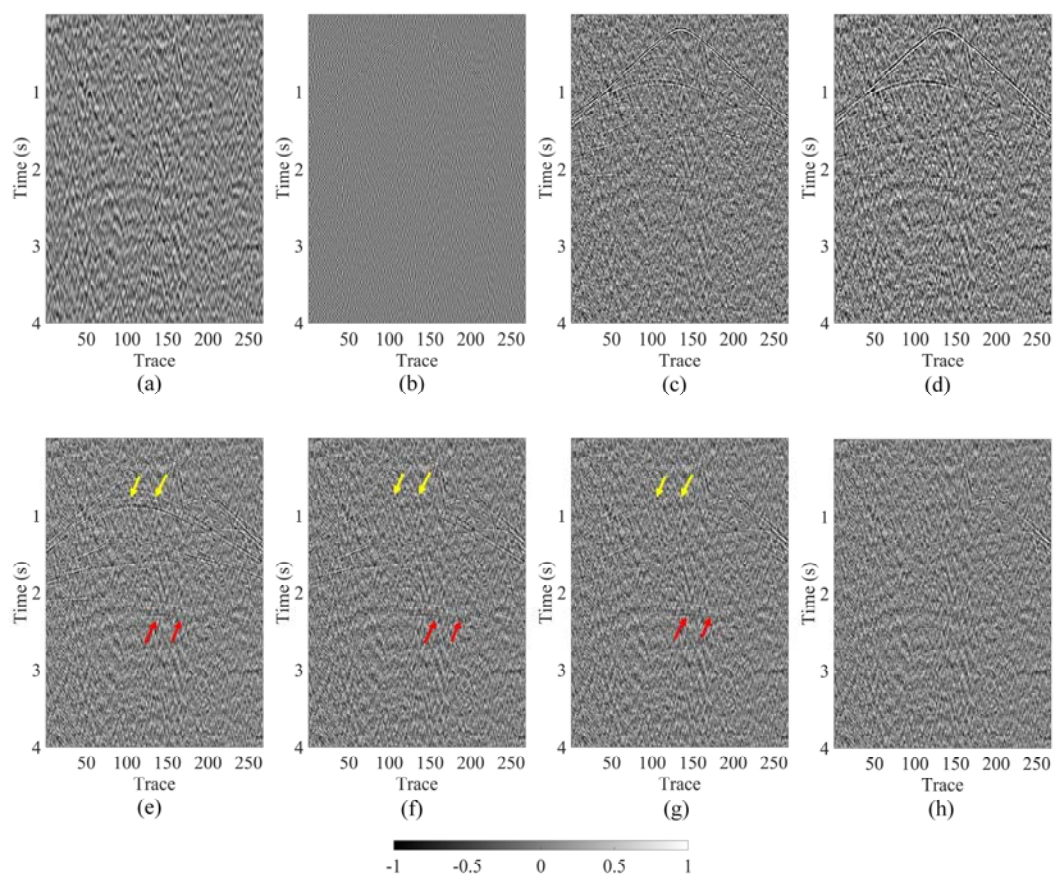


图 4.7 不同方法的差值结果 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

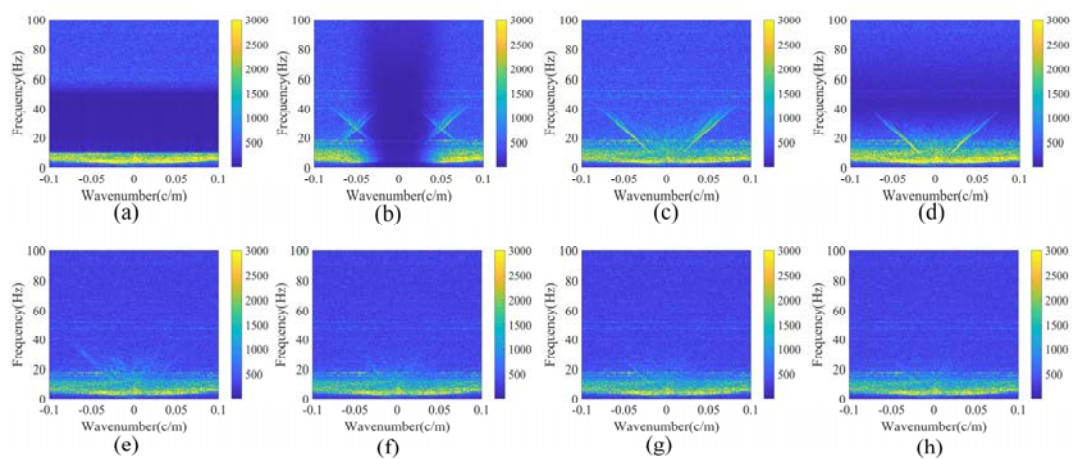


图 4.8 不同方法的差值结果的 F-K 谱图 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

因此, 图 4.6 中同时展示了不同去噪结果的 F-K 谱用于进一步的详细比较, 以研究各种方法的噪声衰减能力。图 4.8 中则展示了不同去噪方法的差值的 F-K 谱。

表 4.4 不同去噪方法去噪后结果的信噪比

原纪录	去噪后结果的 SNR							
SNR	BPF	WTF	WNNM	VMD	DnCNN	U-Net	MIRNet	SDT
-0.0394	5.4917	1.5503	2.9496	1.0912	7.2091	9.7852	10.6081	12.0818
-5.4163	0.2682	-2.6735	2.1312	-4.2569	7.0336	9.3396	10.1625	11.9063
-9.1580	-3.4367	-6.0921	1.1007	-8.0936	6.9280	8.6335	9.2174	10.2878
-15.1786	-9.4371	-11.9245	-1.9368	-14.1661	6.2548	7.4907	7.8958	8.5755
-18.7004	-12.9552	-15.4105	-4.3263	-17.6936	5.3469	6.1965	6.4682	6.8931
-20.0394	-14.2399	-16.7418	-5.3484	-19.0337	4.3407	5.1652	5.3628	5.3666

表 4.5 不同去噪方法的去噪后的 RMSE 对比

原纪录	去噪后结果的 RMSE							
RMSE	BPF	WTF	WNNM	VMD	DnCNN	U-Net	MIRNet	SDT
0.0174	0.0092	0.0145	0.0124	0.0153	0.0076	0.0056	0.0051	0.0043
0.0324	0.0168	0.0236	0.0136	0.0284	0.0077	0.0059	0.0054	0.0044
0.0499	0.0258	0.0350	0.0153	0.0441	0.0078	0.0064	0.0060	0.0053
0.0997	0.0515	0.0686	0.0217	0.0887	0.0085	0.0073	0.0070	0.0064
0.1496	0.0772	0.1024	0.0286	0.1332	0.0094	0.0085	0.0082	0.0078
0.1745	0.0900	0.1194	0.0322	0.1554	0.0105	0.0096	0.0094	0.0094

在对含噪数据进行分析时，如图 4.4（e）和 4.4（f）中所示，在频域中已经可以观察到信号和噪声分量之间有着明显的频谱重叠。由于混叠现象，BPF 无法从混叠的共享频带中提取出全部的信号。WTF 去噪后的差值图中也能够明显观察到大量的信号残留。WNNM 和 VMD 也是面临着同样的信号泄露问题。与在时域中观察的结果类似，基于深度学习的方法在噪声衰减方面更有效，从频域中能够更加清晰地观察到信号和噪声的明显区分。然而，在 DnCNN 恢复信号的 F-K 谱中仍然观察到噪声分量。对 U-Net 和 MIRNet 来说，其结果中基本观察不到噪声分量。仅从频域分析来说，它和 SDT 基本效果一致，对这几种深度学习去噪方

法的对比应主要聚焦于后续定量指标的比较上。

表 4.6 不同去噪方法的去噪后的 SSIM 对比

原记录	去噪后结果 SSIM							
SSIM	BPF	WTF	WNNM	VMD	DnCNN	U-Net	MIRNet	SDT
0.6352	0.8507	0.7219	0.9019	0.6780	0.8903	0.9338	0.9464	0.9604
0.3386	0.6536	0.4679	0.8315	0.3917	0.8862	0.9271	0.9411	0.9590
0.1708	0.4589	0.2851	0.7485	0.2096	0.8834	0.9141	0.9266	0.9417
0.0404	0.1759	0.0882	0.5304	0.0535	0.8662	0.8917	0.9033	0.9171
0.0156	0.0812	0.0370	0.3715	0.0211	0.8405	0.8614	0.8720	0.8844
0.0107	0.0586	0.0260	0.3120	0.0146	0.8083	0.8329	0.8427	0.8459

表 4.7 不同方法的计算时间及提升的信噪比

	BPF	WTF	WNNM	VMD	DnCNN	U-Net	MIRNet	SDT
训练时间	-	-	-	-	8h 36min	9h 15min	26h 02min	40h 38min
处理时间 (秒)	0.23	0.41	271.66	440.89	1.29	0.98	1.56	2.57
平均信噪 比提升 (dB)	5.72	3.75	10.34	1.08	18.50	20.03	20.55	21.53

(3) 定量指标与分析

利用信噪比 (Signal-to-noise ratio, SNR)、均方根误差 (Root Mean Square Error, RMSE) 和结构相似性 (Structural similarity, SSIM) 三个指标, 对不同方法获得的去噪结果进行定量评估。其中, RMSE 是 MSE 的平方根。由于定义 SNR, MSE 和 SSIM 的方程式在上一章中均已经介绍, 在此均不再赘述。

为了充分比较不同方法在各个信噪比下的去噪性能, 计算了不同方法对不同信噪比数据的去噪结果的 SNR、RMSE 和 SSIM, 相应的结果列于表 4.4、4.5 和 4.6 中。从中可以看出, SDT 去噪后的结果有着最小的 RMSE、最高的 SNR 以及最高的 SSIM, 从而表明了 SDT 具有最优异的幅度保持和噪声抑制能力。

此外，为了评估各种去噪方法在不同区域内的效果，选择局部信噪比作为一种度量，用于量化局部数据的保真度。通过采用尺寸为 5×5 、步长为 1 的移动窗口对地震数据进行分割，在移动窗口中计算数据的信噪比作为局部信噪比。通过在移动窗口中对数据进行分析，允许细致地量化数据的每一个小区域，从而得到更为精细化的去噪效果评估，同时也能得到局部视角下地震数据保真度的直观度量。局部信噪比的公式可参考上一章节的公式 (3.13)。各种去噪方法的结果的局部信噪比图如图 4.9 所示。这里，横坐标表示地震道的索引，纵坐标对应于数据采样点数，颜色条内的数值表示以分贝 (dB) 为单位的信噪比值。图 4.9 (h) 展示了 SDT 处理后的局部信噪比结果。值得注意的是，在存在信号的区域内，局部信噪比的增强变得明显，如图 4.9 (h) 中的白色箭头所示。与其他方法相比，所提出的 SDT 方法对地震记录的局部信噪比的改善更为显著。

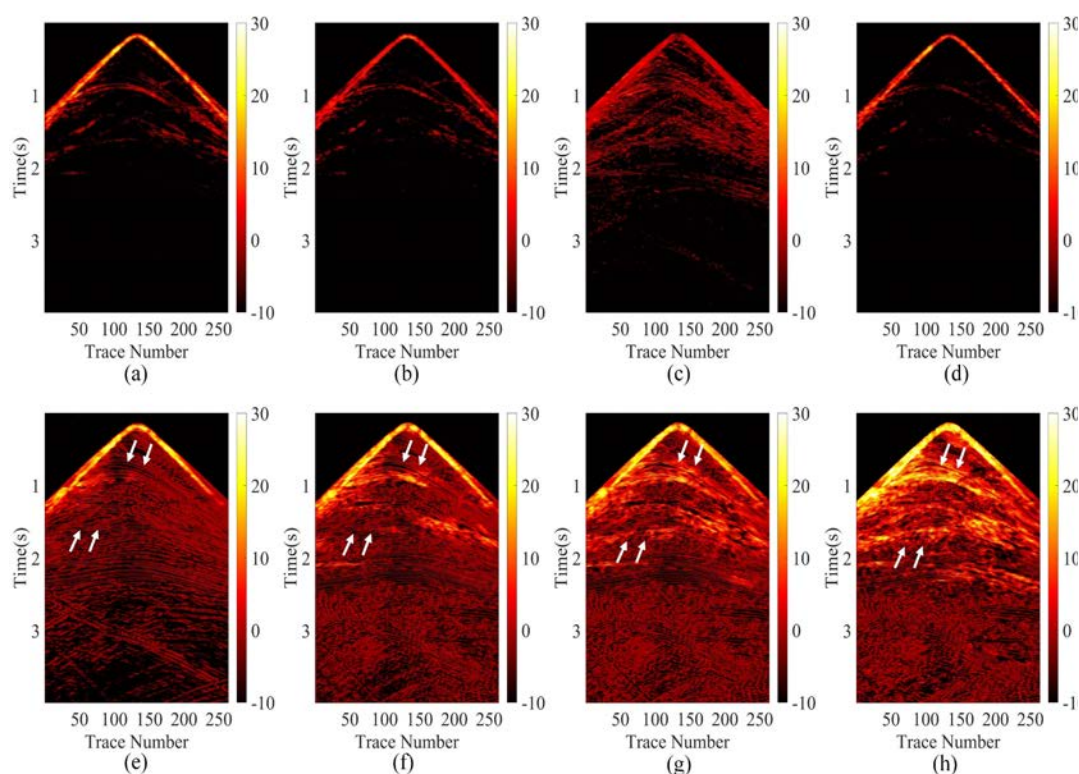


图 4.9 不同方法的结果的局部信噪比分析 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

为了更加综合全面的评估，使用不同的方法处理 50 个具有不同信噪比的合成记录，以更加公平地量化评估去噪性能。表 4.7 列出了相应的信噪比提升、训练时间和处理时间。结果表明，基于深度学习的方法在去噪能力上显著优于传统

方法,例如,SDT 的信噪比提升为 21.53dB,而 BPF 和 WT 的信噪比提升仅为 5.72dB 和 3.75dB。同时 SDT 的性能也明显超越其他深度学习方法。

计算效率是评估深度学习方法应用的一个关键指标,尤其是在大规模数据处理任务中。研究了与不同去噪方法相关的计算时间。基于 CNN 的方法使用类似 SDT 的超参数设置。各个方法的运行时间如表 4.7 中所示。从中可以看出,SDT 的训练时间和测试时间都要略高于 DnCNN、U-Net 和 MIRNet。这可能归结于自注意力运算相比于卷积运算要更耗时且难训练。从测试时间来看,三种深度学习方法相差不大,都可以在很短的时间内完成。训练时间虽然很长,但是一旦训练完成,测试时间便可以忽略不计。基于深度学习的方法在训练过程中可以自适应地优化网络参数,避免手动微调过程,从而显著提高去噪方法的智能化水平。训练后的模型具有很高的泛化能力,可以应用于处理在类似环境中采集的地震数据。因此,基于深度学习的方法的高计算成本训练是可以接受的。

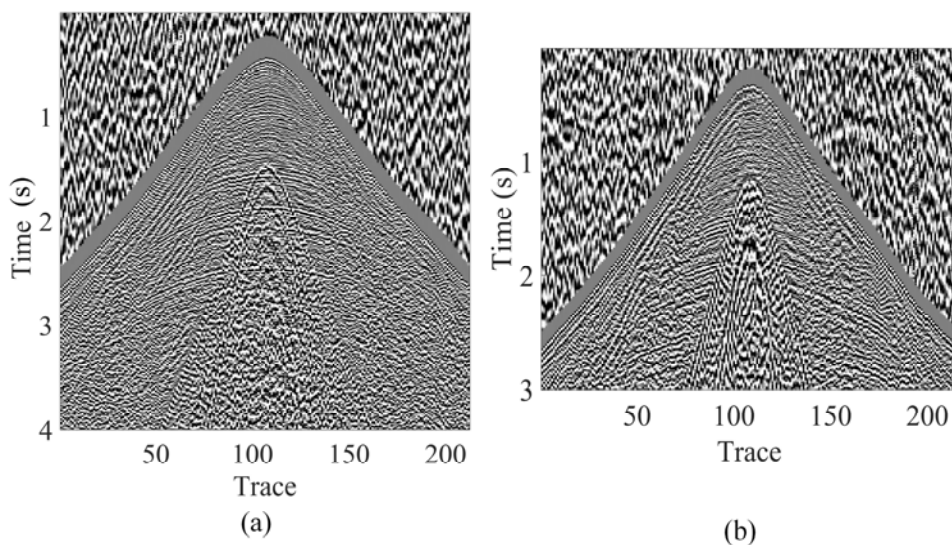


图 4.10 野外采集的含噪实际记录 (a) 沙漠地区勘探地震记录, (b) 同一测区不同测线获得的另一记录

4.3.3 实际实验结果与分析

使用训练好的 SDT 对图 4.10 中的两幅实际数据进行了处理,以进一步研究所提出的网络的去噪能力。网络参数没有进行重新微调,而是使用在处理模拟数据时的权重。图 4.10 (a) 显示了用于处理的第一幅记录,该数据采集自我国塔

里木地区，由 212 条记录道组成，记录道间距为 10 m，数据的采样频率为 500 Hz。

经过仔细观察可以看出，实际记录受到各种复杂干扰的影响，包括随机噪声

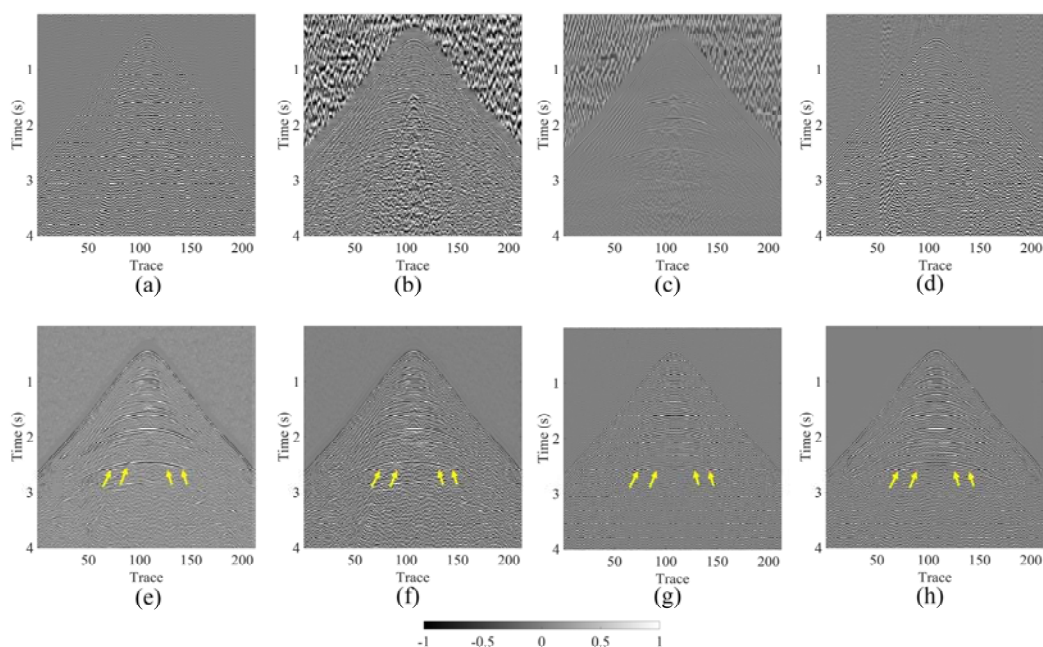


图 4.11 不同方法对实际记录的去噪结果 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

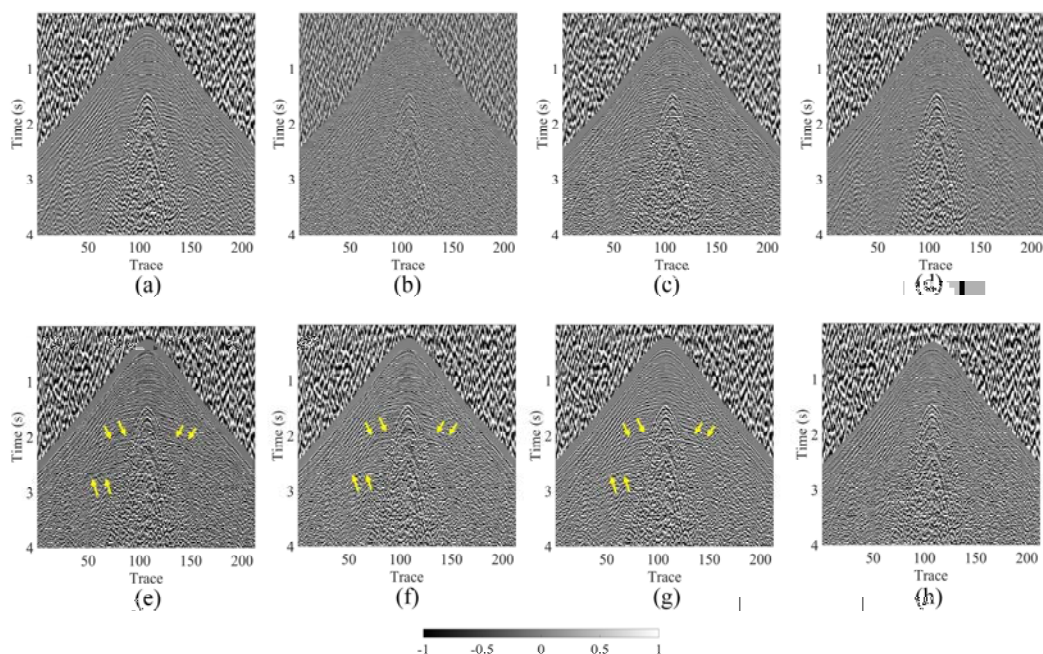


图 4.12 不同方法对实际记录去噪后的差值 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

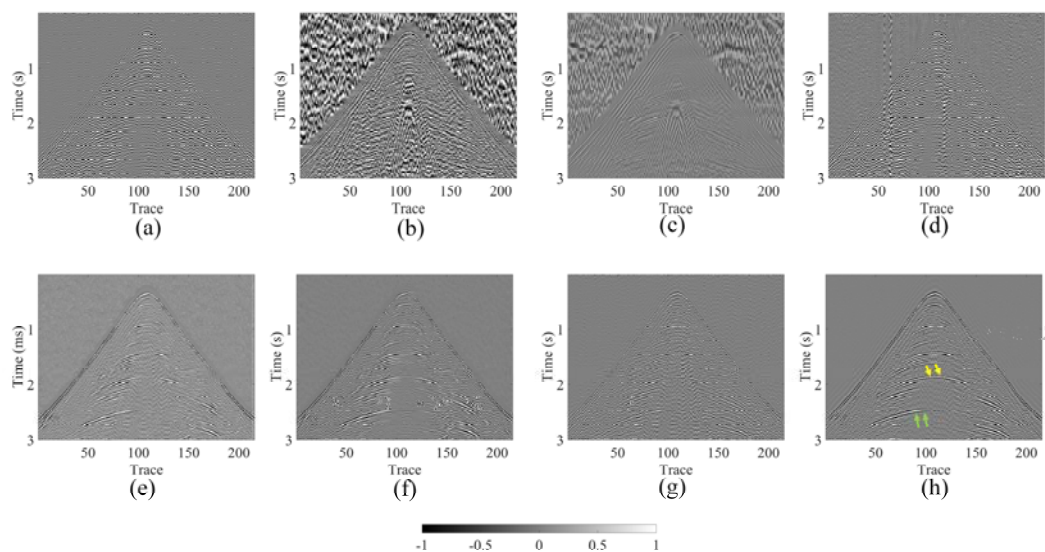


图 4.13 不同方法对实际记录的去噪结果 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

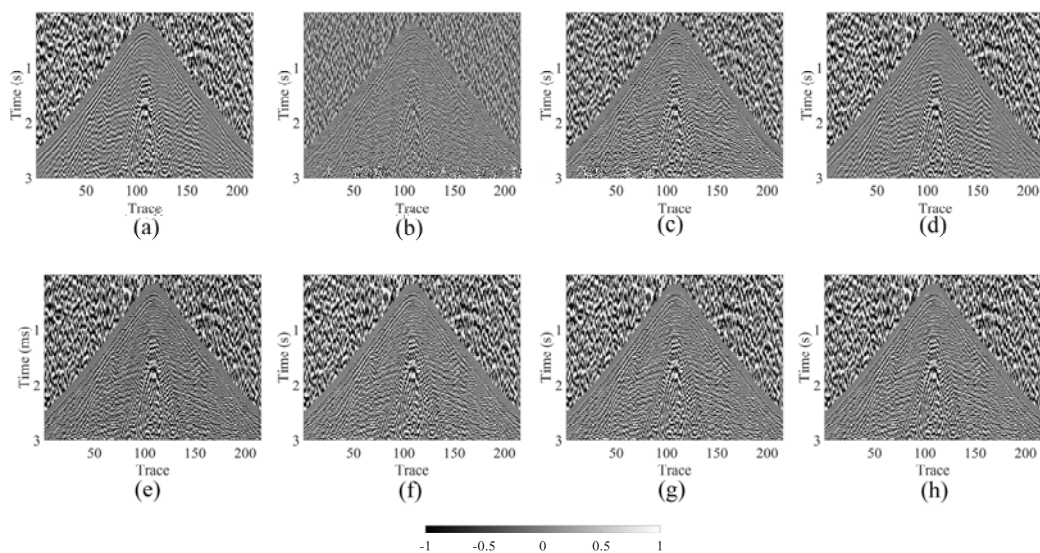


图 4.14 不同方法对实际记录去噪后的差值 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

以及面波影响。在浅层和深层中,可以观察到弱信号淹没在其中。在不对网络参数进行重新微调的情况下,直接应用训练好的 SDT 模型处理这些实际数据,是对模型泛化能力的一大考验。通过这种方式,可以检验 SDT 模型在处理实际低信噪比、复杂噪声干扰的地震数据时的去噪效果。

应用不同的去噪方法来处理复杂的噪声,并恢复被噪声污染的信号。相应的结果如图 4.11 所示,各种方法处理后的差值如图 4.12 所示。BPF 只能机械式的

过滤固定频带的噪声。WTF 抑制背景噪声的能力有限，其结果中观察到大量的背景噪声仍然存在。WNNM 和 VMD 虽然有一定的去噪效果，但是从差值图中依旧能够观察到比较多的信号残留。与传统方法相比，图 4.11 (e) - (h) 中基于深度学习的去噪方法则表现出更好的性能，大多数的强噪声干扰都被消除了，恢复的信号也能被清晰地识别出来。总的来说，如图 4.11 (h) 所示，SDT 恢复的信号具有最好的连续性。借助更多的远程特征的帮助，SDT 也倾向于恢复更多被噪声覆盖的信号。

为了继续测试网络的效果，在采集自同一测区其他测线的记录上继续进行实验。图 4.10 (b) 显示了同一区域在另一测线采集的现场记录图。图 4.13 为不同方法对该记录的处理结果，图 4.14 为不同方法处理该记录后的差值。通过比较不同方法的去噪性能，BPF, WTF, WNNM 和 VMD 等方法恢复的信号仍然受到较多残余噪声的影响。相反，深度学习类方法拥有更好的噪声压制能力。在这其中，如图 4.13 (h) 中由黄色和绿色箭头标记的区域所示，SDT 具有最显著的信号恢复能力，并且恢复的同相轴信号可以保持高连续性。实验结果表明，SDT 是一种较完善的复杂噪声衰减方法，能够极大地恢复被噪声覆盖的有效信号。

4.4 算法分析讨论

本节通过对网络进行通用性验证测试该方法在其他类型的地震记录中是否依旧能够保持去噪性能，通过归因分析方法解析所提 SDT 算法在地震数据去噪任务上的有效性的原因，通过消融实验验证各个部件的作用，同时也通过对不同网络的预训练策略展开了分析研究。

4.4.1 网络的通用性验证

分布式光纤传感技术 (Distributed Acoustic Sensing, DAS) 是一种新兴的采集技术^{[120]-[125]}，已应用于垂直地震剖面采集 (Vertical Seismic Profile, VSP)。其数据中往往含有多种类型的噪声污染信号，且特殊的性质使得传统方法往往无效。本研究中通过重新正演，组织了新的训练集，正演的参数如表 4.8 所示，部分训练集中的数据切片如图 4.15 所示，在此训练集上对 SDT 进行了重新的训练。

表 4.8 DAS-VSP 数据正演参数设置

参数	含义	具体设置
震源子波类型	地震信号的小波函数种类	雷克子波
子波频率	地震信号的频率范围	30-70 Hz
波速范围	地震波在介质中传播的速度范围	1200-4000 m/s
密度范围	介质中密度设置的范围	1300-4000 kg/m ³
采样时间间隔	地震数据记录中连续采样数据的时间间隔	4×10^{-3} s
接收器数量	地震数据采集集中使用的接收器的数量	2000
最大传播时间	地震波从震源到接收器的最长传播时间	3 s

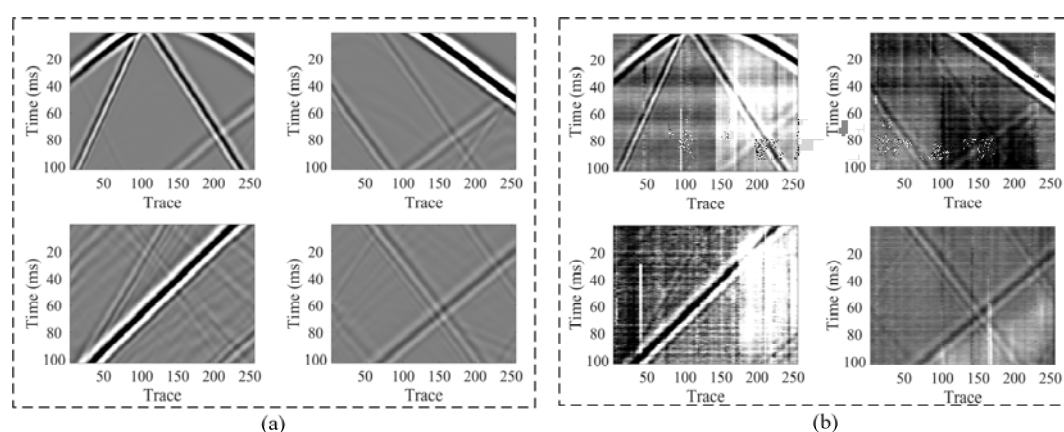


图 4.15 训练集中 DAS-VSP 数据样本 (a) 纯净信号切片, (b) 含噪信号切片

使用训练后的模型对图 4.16 中的现场 DAS 数据进行了处理, 以进一步研究所提出的网络的去噪能力。DAS 记录由 2560 条记录道组成, 记录道间距为 1 m, 现场数据的采样频率为 2500 Hz。与传统的电检波器不同, DAS 可以提供高分辨率的地震数据, 同时更灵敏的传感器也受到了更多噪声的困扰^[122]。从图 4.16 所示的 DAS-VSP 数据中, 可以观察到数据受到严重的噪声污染, 有效信号如入射下行波 (标记为蓝线) 和反射上行波 (标记为黄线) 几乎被噪声覆盖。这些噪声类型复杂多样, 包括随机噪声 (如橙色虚线区域所示)、光学噪声 (如紫色虚线

区域所示)、衰落噪声(如黑色虚线区域所示)、水平噪声(如绿色虚线区域所示)。众多关于 DAS-VSP 记录的研究都认为同时压制这些噪声是一种挑战^{[120]-[125]}。下面对这些噪声的特性和成因进行简要的描述。

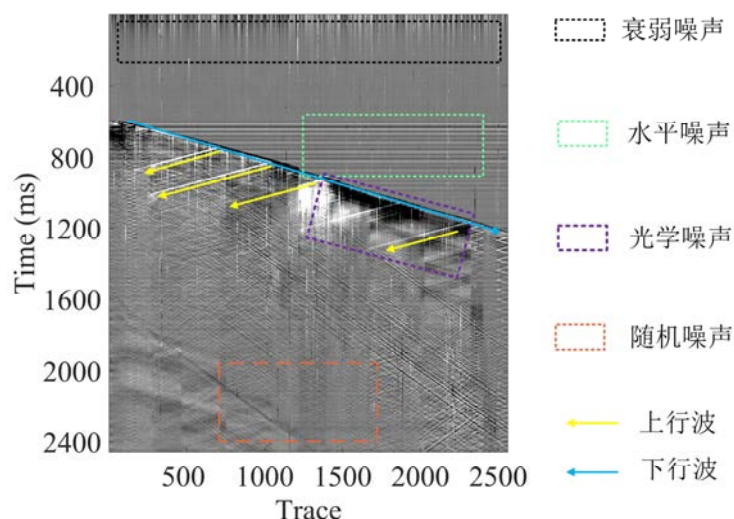


图 4.16 实际环境中采集的 DAS-VSP 记录

随机噪声是 DAS-VSP 数据中常见的背景噪声,具有宽频带和强能量,主要由数据采集过程中的振动引起。噪声来源可能包括机器振动、地下随机振动、电流干扰等。随机噪声通常分布较为均匀,是影响 DAS-VSP 数据质量的主要因素。

光学噪声具有较低的主频和非常强的能量,可能由井下温度异常引起。光学噪声倾向于在记录中大面积出现并且振幅高。在光学噪声存在的情况下,有效信号完全被覆盖,严重影响 DAS-VSP 数据的质量。

衰落噪声由随机间隔的后向散射光的相位抵消干扰引起。衰落噪声主要表现为长周期和高振幅。通常,衰落噪声出现在记录的最上方,对初至波以下的有效信号影响较小。它容易与水平噪声混合,生成新的噪声类型。

水平噪声由光学测量过程中的振动引起。电子设备的潜在泄露也可能是水平噪声的另一个成因。水平噪声通常表现为短暂时间的水平带,在所有道中几乎具有相同的相位,并且随时间振幅逐渐递减。

应用 SDT 来处理这四类复杂的噪声,同时,选择一系列对比方法并行地来处理这个记录,以与 SDT 方法进行对比。

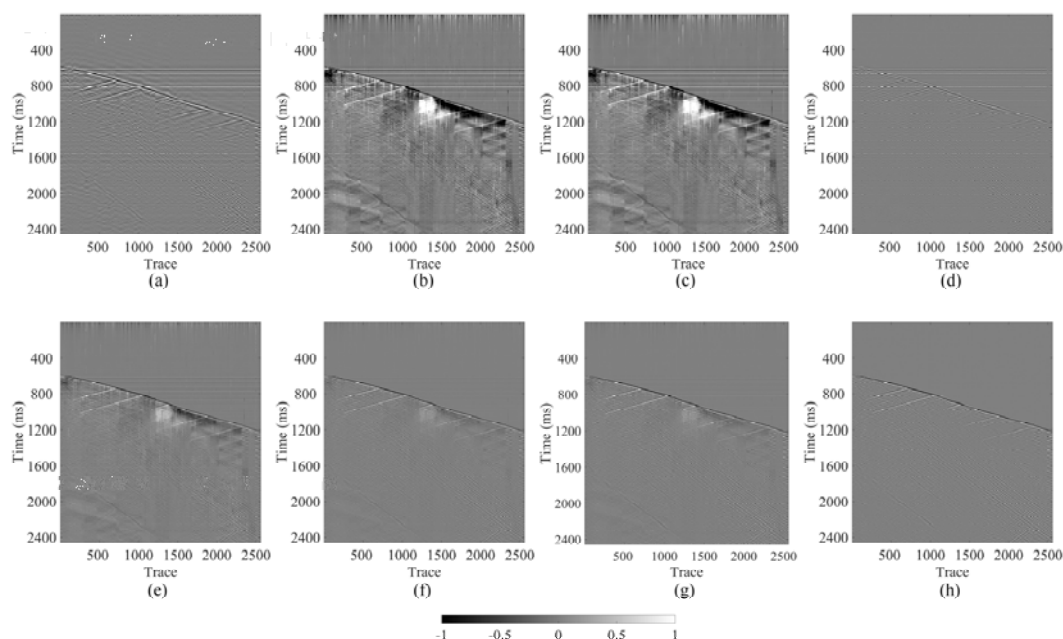


图 4.17 不同方法对现场 DAS-VSP 地震记录的去噪结果 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

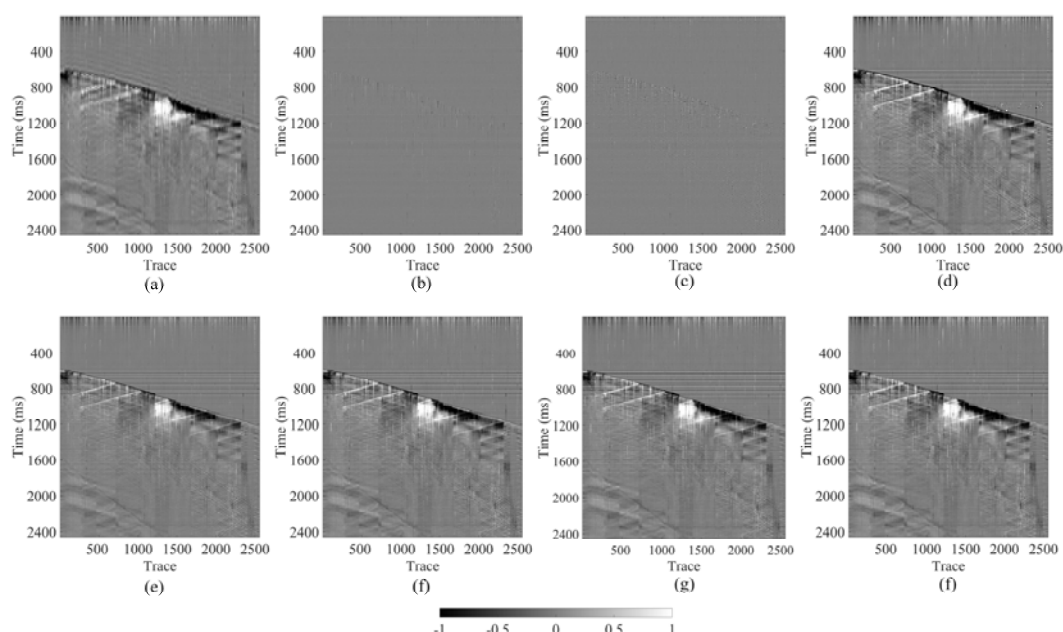


图 4.18 不同方法对 DAS-VSP 地震记录去噪后差值 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

从结果中可以看出, BPF 和 VMD 抑制 DAS 中噪声的能力有限, 观察到大量的噪声仍然存在, 尤其是对于一些特定形态的噪声如水平噪声, 图 4.17 (a)

和 (d) 中的水平噪声依旧还有大量的剩余。WTF 和 WNNM 对信号的保幅能力较弱, 在其结果即图 4.17 (b) 和 (c) 中很难看出有效信号。与传统方法相比, 基于深度学习的方法对于 DAS 记录这种多类型噪声记录能够表现出更好的去噪性能。多种噪声往往会被同时被消除, 恢复的信号也能被清晰地识别出来。相比于 DnCNN、U-Net 和 MIRNet 三种方法, SDT 恢复的信号具有良好的连续性, 且从全局来看, SDT 恢复的信号具有更好的全局一致性。

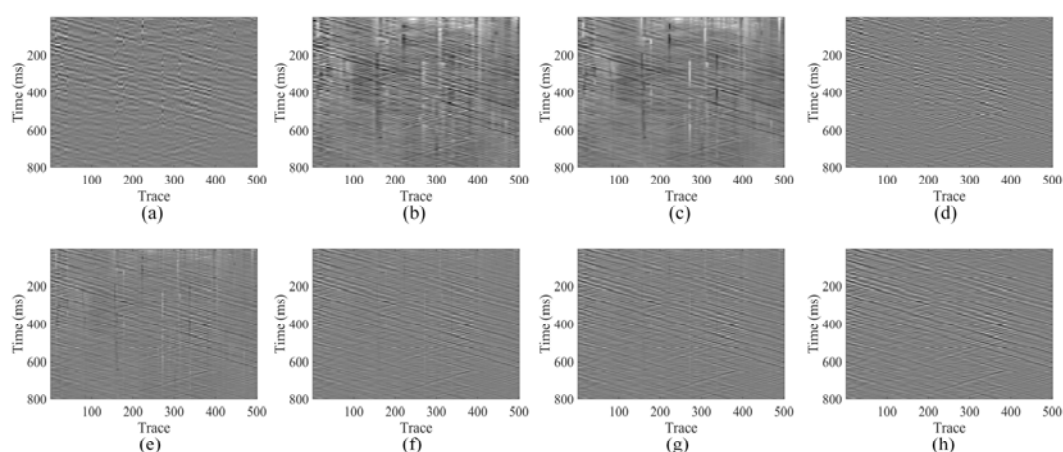


图 4.19 不同网络模型的去噪结果的放大图 (a) BPF, (b) WTF, (c) WNNM, (d) VMD, (e) DnCNN, (f) U-Net, (g) MIRNet, (h) SDT

特别是对于如图 4.16 中被光学异常噪声污染的弱上行信号区域, 通过放大这一区域中的结果以进行详细分析。可以观察到, 由 DnCNN、U-Net 和 MIRNet 重建的同相轴质量较差, 许多残余的噪声仍然影响去噪质量。相比之下, 只有 SDT 准确地恢复淹没在强光学异常噪声中的弱的同相轴信号。所有这些实验结果表明, SDT 在应用于复杂 DAS 背景噪声衰减时依旧是一个有效的去噪方法, 且在恢复长范围信号、处理具有复合噪声模式的多噪声类型记录方面有着更好的效果。

4.4.2 去噪网络的归因分析

深度学习网络在地震去噪方面的应用已经取得了显著的成果, 但其内部工作机制的可解释性仍然是一个挑战, 主要是因为深度学习模型通常被视为“黑箱”, 其决策过程复杂且不透明。为了揭示深度学习网络在回归类任务中的工作原理, 一部分研究者们已经尝试利用诊断工具如局部归因图 (Local Attribution Maps,

LAM)^[145]来对网络进行解释,这是一种有效的归因方法。LAM 的基本思想是为深度学习模型提供一种视觉化的解释,通过这种方式,可以直观地了解模型是如何对输入数据做出决策的。在地震数据去噪的场景中,LAM 可以帮助理解模型是如何识别和处理噪声的。具体来说,LAM 通过生成与输入数据同尺寸的归因图来实现这一点。这些归因图显示了每个采样点对模型输出的贡献度,让研究者能够清楚地看到模型在数据的哪些区域上投入了更多的关注,这些区域通常是噪声较多或数据信息较为关键的部分。使用 LAM 的一个主要优势是其能够揭示模型决策过程中的细节,这对于改进模型设计和解释模型行为至关重要。例如,如果发现模型在去除某种特定类型的噪声时表现不佳,LAM 可以帮助研究者定位问题,并指导研究者对模型结构或训练过程进行调整。此外,这种方法还可以帮助验证模型是否按照预期工作,或者是否存在过度依赖于无关特征的情况。因此,LAM 为深度学习网络的去噪机制提供了一种可视化和直观的理解方式,有助于揭示模型的决策过程,从而为模型的优化和改进提供指导。

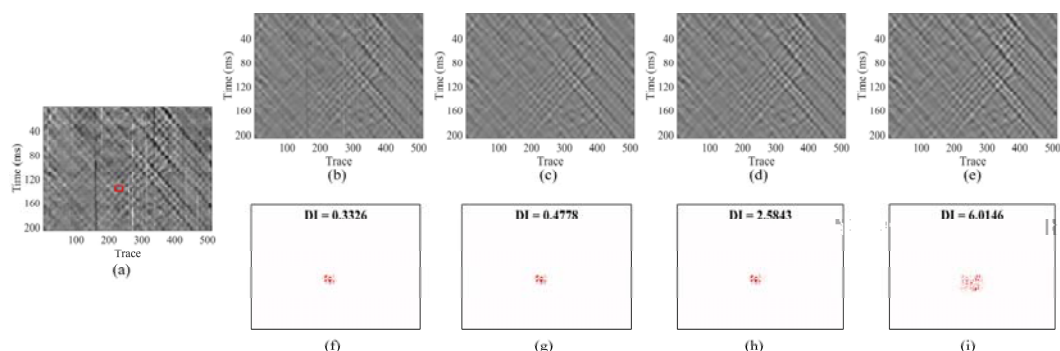


图 4.20 不同的网络模型的局部归因图 (a) 原始记录, (b) DnCNN 结果, (c) U-Net 结果, (d) MIRNet 结果, (e) SDT 结果, (f) DnCNN 的 LAM 图, (g) U-Net 的 LAM 图, (h) MIRNet 的 LAM 图, (i) SDT 的 LAM 图

本研究中尝试使用诊断工具 LAM 来分析几个不同的深度学习模型的性能差异。LAM 结果中的红点表示用于重建去噪图像中用红框标记的区域时所使用的采样点,并且计算扩散指数 (DI)^[145]以反映所涉及的采样点的范围。用于恢复特定输入区域的采样点越多,LAM 中红点的分布就越宽,DI 就越高。使用 LAM,既可以判断哪些输入采样点对所选区域的贡献最大,又可以借此评估去噪模型利用远程信息的能力。针对如图 4.19 所示的结果,不同深度学习去噪方法的 LAM

分析图如图 4.20 所示,红色标记点是有助于重建的采样点。一般来说,利用的信息越多,就可以获得更好的性能。从图中可以观察到,相比于 DnCNN、U-Net 和 MIRNet,SDT 的红色区域要明显更大。其次,由于使用的采样点范围有限,DnCNN 可能会恢复出错误的结构,如一些垂直方向的同相轴。U-Net 和 MIRNet 虽然利用了比 DnCNN 更多的输入采样点,但整体范围依旧不大。对比发现,SDT 可以使用最大范围的采样点,并恢复正确清晰的同相轴。对于定量的度量,SDT 也获得了比 DnCNN、U-Net 和 MIRNet 三种方法更高的 DI 值。这些结果表明,SDT 使用了更多的远程特征来重建去噪数据。因此,SDT 对信号的恢复结果往往具有更好的连续性。

可以从两种模型的计算机制上去理解这种差异:CNN 中的卷积操作和 Transformer 中的自注意力机制都可以用来融合不同位置的特征,以捕捉地震数据中尽可能多的远程信息。在 CNN 中,卷积操作通过滑动卷积核提取不同位置的特征,并将这些特征在空间上进行局部的加权和融合,从而得到具有位置不变性和层次结构的特征表示。而在 Transformer 中,自注意力机制则是通过计算不同位置之间的相似度,来融合不同位置的信息。具体地,通过计算每个位置与其他位置的相似度得到一个注意力分布,再将该注意力分布作为权重,对其他位置的特征进行加权和融合,从而得到一个综合的表示。相比于 CNN 结构,Transformer 的特征提取要更加灵活,其特征提取不在空间上受到局部限制,从而能够更好地提取地震数据中的远程信息。因此,在使用 LAM 分析结果时,能够看到 SDT 相比于三种基于 CNN 的网络模型有着明显更大的红色范围。

需要强调的是,作为一个辅助工具,LAM 方法也有其局限性。归因图的解释需要依赖于一定主观判断,这可能导致解释的不一致性。而且 LAM 并不能完全揭示模型内部的所有复杂交互作用,其仅仅是一种理解深度学习模型的手段。在深层网络中,完整的解释机制依旧是个悬而未决的问题。

4.4.3 去噪网络的消融实验分析

在深度学习网络去噪的研究中,消融研究通过去除某些组件来研究模型的性能,以了解组件对整个系统的贡献。本节通过消融研究详细分析 SDT 各成分的

作用。选择 SNR 提升量和 SSIM 提升量这两个指标作为度量来全面衡量测试集上的模型性能。消融结果见表 4.9。首先，用基于卷积的残差块替换 SDT 中的所有 SW-MSA 块，同时保持所有其他块不变。可以观察到，具有 SW-MSA 的 SDT 能够实现更好的性能，并且优于将 SW-MSA 替换为残差块的神经网络。这项研究表明了 SW-MSA 与卷积相比在网络结构中更具有有效性。表 4.10 还显示了 FFN 的消融结果。可以观察到，去除 FFN 后 SDT 的去噪效果也出现了下降，这表明 FFN 对于提升地震数据的去噪也具有一定的作用。本研究中也进行了消融实验来验证所提出的 CAB 的有效性。如表 4.9 所示，与没有 CAB 的模型实现的 19.81dB 的增益相比，有 CAB 的 SDT 模型实现了 21.53dB 的性能增益，从而验证了 SDT 网络中加入通道注意力对于去噪效果的提升也产生了正向作用。

表 4.9 SDT 模型各部件消融研究

W-MSA	×	✓	✓	✓
CA	✓	×	✓	✓
FFN	✓	✓	×	✓
平均信噪比 提升 (dB)	18.76	19.81	20.04	21.53
平均 SSIM 提升	0.7491	0.7545	0.7612	0.7709

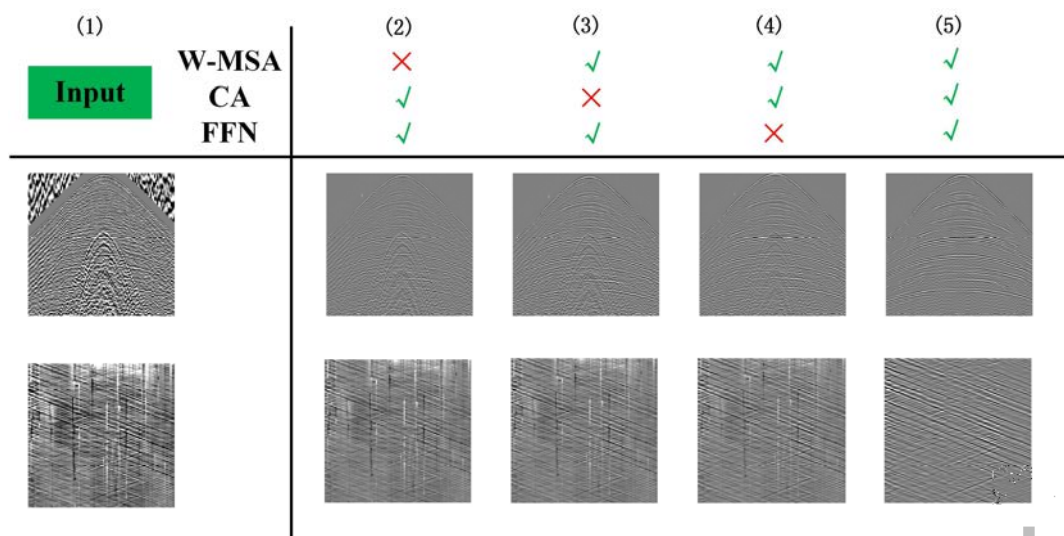


图 4.21 对 SDT 方法的消融研究结果

4.4.4 预训练对结果的影响

与传统的深度学习网络（例如多层感知器、卷积神经网络、循环神经网络）相比，Transformer 的架构很容易加深，且具有更小的归纳偏置，更适合于“预训练-微调”的训练范式。典型的“预训练-微调”范式是指，首先（通常以自监督方式）在大量训练数据上对模型进行训练，然后针对下游任务在较小的数据集上进行微调。预训练有助于模型学习通用表征，进而有利于所需要的任务。为了充分研究预训练策略对不同网络的影响，本研究中对四个网络都进行了预训练：DnCNN、U-Net、MIRNet 和 SDT，结果如表 4.10 所示。可以观察到，所有四个网络都可以从预训练中受益，这表明了所提出的预训练策略的有效性。同时，可以观察到，对于网络容量越大的网络，预训练的性能增益就越多。其中，SDT 从预训练中获得了更大的性能增益，这表明基于 Transformer 的模型需要更多的数据来释放模型的潜力，同时这也表明了对于 SDT 模型预训练策略的必要性。在满足了大模型和大规模数据的条件下，SDT 在地震数据去噪任务的性能上限得到了显著扩展。

表 4.10 预训练流程对不同网络的影响

是否应用 预训练	去噪后的平均信噪比提升（dB）			
	DnCNN	U-Net	MIRNet	SDT
×	18.50	20.03	20.55	20.95
✓	18.71	20.14	20.63	21.53

4.5 本章小结

现有的多数深度学习类去噪算法多是基于卷积神经网络的，尽管在处理地震数据去噪任务中对噪声问题表现出了一定的有效性，但该类方法基础架构是建立在卷积神经网络之上的，因此在捕获远程特征的能力上仍存在一定的不足，在低信噪比情景下，该方法对信号的恢复能力往往较差。针对这些局限性，本章基于 Transformer 架构提出了一种新型地震数据去噪算法——SDT 算法，该方法通过

自注意力机制学习地震信号的内在特征和结构，有效地捕捉到关键信息。此外，为了保证模型在实际数据中的泛化性，本研究中提出一种结合自监督预训练和微调训练的双阶段训练策略，首先模型通过修复掩码后的未标注地震数据，自动学习和提取关键特征，在这一过程中模型能够生成高质量的特征表示。之后，在微调阶段中，SDT 模型在特定的、设计好的数据集上进行调整，并适应特定的地震信号特征和噪声模式。为了验证 SDT 算法的通用性，在设计好的数据集上进行了广泛的实验。这些数据集包括多种类型的地震记录，既有传统的地面地震记录，也有新型传感方式的 DAS-VSP 记录。实验结果显示，SDT 算法在处理这些地震记录方面，性能显著优于当前最先进的地震信号处理方法。此外，本研究中还特别采用了局部归因图分析，来评估模型在捕捉远程特征方面的能力。局部归因图分析结果进一步证实了 SDT 算法在提取和理解地震数据中的远程特征，尤其是在处理那些具有长距离依赖性的地震同相轴信号方面的强大能力，这不仅验证了 SDT 算法在低信噪比地震数据去噪方面的高效性，也展现了其成为通用去噪模型的潜力。

第5章 基于Transformer的地震速度模型反演算法

5.1 引言

除了遭遇严重的噪声问题外,地震数据处理任务还面临着快速精确地获取速度模型的需求。第3章和第4章提出的 DnDenseNet 和 SDT 算法可以有效缓解地震数据中噪声的影响,但是这仅仅是一个预处理的步骤。地震勘探通过人工激发、观测和分析地层中传播的地震波来查明地下地层形态和性质。其中,地震波在地层中的传播速度是地震勘探中最重要的参数之一,其对于高精度的地震成像及预测储层岩性具有重要意义。地震信号处理后续的一系列步骤高度依赖于通过反演获取精确的速度模型,比如,叠前深度偏移成像极大依赖于地层速度模型的精度^[64]。相对于正演任务,反演任务的定义如图5.1所示。随着地震勘探环境与条件的复杂化,对速度估计的精度和速度都提出了越来越高的要求。如何快速而精确地获得速度模型成为一个重要的问题。

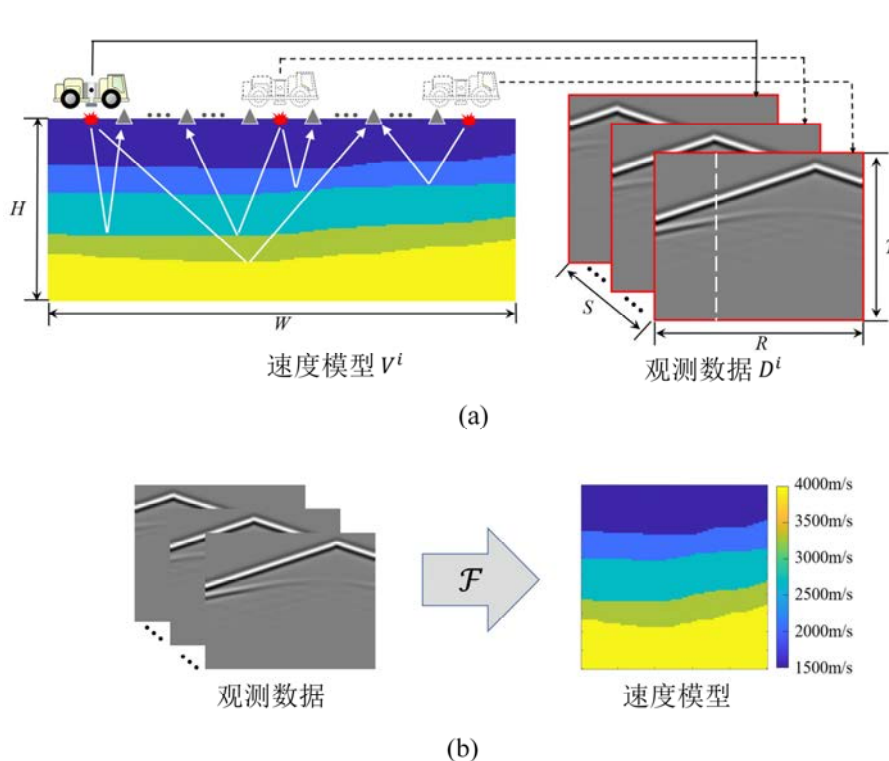


图 5.1 正反演任务定义^[88] (a) 正演任务, (b) 反演任务

鉴于速度模型的重要性,人们开发了很多用于求解它的反演方法。传统反演方法中,偏移速度分析方法以及走时层析反演方法只能得到宏观速度场,也就是低频部分,高分辨率的速度模型则是无法获得,这严重限制了其实际应用^{[67],[68]}。地震全波形反演(Full Waveform Inversion, FWI)是近年来备受关注的方法,其利用地震波场的运动学和动力学信息建立地层速度模型,能够精确刻画模型细节,成为研究的热点。具体来说,FWI通过迭代来最小化地震数据和合成数据间的差异来重建速度模型,其使用梯度求解器来求解。但是,全波形反演方法严重依赖于初始模型来避免局部最小值。同时,由于噪声污染等因素,FWI方法有着极大的非唯一性,且存在局部极小问题。对于FWI的局部极小问题,诸多的改进策略被提出以解决它,例如转换到其他域的处理^{[73]-[76]}。另外,FWI也在多个方面得到拓展,如应用于粘弹性介质^[69]和联合反演^[79]中。但是,依赖初始模型的问题依旧需要新的方法去解决。同时,迭代优化带来巨大的计算量,使得FWI用时较多,难以广泛应用到实际,这也成为FWI方法的一个值得注意的问题。

近年来,随着硬件计算能力的快速增长,在大规模数据集上学习的深度学习模型无论在地震记录的去噪^[80]、反演^[81]、解释^[82]等任务上都展现出优于传统方法的有效性。随着模型的表示能力的不断提高,数据驱动方法为速度模型反演提供了新思路,使得这部分的研究更受研究者瞩目。现有的方案多是基于CNN,尽管有着非常好的效果,CNN应用于反演时依旧存在潜在问题,本研究更加聚焦于解决这些问题并寻求新的解决方案。多数的CNN方法会在提取地震记录中信号特征的过程中提供适当的归纳偏置(Inductive Biases),详细点说,CNN中的卷积运算使用了两个重要的空间约束以提高特征的学习和提取能力:一是权重共享机制,使得卷积层所提取的特征具有平移不变性,模型更在乎对于目标任务决定性的特征是否存在,这样的特性使得其在回归、识别等任务中可以产生极好的作用。二是卷积算子,卷积的特征图具有局部敏感性,这使得其具有强大的特征提取能力,但随之而来的弊端是每次卷积操作只提取到局部信息。CNN的归纳偏差使得它缺乏对输入数据的整体把握,难以提取全局数据之间的长距离特征。以上特点成为CNN应用于速度反演任务时的潜在的问题。

具体来说,首先,空间不对应性是深度网络在反演任务中遇到的问题之一。

如图 5.2 所示, 地震记录和速度模型之间存在着极弱的空间对应关系, 即地震剖面上存在反射波的位置和在速度模型的存在对应界面的位置不具有空间对应性。而卷积操作被认为是位置强相关, 即卷积之后的特征图与原来的目标有着极大的位置关联。这一点使得卷积神经网络在用于速度反演这种空间不对应任务极大受限。第二, 地震记录中波场的衰减特性也给在整个空间维度上共享同一权重的 CNN 结构带来巨大挑战, 使得 CNN 对速度模型中的深层精细结构难以恢复。第三, 面对反演任务中的非唯一解问题, 依赖局部信息的 CNN 存在天然弱势。对于复杂的包含各种地下结构的速度模型来说, 接收器从一个源接收的反射信号, 引起反射的对应界面是不确定的。由于卷积的固有性质, 一个潜在限制是速度模型重建只能依靠在局部视图中学习的特征。然而, 当执行速度模型的重建时, 获得整个数据中的全局感知是很重要的, 这能够应对非唯一解的挑战。CNN 也可以通过不断地堆积卷积层来完成从局部信息到全局信息的提取, 研究者们往往通过扩大卷积核、增加通道数等手段来捕捉远程特征, 但是这种方式往往会使模型陷入训练崩溃而导致模型无法深入训练。

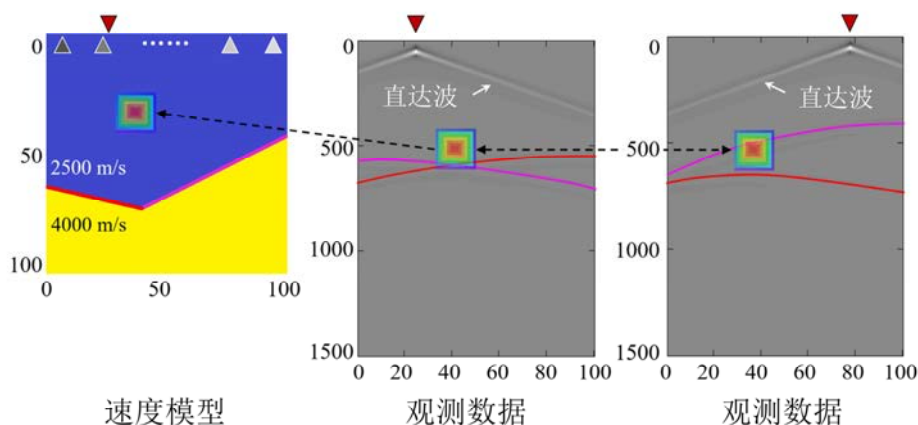


图 5.2 反演的空间不对应性示意图^[88]

CNN 的这一类问题也引起了研究者的广泛关注, Transformer 就是在此基础上提出的一种模型。与 CNN 专注于学习局部移位不变特征并通过堆叠卷积层逐步扩展感受野不同, Transformer 模型通过自注意力运算在更大的视图中提取特征。作为一种新型的深度学习模型, 其突出特点是每个特征学习层中的隐藏单元都能够接收输入的更多上下文信息。这种特性使得 Transformer 更加有利于

速度模型的重建,因为在更大的视图中学到的特征对于精确恢复速度模型更为重要。为了应对 CNN 反演任务中的多种弊端,本文中提出一种新的思路,提出了 Seismic Vectomy Inversion Transformer (SVIT) 模型来解决反演问题。该模型是一种基于自注意力运算的 Transformer 编码器-解码器架构,通过自注意力机制在多尺度分辨率下的特征图中广泛提取输入道中的特征,并在此基础上推理速度模型。SVIT 的优势在于利用自注意力的方式来捕获远程特征,这些远距离的特征对于重建速度的不同区域更为重要,从而使得反演网络模型泛化能力大大加强。

本章中对 SVIT 的训练细节进行了详细的辩证讨论,包括损失函数的选择以及带来的影响、各类超参数的设置等。为了解决 Transformer 计算效率低的问题,本研究中采用窗口自注意力机制和带卷积的前馈网络相结合的方案,以提高计算效率,并借助这种设计取得了较好的效果。同时,本章也展示了制作训练集的方案,并通过使用这些数据对 SVIT 实施训练,接下来在提出的测试数据集上进行了实验。从结果可以看出,相比于传统反演方法 FWI 和基于 CNN 的深度学习反演方法,SVIT 具备有更好的效果,其反演结果无论在速度值、地下结构、地质界面等方面更加符合真实值。同时,本研究中也对方法的抗噪性、泛化性等做了论证,并还对深度学习反演的未来前景做了探讨。深度学习反演拥有很快的速度,且可以实现和传统计算密集型反演方法相竞争的效果,是一种不可忽视的反演方法。此外,深度学习方法估计的速度模型也可以用作初始模型,进而继续使用迭代类的反演方法产生更高分辨率的模型,这些都有可能成为未来潜在的发展方向。

5.2 地震速度模型反演 Transformer 算法

本节中从反演问题定义、数据集的构建、SVIT 的网络结构以及基于 Transformer 的反演原理等方面详细描述了地震速度模型反演 Transformer 方法。

5.2.1 深度学习速度模型反演问题定义

地震勘探通常使用人工震源来激发地震波场,并使用地面上的地震检波器来记录地震波,以获得有关地下的信息。等密度二维声波方程表示为

$$\frac{\partial^2 u(x, y, t)}{\partial t^2} = v(x, y)^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + s(x, y, t) \quad \dots\dots\dots (5.1)$$

其中 v 表示波速, u 表示声波场, $s(x, y, t)$ 表示源的函数, (x, y) 表示空间位置, t 表示时间。通过设置检波器并记录各点处的波场, 可以获得观测数据为 d 。地球物理学家在观测数据和速度模型之间建立关系, 即 $d = F(v)$ 。对于速度模型反演, 可以从相应的观测数据 d 中推断出速度模型 v , 这是一个反问题。该过程一般描述如下:

$$v = F^{-1}(d) \quad \dots\dots\dots (5.2)$$

以经典反演算法 FWI 为例, 根据目标函数推导的梯度对速度模型进行迭代优化。目标函数通常被定义为速度模型正演的数据与观测数据之间的幅度和相位的误差。在迭代优化的过程下, 结果往往收敛到局部最小值。与之对应的是, 深度学习类反演方法将速度模型反演定义为一个回归问题。从观测数据到速度模型的非线性映射是通过不断优化具有巨量参数的端到端模型来学习的。深度神经网络训练的本质是损失函数数学期望的最小化, 传统地球物理反演的本质是目标函数的最小化, 这两个范式表达是相似的。不同之处在于, 传统的反演目标函数是对估计的速度模型或者观测数据的 L2 范数或其正则化改进函数, 且目标函数最小化的计算成本较高。对于基于深度学习的反演方法, 训练后的网络直接获得反演结果, 用于反演时没有迭代过程。对于基于深度学习的反演方法, 反演结果的好坏在很大程度上取决于完整的数据集和端到端网络设计。在下文中, 将详细描述这两个部分。

5.2.2 深度学习速度反演任务数据集构建

在深度学习领域, 数据集的质量和规模是决定模型性能的关键因素之一。尤其是在采用有监督学习方法时, 模型的训练过程高度依赖于数据的质量和多样性, 通过学习大量的样本数据, 模型能够抓住数据背后的复杂模式和关系, 从而在相关任务上表现出色。充足的训练数据不仅能帮助模型更好地理解数据, 还能增强其泛化能力, 使模型能够适应更多未见过的场景。在执行速度模型反演任务时, 本研究遵循有监督学习的范式, 所以大量带有标签的数据是训练模型不可缺少的

一部分。受数据驱动方法的特性，深度学习方法的性能受到训练时使用的数据影响。一般来说，合理而充分的速度模型设计可以帮助训练好的网络应用于更多的现实情况。为训练一个更加高效而准确的网络，需要创建一个完备的大型数据集。尤其对于 Transformer 来说，大数据集对于其训练不可或缺。相比 CNN，Transformer 对训练集数量的要求更高，因为 Transformer 模型更不容易收敛。本节中，将详细介绍准备数据集的过程。

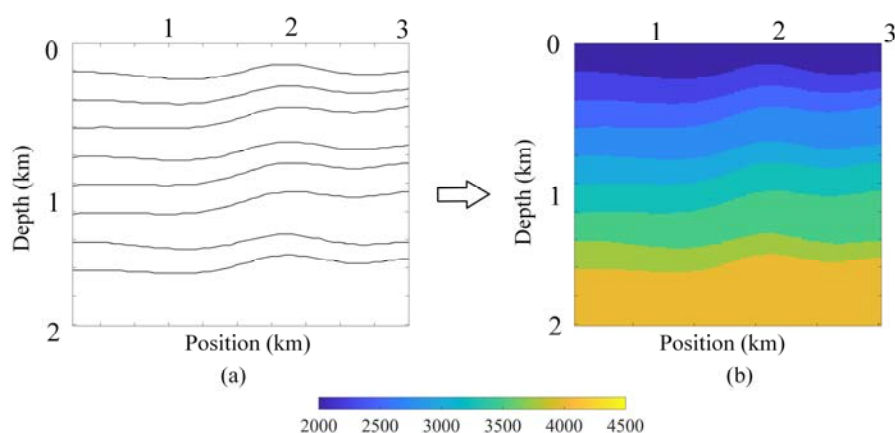


图 5.3 生成界面曲线和相应的致密层速度模型的步骤示意图
(a) 随机生成界面曲线，(b) 填充成为对应速度模型

训练集的设计主要以考虑致密层为主，为了力求在有限的深度和偏移距下，保证各层的连续性和可变性，尽可能多地添加地下介质的数量，设置每个模型有 5-12 层作为背景速度。采用如下方式去生成致密层速度模型：

(1) 生成多条曲线作为模型介质间的界面

使用一个由三角函数和线性函数组成的函数，用于生成连续的波动曲线来模拟地层界面。方程如下：

$$c_k = d_k + r_k \times x + \sum_{i=1}^3 a_i \left(\sin \left(\frac{x}{2\pi T_i} + \theta_i \right) \right)^i, i = 1, 2, 3 \dots \dots \dots (5.3)$$

其中， d_k 和 r_k 控制第 k 个界面的深度和倾斜程度。在创建速度模型的步骤中，使 d_k 的值增加，并且每个界面的 r_k 都有一个稳定的取值空间， a_i 、 T_i 和 θ_i 是不同三角方程的参数，它们分别控制三角函数的振幅、周期和相位。设置 $a_1 > a_2 > a_3$ 和 $T_1 < T_2 < T_3$ 。 θ_i 参数是随机设置的。通过这种方式，变量 a_1 成为曲线的主导部分。

为了使两个相邻地层界面之间的走向趋势相似,与前一个界面曲线相比,调整了每个界面曲线的控制参数值,包括 r_k 、 a_i 、 T_i 和 θ_i 。

(2) 为了削弱曲线的平滑度,增加其真实性,对这些函数进行离散化后重新连接形成相对更加曲折的新界面曲线。

(3) 确定好界面后,自上而下设置各层的纵波速度值,取值范围为(1500, 4500),根据实际情况下地球介质的固结效应,各层速度值设定随层数增加而增加。本研究中设置介质的速度随地层深度均匀增加,整个过程图示如图5.3所示。

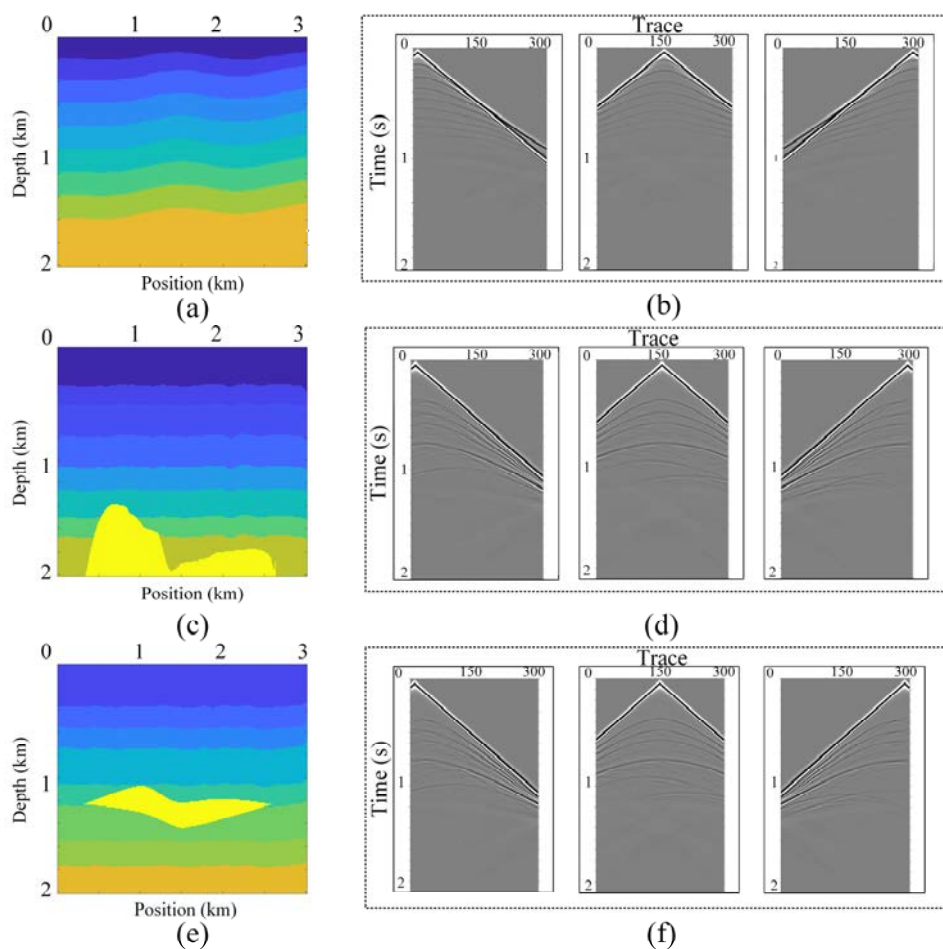


图 5.4 训练集中具有代表性的速度模型及其相应的模拟数据 (a) 致密层速度模型, (b) 致密层速度模型的观测记录, (c) 底层盐体速度模型, (d) 底层盐体速度模型的观测记录, (e) 中间层盐体速度模型, (f) 中间层盐体速度模型的观测记录

同时,为了保证数据的完备性,还有一部分盐丘模型被加入到数据集中,这些盐丘数据来自开源数据库^[86]。在这部分中,一个任意形状和位置的盐丘被嵌入到每个模型中。数据集中共有 2400 个不同的速度模型和对应的合成数据对,其

中包含 1200 个按照自定义规则生成的数据，1200 个来自盐丘数据。图 5.4 中展示了三种类型的速度模型，其中图 5.4 (a) 是一种典型的致密层速度模型，图 5.4 (c) 和图 5.4 (e) 则是两种盐丘速度模型。所有用于反演的速度模型的大小为 301×201 。

在准备完速度模型后，对应生成每个速度模型的炮点记录，通过求解声波方程得到每个速度模型对应的观测记录，声波方程的公式在前述章节已经介绍，在此不再赘述。使用时间域的交错网格有限差分方法求解方程，在空间方向上使用二阶，在时间方向上使用八阶，对于每个速度模型，均匀放置 29 个震源以模拟炮点道集。检波器阵列有 301 个接收器组成，它们都以均匀的空间间隔均匀放置。正演建模的各个参数如表 5.1 所示，使用完全吸收层边界条件^[146]来吸收左右和底部边缘的反射。图 5.4 的右侧分别展示了与左侧的速度模型对应的 3 个地震记录。

表 5.1 正演参数设置

参数	具体设置
震源数量	30
接收器数量	300
速度模型大小	(301, 201)
网格大小	10 m $\times$ 10 m
采样间隔	1×10^{-3} s
子波类型	Ricker
子波频率	25 Hz
边界条件	PML
最大时间	2 s

其中，记录中第一个明显的波为直达波，即震源直接传播至检波器的波，下方各类反射波代表着个各层地下结构的信息，但是振幅要小于直达波，随着时间的推移，地震记录中的波的幅度会由于弥散而逐渐减小。有研究者指出，这种波

场衰减的性质对于在整个空间维度上共享同一权重的 CNN 类反演架构带来巨大挑战，也是 CNN 类方法在反演任务中，对速度模型中的深层结构难以恢复的主要原因^[88]。

接下来按照表 5.2 中的比例将所有速度模型-观测数据数据对分成训练集、验证集和测试集，它们被分别用于训练网络、验证训练过程中的网络性能以保存最佳参数、测试网络精度。训练集、验证集和测试集是彼此独立的。在处理时，为缩减数据量，将单次地震数据的维度下采样到 200×301 。同时对每个观测记录和速度模型都进行值归一化处理。

表 5.2 本实验所用数据集分布

数据集类型	致密层速度模型数量	盐体速度模型数量
训练集	1200	1200
验证集	300	300
测试集	100	100

5.2.3 速度模型反演网络 SVIT

本研究中使用 SVIT 来进行从观测记录到速度模型的反演任务，如图 5.5 所示，整个模型共分为混合震源编码器、特征编码器、特征转换器、特征解码器、速度模型重建器五个部分。使用混合震源编码器，用于从多炮点记录中提取底层特征，接下来，特征被切割为多个“键”，在特征编码器阶段由多个编码器处理，每个编码器包含多个局部窗口 Transformer 模块与一个下采样层，这样使得各个编码器在不同的分辨率下进行编码。其中，局部窗口自注意力模块利用自注意力机制来捕获远程依赖关系。解码器与编码器一样包含多个阶段，每个阶段包含一个上采样层与多个 Transformer 模块，特征解码器和特征编码器一一对称。特征编码器和特征解码器在接下来会详细介绍。其中，在特征编码器和特征解码器之间加入了一个特征转换器阶段，用于获取全局特征（此时执行自注意力运算的窗口大小与特征大小相同），从而进行从观测记录特征到速度模型特征的转换。在解码阶段结束后，使用速度模型重建器来重建速度模型。各部分的详细介绍如下。

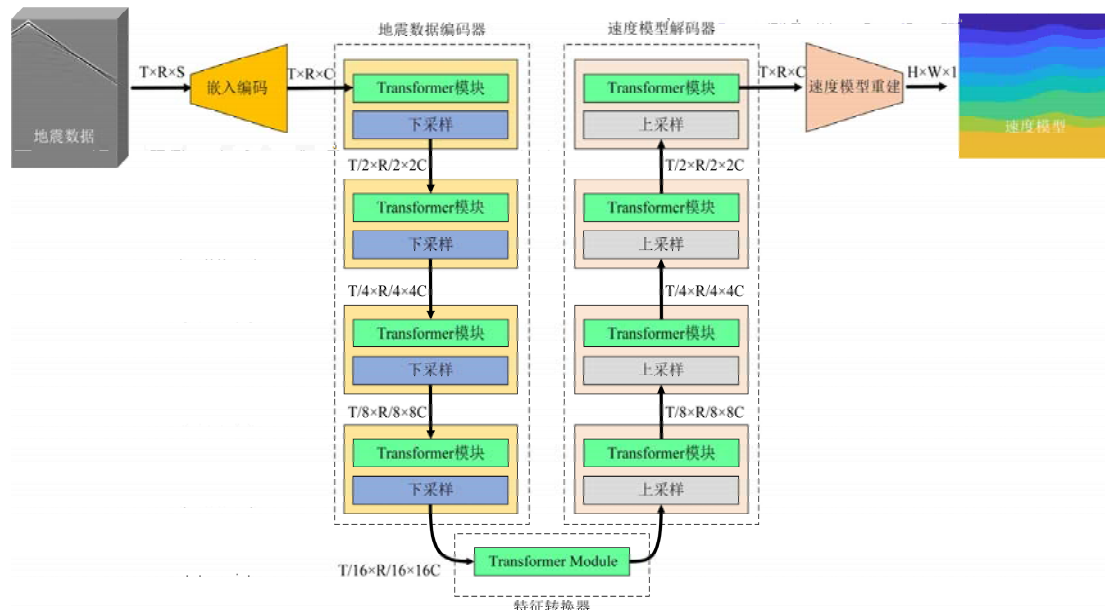


图 5.5 基于 Transformer 的速度反演方法的总体结构

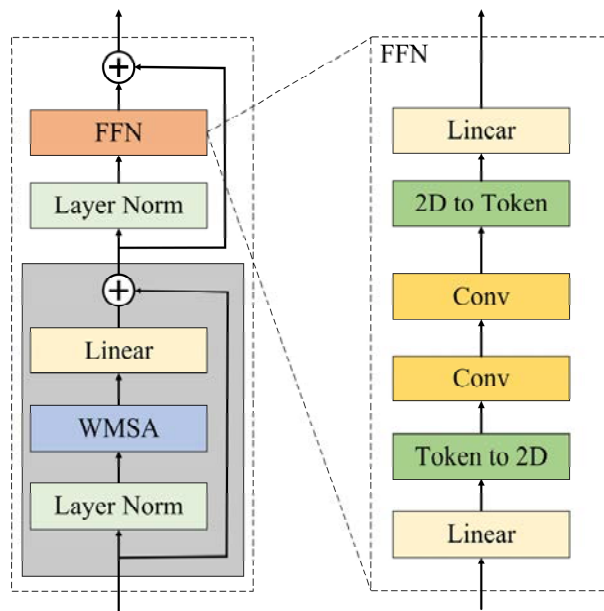


图 5.6 Transformer 模块的示意图

(1) 混合震源数据编码器

混合震源数据编码器采用一个简单网络构建, 假设输入的共炮点记录集表示为 $I \in \mathbb{R}^{T \times R \times S}$, 其中 T 、 R 、 S 分别代表输入地震数据在时间轴方向的长度、检波器的数量、震源的数量, 经过混合震源编码器后生成特征 $X_0 \in \mathbb{R}^{T \times R \times C}$, 在本研究中, C 默认设置为 64。

表 5.3 网络中各层的具体说明

层名称	描述	功能
Linear	线性变换层	$y = xA^T + b$
Conv	卷积层	$y = W * x + b$
BN	批归一化层	对批中所有的样本进行归一化
Layer Norm	层归一化层	对层中所有的样本进行归一化
ReLU	非线性激活层	$y = \max(0, x)$
Leaky ReLU	弱非线性激活层	$y = \max(0, x) + leak \times \min(0, x)$
GELU	GELU 激活函数	$y = xP(X \leq x), x \sim N(0, 1)$
SoftMax	Softmax 激活函数	$SoftMax(i) = \frac{e^i}{\sum_{j=1}^n e^j}$
Token to 2D	特征一维转二维层	一维特征转为二维特征
2D to Token	特征二维转一维层	二维特征转为一维特征

(2) 特征编码器

特征编码器是用于对混合震源编码器生成的特征进行进一步特征提取的部分，在完成上一步的特征编码后，特征图 X_0 会随后进入编码器阶段。

在特征编码器中，每个阶段都包含多个 Transformer 模块和一个下采样层，允许各个编码器在不同的尺度下对特征进行编码^[147]。Transformer 模块利用自注意力机制来捕获长程依赖关系，此外也利用卷积算子来捕获有用的局部上下文，其中包含两个核心设计：基于移位窗口的多头自注意力机制（Shifted Windows Multi-Head Self-Attention, SW-MSA）和前馈网络（FeedForward Network, FFN）。它们将在本节最后进行详细描述。需要注意的是，与由 MSA 和 W-MSA 交替组成的原始 Swin Transformer 不同，本研究中的 SVIT 模型使用了完全由 SW-MSA 组成的结构，这允许所有窗口之间的完全交互。做出这一设计选择是为了确保方法在不同尺度上充分利用自注意力的优势。在下采样层中，首先将展平的特征重塑为二维空间特征图。然后，对特征图进行下采样，使用步长为 2 的 4×4 卷积提

取特征，将特征的大小减小一倍，并将通道加倍。具体而言，给定第 $(l-1)$ 编码器阶段的特征，编码器的计算过程由 Transformer 模块和下采样层依次执行。Transformer 模块中的操作可以表示为

$$X_l' = \text{SWMSA}(\text{LN}(X_{l-1})) + X_{l-1} \dots\dots\dots (5.4)$$

$$X_l'' = \text{FFN}(\text{LN}(X_l')) + X_l' \dots\dots\dots (5.5)$$

下采样层中的操作可以表示为

$$X_l = \text{DW}(X_l'') \dots\dots\dots (5.6)$$

其中 l 表示解码器的次序，SWMSA 表示基于窗口的多头自注意模块，FFN 表示前馈前向网络，DW 表示下采样层。LN 表示层归一化， X_l' 、 X_l'' 和 X_l 分别是 SWMSA 模块、FFN 模块和 DW 模块的输出。

(3) 特征转换器

在特征转换器阶段，仅使用一次 Transformer 模块和一次前馈网络进行一次处理。此时执行自注意力运算的窗口大小与特征图的大小是相同的。受益于层次化特征的结构，窗口自注意力机制可以在此时的特征图中捕获全局的信息，实现从共炮点数据特征到速度模型特征的转换。

(4) 特征解码器

特征解码器部分采用了与特征编码器类似的架构，其接收特征转换器的输出作为输入，特征解码器部分与特征编码器部分一一对应，呈现“U 型”结构，并将下采样替换为上采样。采用步长为 2 的转置卷积用于特征上采样，这样使特征图的大小加倍，并减少一半的特征通道数。解码器接收特征转换器的输出特征作为输入。在这个阶段，特征在上采样层中进行处理，然后输入到其中的 Transformer 模块中。在特征解码器阶段中，速度模型的特征被连续地重构。具体来说，给定第 $(l-1)$ 个解码器阶段的特征，解码器的计算过程描述如下：

$$X_l' = \text{UP}(X_{l-1}) \dots\dots\dots (5.7)$$

$$X_l'' = \text{SWMSA}(\text{LN}(X_l')) + X_l' \dots\dots\dots (5.8)$$

$$X_l = FFN(LN(X_l')) + X_l'' \dots\dots\dots (5.9)$$

其中 l 表示解码器的次序, UP 表示上采样层, $SWMSA$ 表示基于窗口的多头自注意力模块, FFN 表示前馈网络。 LN 表示层归一化, X_l' 、 X_l'' 和 X_l 分别是上采样层、 $SWMSA$ 模块和 FFN 模块的输出。

(5) 速度模型重建器

为了使用前述阶段编解码后的特征来重建速度模型, 搭建了一个速度模型重建器用来重建速度模型, 其使用多个卷积层来构建, 在最后一层, 特征图截断为与期望的速度模型相同的大小, 且通道被缩减为 1, 从而将特征图映射到精确的速度模型上。

(6) Transformer 模块

以上是反演网络的整体组成, 接下来, 将详细描述 Transformer 模块中的重要基本组件。Transformer 模块利用自注意力机制来捕获长程依赖关系。使用两个部分构建该模块: 基于移位窗口的多头自注意力和前馈网络, 这两个部分的细节如下所述。

在反演任务中, 需要同时考虑局部特征和全局特征。如果在所有键之间进行全局自注意力运算, 那么将产生巨大的计算消耗。作为权衡, 本研究中采用非重叠局部窗口内进行自注意力计算的策略, 即基于窗口的多头自注意力, 可以有效地降低计算成本。移位窗口方案通过将自注意力计算限制在不重叠的局部窗口, 同时允许跨窗口计算, 从而减少计算量。此外, 本研究中在 SVIT 中引入了移位窗口自注意力计算, 以解决窗口之间缺乏信息交互的问题。在两个相邻的 Transformer 块中, 后者对窗口向左上方向执行步长为 2 的循环移位^[126]。在移位之后, 一个窗口可能由几个在特征图中不相邻的子窗口组成, 因此采用了掩码机制来将自注意力计算限制在这些子窗口内。通过循环移位, 此时窗口的数量与不移位时常规窗口的数量保持相同。

在将前述特征送入到 Transformer 结构之前, 需要将输入特征拆分成块, 每个块称之为“键”。具体来讲, 输入特征 $X_l \in \mathbb{R}^{T \times R \times C}$ 将被拆分为一系列块 $X^i \in \mathbb{R}^{M^2 \times C}$, 并将它们展平且进行嵌入编码, 其中 C 表示特征块的数量, M 表示特征块尺寸。本研究中设置 M 大小为 5。然后, 对得到的特征进行自注意力计

算。自注意力机制允许每个键之间相互交互，从而找到需要更多关注的输入。自注意力头是自注意力计算的执行对象，假设自注意力头编号为 h_G ，自注意力头的维度为 $d_h = C / h_G$ 。由于更深的计算阶段往往需要包含更多的头，以捕捉更多的远程特征，因此在特征编码器阶段中，自注意力头的数量分别设置为 1、2、4 和 8。在特征转换器阶段，自注意力头的数量被设置为 16，以广泛地提取远程特征依赖关系。在特征解码器阶段中，头的数量分别设置为 8、4、2 和 1。接下来，在每个窗口特征上执行自注意力。第 h 个头的计算过程可以描述如下：

$$X = \{X^1, X^2, \dots, X^N\}, N = \frac{TR}{M^2} \dots\dots\dots (5.10)$$

$$Y_h^i = \text{Attention}(X^i W_h^Q, X^i W_h^K, X^i W_h^V), i = 1, \dots, N \dots\dots\dots (5.11)$$

$$\hat{Y}_h = \{Y_h^1, Y_h^2, \dots, Y_h^M\} \dots\dots\dots (5.12)$$

其中 X 表示该组键的集合， $\hat{Y}_h$ 是第 h 个头的输出， $X^i W_h^Q, X^i W_h^K, X^i W_h^V \in \mathbb{R}^{C \times d_h}$ 分别表示第 h 个头“查询向量”、“关键向量”和“数值向量”的投影矩阵。自注意力计算的输出可以被视为键之间的相关系数。在注意力模块中引入了相对位置编码^[138]，以保证每个键的相对位置不会变化，自注意力计算的过程可以表示成如下形式：

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d_h}} + B\right)V \dots\dots\dots (5.13)$$

其中 Q, K, V 是自注意力计算的输入， B 是相对位置偏移^[138]。

由于速度模型中的精细结构依赖于要恢复的局部特征，因此局部上下文信息对于准确预测速度模型的精细结构至关重要。许多的研究工作表明，传统的 Transformer 在捕捉局部上下文信息方面存在局限性^[61]，对于地震记录，空间中的邻域采样点是强相关的，CNN 可以通过局部感知和共享权重来捕捉这些局部特征，而传统的 Transformer 缺乏这些内在特性。因此，利用卷积加入到前馈网络中来捕获一定的局部上下文信息，从而构成了一个新的前馈网络结构。基于卷积的前馈网络结构如图 5.6 所示，在该结构中，采用线性投影层提升每个键的维度，然后将一维的键重置为二维特征并采用两层卷积捕获其中的局部信息，然后，

将上述特征展平，并通过线性层缩小尺寸以重新匹配输入时的大小。同时，将 GELU 激活函数^[135]应用于卷积后的特征图以增加其非线性。GELU 的具体说明可见表 5.3。

以上介绍的移位窗口的多头自注意力模块和前馈网络模块为 Transformer 模块的两个重要组成部分，它们也是 SVIT 能够实现反演任务的重要依靠。

5.2.4 速度模型反演网络 SVIT 方法原理

本研究中将反演任务定义为一个回归问题。具有巨量参数的 SVIT 被训练直接学习从地震数据到地震速度模型的映射，如公式（5.14）所示，

$$v = SVIT(d, \Theta) \dots\dots\dots (5.14)$$

其中 $SVIT$ 表示地震速度反演 Transformer 拟合的函数， Θ 表示其参数。将整个步骤分为两个阶段：训练阶段和预测阶段。在训练阶段，网络在监督学习下学习相应的映射，需要使用由观测数据和对应的速度模型一一对应组成的配对数据集。设计合理的损失函数是优化监督学习算法的先决条件。损失函数用于计算速度模型在训练集上的真实值和预测值的偏差。不同的参数一般对应于损失函数的不同值，数据驱动反演问题的目标是找到一组使损失函数最小化的参数。一般来说，损失函数决定了模型优化的目标，甚至决定了优化的空间。对于损失函数的选择，本研究中评估了三个损失函数的性能：L1 损失、L2 损失和 Charbonnier 损失^[148]。假设 $\hat{v}$ 是预测速度， v 是真实值，三个损失函数的定义如下：

$$L_1 = \sum |\hat{v} - v| \dots\dots\dots (5.15)$$

$$L_2 = \sum \|\hat{v} - v\|_2^2 \dots\dots\dots (5.16)$$

$$L_{Charbonnier} = \sum \sqrt{\|\hat{v} - v\|_2^2 + \epsilon^2} \dots\dots\dots (5.17)$$

其中 Σ 表示求和符号， ϵ 在实验中是常数。一般来说，对于回归问题，L1 损失和 L2 损失是最常用的指标，其在速度反演研究中也广泛被使用，然而这两个指标倾向于去单独处理数据的每个点，使得网络难以捕捉地下介质的局部结构和细节，而局部结构和细节又是重建地震速度模型时需要考虑的重要因素。

利用均方误差 (Mean-square error, MSE) 和结构相似性 (Structural similarity, SSIM) 两个指标, 测量了验证集上不同损失训练的模型的性能, 具体结果如表 5.4 所示。可以观察到, 用 L2 损失训练的模型的结果往往有着较低的 MSE, 但用 Charbonnier 损失训练的模型的结果将倾向于较高的 SSIM, MSE 则略低于 L2 损失训练的模型。对于整个速度模型来说, 一些局部结构和细节可以被视为异常值。由于常数较小, Charbonnier 损失可以更好地处理异常值, 并提供更平滑的梯度。它具有 L1 损失抗干扰性高和 L2 损失对异常值敏感的优点。因此本研究最终选择了 Charbonnier 损失作为最终使用的损失函数。

表 5.4 验证集上不同损失训练的模型的性能对比。

量化指标	Charbonnier 损失	L1 损失	L2 损失
MSE	0.0149	0.0189	0.0147
SSIM	0.8637	0.7965	0.8587

表 5.5 网络超参数设置

超参数	具体设置值
批大小	32
迭代周期数	200
学习率设置范围	$[1 \times 10^{-4}, 1 \times 10^{-3}]$
优化方法	Adam

在网络训练时, 通过计算预测模型和实际模型之间的损失, 以导出梯度并更新网络参数。在训练阶段使用反向传播方法和梯度下降法来优化网络。具体而言, 使用的是 Adam 算法^[98]。Adam 不同于经典的随机梯度下降方法。经典的随机梯度下降方法对所有参数更新都保持单一的学习率。相反, Adam 根据梯度的一阶动量和二阶动量计算不同参数的自适应学习率。与经典的随机梯度下降算法相比, Adam 算法可以快速逃离鞍点, 并使模型更快地收敛。在多次周期迭代之后, 损失将收敛到足够小的值, 届时便可以确定 Transformer 的参数。通过输入大量的

训练数据来迭代更新网络参数，并沿着梯度方向不断减小损失函数的值，最终得到一组使损失函数值最小的网络参数。在这一状态下，可以认为模型已经收敛到具有可接受的损失函数值的状态，即模型已经学习到了期望的映射。在测试阶段，网络可以在输入地震记录时可以直接预测速度模型。此外，网络超参数的设置是实现准确预测的重要组成部分。超参数是基于现有经验人工配置的设置，如学习率、批处理大小和迭代次数等。通常，为网络选择一组最优的超参数可以充分提高学习的性能和有效性，避免训练不足。本研究在连续实验的基础上，选择了合适的网络超参数来提高网络性能，具体训练细节如表 5.5 所示。

5.3 算法性能验证

本节通过实验验证所提的 SVIT 算法对于不同速度模型的反演性能。具体内容构建的实验数据集，相关的实验设置，实验结果和对结果的分析，以及与本领域先进的反演算法性能对比。

5.3.1 实验设置

在本实验中，对本研究中提出的基于 Tranformer 的反演方法即 SVIT 和基于 CNN 的反演方法——全卷积速度反演网络（Fully Convolutional Neural network for Velocity-Model Building, FCNVMB）^[86]以及常规反演方法——全波形反演进行了比较。FWI、FCNVMB 都是当前相对前沿的速度反演技术。FWI 作为一种传统的反演技术，以其能够利用地震波形中的全部信息进行速度模型反演而备受推崇，它通过最小化观测数据与模拟数据之间的差异，迭代更新地下速度模型，从而达到高精度的反演效果。然而，FWI 在处理高维数据时计算成本极高，且容易陷入局部最优解，特别是在初始模型选取不佳或数据噪声较大时的情况下。而基于深度学习的反演方法提出了一种新的思路，通过学习地震数据与地下速度模型之间的映射关系，直接预测地下结构的速度分布。这种方法能够有效降低计算成本，并且在处理复杂地质结构时显示出较好的鲁棒性。本实验中评估了各种反演方法的结果，包括对各种方法反演结果进行了视觉比较和定量分析。用于评估反演性能进行比较的常用指标包括平均绝对误差（Mean Absolute Error, MAE）、

均方误差 (Mean Squared Error, MSE)、结构相似性指数 (Structural Similarity, SSIM) 和平均结构相似性 (Mean Structural Similarity, MSSIM)^[149]。这些指标能够从不同角度量化反演结果与真实地下结构之间的差异, 其中 MAE 和 MSE 关注于预测值与真实值之间的绝对差异, SSIM 和 MSSIM 则更加注重于评估反演结果的结构相似度。实验在配备 Inter Xeon CPU E5-2620 和 4 个 Nvidia V100 GPU 的服务器上运行, 实验运行在 Python 环境中, 借助 PyTorch 深度学习框架^[141]实现。对比方法包括 FWI 和 FCNVMB, 其中, 后者是具有代表性的基于 CNN 的反演算法。所有方法的实验数据通过时域正演建模生成, 以模拟实际地震探测过程中的地震数据。对于 FCNVMB, 实验遵循基于其原始论文^[86]中默认的设置, 为了重点比较架构, 本实验中使用了 FCNVMB 的通用版本, 并基于提出的数据集进行了重新训练。对于 FWI, 在实验中采用了时域反演策略和基于伴随状态法的梯度下降方法。FWI 方法使用的观测数据与输入 SVIT 和 FCNVMB 的地震数据相同。值得一提的是, 初始模型的选择决定了 FWI 方法的有效性。本实验中采用高斯平滑函数平滑后的速度模型作为 FWI 初始速度模型, 以保证 FWI 的收敛结果。震源的频率设定为 20 Hz。对于每个速度模型, 使用 FWI 方法进行了 50 次迭代。因为在 GPU 上编程难度较高, FWI 的数值实验是在机器上利用 CPU 进行的。

5.3.2 实验结果与分析

(1) 整体性能

尽管通过函数模拟可以生成各种多样的速度模型, 但是实际情况下的速度模型依旧有着很高的复杂性, 因此无法合成所有的情况进行训练。但是, 良好的反演网络应该具备很好地适应反演任务中各种情景的能力。为此, 本研究中着重进行了一系列实验以验证模型的泛化能力。通过在独立于训练集以外的速度模型上开展测试, 各种方法的预测结果如图 5.7 所示。其中, 选择了几组致密层模型、盐体模型的结果进行可视化对比。所有的这些子图颜色条刻度相同, 速度范围值为 2000 到 4500 m/s。

接下来，在三组具有代表性的模型下进行了比较分析，分别是图 5.7 (a) 中的致密层模型以及图 5.7 (f) 和图 5.7 (k) 中的盐丘模型。对于致密层模型，图 5.7 (b) 到 (d) 分别为 SVIT、FCNVMB 以及 FWI 的结果。对于 FWI 方法，建立一个合适的初始速度模型影响着其最终反演结果，可以避免 FWI 陷入局部极小值。而对比之下，作为一种深度学习反演方法，SVIT 可以在输入地震记录后直接推断速度模型，而无需迭代。因此，它不依赖于初始模型，便可以推断出每一层的准确速度值。同时，可以观察到，SVIT 实现了与最先进的基于 CNN 的反演方法 FCNVMB 相当的性能。这一结果提供了证据，验证了 Transformer 架构可以在速度反演任务中与卷积神经网络相媲美。对于盐体模型，可以观察到 FWI 很难为盐体模型恢复出精细的结构，如图 5.7 (i) 中所示。而 FCNVMB 虽然可以拟合每一层的速度，但对于恢复盐体模型的细节并不完美，比如在图 5.7 (h) 中，盐体的边缘出现了预测误差。相比之下，SVIT 对于各层速度值的预测和盐体的恢复都是准确的，如图 5.7 (g) 所示。从对中层和底层的盐体模型的恢

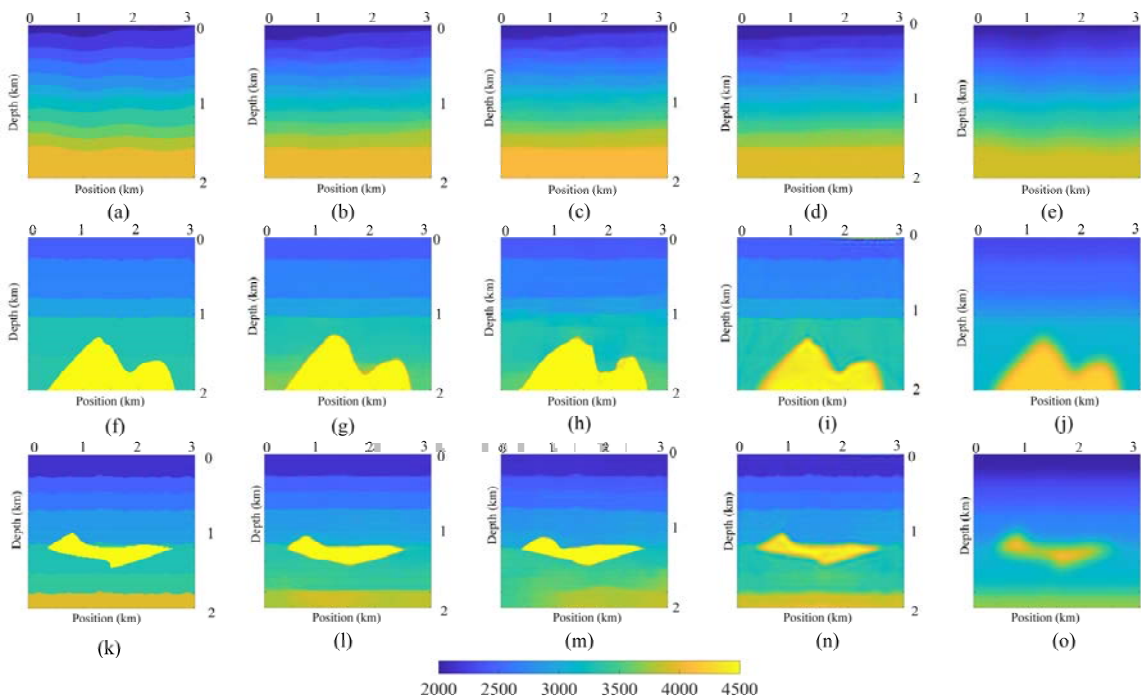


图 5.7 不同方法的反演结果对比 (a) 致密层速度模型真实值, (b-e) SVIT、FCNVMB 和 FWI 的相应预测结果及 FWI 的初始速度, (f) 底层盐体速度模型真实值, (g-j) SVIT、FCNVMB 和 FWI 的相应预测结果及 FWI 的初始速度, (k) 中层盐体速度模型真实值, (l-o) SVIT、FCNVMB 和 FWI 的相应预测结果及 FWI 的初始速度

复效果来看，尽管这三种方法都可以获得盐体的位置、波速和形状，但 SVIT 的结果最接近真实值。

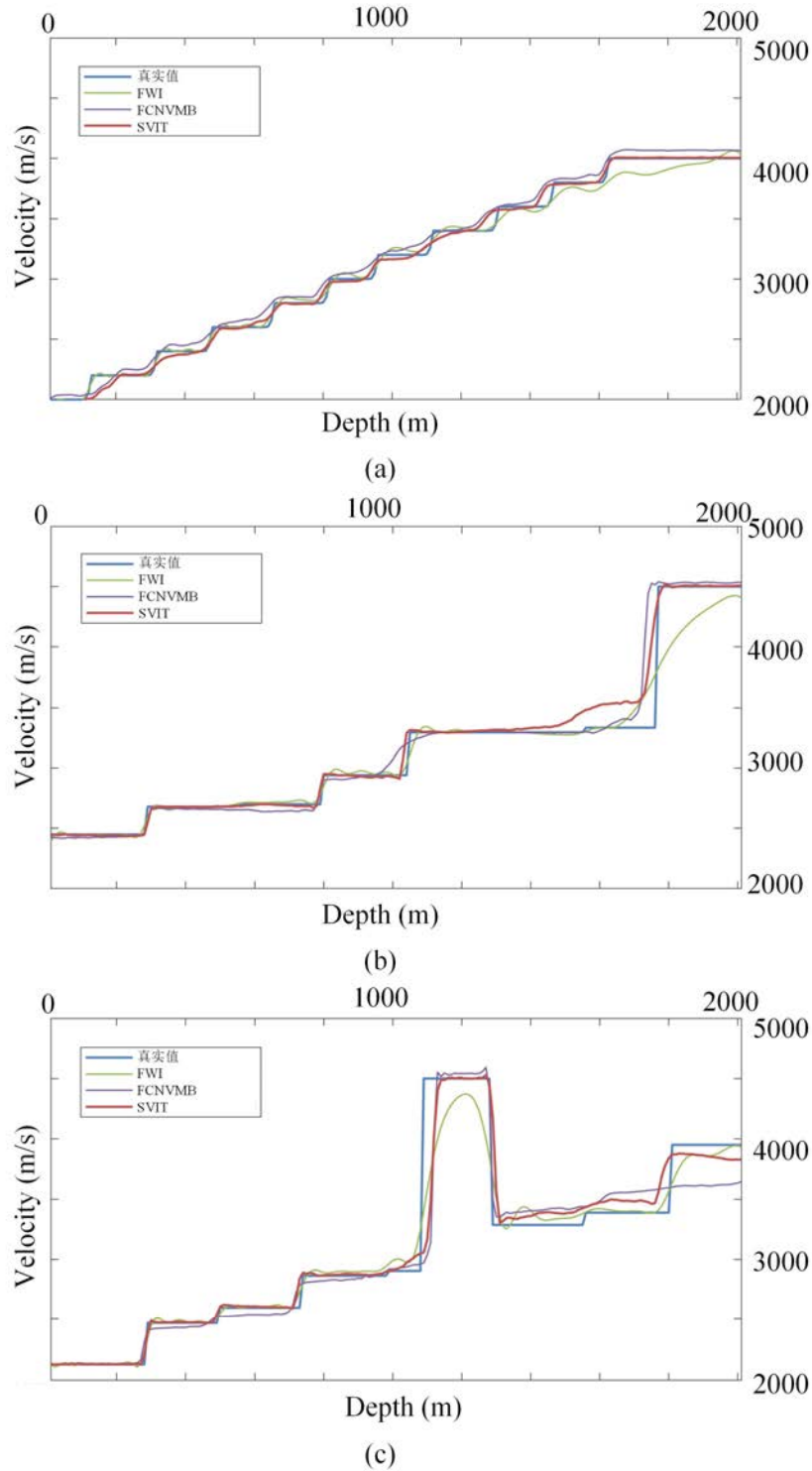


图 5.8 不同方法反演结果的垂直速度剖面对比 (a) 致密层模型, (b) 底层盐体层模型, (c) 中层盐体模型

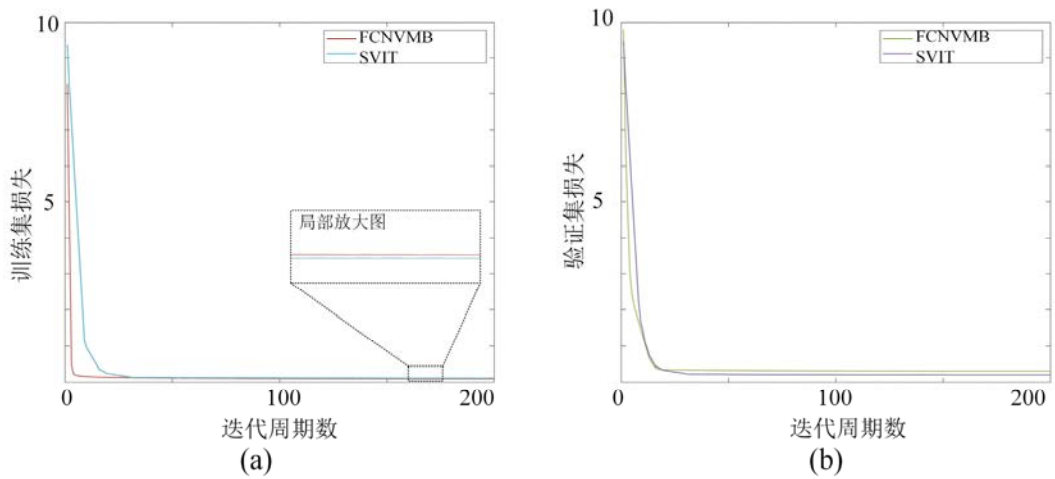


图 5.9 两种深度学习模型的训练损失和验证损失图 (a) 训练损失下降图, (b) 验证损失下降图

表 5.6 不同方法在测试集上反演结果的量化指标对比

量化指标	SVIT 方法	FCNVMB 方法	FWI 方法
MAE	0.0478	0.0682	0.0693
MSE	0.0181	0.0240	0.0189
SSIM	0.8851	0.8185	0.3803
MSSIM	0.9995	0.9994	0.9989

接下来，在图 5.8 中绘制了单道速度曲线，以细化各类方法的差异并进行对比。垂直速度剖面的位置选择在 620m 的水平位置处，在此图中，分别绘制了速度与深度剖面中的真实值（蓝色）、SVIT 预测值（红色）和 FCNVMB 的预测值（紫色）以及 FWI 方法的预测值（绿色）。从单道对比图中可以观察到，相对于其他两种方法，SVIT 的反演结果与真实值的重合度要更高一些。

在对地震数据反演的过程中，底层盐体的恢复较为关键，它直接关系到地下结构解析的准确性和油气资源勘探的成功率。从三种反演方法里对最底层的盐体的恢复可以推断出三种方法特性的不同，因为地震数据只能记录有限的时间，其

中来自较深的层的反射信息较少。另外,因为信号的衰减,底层的信号也比较微弱。这一点使得权重共享的 CNN 网络模型带来极大挑战。从实际效果可以看出,FCNVMB 对于盐体的恢复结果中结构细节往往有很多的误差。这是因为底层信号的强度不足以对速度模型的恢复作出贡献。相比之下,基于 Transformer 的反演方法 SVIT 展现出了对盐体的更好恢复能力。这得益于 Transformer 结构在处理数据时的长距离依赖捕获能力,使得 SVIT 能够有效利用深层微弱信号,从而在盐体等地质结构的反演上取得更好的结果。此外,Transformer 的自注意力机制能够自适应地调整网络对不同层级信号的关注度,这一点对于弱信号的识别和利用尤为重要。

接下来,计算了各种方法在测试集的致密层模型和盐体模型上的反演结果的统计数据,如表 5.6 所示。选择了四个指标包括 MAE, MSE, SSIM 和 MSSIM^[149],并对各种方法在测试集上的结果进行衡量并求得平均。一般来说,更好的反演网络意味着其结果应具有更小的 MAE, MSE 以及更大的 SSIM 和 MSSIM。从这些统计数据中可以观察到,SVIT 网络相比于 FWI 和 FCNVMB 拥有更好的表现。

与同样作为深度学习方法的 FCNVMB 的对比是体现算法优越性的重要部分。为了进一步验证基于 Transformer 反演模型对基于 CNN 反演模型的优势,本研究尝试分析了训练集和验证集相对于历次迭代周期的损失曲线。如图 5.9 所示,可以看出两种模型的损失函数在训练过程中都单调下降,并迅速收敛到一个稳定的点。这其中,验证集的波动会更大一些,因为验证集中可能包含有训练过程中未经遇到的复杂盐体或地质结构。从 SVIT 方法和 FCNVMB 的对比中可以看出,SVIT 在验证集上比 FCNVMB 更加稳定和可靠。图 5.9 (a) 中这些曲线的最小值被标记了出来,通过对比,可以看出 SVIT 网络在最终的收敛值上更具有优越性。

(2) 噪声适应能力及低频缺失适应能力

另外,测量反演方法对噪声的稳定性也是一个重要内容。为了进一步验证所提方法在真实条件中的能力,本研究中展示了地震观测数据受到随机噪声污染时各个方法的反演结果。其中,图 5.10 (b) 中的含噪声数据是通过添加零均值,方差为 5%的高斯噪声生成的,图 5.10 (c) 中的噪声则来自于实际勘探环境中。值得一提的是,对于两种深度学习模型的训练并没有添加噪声训练,在这里更多

还是希望测试噪声对于模型直接反演性能的影响。

图 5.10 中显示了在噪声输入下各个方法的预测结果，其中表明本研究所提出方法相比于 FWI 和 FCNVMB，拥有更出色的性能。同时，也可以观察到，和真实值相比，预测的值大多是接近真实值，但是略低于使用干净数据时反演网络的预测值，这可能是有噪声扰动引起的，毕竟在之前的训练过程中未曾考虑加入过噪声。在未来，本研究会尝试通过训练过程中加入噪声扰动来解决这个问题。整体来看，SVIT 在应对高斯及实际噪声时都具有最好的反演效果。

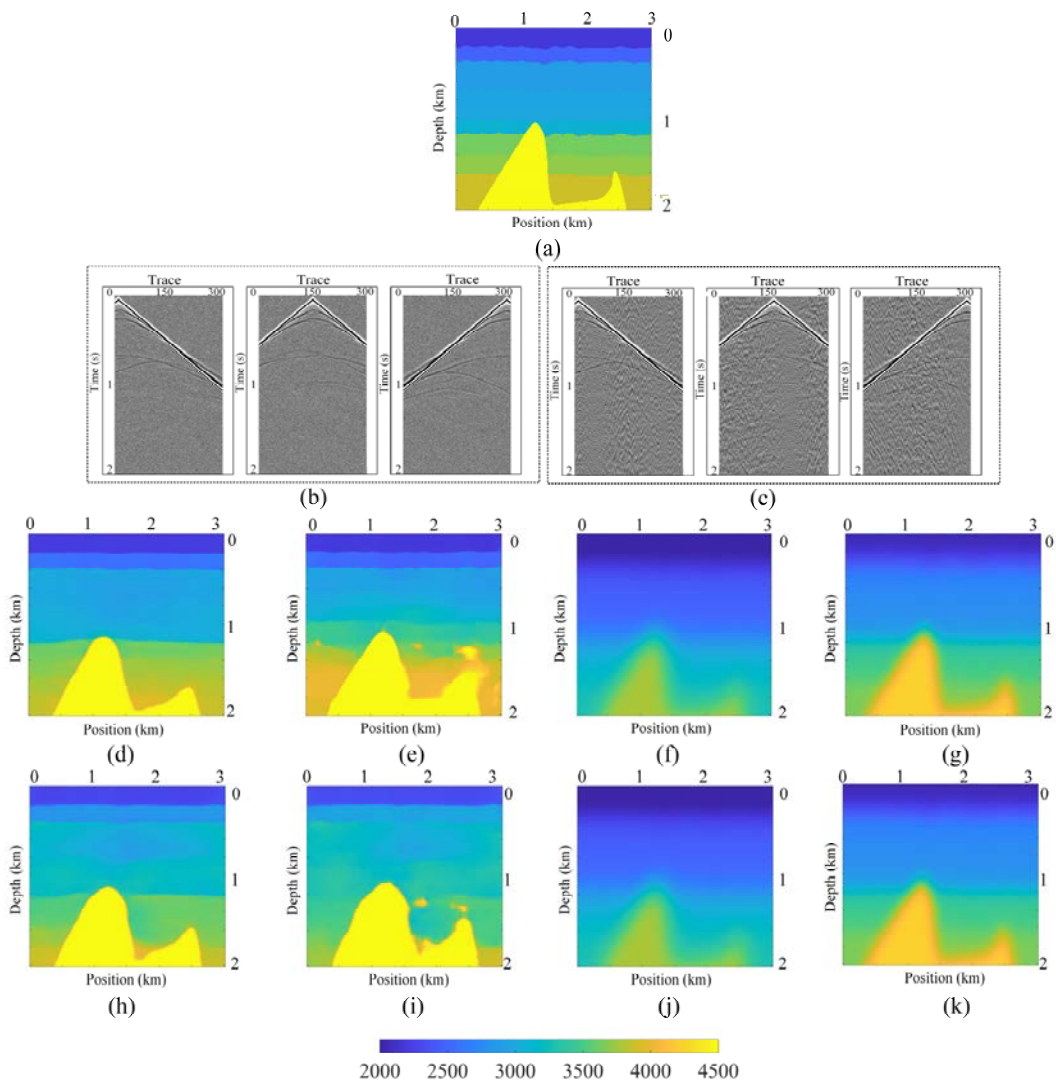


图 5.10 不同方法在观测记录有噪声的情况下的反演结果 (a) 真实值, (b) 加入高斯噪声的含噪观测数据, (c) 加入实际地震噪声的含噪观测数据, (d-g) 添加高斯噪声后 SVIT、FCNVMB 的结果、FWI 初始模型和 FWI 结果, (h-k) 添加实际噪声后 SVIT、FCNVMB 的结果、FWI 初始模型和 FWI 结果

此外，低频缺失是实际勘探记录中的常见情况，也是传统反演方法中遇到的重大问题。低频信息的提取是检验反演方法性能的一个重要指标。因此，本研究中使用带通滤波器来去除观测数据中低于 15 Hz 的低频分量，形成如图 5.11 (a) 所示的观测记录，并在低频缺失的情况下测试了几种方法，结果一并列入图 5.11 中。

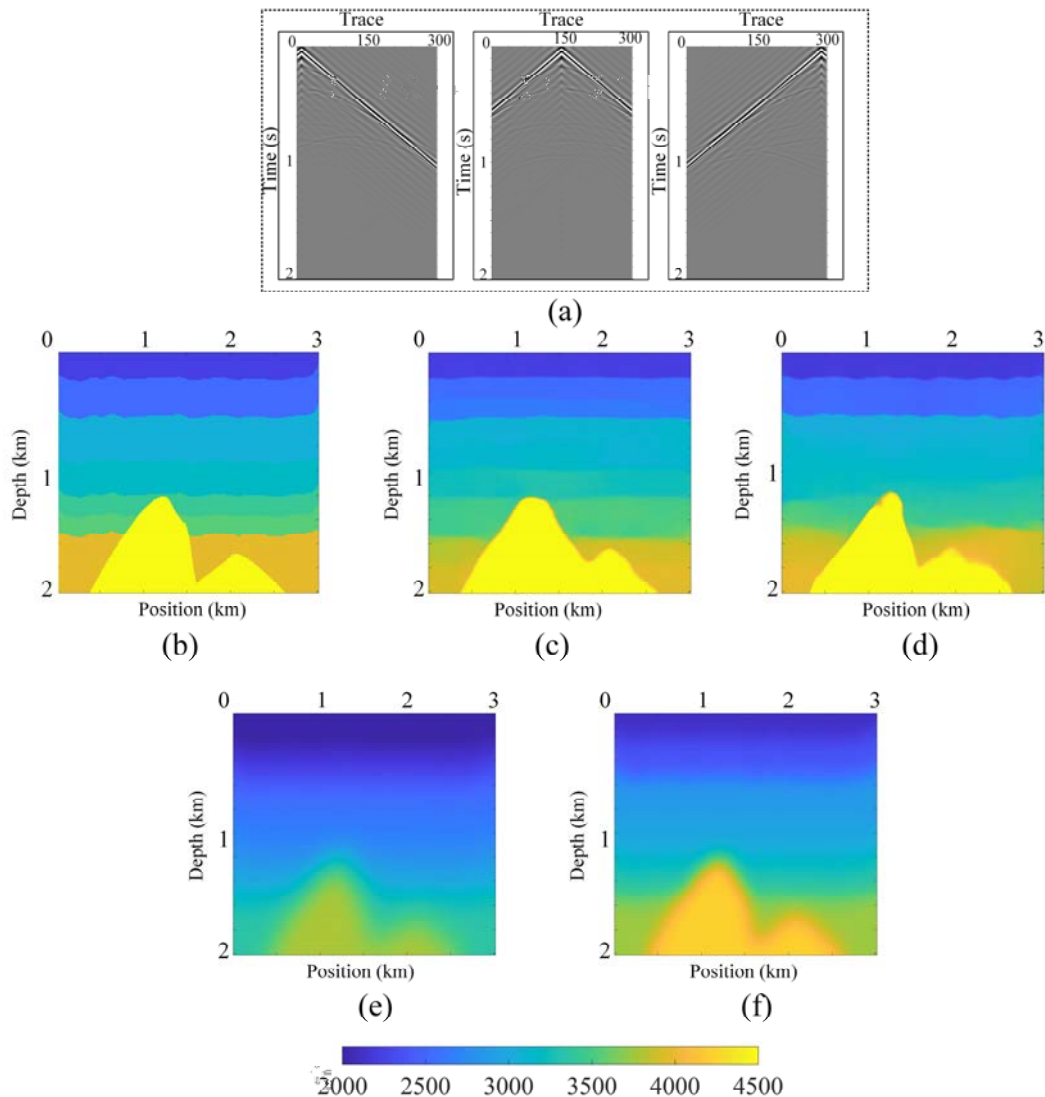


图 5.11 当地震资料缺乏低频分量时不同方法的反演结果 (a) 缺失低频的地震数据, (b) 真实值, (c-f) SVIT、FCNVMB 的结果、FWI 初始模型和 FWI 结果

如图 5.11 (f) 所示，除了对盐体的预测较差外，FWI 在没有低频的情况下对致密层的重建也较差。可以看出，FCNVMB 因为使用的 U-Net 类的结构，该模型可以通过利用特征图的缩放提供的更大的感受野来克服这种情况下的挑战。

但由于缺乏低频信息，该模型的结果中可以看到一些地层变得模糊。相比之下，图 5.11 (c) 中的 SVIT 可以恢复大多数地层和盐体的形状，这是研究所提出的方法相对于其他方法的显著优势。这种性质与自注意力机制的长距离特征提取有关。接下来，还使用 MSE 和 SSIM 测量了不同方法的模型误差。SVIT 的反演结果具有 0.0135 的 MSE 和 0.7941 的 SSIM，FCNVMB 的结果具有 0.0419 的 MSE 和 0.7305 的 SSIM，FWI 结果具有 0.0530 的 MSE 和 0.7178 的 SSIM。与 FCNVMB 和 FWI 相比，SVIT 结果具有最低的 MSE 和最高的 SSIM，这意味着 SVIT 在观测数据低频缺失的情况下表现出最好的性能。

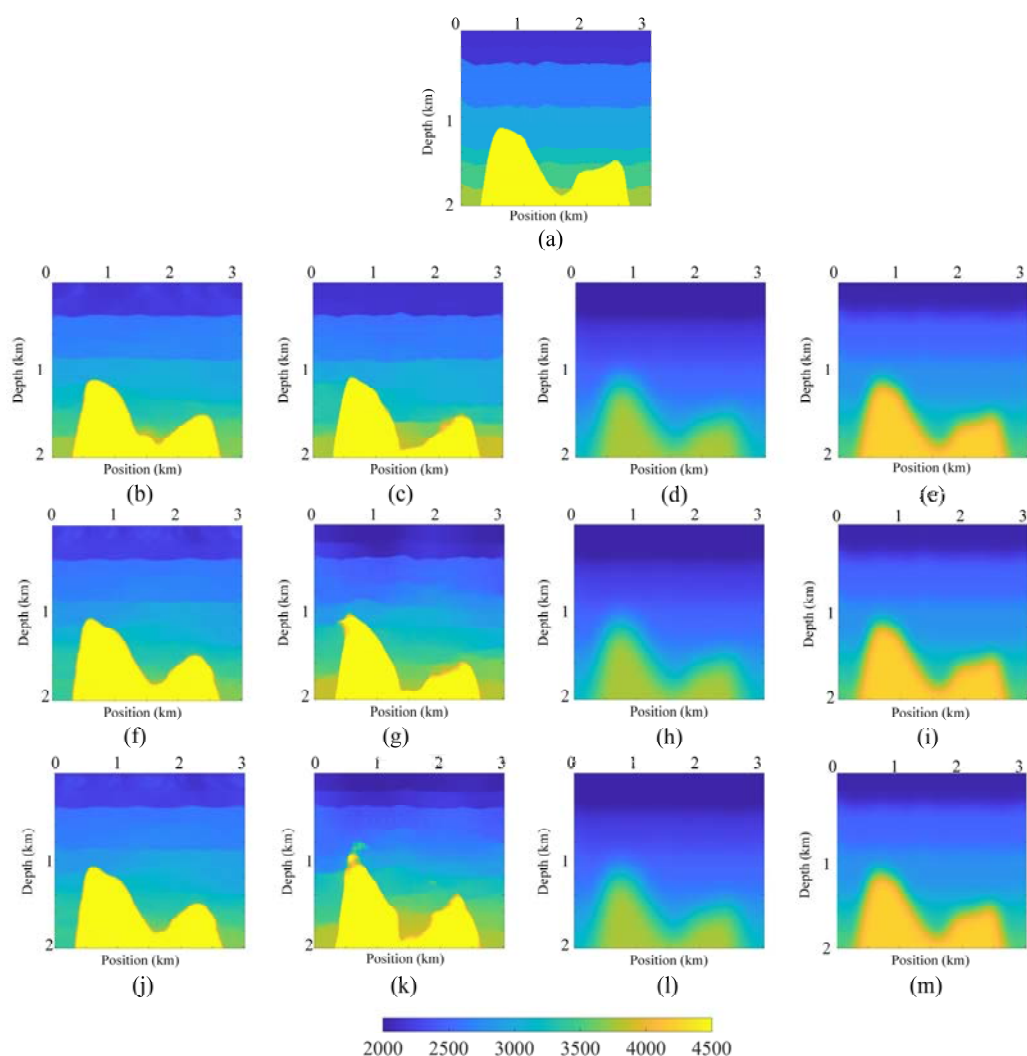


图 5.12 观测数据缺失三个、五个及十个记录时不同方法的反演结果 (a) 真实值, (b-e) 缺失三个记录时 SVIT、FCNVMB 结果、FWI 初始模型和 FWI 结果, (f-i) 缺失五个记录时 SVIT、FCNVMB 结果、FWI 初始模型和 FWI 结果, (j-m) 缺失十个记录时 SVIT、FCNVMB 结果、FWI 初始模型和结果

(3) 观测系统适应性

除了对噪声和低频缺失的适应能力,训练网络在不同观测系统设置下的适应性也是一个值得讨论的问题。在获取地震数据的过程中,考虑到地表的复杂情况,往往很难在所有规划的位置激发炮点,因此,反演方法在缺少对应炮记录的情况下对速度模型的预测也应该精准。为了进一步验证基于 Transformer 的反演方法对基于 CNN 的反演方法的优势,尤其是鲁棒性方面的优势,本研究中评估了在丢失一部分炮集记录以及丢失一部分数据道时的预测结果。图 5.12 展示了在三个缺失炮点记录、五个缺失炮点记录和十个缺失炮点记录的数据反演结果,其中图 5.10 (a) 给出了真实值,图 5.10 (b)、(f)、(j) 分别给出了 SVIT 在几种情况

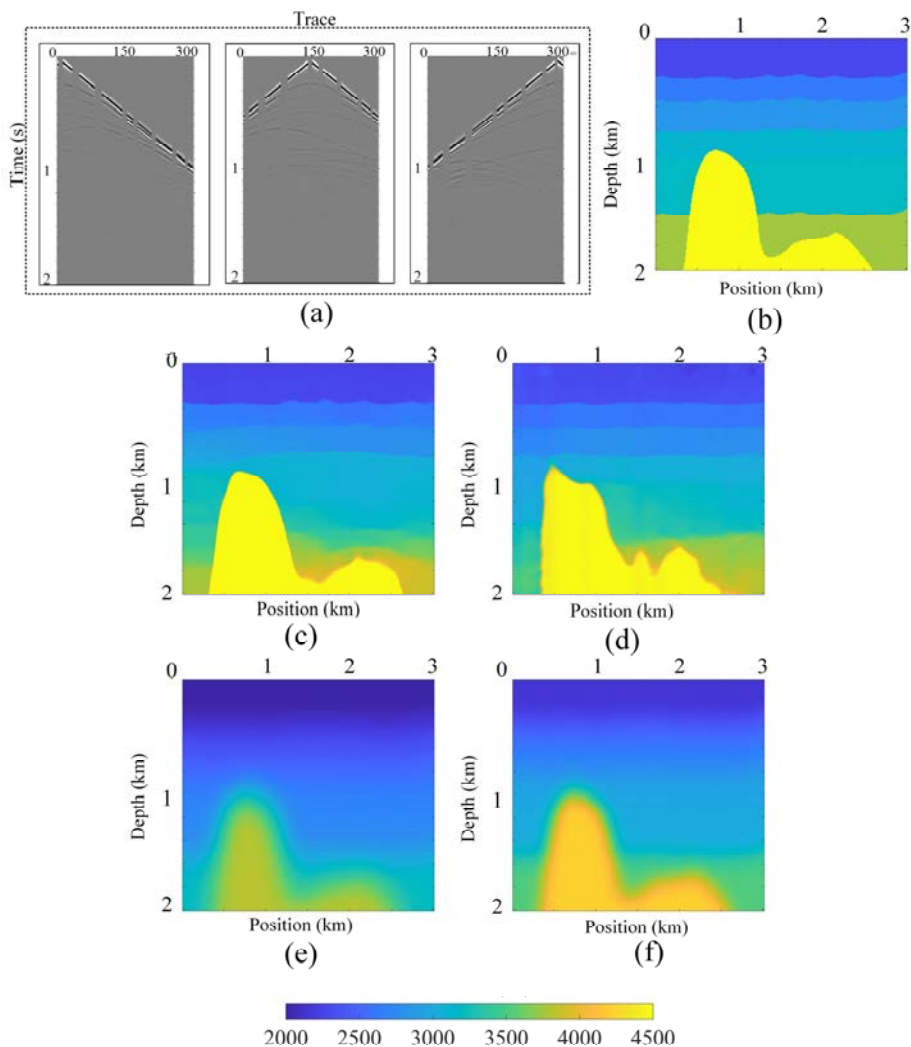


图 5.13 在观测数据中丢失一部分检波器数据时不同方法的反演结果 (a) 缺失道的观测数据, (b) 真实值, (c-f) SVIT、FCNVMB 的结果、FWI 初始模型和 FWI 结果

下的结果,图 5.10 (c)、(g)、(k) 给出了在 FCNVMB 的预测结果。尽管缺少了几个炮集,两种方法依旧可以很好地工作,但是随着缺失炮数目的增加,两种方法的预测结果都变得不准确,尤其是对于盐体的预测。相比于 FCNVMB,SVIT 方法的反演质量在一定程度上能够得以保持,表现出更强的鲁棒性。

接下来,在测试过程中随机舍弃一部分接收器的数据来进行另一次实验。实验结果如图 5.13 所示,在随机丢失多个地震道后,SVIT 方法仍然提供了可接受的结果。相比之下,FCNVMB 的预测结果如图 5.13 (d) 所示变得不可接受。对于反演方法来说,全局信息中包含着一些局部地震道没有的信息,可能会有利于速度模型的重建。一般来说,全局的上下文信息可以从整个地震剖面中获取,它可能包含着速度分布和界面数量等信息。对于 SVIT 来说,由于其全局运算的特性,使得其推广性和适用性大大增加,编码器可以从全局提取速度模型的有效信息,并利用解码器重建速度,使得模型在缺失部分炮集记录和缺失部分接收器数据的情况下,依旧可以准确预测盐体和地层界面。这也表明了 SVIT 方法在不同的观测系统设置下仍然具有泛化能力。

(4) 时间消耗

对计算成本的讨论也是比较不同反演方法的重要组成部分。传统的反演方法往往耗时且计算效率低。在模拟速度模型上计算的 FWI 方法耗时 6 小时 27 分钟,而 SVIT 和 FCNVMB 的预测分别仅耗时 3.84 秒和 2.14 秒。对于数据驱动的反演,主要的计算成本主要发生在训练阶段,在模型建立过程中只进行一次训练。经过训练后,预测成本则可以忽略不计。尽管 FWI 和两种基于深度学习的数据驱动反演方法分别是在 CPU 和 GPU 上部署实现的,缺乏对比的一致条件,但两种方法在计算时间上的显著差异可以推断出,就计算效率而言,两种数据驱动反演方法比 FWI 具有优势。作为一种传统的反演方法,FWI 需要对每个速度模型进行单独的计算。相反,SVIT 和基于 CNN 的方法 FCNVMB 的计算成本主要发生在训练阶段,而预测阶段时间成本则较小。由此可以得出结论,深度学习方法比传统的反演方法在实时反演计算方面更有效率优势。

同时,两种基于深度学习的反演方法在参数数量上存在一定差异。在描述深度学习方法时,除了其准确性之外,通常还根据模型的参数数量来描述模型大小。

从表 5.7 中, 可以看到, SVIT 方法比基于 CNN 的模型 FCNVMB 有着稍小一些的参数规模, 但是却实现了更优的性能。因此, 可以得出结论: 对比基于 CNN 架构的 FCNVMB 方法, SVIT 可以在准确性和较少的参数量方面胜出。

此外, 本研究中还对硬件上方法的真实计算速度进行了总结, 由于理论推导并不能很好地预测计算时间, 只能通过实际训练和测试过程中记录获得这些结果。为此, 本研究对这两种方法进行训练时间和推理时间的计时。具体的结果也一并列在表 5.7 中。从中可以看到, SVIT 的训练时间要高于 FCNVMB, 这应当归因于自注意力计算带来的巨大计算开销, 同时因为缺少一定的归纳偏置, 基于 Transformer 的模型收敛速度普遍慢于基于 CNN 的模型, 这也可能造成 SVIT 训练时间的延长。在未来的研究中, 本研究也将继续优化 SVIT 来解决这一点。

表 5.7 SVIT 和基于 CNN 的方法的计算时间及模型参数量对比。

方法	模型参数量	训练时间	测试时间 (秒)
SVIT	2.754×10^7	53 h 53 m	3.84
CNN	3.1×10^7	36 h 14 m	2.14

表 5.8 消融研究

消融项	SW-MSA	FFN 中的卷积层	MSE
	✓	×	0.0309
SVIT 方法	×	✓	0.0453
	✓	✓	0.0181

(5) 消融研究

消融研究对于深度学习反演算法的研究有着重要意义, 消融研究通过移除某些组件来研究网络模型的性能, 以了解组件对整个网络模型的贡献。在本节中,

通过消融研究详细分析了 SVIT 各组成部分的效果。本研究中选择 MSE 作为度量来全面衡量消融各个组件后网络在测试集上的性能。消融结果如表 5.8 所示。首先,用卷积残差块^[51]替换 SVIT 中的所有 SW-MSA 块,同时保持所有其他构建块不变。可以观察到,具有 SW-MSA 的 SVIT 实现了更好的性能,并且优于具有卷积残差块的网络。这项研究表明了速度模型反演任务中,SW-MSA 与卷积运算相比的有效性。一般来说,自注意力运算相比于卷积运算,能够提取更多的远程特征,这将有益于网络执行反演任务。此外,表 5.8 还显示了在前馈网络中对卷积层的消融结果。可以看到,消融掉卷积层后,SVIT 的反演性能出现了下降,这验证了将卷积层引入前馈网络确实可以获得更好的性能,卷积层的加入一方面可以发挥其提取局部特征的优势,另一方面,其对于速度反演过程中一些细节结构的恢复发挥着重要作用。因此,将卷积引入前馈网络中是适合速度反演任务的选择。

(6) Sigsbee 案例研究

尽管 SVIT 方法在合成数据集上表现良好,但其在更真实通用的速度模型上的性能仍旧未经测试。因此,选择如图 5.14 (a) 所示的 Sigsbee 模型的一部分作为测试数据进行了测试。因为这部分速度模型与训练集中的数据的隐含概率分布大致类似,可以测试不同的反演方法对盐体速度的推断能力。Sigsbee 模型的速度值范围为 1500 至 4500 m/s。在 Sigsbee 速度模型上设置了 29 个震源和 301 个接收器,它们均匀地分布在表面。使用主频为 25 Hz 的雷克子波作为源。在模拟生成该模型对应的观测数据时,正演建模的设置与先前模拟测试集时是相同的。采用先前训练的 SVIT 模型直接从这些地震数据中估计速度模型,需要说明的是,没有根据这些 Sigsbee 模型生成的数据重新调整网络。不同方法的反演结果如图 5.14 所示,可以看出对于 Sigsbee 模型 SVIT 仍然可以准确地揭示其中大部分地层,并准确地捕捉盐体。从 FCNVMB 和常规反演方法 FWI 在 Sigsbee 数据上的结果,可以观察到,SVIT 方法要略优于 FCNVMB 方法,尤其是对一些边界的连续性的保持上。但是,直接推断的反演结果中仍然缺少一部分信息,这可能归因于此测试集的数据分布仍与训练集存在一定不同。同时,从结果中可以观察到两种深度学习反演方法的结果要明显优于 FWI 方法,这也展现了将本研究中提出

的数据驱动方法应用到真实数据上进行反演的潜力。

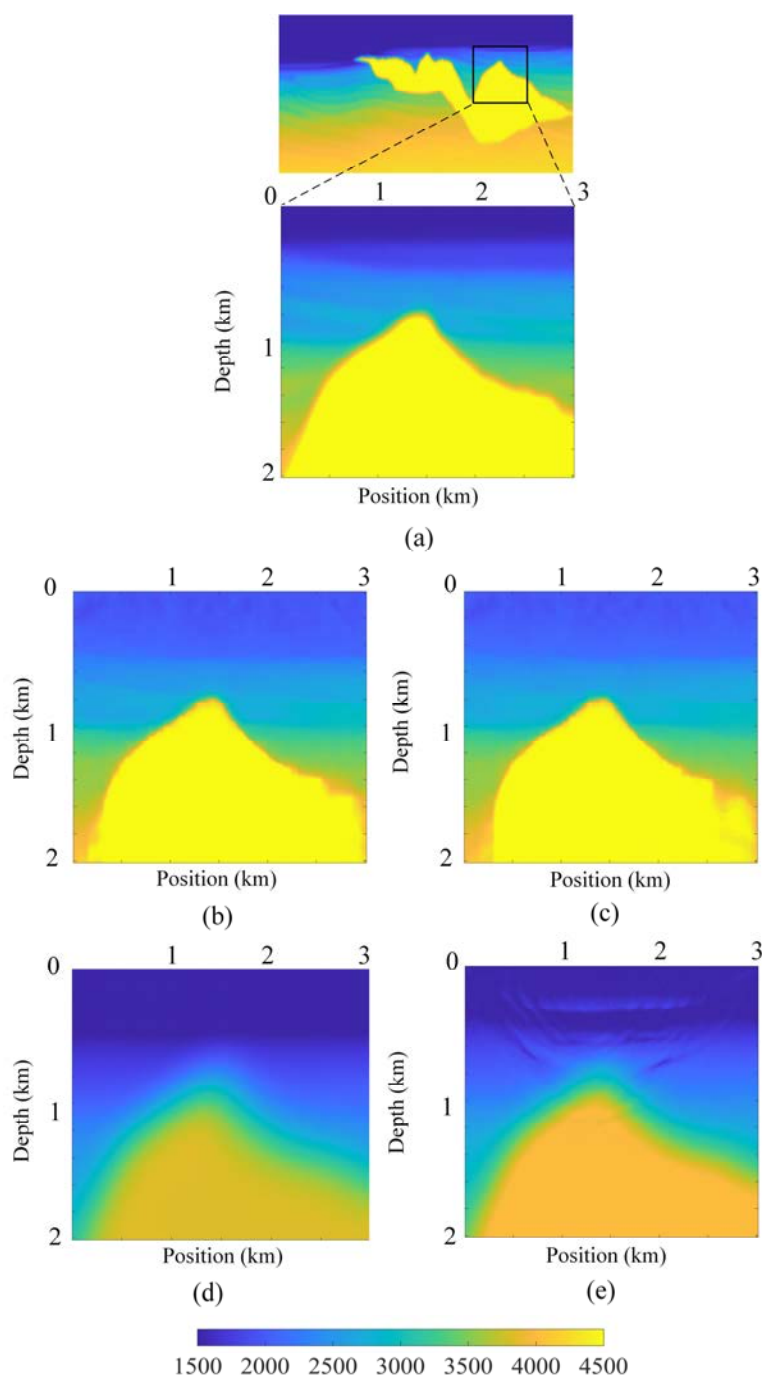


图 5.14 不同方法对 Sigsbee 模型的反演结果 (a) 真实值, (b) SVIT 预测结果, (c) FCNVMB 预测结果, (d) FWI 初始速度模型, (e) FWI 预测结果

尽管本次测试仅基于 Sigsbee 模型的部分数据进行, 而非真实的地震观测数据, 但实验结果对于验证反演网络的推断能力依旧有着重要的意义。首先, Sigsbee 模型作为一个复杂且综合的速度模型, 其内部结构的多样性和复杂性在

一定程度上模拟了真实地质结构的特点。因此,使用此类速度模型作为测试数据,能够为验证和评估不同地震数据反演方法的有效性和适用性提供了一个相对真实的环境。

此外,通过在这种高度控制且已知的环境中测试,可以更加精确地评估不同反演方法对于不同地层界面、特别是对于复杂地质结构如盐体的反演能力。这不仅能够展示数据驱动方法在地震数据反演领域的应用潜力,同时也为理解反演网络的泛化能力提供了一定的经验。从实验中能够看出,基于深度学习的反演方法对比传统的全波形反演方法最大的一个优势便是对于高速度地质体结构的精确恢复,这主要是因为深度学习方法能够从大量的数据中学习到复杂地质体的特征和模式,而这些特征和模式在传统的全波形反演方法中往往难以直接识别和处理。深度学习方法通过其强大的非线性拟合能力,能够有效地捕捉到地震数据中蕴含的细微变化,从而在恢复高速度地质体结构时表现出更高的精度和灵敏度。尽管最终的目标是将这些方法应用于真实的地震数据上,但是在那之前,基于复杂速度模型的测试也能为本研究提供一条评估方法效果和指导后续研究方向的重要途径。

(7) 与更多先进 CNN 架构进行对比

为了进一步强调所提出架构在地震数据处理中的优势,进行了与其他先进 CNN 架构的深入对比,选择了 VMB-Net^[89]作为对比方法。VMB-Net 是一种对传统全卷积网络的改进,专门针对地震数据与其对应速度模型之间的弱空间对应问题进行优化。其主要创新在于通过全连接网络来改善空间对应关系,从而提升了反演精度。在实验中,VMB-Net 遵循其于 2020 年发表的原论文的实验设置^[89]。使用与 SVIT 相同的数据集对 VMB-Net 进行训练,并在相同的测试集中进行评估。测试结果如图 5.15 所示,SVIT 方法在 MSE 和 SSIM 方面均表现优异,分别为 0.0201 和 0.8802。相比之下,改进后的 VMB-Net 虽然在一定程度上提高了性能,其 MSE 为 0.0205,SSIM 为 0.8777,但仍未能超越 SVIT。

从上述结果可以看出,尽管 VMB-Net 在改善地震数据和速度模型的空间对应关系方面取得了一定成效,但其性能仍受限于卷积结构的局部特性。这表明,在特征提取过程中,SVIT 所采用的改进架构在整体上更具优势。这项研究不仅

验证了 SVIT 在地震数据处理中的潜力,也为未来的网络设计提供了宝贵的参考。

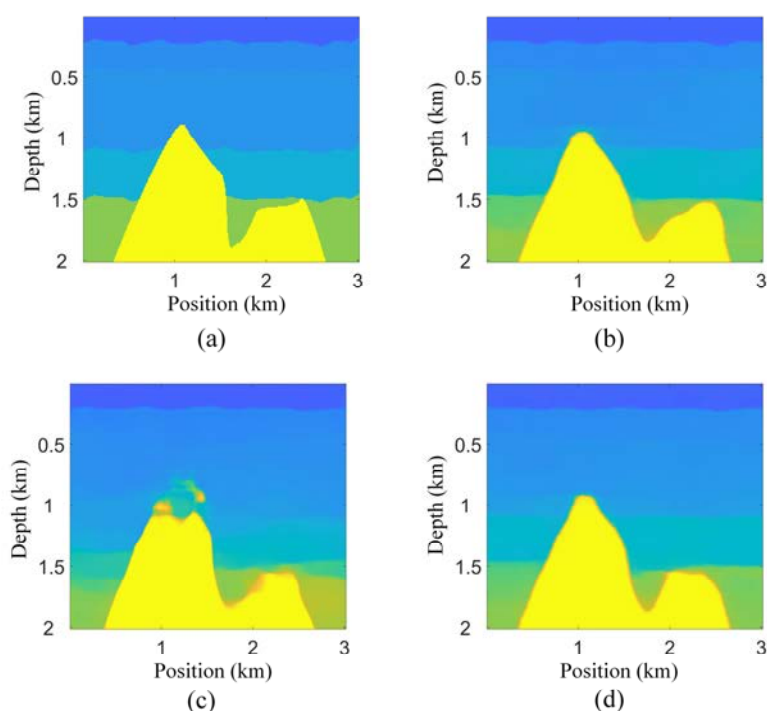


图 5.15 SVIT 和更多先进卷积架构的反演结果 (a) 真实值, (b) SVIT 预测结果, (c) FCNVMB 预测结果, (d) VMB-Net 预测结果。

5.3.3 算法分析讨论

(1) 进行基于深度学习的速度反演算法研究时新的网络架构的研究意义

之前的实验已经验证了直接以地震炮集作为原始输入进行速度反演的方法是可行的,而且此类方法不需要初始模型,这一点使得该类型方法不需要任何预先的假设。而且前述的实验也表明在噪声扰动和观测系统缺失等一系列复杂情况下,反演网络依旧拥有能力有效地去完成反演任务。且与传统方法相比,该类方法计算时间更短。传统方法 FWI 的计算成本主要都在迭代找最优解的过程中,而相比之下,反演网络的计算成本主要在训练阶段,在一次训练结束后,预测成本可以忽略不计,能够极大提高计算效率。

尽管近年来的研究表明,如果能够接受这些计算代价,深度学习的反演方法能够明显优于其他的传统反演方法,但是这些实验都是在模拟数据上进行的,换

言之，研究者们主动创造出数据完备的这一前提条件，而这个条件在实际勘探场景中并不能轻而易举地被满足。对于深度学习方法，在数据角度还是模型角度进行权衡，即到底是数据优先还是模型优先，一直是饱受争论的问题。就目前的实验结果看来，数据和模型同样重要。

首先，基于深度学习的方法的局限性是其反演能力依赖于训练集，训练集决定了方法的性能上限。一般来说，要训练的数据集中应当含有和待预测的速度体相似特征或相似分布的地质结构，才能在测试中产生好的效果。大多数情况下，更加完备、更加多样、更加复杂的训练样本集会训练产生更强大性能的反演网络。

除了多样化的数据构建方案是研究的重点外，也应当意识到，对于反演这样一个低特征对齐、不确定性大的“跨模态”任务来说，对反演网络的架构研究是同等重要的。研究表明，假设在拥有充足数据的情况下，不同的网络结构在实现从观测数据到速度体的映射推理时存在着明显的性能差异。本文的假设前提是在数据充足的情况下，尽可能提升反演网络的能力。SVIT 的表现验证了借助先进的架构，反演网络的性能可以在同一个数据集上训练实现更大的性能提升，这将有助于在更大更广泛的反演数据集上进一步拓展发挥深度学习模型的反演性能。因此，在未来，研究新的反演网络架构依旧是该领域可以继续探究的方向。

（2）网络学习到特征的稳健性分析

对深度学习反演方法的一个疑虑，便是对网络学习的特征的稳健性的怀疑。在本研究中，可以深刻意识到一点，基于深度学习的反演并没有取代也不可能完全取代当前的反演方法。这个性能问题凸显了监督学习方法的一个关键性问题：对模型的泛化性的担忧。究其原因，一方面，这是缺乏标注数据所导致的。深度学习反演方法的准确性与用于训练网络权重参数的一一匹配的“观测数据-速度模型”数据对有着直接关系。因此，当仅仅使用少量精心准备的数据来训练模型并意图使其作为通用速度反演解决方案时，就会引发使用者对泛化和统计可靠性的担忧。另一方面，深度学习反演方法推广到其他问题时是否保持有效，也是一个难以回避的问题。一般来说，深度学习反演方法很难直接使用训练好的网络去测试更加复杂的速度体，因为不同的来源的速度体往往分布不相似。解决这个问题，一方面可以让模型在多种类型速度体的训练集中去训练，尽可能地去熟悉

多种分布。在这个过程中，模型不仅仅是面向某个单一分布下数据集的特征，而是学习到更多的是通用特征。另一方面，将当前基于先验的反演方法融入到反演网络的训练之中，也是一种解决思路。通过结合传统地球物理学的先验知识和深度学习的数据驱动优势，可以在提高反演精度的同时，降低对大量标注数据的依赖。通过这些方式，可以有效地缓解对深度学习反演方法稳健性和泛化性的担忧。综合来说，现阶段的深度学习反演研究中网络学习到的特征在鲁棒性方面依旧较差，在这一方向上的研究未来仍有较长的路要走。

(3) 传统反演方法和深度学习反演方法的关系

物理建模（理论驱动）与深度学习建模（数据驱动）过去往往被认为是两个领域，具有不同范式。但其实两种方法可以相互补充的，前者外推能力强，后者更能灵活地发现新规律，二者是可以相互促进的，因此研究深度学习为基础的反演方法有着重要的意义。从对常规反演方法的介绍之中，可以看出常规反演方法和深度学习方法之间存在着相似的关系和机理——即“直接对结果多次迭代”和“通过迭代求取一个模型来实现对结果的映射”有着某种共通的原理。也同样应该意识到，深度学习反演和全波形反演不是对立，而恰恰可能是互补的，深度学习反演方法可以为全波形反演提供高效精确的初始模型，改善其初始模型“选择困境”，而全波形反演能够在其基础上刻画更加细节的结构，对深度学习反演方法的结果“不断改善”。这也为实际情况下选择最合适的反演策略提供了新的方案。在下一章中将详细介绍这一思路的研究。

5.4 本章小结

除了地震数据去噪任务以外，速度模型反演也是地震勘探中相对重要的任务。近年来，在该任务领域出现了一部分利用深度学习技术直接反演速度模型的研究思路，但是其架构设计多是基于 CNN，该类方法在捕获长距离特征方面受限。此外，地震数据的时变特性也不利于权重共享的 CNN 架构充分提取特征。针对以上问题，在本章中，本研究中提出了一种用于地震反演任务的深度学习模型即 SVIT 用于推理地震数据速度模型。与现行的基于 CNN 的结构相比，SVIT 主要依赖自注意力计算提取特征，可以有效捕获远程依赖信息，从而为重建速度模型

提供更充分的特征表示。通过建立一个一一配对的“观测数据-速度模型”的数据集来对 SVIT 进行训练，使其能够在共炮集数据中接收特征，实现对速度模型的直接的端到端推理。实验阶段构建了包含致密层及盐丘模型的观测数据组成测试集来验证本章算法的性能。通过与常规反演方法 FWI 和基于 CNN 的反演方法 FCNVMB 进行对比比较，实验凸显了所提出 SVIT 算法在反演任务上的优势。在大量的实验中，SVIT 获得了比 FWI 和 FCNVMB 更有竞争力的结果，且在多种极端数据情况及观测系统设置下比基于 CNN 的反演网络 FCNVMB 表现出更加出色的性能。

第6章 基于特征对齐 Transformer 地震速度混合反演策略

6.1 引言

第5章提出的SVIT反演算法虽然在理论上能够快速获得地震速度模型,但在实际应用中,深度学习反演方法却面临着适应实际地震数据的挑战。实际地震数据常常具有复杂的噪声和不确定性,这些因素使得深度学习算法的性能受到限制。尽管深度学习反演算法在地震速度模型的获取方面具有潜力,但在实际应用中仍存在挑战。速度模型反演是油气勘探、地热能、碳捕获和储存等应用中地下成像的关键步骤。速度模型表示地震波速度在地下的分布,对于解释地震数据和预测地下特性至关重要。全波形反演(Full Waveform Inversion, FWI)是一种广泛使用的通过反演整个记录的地震波形来估计速度模型的方法。FWI通过迭代调整速度模型来最小化观测数据和模拟数据之间的失配,并以此为依据不断迭代优化速度模型。FWI方法具备反演得到高分辨率速度模型的潜力,但计算成本高,且对初始模型的选择和噪声敏感。近年来,深度学习反演技术作为一种全新的思路近年来受到研究者的广泛关注。人们对将深度学习反演方法与FWI相结合以克服传统FWI方法的一些局限性并提高地下速度反演的准确性和效率越来越感兴趣。

深度学习反演和全波形反演的集成对地震数据处理领域来说是一个前沿的研究方向,但这一过程也面临着几项挑战:首先,深度学习算法训练有着大量标注数据需求。在实际应用中,尽管目前已有研究通过使用数值模拟技术或是数据增强手段来生成合成数据,以此来克服数据不足的问题,但这些方法往往无法完全替代真实实际地震数据的多样性和复杂性。真实的地震数据不仅稀缺,而且往往具有高度的复杂性,包括各种噪声和非线性特征,这对数据的采集和预处理提出了较高要求。因此,如何获取完备且具有代表性的观测数据,以便训练出能够准确反演速度分布的深度学习模型,仍是一个待解决的关键挑战。其次,如何设计有效的方案将深度学习和全波形反演技术有效结合,以充分发挥两者各自的优势,也是一个重要的研究问题。全波形反演作为一种高精度的地震资料解释方法,

能够提供地下介质的详细信息，但其计算成本高昂，且对初始模型敏感。而深度学习由于其强大的特征提取能力，可以有效地从地震数据中识别出复杂的地质特征，但可能缺乏足够的物理约束来确保解的准确性。尽管深度学习反演和全波形反演有着不同的起源和方法，但它们在基本原理上有一些共同点。这两种方法都旨在通过迭代调整一组模型参数来最小化观测数据和建模数据之间的不匹配。在深度学习反演的情况下，这些参数是神经网络的权重和偏差，而在 FWI 中，它们是地下的地震速度分布。此外，这两种方法都可以受益于正则化技术的使用，以避免过拟合并提高学习模型的泛化能力。此外，深度学习反演和 FWI 也有独特的优势和劣势，使它们相互补充。例如，深度学习反演可以处理输入和输出数据之间的复杂和非线性关系，它可以从大型数据集中学习，以捕捉数据集中速度体的统计分布。另一方面，FWI 则适合于对波传播的物理建模，并且可以提供速度模型的高分辨率估计。通过结合这两种方法的优点，使用深度学习和 FWI 的混合地震反演可能比单独使用任何一种方法都能实现更准确、更高效的地下速度反演。因此，如何将深度学习的数据驱动特性与全波形反演的物理驱动特性相结合，从而设计出既能保证高精度也能提高效率的混合地震反演方法，是本章中着重解决的问题。此外，对深度学习速度反演方法也需要进一步改进以适应实际地震数据反演的需求。在实际场景下，网络接受多个叠加的共炮点数据作为输入并被要求直接推理速度模型，并面临两个主要挑战：特征对齐难题和反演区域大小限制问题。

特征对齐难题是涉及到利用深度学习对地震数据进行反演时尤为显著的问题。这个问题发生在从不同震源位置激发并收集到的共炮点数据被整合作为多个独立通道输入到神经网络中时。尽管这些数据可能来源于相同或相近的地质界面反射，但由于震源点和测线的相对位置变化会导致严重的特征不对齐问题。具体而言，即使是源自同一地质界面的反射信号，在通过不同的地层路径传播并被不同位置的接收器记录时，其在数据层面上的表现形式（例如，波形特征、到达时间等）可能会因为路径差异而在观测记录中出现空间位置上的偏差。这种位置上的偏差在多个数据通道叠加后直接输入神经网络进行处理时，会给网络的学习过程带来显著的困难，因为网络需要从这些不对齐的特征中学习准确的地下结构

信息,这无疑增加了模型训练的复杂度。解决特征对齐问题是提升深度学习反演精度的关键。

此外,深度学习速度反演方法通常需要在训练阶段定义一个固定大小的输入窗口或反演区域,以便网络能够从中学习到地下速度结构的特征。这种设计在一定程度上简化了模型的训练过程,使得训练变得可行。然而,这也意味着模型在应用于实际数据时,只能处理与训练阶段相同大小的反演区域。在实际地震勘探任务中,地下结构的复杂性和多样性要求反演方法具有高度的灵活性和适应性,能够处理不同尺寸和形状的反演区域。因此,固定大小的反演区域设置限制了深度学习方法的普适性和灵活性,难以满足实际场景中对于广泛应用的需求。

针对以上两个问题,本章提出一种基于偏移卷积的特征对齐 Transformer 反演网络和滑动反演策略,来有效解决深度学习模型在反演过程遇到的特征对齐难题和反演区域大小限制问题。首先,在 Transformer 架构引入具有自适应对齐能力的偏移卷积,允许网络动态调整卷积核的采样位置,以使模型能够在提取特征时自动识别和调整共炮点数据在空间上的不一致性,从而提高从复杂地震数据中准确提取地下速度模型信息的能力。其次,采用分块处理和滑动窗口技术,具体来说,将整个待反演区域划分为多个固定大小的子区域,并逐一对这些子区域进行深度学习反演,然后通过平均策略整合各子区域的反演结果,以获得整体记录。

另外,本研究中开发了一种混合地震反演技术,该技术结合了深度学习反演方法与全波形反演方法的优势,显著提升地下速度模型反演的精度。首先采用基于特征对齐的 Transformer 架构的深度学习模型,利用其强大的特征提取能力,从地震数据中学习统计分布,生成高质量的初始速度模型。特别地,通过引入偏移卷积技术,网络能够有效解决共炮点数据输入时的特征对齐难题,使得模型能够从多个炮集数据中更准确地提取有用信息。此外,采用滑动推理策略,即在固定大小的反演区域内应用深度学习模型,并通过滑动该区域覆盖整个探测范围,并对每一部分的反演结果进行平均,从而克服深度学习反演方法因反演区域尺寸限制而难以广泛应用的问题。这种策略不仅显著提高了初始速度模型的估计精度,也为后续的全波形反演提供了更为可靠的起点。进一步地,利用全波形反演技术对初始速度模型进行迭代优化,精细调整速度模型的细节,以获得更为精确的地

下速度分布。同时, FWI 的引入为深度学习生成的速度模型提供了进一步的细化和验证, 确保了最终结果的准确性和可靠性。通过这种深度学习与 FWI 的结合, 混合反演方法不仅克服了各自方法的局限性, 还发挥了二者的互补优势, 从而实现了在地下速度模型反演任务中的显著性能提升。在实验中, 本研究提出的混合地震反演方法在 SEAM 案例和墨西哥湾海洋地震实际数据的反演中得到了较好的结果, 展示了该方法在实际案例速度模型精确反演方面的巨大潜力。

6.2 特征对齐 Transformer 混合反演策略

在本节中, 从数据集的构建、特征对齐 Transformer 的网络结构以及混合反演的原理等三个方面详细描述了该方法。首先, 介绍了本方法使用的数据集制作步骤及参数设置, 以及数据集分布情况。之后, 详细介绍了特征对齐 Transformer 的网络结构及细节。最后, 将详细介绍混合反演算法的流程步骤及原理。

6.2.1 深度学习速度反演任务数据集构建

对于深度学习反演部分, 本章研究依旧遵循有监督学习的范式, 所以大量带有标签的数据是训练模型不可缺少的部分。由于数据驱动方法的特性, 深度学习方法的性能受到训练时使用的数据影响。一般来说, 合理而充分的模型设计可以帮助训练好的网络应用于更多的现实情况。为训练一个更加高效而准确的网络, 选择借助开源的 SEAM 模型来构建一个完备的大型速度模型数据集。本节中, 将详细介绍准备数据集的过程。

为了给混合地震反演方法构建数据集, 本研究通过在 SEAM 速度模型上进行正演建模来模拟实际地震波的传播。通过借助 SEAM 速度模型, 使用有限差分法求解波动方程, 获取每个速度模型所对应的地震观测数据, 具体方法已在之前的章节有所介绍, 故此处不再详细说明。本研究采用时间域的交错网格有限差分法来求解该方程, 其中时间方向采用二阶精度, 而空间方向则采用八阶精度。为了模拟地震震源, 在每个速度模型中自左到右布置 64 个震源点, 震源间隔为 37.5 米, 并部署了 480 个检波器, 与震源最近和最远间隔分别为 198 米和 6 公里, 检波器间隔为 12.5 米。它们被均匀地分布以捕捉地震波的传播数据。为了获得实际的地震记录, 使用 25 Hz 的雷克子波作为震源来生成通过地下传播的地震

波，并由检波器阵列记录。正演模拟采用的参数详见表 6.1，而为了减少边界处的反射影响，采用完全吸收层^[146]作为模型的边界条件。图 6.1 展示了一个速度模型及其产生的三份地震记录样本。所得到的地下速度模型以及相应的合成地震记录被用于构建混合地震反演方法的数据集。根据表 6.2 的划分比例，所有的速度模型和观测数据被分为训练集、验证集和测试集三部分，分别用于网络的训练、验证网络训练过程中的表现以及评估网络的准确度。这三个数据集是相互独立的。在进行深度学习反演步骤时，为了减少处理的数据量，对每份观测数据的维度进行了降采样，使之变为 400×480 的规模。同时，对所有观测记录和速度模型数据进行了归一化处理。

表 6.1 正演参数设置

参数	具体设置
震源数量	64
接收器数量	480
速度模型大小	(1344,401)
网格大小	6.25 m × 6.25 m
采样间隔	1×10^{-3} s
子波类型	Ricker
子波频率	25 Hz
边界条件	PML
最大时间	10 s

表 6.2 本实验所用数据集分布

数据集类型	SEAM 致密层模型数量	SEAM 盐体模型数量
训练集	2000	2000
验证集	100	100
测试集	100	100

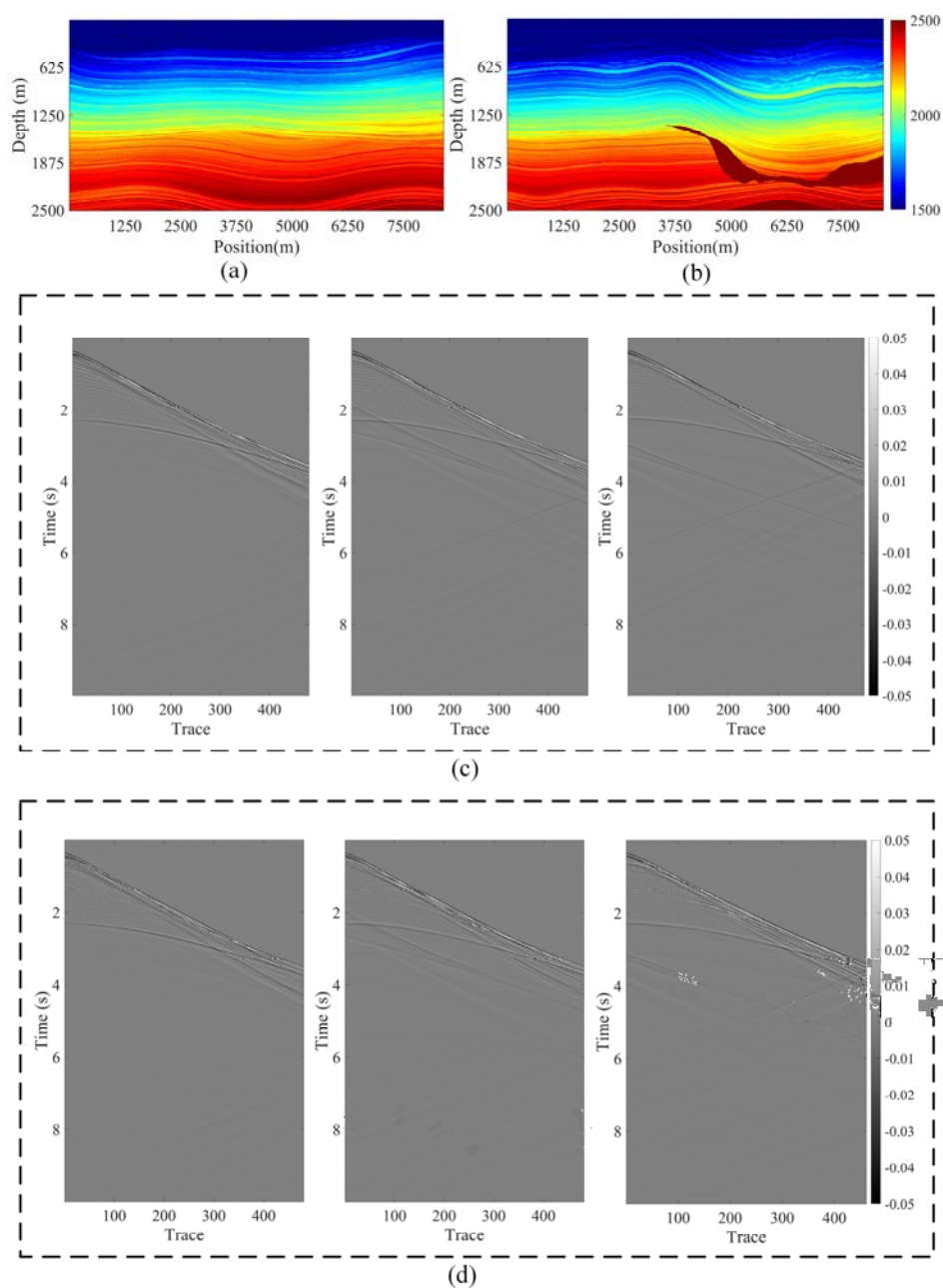


图 6.1 训练集中具有代表性的速度模型及其相应的模拟数据 (a) 致密层速度模型, (b) 盐体速度模型, (c) 致密层速度模型的三个有代表性的观测记录, (d) 盐体速度模型的三个有代表性的观测记录

6.2.2 特征对齐 Transformer 网络结构

本研究中建立了一个具备特征对齐特性的基于 Transformer 的反演网络。为了降低计算复杂度,网络采用 U 型结构。为叙述方便,将提出的模型称为特征对齐反演 Transformer (Feature Alignment Inversion Transformer, FAIT),其结构如图 6.2 所示。在以下部分中,首先描述 FAIT 的总体架构。

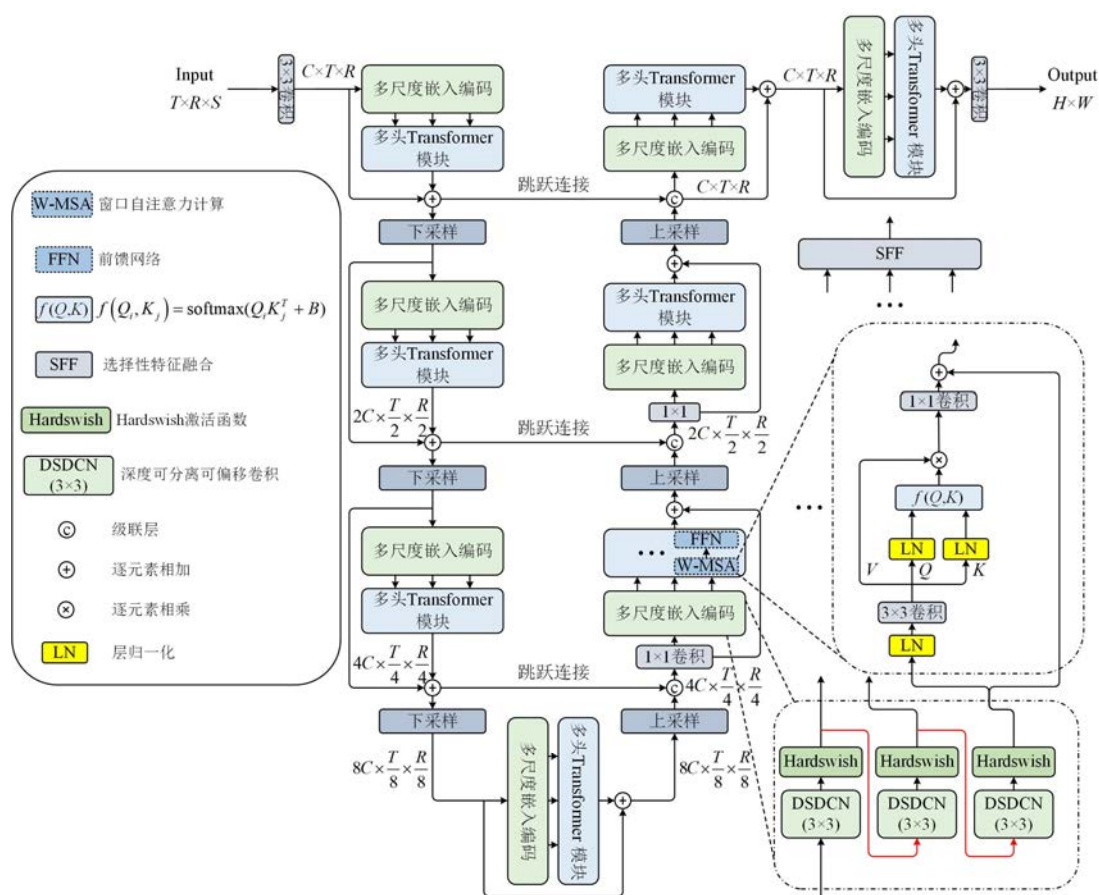


图 6.2 特征对齐的 Transformer 的速度反演网络结构

给定输入 $I \in \mathbb{R}^{T \times R \times S}$ ，应用卷积进行浅层特征提取，生成特征 $F_0 \in \mathbb{R}^{T \times R \times C}$ 。其中 T 、 R 、 S 分别代表输入地震数据在时间轴长度、道数、炮点数。随后，使用四阶段的编码器-解码器网络进行深度特征提取。每个阶段都有一个包含多尺度嵌入编码块和多分支 Transformer 块。使用多尺度嵌入编码以生成多尺度的特征“键”，并将它们分别馈送到多个 Transformer 分支中。

每个 Transformer 分支包含多个 Transformer 编码器。通过在多分支 Transformer 块的末端应用选择性核特征融合模块 (Selective kernel feature fusion, SFF) [160] 来熔断不同分支产生的特征，这是一种利用自注意力机制对多分辨率的融合特征进行非线性处理的过程。使用跳跃连接聚合来自编码器和解码器的信息，并使用 1×1 卷积层对聚合后的特征进行降维（第一阶段除外）。在编码器-解码器结构之后加入一个重建层来恢复速度体中精细的结构，重建层由包含卷积层的一个残差块组成。最后，应用一个 3×3 卷积层来减少通道，并根据速度模型的大小进行截断，从而输出最终反演结果 $I \in \mathbb{R}^{H \times W \times 1}$ ，即估计的初始速度模型。其中，

H 和 W 分别代表速度模型的高和宽。为了压缩计算量和参数,在模型中应用的卷积主要是深度可分离卷积。

接下来,介绍两个核心模块:多尺度嵌入编码、自注意力计算模块。

(1) 多尺度嵌入编码

观测地震数据中记录的波场的尺度变化很大,为了解决这个问题,本研究中设计了一种新的多尺度嵌入编码,它具有捕获多层次语义信息和感受野灵活可变两种特性。具体而言,如图 6.2 中所示,通过递进设计具有不同大小卷积核的多个可变形卷积(Deformable Convolution, DCN)^[161],进而形成多尺度的嵌入编码并生成粗略和精细的特征表示,形成用于自注意力运算的特征“键”。可变形卷积层由几个递进的具有小核的可偏移卷积层堆叠形成,而不是单个具有大核的可偏移卷积层。这不仅能增加网络的深度,也提供了多层次的语义信息,有助于减少参数和计算负担。所有可偏移卷积层后都使用 Hardswish 激活函数^[162]。

考虑到观测数据中信号是局部相关的,嵌入编码时用于形成“键”捕获的特征图多数还是集中在局部,通过偏移卷积也可以同时解决这一问题。设置每一个偏移卷积的偏移量范围为 $[-3, 3]$ 。网络能够根据特征的形状自动通过学习来选择感受野大小。对于偏移卷积来说,其上限为 9×9 ,相当于扩张率为 4 的扩张卷积,下限为 1×1 。当递进设置多尺度嵌入编码时,不同分支的感受野大小逐渐在前一支基础上递增,进而实现对特征的多尺度嵌入编码。

(2) 多头自注意力机制

通过浅层特征提取模块提取的特征在送入到 Transformer 结构之前,需要将输入特征拆分成块,每个块称之为“键”。具体来讲,输入特征 $F_I \in \mathbb{R}^{T \times R \times C}$ 将被拆分为一系列块 $F^i \in \mathbb{R}^{M^2 \times C}$,并将它们展平后进行嵌入编码,其中 C 表示特征块的数量, M 表示特征块尺寸。本研究中设置 M 大小为 1。然后,对编码后的“键”进行自注意力计算。此外,还采用了多头结构,其中头的数量从水平的顶部到底部逐渐增加。假设自注意力头编号为 h_G ,自注意力头的维度为 $d_h = C / h_G$ 。接下来,在每个窗口特征上执行自注意力。第 h 个头的计算过程可以描述如下:

$$F = \{F^1, F^2, \dots, F^N\}, N = \frac{TR}{M^2} \dots\dots\dots (6.1)$$

$$Y_h^i = \text{Attention}(F^i W_h^Q, F^i W_h^K, F^i W_h^V), i = 1, \dots, N \dots\dots\dots (6.2)$$

$$\hat{Y}_h = \{Y_h^1, Y_h^2, \dots, Y_h^M\} \dots\dots\dots (6.3)$$

其中 x 表示该组键的输入集, $\hat{Y}_h$ 是第 h 个头的输出, $X^i W_h^Q, X^i W_h^K, X^i W_h^V \in \mathbb{R}^{C \times d_h}$ 分别表示第 h 个头“查询向量”、“关键向量”和“数值向量”的投影矩阵。在注意力模块中引入了相对位置编码^[138], 以保证每个键的位置相对不会变化, 注意力计算过程可以表示成如下形式:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d_h}} + B\right)V \dots\dots\dots (6.4)$$

其中 Q, K, V 是自注意力计算的输入, B 是相对位置偏移。

6.2.3 混合反演策略研究

本研究中提出的混合反演方法将深度学习反演和全波形反演相结合, 整体结构包括使用 FAIT 来获得初始速度模型, 并使用全波形反演进行细化, 其整体流程如图 6.3 所示。具体来说, 首先通过对速度模型进行正演建模来构建大规模数据集, 以模拟实际地震波的传播, 并通过求解波动方程来获得地震记录。然后, 使用数据集对 FAIT 进行训练, 以借助网络获得比传统方法更准确的初始速度模型。FAIT 被训练直接学习从地震数据到地震速度模型的映射, 如方程所示

$$v = FAIT(d, \Theta) \dots\dots\dots (6.5)$$

其中 $FAIT$ 表示特征对齐反演 Transformer 所拟合的函数, Θ 表示其参数。对于 FAIT 来说, 整个步骤分为两个阶段: 训练阶段和预测阶段。在训练阶段, 网络在监督学习下学习相应的映射。损失函数用于计算速度模型在训练集上的真实值和预测值的偏差。不同的参数 Θ 一般对应于损失函数的不同值, 训练的目标是找到一组使损失函数最小化的参数。在损失函数的选择上选择使用 Charbonnier 损失, 定义如下:

$$L_{Charbonnier} = \sum \sqrt{\|\hat{v} - v\|_2^2 + \epsilon^2} \dots\dots\dots (6.6)$$

其中 $\hat{v}$ 是预测速度, v 是真实值, Σ 表示求和符号, ϵ 在本实验中被设置为常数。在网络训练时, 通过计算预测结果和真实值之间的损失, 以导出梯度并更新网络

参数。通过输入大量的训练数据来迭代更新网络参数，并沿着梯度方向不断减小损失函数的值，最终得到使损失函数值最小的网络参数。在训练阶段使用反向传播方法和梯度下降法来优化网络。具体地，使用的是 Adam 算法^[98]。在多个迭代周期之后，损失将收敛到足够小的值，标志着模型的训练阶段完成，并确定 FAIT 的参数 Θ 。此时，模型已经具备了对训练数据中特征的理解，并能够根据这些学习到的特征进行反演推断。更多训练的细节如表 6.3 所示。

表 6.3 网络超参数设置

超参数	含义	具体设置
批次大小	在训练神经网络时一次输入的样本数量	32
轮次	神经网络在整个训练数据集上反复迭代的次数	1000
学习率范围	控制神经网络权重更新幅度的参数	$[1 \times 10^{-6}, 1 \times 10^{-4}]$
优化器	用于调整模型权重以最小化损失函数的算法	Adam 算法
自注意力头数	多头自注意力机制中的注意力头数	1,2,4,8,16
窗口大小	执行窗口自注意力的窗口大小	16×16
“键”大小	用于形成进行自注意力计算的每个“键”覆盖的范围	1×1

接下来，模型可以进入测试阶段，其性能和泛化能力将在此环节得到验证。在测试阶段，模型将面对之前未曾见过的数据，并利用之前学习到的模式进行预测推断。在测试阶段，当输入观测记录和训练集中观测记录大小一致时，预测速度结果可表示为

$$\tilde{v} = FAIT(d, \Theta) \dots\dots\dots (6.7)$$

然而，实测数据的大小往往要大于训练阶段使用的数据，因此使用滑动窗口技术来处理这类场景，在这种情况下，观测记录被分割成许多的与训练集数据具

有相同输入通道数量的片段, 然后 FAIT 模型分别对每个片段进行处理。最后, 所有片段的处理结果被合并成最终的速度模型的处理结果。假设原始观测记录为 x , 它被分割成 n 个可能重叠的片段即 $X = [X_1, X_2, \dots, X_n]$, 每个片段的通道数与训练模型时使用的通道数相同。对于每个片段 X_i , 模型 FAIT 产生一个输出 Y_i 。设 Y 为最终合并后的速度模型反演结果, 可以表示为

$$Y = \frac{1}{n} \sum_{i=1}^n Y_i \dots\dots\dots (6.8)$$

其中, 对于 Y 之间有重叠的部分, 在计算合并结果 Y 时, 重叠区域的值为所有重叠片段值的平均。步长 b 决定了片段之间的重叠程度, 步长较小意味着重叠更多, 这可以减少边缘效应, 但同时也会增加计算量。通过这一过程, FAIT 能够处理与它在训练时规模不一致的观测记录, 同时也能尽量减少边界效应对最终结果的影响。本研究中设置 b 为 5。

获得初始速度模型 Y 后, 如图 6.3 中紫色箭头标注部分所示, 使用全波形反演方法对其进行进一步优化, 通过进一步最小化观测数据和预测速度模型正演数据之间的差异来迭代更新模型。

为了实现全波形反演, 首先将速度模型离散为网格单元。然后, 使用有限差分法对波动方程进行离散, 并在离散模型中模拟波动的传播。在正向建模步骤中, 在每个源位置应用源小波以生成模拟波场, 并在对应的接收器位置进行接收记录。在反向建模步骤中使用记录的波形来计算失配相对于速度模型的梯度, 然后在下一次迭代中使用该梯度来更新速度模型。对于 FWI 的设置, 使用与用于生成时域前向建模的训练地震数据相同的参数设置。采用多尺度频域反演策略^[166], 在实施 FWI 时选择的反演频率为 2.5 Hz、5 Hz、10 Hz、15 Hz 和 21 Hz。同时, 采用基于伴随状态法的梯度下降方法。FWI 的观测数据与先前用于 FAIT 预测的地震数据相同。由于大量的迭代和在每次迭代时存储波场的需要而需要大量的存储器。为了解决这个问题, 使用低秩近似技术来减少计算成本和内存使用。低秩近似使用奇异值分解将波场分解为低秩矩阵, 这些矩阵可以有效地计算和存储。本研究中, 使用 L-BFGS 算法通过逼近 Hessian 矩阵来加速优化过程。

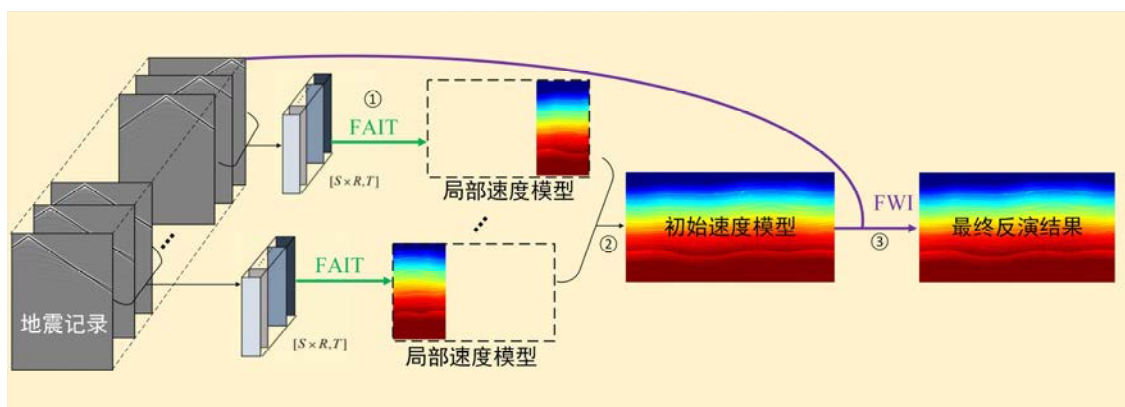


图 6.3 特征对齐的 Transformer 的混合反演框架

6.3 算法性能验证

（1）整体性能

实验中，通过使用 SEAM 模型构建的测试集对基于 FAIT 的混合反演策略进行了检验。实验在配备 Inter Xeon CPU E5-2620 和 4 个 Nvidia V100 GPU 的服务器上运行，实验运行在 Python 环境中，借助 PyTorch 深度学习框架^[141]实现。同时，使用 FCNVMB 和 FWI 作为对比方法。其中，FCNVMB 依旧选择使用通用版本，并遵循其原始论文^[86]中的超参数设置，同时，其模型重新在 SEAM 数据集上进行了训练。对于 FWI，采用时域反演策略和基于伴随状态法的梯度下降方法，并采用高斯平滑函数平滑后的速度模型作为 FWI 初始速度模型。FWI 中震源的频率设定为 20 Hz。FWI 的数值实验在 CPU 上运行，对于每个速度模型，使用 FWI 方法进行 50 次迭代。图 6.4 展示了不同方法的预测结果，其中特意挑选了致密层模型与盐体模型进行比较，它们的速度值范围设定为 1500 至 2500 m/s 之间。在致密层模型的案例实验中，图 6.4（a）展示了真实值，图 6.4（c）展示了混合反演的结果，图 6.4（e）展现了纯 FAIT 的结果，图 6.4（g）展现了 FCNVMB 的结果，而图 6.4（i）则是 FWI 的输出，图 6.4（k）则是 FWI 的初始速度模型。在含有盐丘模型的案例实验中，图 6.4（b）展示了真实值，图 6.4（d）展示了混合反演的结果，图 6.4（f）展现了纯 FAIT 的结果，图 6.4（h）展现了 FCNVMB 的结果，而图 6.4（j）则是 FWI 的输出，图 6.4（l）则是 FWI 的初始速度模型。

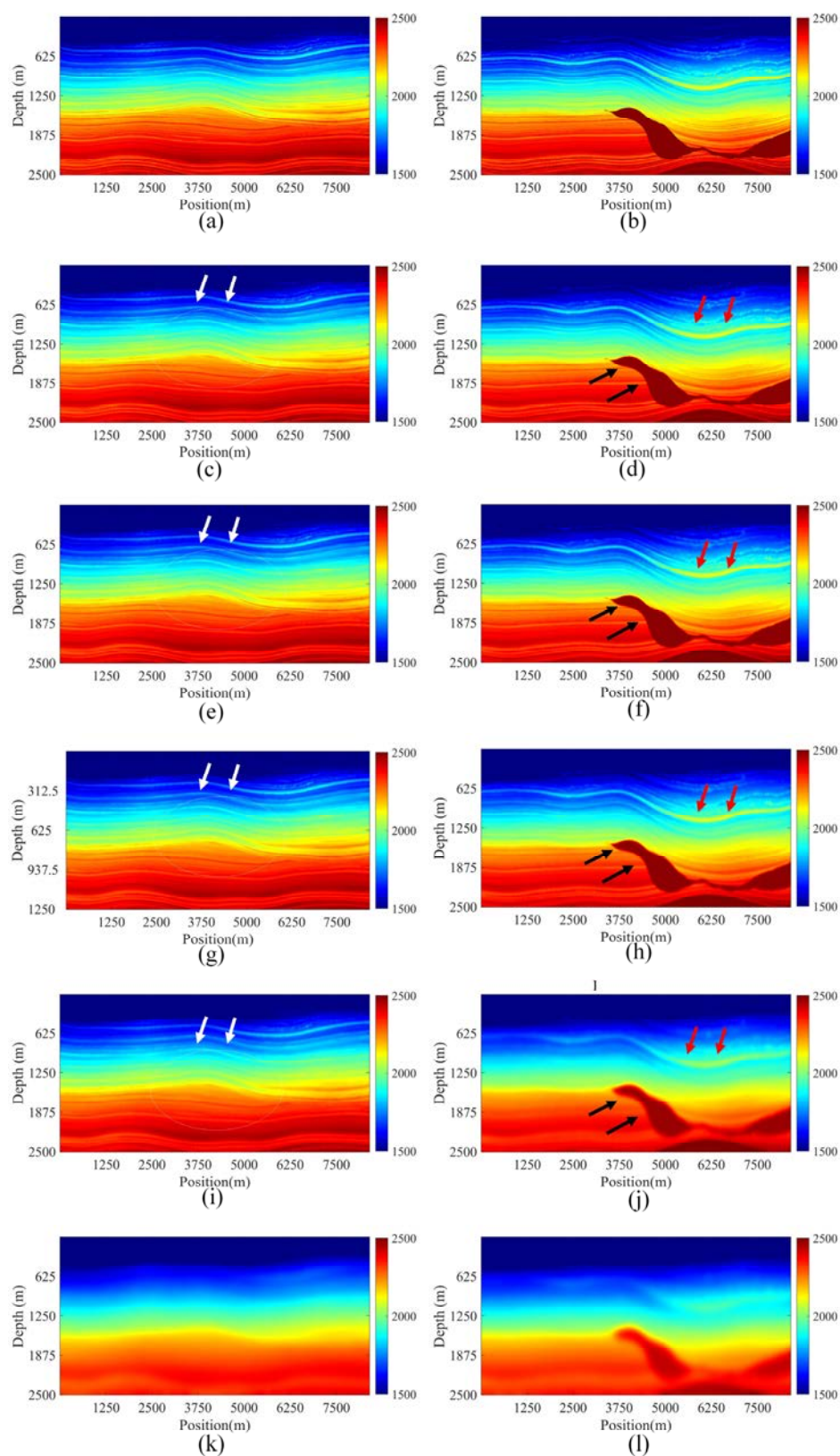


图 6.4 不同方法的反演结果对比 (a) 致密层模型真实值, (b) 盐体模型真实值, (c-d) 混合反演的预测结果, (e-f) FAIT 预测结果, (g-h) FCNVMB 预测结果, (i-g) FWI 预测结果, (k-l) FWI 的初始速度

在致密层速度模型的结果中，可以看出 FAIT 能够基于输入的地震数据直接推导出速度模型，但是其中地层的层位细节往往不清晰。而混合反演方法的结果中各个层位的细节和边缘相对来说要更好一些，且能准确推断出各个地层的速度。观察结果显示，FAIT 在性能上与基于 CNN 的反演技术 FCNVMB 近似，在一些结构细节上略好于 FCNVMB。在盐体速度模型的结果中，两种深度学习方法即 FAIT 和 FCNVMB 都能够很好地恢复出盐体。相比之下，FAIT 的预测结果在速度值以及盐体的恢复方面表现出色。而此时混合反演的结果与只使用 FAIT 获得的结果相差不大。虽然 FCNVMB 能够较好地拟合各层速度，但在对盐体的恢复上效果略差于 FAIT，尤其是在图 6.4 (h) 中，盐体边缘容易出现预测误差，呈现比较模糊的效果。就盐体模型而言，可以注意到在图 6.4 (j) 中 FWI 难以准确恢复盐体的精细结构。由此可以总结出，在两种类型的速度模型案例上，混合反演方法所得结果与真实值是最为接近的。

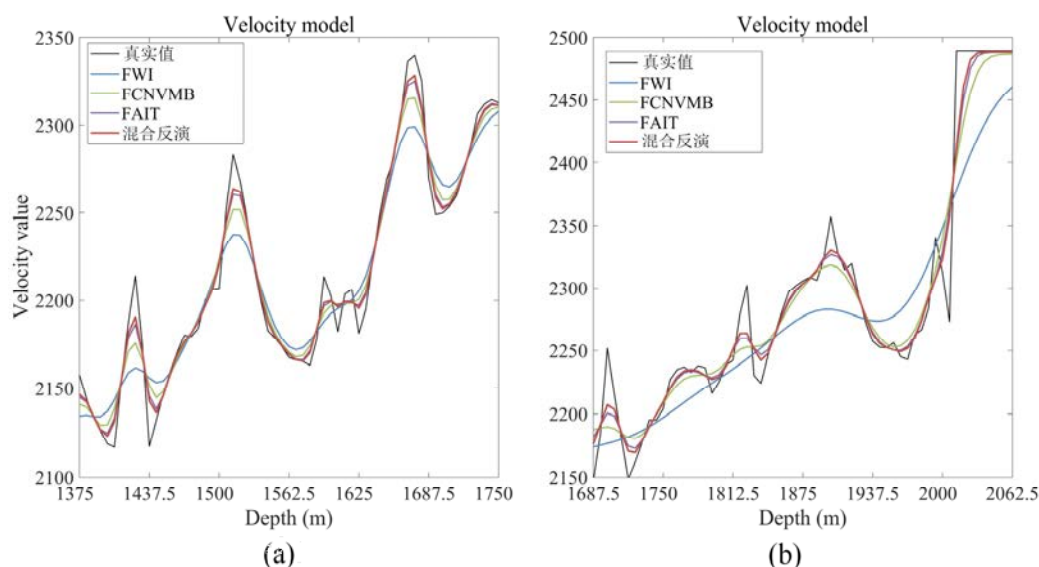


图 6.5 不同方法反演结果的垂直速度剖面对比 (a) 致密层速度模型, (b) 盐体速度模型

进一步地，在图 6.5 中通过绘制单道速度曲线，以加深了对各种方法之间差异的比较。垂直速度剖面的位置选择在 7500m 的水平位置处。致密层模型的结果如图 6.5 (a) 所示，盐丘模型的结果如图 6.5 (b) 所示。在两个图中，展现了深度剖面的真实速度值（以蓝色标出）、混合反演的预测速度值（以红色表示）、FAIT 预测值（以橙色显示）、FCNVMB 预测值（以紫色显示）以及 FWI 预测值

(以绿色表示)。从中可以看出,混合反演方法的单道曲线与真实值是最为接近的。其次最接近的是 FAIT,其接近程度要高于 FCNVMB。

接下来,对测试集中的反演结果进行了统计分析,选择了 MAE、MSE、SSIM 和 MSSIM 四项指标进行评估。更低的 MAE 和 MSE 值,以及更高的 SSIM 和 MSSIM 值意味着方法具有更好的反演性能。量化结果如表 6.4 所示。从这些统计分析结果来看,提出的混合反演方法在所有指标上都有着最好的效果。同时,也可以观察到,与 FWI 和 FCNVMB 的比较中,FAIT 也展现了更优异的性能,这验证了特征对齐后的 Transformer 架构对比以 CNN 为基础的深度学习反演架构,在反演方面有着更高的性能。由于地震数据仅能记录有限的时间范围,较深层次的反射信号较为稀少。同时,信号衰减的影响使得底层信号较弱。这对于采用权重共享机制的 CNN 算法构成了巨大挑战。实际效果表明,CNN 算法对于盐体恢复的结构细节通常存在较大误差。与 CNN 相比,FAIT 能够提取更深层次的较弱信号,进而恢复更深的结构。同时,因为加持有特征对齐特性,网络能够提取更具鲁棒性的特征用于反演,在整体反演性能方面更具优势。

表 6.4 不同方法在测试集上反演结果的量化指标对比

量化指标	混合反演方法	FAIT 方法	FCNVMB 方法	FWI 方法
MAE	0.0387	0.0455	0.0541	0.0614
MSE	0.0102	0.0121	0.0221	0.0130
SSIM	0.8516	0.8279	0.8079	0.4081
MSSIM	0.9989	0.9987	0.9981	0.9978
训练时间	62h 43min	62h 43min	40h 45min	-
测试时间	23h 34min	6.69s	3.56s	23h 52min

（2）计算时间

在地震数据反演技术的比较研究中,探讨计算效率也是不可或缺的一环。不同方法的计算时间如表 6.4 所示。传统的全波形反演技术因其需对每一数据点进行独立计算而耗时较长,在以上的致密层案例中,全波形反演方法可能需要超过二十个小时的处理时间。对比之下,结合了深度学习反演技术和全波形反演的混合反演方法,计算消耗主要在深度学习模型的训练阶段及后续的全波形反演阶段。深度学习反演方法如 FAIT 和 FCNVMB,主要将计算成本集中在模型训练阶段,而在训练完成后的预测阶段,时间成本大幅降低。先后使用深度学习反演和 FWI 方法的混合反演策略能提高预测的准确性,因此,对于计算时间的消耗是可以接受的。通过在初期采用深度学习方法快速获得一个较为准确的地震数据反演模型,再利用 FWI 进行细节上的优化和调整,这种策略兼顾了深度学习的高效率和 FWI 的高精度优势。

此外,比较 FAIT 和 FCNVMB 这两种深度学习方法,可以发现在模型性能上存在差异。对比 FCNVMB,FAIT 能够实现更佳的性能表现,这归功于其高效的自注意力机制及通过使用可形变卷积获得的特征对齐的能力。然而,这种复杂的机制也使得 FAIT 的训练时间要长于单纯基于 CNN 架构的 FCNVMB,一方面因为自注意力机制的计算复杂性更高,使得模型收敛速度相对较慢,另一方面,可形变卷积也增加了一定的计算量,使得整体计算的复杂度进一步上升。

（3）消融研究

在深度学习反演算法的评估中,消融研究起着重要的作用,其通过剔除网络中的特定部件,进而揭示这些组件对模型整体表现的影响。本研究深入探讨了 FAIT 模型中不同构件的作用,特别是自注意力机制和特征对齐模块的贡献。为此,采用 MSE 和 SSIM 这两指标,对比分析了在测试集上去除某些关键组件后模型性能的变化。结果如表 6.5 所示。

首先,本研究聚焦于自注意力机制的消融实验。通过将 FAIT 模型中的自注意力机制模块替换为标准的卷积残差块,同时保留模型的其他部分不变,评估自注意力机制的作用。结果表明,含有自注意力机制的 FAIT 模型明显胜过仅使用卷积残差块的模型,这一对比验证了自注意力机制在提高模型反演性能方面的重

要性。其次，本研究也对特征对齐模块进行了消融实验，通过从模型中移除特征对齐模块，可以观察到模型性能有所下降，这验证了特征对齐模块在优化速度反演任务中的有效性。这充分说明了特征对齐模块可以通过调整不同层次的特征，更好地配合反演任务的需求，进而显著提升模型的整体效能。

表 6.5 消融研究

消融项	SW-MSA	特征对齐模块	MSE	SSIM
FAIT 方法	✓	✗	0.0231	0.7895
	✗	✓	0.0195	0.8005
	✓	✓	0.0121	0.8279

综上所述，通过对自注意力机制和特征对齐模块的消融研究，验证了这两个模块在提升地震数据反演精度方面的关键作用。

（4）实际数据反演结果

在验证混合反演方法的实用性和鲁棒性方面，将算法应用于真实数据是至关重要的一步。接下来，将提出的混合反演方法应用于海洋调查记录的拖缆数据，该数据集使用了 192 次放炮，放炮间隔为 37.5 米。时间采样间隔为 2 ms，记录时间为 10 s，并且每个记录有 480 个水听器数据。水听器间隔为 12.5 米，最小和最大震源-接收器偏移距离分别为 198 米和 6 公里。在将混合反演方法应用于该海洋数据之前，应用了三个数据预处理步骤：

- 1) 将观测数据从三维格式校正为二维格式，并对观测到的地震记录进行振幅缩放和相位校正。
- 2) 对初至前的噪声进行去噪处理。
- 3) 以 1 ms 的采样间隔重新采样数据。

部分处理后的数据如图 6.6 所示，不同方法的反演结果如图 6.7 所示。选择 FWI 作为对比方法，图 6.7（d）中所示的模型被用作 FWI 的初始模型。图 6.7（b）中所示的模型是 FAIT 通过滑动窗口推理的速度模型进行平均后的结果，被

用作混合反演方法中 FWI 阶段的初始模型。在进行真实案例的测试中，并未对 FAIT 模型进行针对实际数据的特别调整，而是直接应用了之前训练好的模型来估计速度模型。通过这种方式，能够评估模型对于复杂的地质结构的泛化能力。测试结果显示，FAIT 模型能够有效地反演出实际模型中的主要地层结构，这验证了 FAIT 本身已经具备出色的反演能力。

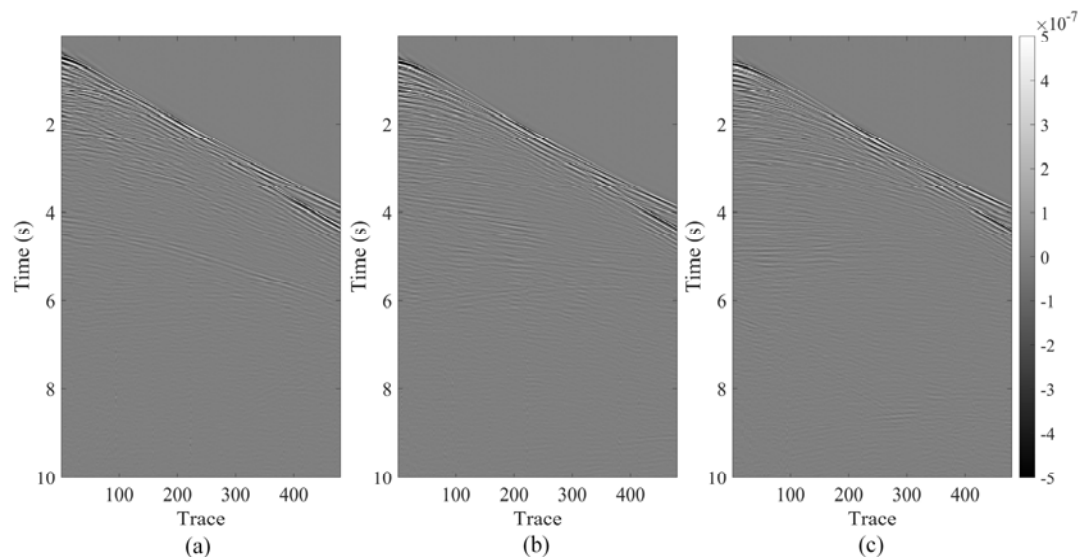


图 6.6 墨西哥湾拖揽数据部分处理后数据 (a) 第 10 炮, (b) 第 60 炮, (c) 第 110 炮

为了继续测试混合反演方法相对于 FWI 的优势，这些初始模型被离散为 401×2112 个网格点，网格间距为 6.25m。图 6.7 (a) 和 6.7 (c) 分别描绘了混合反演方法的结果和 FWI 的结果。与初始速度模型相比，FWI 的结果具有更高的分辨率。混合反演的结果也比 FAIT 结果在细节结构上更加丰富。此外，可以观察到，混合反演结果的分辨率略高于 FWI 的结果，在很多的细节上要更加丰富。

结果表明，混合反演方法在恢复地下结构方面表现要优于 FWI 方法。这一发现验证了深度学习方法在处理真实地震数据时的高效性和准确性。尽管测试的数据分布与训练集不完全相同，导致 FAIT 直接反演结果依旧存在有一定程度的信息缺失，但这些初步成果已经充分展示了通过与全波形反演方法进行联合，数据驱动反演方法在真实地震数据反演中能够起到非常重要的辅助作用。因此，将深度学习反演技术与全波形反演技术进行整合，可能是一条让深度学习技术更有

效投入到实际数据反演的技术路线。

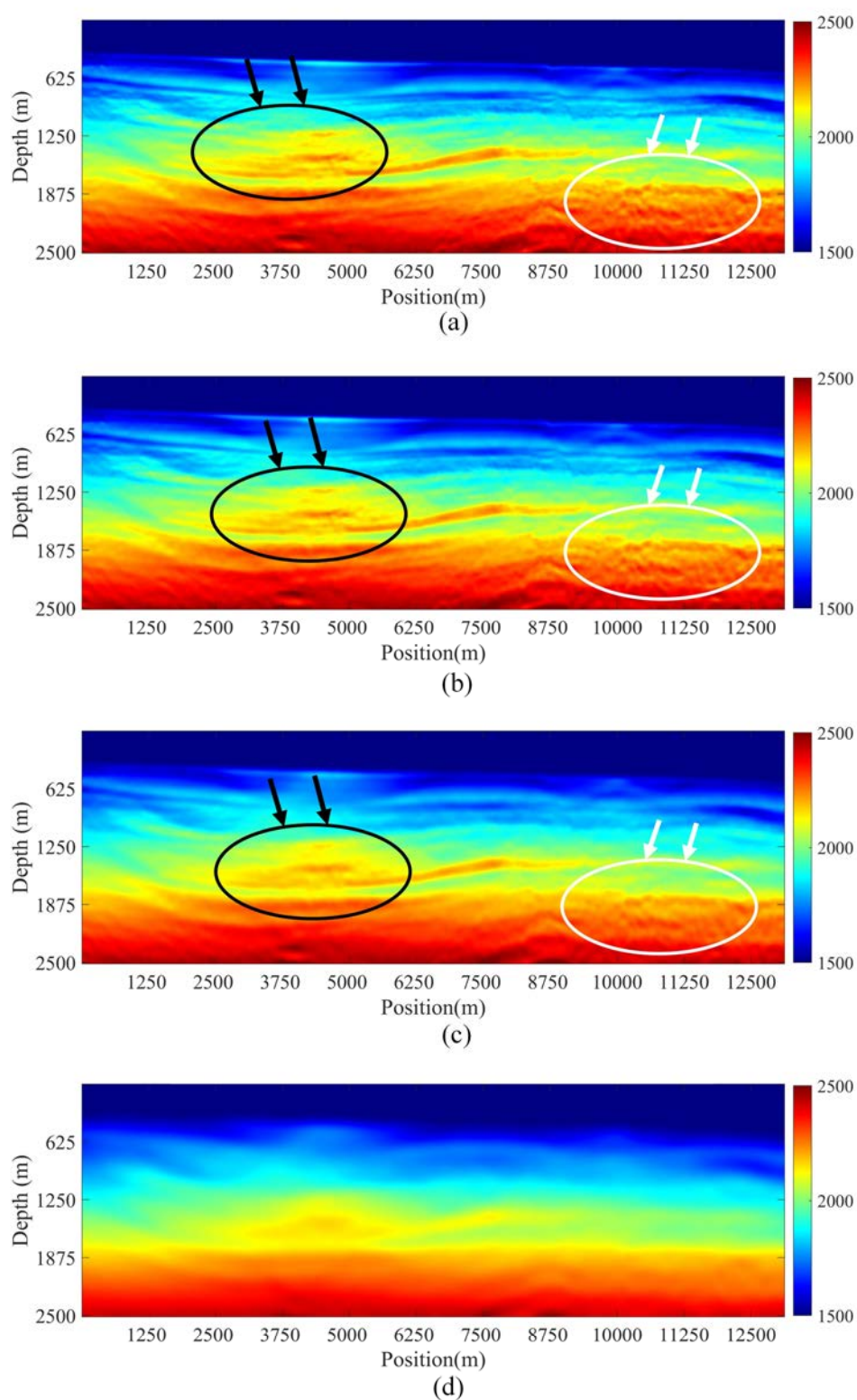


图 6.7 不同方法对墨西哥湾拖揽数据的反演结果 (a) 混合反演方法, (b) FAIT 预测结果, (c) FWI 结果, (d) FWI 初始速度模型

6.4 本章小结

本章综合考虑了深度学习反演算法与传统的全波形反演方法在地震速度模型反演方面的实际应用挑战。尽管深度学习方法在处理复杂地质结构的速度模型反演中展现出巨大潜力,但实际地震数据的复杂性、噪声和不确定性给算法的性能带来了限制。为了解决这些问题,本章提出了一种将深度学习与全波形反演结合的混合反演策略,以求利用两种方法的互补优势提高地下速度模型反演的准确性和效率。

本章中首先分析了深度学习反演面临的主要挑战,包括大量标记数据的需求、实际地震数据的多样性和复杂性,以及如何有效结合深度学习和全波形反演技术。针对特征对齐难题和反演区域大小限制问题,本章分别提出了基于偏移卷积的Transformer反演网络和滑动反演策略。这些策略不仅提高了深度学习网络从复杂数据中准确提取地下速度模型信息的能力,更解决了深度学习模型在处理大范围地震数据的反演任务时遇到的挑战。此外,本章还介绍了一种混合地震反演技术,通过结合深度学习的强大特征提取能力和全波形反演的高精度物理建模优势,显著提升对地下速度模型反演的精度与效率。这种混合策略不仅克服了单独使用深度学习或全波形反演方法的局限性,还能发挥二者的互补优势,为地震数据处理领域提供一种新的解决方案。实验结果表明,本章提出的混合地震反演策略在SEAM案例和墨西哥湾海洋地震实际数据的反演中均取得了较好的成果。通过这一混合反演策略,有效提高了速度模型反演的精度,也为未来数据驱动地震数据反演和物理驱动地震数据反演如何有效结合提供了一种研究方向。

第7章 总结与展望

在探索地球内部结构和搜索地下油气资源的过程中,地震勘探技术扮演着重要的角色。通过分析人工激发的地震波在地下的传播和反射,研究者们能够揭示地层结构的细节。然而,这一过程面临着重大挑战,尤其是地震数据本身经常被各种噪声所污染。虽然传统的去噪技术如频带滤波在一定程度上有效,但它们往往难以应对更复杂的噪声问题,特别是当噪声与信号属性相近时。除了去噪,速度模型反演也是地震数据处理的关键环节,这一过程涉及到通过地震数据来推断地下介质的速度分布。常规的反演方法通常基于迭代优化过程,不仅计算量大,而且容易陷入局部最优解。深度学习模型以其在识别复杂模式和特征上的强大能力,可以被训练来消除噪声并保留尽可能多的有用信号,同时也能被训练学习直接从地震记录推断速度模型。相对于传统基于迭代优化的方法,深度学习技术能够显著减少计算时间并提高反演的准确性。尽管深度学习为地震数据的处理提供了新的可能,但在实际应用中仍然面临着挑战。例如,训练深度学习模型需要大量的标记数据,这在地震数据处理领域往往难以获取。此外,地震数据的复杂性和多样性要求模型具有广泛的特征捕获能力,同时对模型的训练还需要保证其在实际数据中的泛化性。为了解决以上问题,本文基于深度学习开展对地震数据去噪和速度模型反演的研究,提升模型在实际数据上的泛化能力,以应对地震数据处理中的挑战。本文的主要研究工作和贡献如下:

(1) 在地震数据处理领域,提高记录的信噪比是一项关键任务。地震资料中噪声和信号的分布通常差异性大,不同区域的噪声水平可能存在显著区别,且信号强度变化范围大。常规的 CNN 在处理这类记录时往往性能不佳。本研究中通过从微观和宏观两个层面上引入复杂的密集连接,充分加强网络对特征的复用性。这种架构通过将每一层的特征传递至后续层,实现了低级和高级特征的有效组合,优化了地震信号的重建。此外,这些密集连接还为网络内部提供了短路径,有效缓解了梯度消失的问题,有利于深层网络的训练过程。经过在“含噪数据-纯净数据”配对训练集上的训练,该方法在处理模拟与实际地震数据方面都展现出优异的去噪效果,在模拟地震记录上该方法能够实现 24.0578 dB 的信噪比提升,

且在实际 OBC 数据的处理中,该方法能有效压制噪声,恢复不同噪声水平区域的同相轴信号。

(2) 在极低信噪比条件下执行地震数据去噪任务时, CNN 往往因难以充分提取远程特征而对弱信号的恢复效果较差。针对 CNN 在捕获地震数据远程相关性和全局一致性方面的局限,提出了基于 Transformer 架构的去噪模型,通过自注意力机制,该模型有效解决 CNN 模型在特征提取方面的限制,有效捕获信号的长距离依赖关系。为了充分发挥模型在实际地震数据处理中的作用和潜力,提升该方法在实际数据去噪中的泛化性能,采用了掩码自监督预训练和微调训练相结合的训练方法,使用无标注的大规模地震数据集,通过掩码自修复的训练方式使得模型在预训练阶段有效学习实际地震数据的分布,后续通过含噪与纯净配对数据的微调,进而提升模型的去噪性能与在实际数据中的泛化性能。通过在模拟数据上进行测试,该方法在极低信噪比的模拟地震记录上能够实现平均 21.53 dB 的信噪比提升。在多种实际记录的处理中,该方法在降噪和维持信号完整性方面也具有优越性。

(3) 除了应用于去噪任务外,深度学习方法还可以直接以原始地震数据为输入直接推断速度模型,但是多数基于 CNN 的方法在捕获长距离特征方面受限。此外,地震数据的时变特性也不利于权重共享的 CNN 架构充分提取特征。本研究中提出一种基于 Transformer 的新型反演模型,网络被设计为编解码结构,并通过自注意力机制有效捕获远程相关性,为速度模型构建提供了更加丰富的信息。通过数值实验验证了该方法的有效性,经过合成数据的测试,验证了该方法在提高速度估计准确性、恢复地下结构细节等方面,相较于传统全波形反演和基于 CNN 的深度学习反演方法有显著的提升。

(4) 传统地震反演方法,如全波形反演,通常高度依赖于初始速度模型的准确性,而基于深度学习的算法已经展现了在生成初始速度模型方面的潜力。本研究中提出了一种结合深度学习(数据驱动)与全波形反演(物理驱动)的混合反演方法。在结合两种方法的过程中,针对特征不对齐问题及反演区域限制的问题,本研究采用可变形卷积构建了特征对齐的 Transformer 网络,通过滑动窗口局部反演策略,推断原始地震数据的初始速度模型,并结合全波形反演进行迭代

优化,该策略有效整合了两种方法的优势,显著提高了反演准确度,尤其是在地质结构复杂的区域。在合成数据的应用验证了该混合策略的有效性和准确性,同时,该方法也成功拓展应用至实际海洋拖揽数据的速度反演中。

综上,本文采用递进的方式,实现了基于深度学习对地震数据的去噪及速度模型反演算法研究,并在实际地震数据的去噪和反演中都成功应用。然而,地震数据处理及反演研究中仍有一些研究有待展开。基于本文的研究进展,拟进一步开展的研究工作如下:

(1) 基于深度学习的去噪与反演的联合研究。在本研究中,对去噪任务和反演任务分别作为独立的任务进行了详细研究,但是未对两者之间的联合展开深入研究。去噪与反演在地震数据处理中是相互依赖、紧密相关的两个环节,良好的去噪处理是进行准确反演的前提,而有效的反演方法又能进一步验证去噪成果的有效性。因此,后续研究将尝试开发一种集成式深度学习框架,实现去噪与速度模型反演的联合优化。此外,还将探讨不同类型的深度学习模型在联合处理框架中的应用效果,以及如何通过调整网络结构和参数设置来最大化去噪和反演任务的协同效应,从而提高整体处理流程的效率和准确性。

(2) 更为高效和自动化且减少人工干预的数据预处理方法。地震数据的质量直接影响深度学习模型的去噪效果和反演结果的准确性。然而,高质量的地震数据获取成本高昂,且数据常常包含复杂的噪声成分,这对数据预处理提出了更高的要求。未来的研究需要开发更为高效和自动化的数据预处理方法,以减少人工干预,提高数据处理效率。例如,可以利用无监督学习算法,自动识别并去除噪声,提取有用信号。

(3) 在大数据背景下,解决模型面对新数据时的泛化问题。深度学习模型在特定数据集上可能表现优异,但将其应用于不同地质条件下的数据时,模型如何在实际数据上保持泛化性能仍然是一个挑战。提高模型的泛化能力需要大量多样化的地震数据进行训练,这在实际中往往难以实现。因此,探索新的训练策略也是能够缓解这一问题的办法。在本研究中,使用的方法包括有自监督预训练及微调训练等。未来将探索更多新的训练策略,如提示学习等,使模型能够在有限的训练样本上学习到更广泛的数据分布,从而提高在新数据中的泛化能力。同时,大语言模型领域在近几年的辉煌研究成就也启示着研究者将“语言/文本”这一

项凝结着人类经验与智慧的要素融入到当前的研究中,去极大释放深度学习模型在地球物理研究中的“通用智能”。

(4) 探寻模型运行的解释性方法。深度学习模型尽管在地震数据处理中取得了良好的效果,但其“黑盒”特性使得模型的决策过程缺乏透明性。通过结合模型可视化技术和解释性建模方法,可以帮助更多研究者更好地评估模型预测的可靠性。因此,提高模型的解释性,使研究者能够理解模型的决策依据,也是未来研究的重要方向。

参考文献

- [1] 何樵登. 地震勘探原理和方法[M]. 北京: 地质出版社, 1980.
- [2] 陆孟基. 地震勘探原理[M]. 北京: 石油工业出版社, 1993.
- [3] 李录明, 李正文. 地震勘探原理、方法和解释[M]. 北京: 地质出版社, 2007.
- [4] 李庆忠. 走向精确勘探的道路 高分率地震勘探系统工程剖析[M]. 北京: 石油工业出版社, 1993.
- [5] 牟永光. 地震勘探资料数字处理方法[M]. 北京: 石油工业出版社, 1982.
- [6] 熊翥. 地震数据数字处理应用技术[M]. 北京: 石油工业出版社, 1993.
- [7] SIRGUE L, BARKVED O I, DELLINGER J, et al. Thematic set: Full waveform inversion: The next leap forward in imaging at Valhall[J]. First Break, 2010, 28(4).
- [8] 陈彦虎. 地震波形指示反演方法, 原理及其应用[M]. 科学出版社, 2021.
- [9] 王家映. 地球物理反演理论[M]. 高等教育出版社, 2002.
- [10] 张超. 微弱地震信号时频消噪方法研究[D]. 长春: 吉林大学, 2016.
- [11] 李光辉. 陆地地震勘探随机噪声建模与分析[D]. 长春: 吉林大学, 2016.
- [12] 吴宁. 基于 Trace 变换理论的地震勘探资料面波及随机噪声消减算法研究[D]. 长春: 吉林大学, 2011.
- [13] 张超. 微弱地震信号时频消噪方法研究[D]. 长春: 吉林大学, 2016.
- [14] MA H T, QIAN Z B, LI Y. Noise reduction for desert seismic data using spectral kurtosis adaptive bandpass filter[J]. Acta Geophysica, 2019, 67(1): 123-131.
- [15] TIAN Y N, LI Y, YANG B J. Variable-eccentricity hyperbolic-trace TFPF for seismic random noise attenuation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(10): 6449-6458.
- [16] ZHAO X, LI Y, ZHUANG G H, et al. 2-D TFPF based on contourlet transform for seismic random noise attenuation[J]. Journal of Applied Geophysics, 2016, 129: 158-166.
- [17] 李月, 林红波, 杨宝俊, 等. 强随机噪声条件下时窗类型局部线性化对 TFPF 技术的影响[J]. 地球物理学报, 2009, 52(7): 1899-1906.
- [18] GOMEZ J L, VELIS D R. A simple method inspired by empirical mode decomposition for denoising seismic data[J]. Geophysics, 2016, 81(6): V403-V413.
- [19] HAN J, VAN DER BAAN M. Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding[J]. Geophysics, 2015, 80(6): KS69-KS80.
- [20] Li J, Li Y, Li Y, et al. Downhole microseismic signal denoising via empirical wavelet transform and adaptive thresholding[J]. Journal of Geophysics and Engineering, 2018, 15(6): 2469-2480.
- [21] YU S, MA J. Complex variational mode decomposition for slop-preserving denoising[J].

- IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(1), 586-597.
- [22] ZHOU Y, ZHU Z. A hybrid method for noise suppression using variational mode decomposition and singular spectrum analysis[J]. Journal of Applied Geophysics, 2019, 161: 105-115.
- [23] MOSTAFA MOUSAVI S, LANGSTON C A. Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding[J]. Bulletin of the Seismological Society of America, 2016, 106(4): 1380-1393.
- [24] VARGAS R N, CLAUDIO PASCHOARELLI VEIGA A. Seismic trace noise reduction by wavelets and double threshold estimation[J]. IET Signal Processing, 2017, 11(9), 1069-1075.
- [25] KARBALAALI H, JAVAHERIAN A, DAHLKE S, et al. Seismic channel edge detection using 3D shearlets—a study on synthetic and real channelised 3D seismic data[J]. Geophysical Prospecting, 2018, 66(7): 1272-1289.
- [26] KANGHUI GUO, DEMETRIO LABATE. Sparse multidimensional representation using shearlets[J]. Wavelets XI, 2008.
- [27] SHAN H, MA J, YANG H. Comparisons of wavelets, contourlets and curvelets in seismic denoising[J]. Journal of Applied Geophysics, 2009, 69(2): 103-115.
- [28] HENNENFENT G, FENELON L, HERRMANN F J. Nonequispaced curvelet transform for seismic data reconstruction; a sparsity-promoting approach[J]. Geophysics, 2010, 75(6): WB203-WB210.
- [29] GORZCAYK A, MALINOWAKI M, BELLEFLEUR G. Enhancing 3D post-stack seismic data acquired in hardrock environment using 2D curvelet transform[J]. Geophysical Prospecting, 2015, 63(4): 903-918.
- [30] CAO J, ZHAO J, HU Z. 3D seismic denoising based on a low-redundancy curvelet transform[J]. Journal of Geophysics and Engineering, 2015, 12(4): 566-576.
- [31] FOMEL, Sergey, Ma, et al. Double-sparsity dictionary for seismic noise attenuation[J]. Geophysics Journal of the Society of Exploration Geophysicists, 2016.
- [32] CHENG J, CHEN K, SACCHI M D. Application of robust principal component analysis (RPCA) to suppress erratic noise in seismic records[C]. 86th Annual International Meeting, SEG, Expanded Abstracts, 2015, 4646–4651.
- [33] CADZOW J A. Signal enhancement—a composite property mapping algorithm[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1988, 36(1): 49-62.
- [34] OROPEZA V, SACCHI M. Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis[J]. Geophysics, 2011, 76(3): V25-V32.
- [35] CHEN K, SACCHI M D. Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation[J]. Geophysical Prospecting, 2017, 65(3): 650-668.
- [36] HARRIS P E, WHITE R E. Improving the performance of f-x prediction filtering at low signal-to-noise ratios[J]. Geophysical Prospecting, 1997, 45(2): 269-302.
- [37] NAGHIZADEH M, SACCHI M D. Multicomponent f-x seismic random noise attenuation

- via vector autoregressive operators[J]. *Geophysics*, 2012, 77(2): V91-V99.
- [38] LIU G, CHEN X. Noncausal f-x-y regularized nonstationary prediction filtering for random noise attenuation on 3D seismic data[J]. *Journal of Applied Geophysics*, 2013, 93: 60-66.
- [39] GULUNAY N. Signal leakage in f -x deconvolution algorithms[J]. *Geophysics*, 2017, 82(5): W31-W45.
- [40] MENG F L, LI Y, LIU Y P, et al. Application of the radon-FCL approach to seismic random noise suppression and signal preservation[J]. *Journal of Geophysics and Engineering*, 2016, 13(4): 549-558.
- [41] SABBIONE J I, SACCHI M D, VELIS D R. Radon transform-based microseismic event detection and signal-to-noise ratio enhancement[J]. *Journal of Applied Geophysics*, 2015, 113: 51-63.
- [42] YU S W, MA J W, ZHANG X Q, et al. Interpolation and denoising of high-dimensional seismic data by learning a tight frame[J]. *Geophysics*, 2015, 80(5): V119-V132.
- [43] 唐欢欢, 毛伟建, 詹毅. 3D 高阶抛物 Radon 变换在不规则地震数据保幅重建中的应用[J]. *地球物理学报*, 2020, 63(9): 3452-3464.
- [44] 王升超, 韩立国, 巩向博. 基于各向异性 Radon 变换的叠前地震数据重建[J]. *石油物探*, 2016, 55(6): 808-815.
- [45] WU N, LI Y, YANG B J. Applications of the trace transform in surface wave attenuation on seismic records[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(12): 4997-5007.
- [46] 王思理, 胡天跃, 王润秋. 利用数学形态学滤波方法压制多次波[J]. *地球物理学进展*, 2017, 32(2): 729-736.
- [47] LIU Y, YAN Z. Application of a cascading filter implemented using morphological filtering and time-frequency peak filtering for seismic signal enhancement[J]. *Geophysical Prospecting*, 2020, 68(6): 1727-1741.
- [48] HUANG W, WANG R, ZHANG D, et al. Mathematical morphological filtering for linear noise attenuation of seismic data[J]. *Geophysics*, 2017, 82(6): V369-V384.
- [49] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv:1409.1556*, 2014.
- [50] 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.
- [51] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[J]. *IEEE*, 2016.
- [52] HUANG G, LIU Z, PLEISS G, et al. Convolutional networks with dense connectivity[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2019, 44(12): 8704-8716.
- [53] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J].

- Advances in Neural Information Processing Systems, 2014, 27.
- [54] SONG A, WANG C, ZHANG C, et al. Regeneration-Constrained Self-Supervised Seismic Data Interpolation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-10.
- [55] MENG F, FAN Q Y, LI Y. Self-supervised learning for seismic data reconstruction and denoising[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.
- [56] ZHANG K, ZUO W, CHEN Y, et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.
- [57] ZHAO Y, LI Y, DONG X, et al. Low-frequency noise suppression method based on improved DnCNN in desert seismic data[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16(5): 811-815.
- [58] ZHONG T, CHENG M, DONG X, et al. Seismic random noise suppression by using deep residual U-Net[J]. Journal of Petroleum Science and Engineering, 2022, 209: 109901.
- [59] WANG H, LI Y, DONG X. Generative adversarial network for desert seismic data denoising[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(8): 7062-7075.
- [60] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [61] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [62] DONG X, YU Z, CAO W, et al. A survey on ensemble learning[J]. Frontiers of Computer Science, 2020, 14: 241-258.
- [63] WEISS K, KHOSHGOFTAAR T M, WANG D D. A survey of transfer learning[J]. Journal of Big data, 2016, 3(1): 1-40.
- [64] MITTET R, SOLLIE R, HOKSTAD K. Prestack depth migration with compensation for absorption and dispersion[J]. Geophysics, 1995, 60(5): 1485-1494.
- [65] BAYSAL E, KOSLOFF D D, SHERWOOD J W C. Reverse time migration[J]. Geophysics, 1983, 48(11): 1514-1524.
- [66] AL-YAHYA K. Velocity analysis by iterative profile migration[J]. Geophysics, 1989, 54(6): 718-729.
- [67] CHIAO L Y, KUO B Y. Multiscale seismic tomography[J]. Geophysical Journal International, 2001, 145(2): 517-527.
- [68] LAILLY P, BEDNAR J. The seismic inverse problem as a sequence of before stack migrations[C], Conference on inverse scattering: theory and application. 1983: 206-220.
- [69] TAKOUGANG E M T, BOUZIDI Y. Imaging high-resolution velocity and attenuation structures from walkaway vertical seismic profile data in a carbonate reservoir using visco-acoustic waveform inversion[J]. Geophysics, 2018, 83(6): B323-B337.
- [70] MORA P. Nonlinear two-dimensional elastic inversion of multioffset seismic data[J].

- Geophysics, 1987, 52(9): 1211-1228.
- [71] CHEN Y, CHEN H, XIANG K, et al. Geological structure guided well log interpolation for high-fidelity full waveform inversion[J]. Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, 2016, 207(2): 1313-1331.
- [72] BUNKS C, SALECK F M, ZALESKI S, et al. Multiscale seismic waveform inversion[J]. Geophysics, 1995, 60(5): 1457-1473.
- [73] PRATT R G. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model[J]. Geophysics, 1999, 64(3): 888-901.
- [74] PRATT R G. INVERSE THEORY APPLIED TO MULTI - SOURCE CROSS - HOLE TOMOGRAPHY.: PART 2: ELASTIC WAVE - EQUATION METHOD1[J]. Geophysical Prospecting, 1990, 38(3): 311-329.
- [75] SHIN C, CHA Y H. Waveform inversion in the Laplace domain[J]. Geophysical Journal International, 2008, 173(3): 922-931.
- [76] SHIN C, HO CHA Y. Waveform inversion in the Laplace—Fourier domain[J]. Geophysical Journal International, 2009, 177(3): 1067-1079.
- [77] YANG Y, LI Y, LIU T. 1D viscoelastic waveform inversion for Q structures from the surface seismic and zero-offset VSP data[J]. Geophysics, 2009, 74(6): WCC141-WCC148.
- [78] HU W, ABUBAKAR A, HABASHY T M. Simultaneous multifrequency inversion of full-waveform seismic data[J]. Geophysics, 2009, 74(2): R1-R14.
- [79] KHAN A, CONNOLLY J A D, MACLENNAN J, et al. Joint inversion of seismic and gravity data for lunar composition and thermal state[J]. Geophysical Journal International, 2007, 168(1): 243-258.
- [80] YU S, MA J, WANG W. Deep learning for denoising[J]. Geophysics, 2019, 84(6): V333-V350.
- [81] ARAYA-POLO M, FARRIS S, FLOREZ M. Deep learning-driven velocity model building workflow[J]. The Leading Edge, 2019, 38(11): 872a1-872a9.
- [82] WU X, LIANG L, SHI Y, et al. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation[J]. Geophysics, 2019, 84(3): IM35-IM45.
- [83] RÖTH G, TARANTOLA A. Neural networks and inversion of seismic data[J]. Journal of Geophysical Research: Solid Earth, 1994, 99(B4): 6753-6768.
- [84] ARAYA-POLO M, JENNINGS J, ADLER A, et al. Deep-learning tomography[J]. The Leading Edge, 2018, 37(1): 58-66.
- [85] WU Y, LIN Y. InversionNet: A real-time and accurate full waveform inversion with CNNs and continuous CRFs[J]. arXiv preprint arXiv:1811.07875, 2018.
- [86] YANG F, MA J. Deep-learning inversion: A next-generation seismic velocity model building method[J]. Geophysics, 2019, 84(4): R583-R599.
- [87] WANG L, MENG D, WU B. Seismic inversion via closed-loop fully convolutional residual

- p network and transfer learning[J].
- Geophysics*
- , 2021, 86(5): R671-R683.
-
- [88] LI S, LIU B, REN Y, et al. Deep-learning inversion of seismic data[J].
- IEEE Transactions on Geoscience and Remote Sensing*
- , vol. 58, (3), pp. 2135-2149, 2020.
-
- [89] WANG W, MA J. Velocity model building in a crosswell acquisition geometry with image-trained artificial neural networks[J].
- Geophysics*
- , 2020, 85(2): U31-U46.
-
- [90] ZHANG Z, LIN Y. Data-driven seismic waveform inversion: A study on the robustness and generalization[J].
- IEEE Transactions on Geoscience and Remote sensing*
- , 2020, 58(10): 6900-6913.
-
- [91] CANALES, LUIS L. Random noise reduction[J].
- SEG Technical Program Expanded Abstracts*
- , 1984:329-329.
-
- [92] ANDRZEJ, GÓRSZCZYK, ANNA, et al. Application of curvelet denoising to 2D and 3D seismic data - Practical considerations[J].
- Journal of Applied Geophysics*
- , 2014.
-
- [93] NEELAMANI R, BAUMSTEIN A I, GILLARD D G, et al. Coherent and random noise attenuation using the curvelet transform[J].
- The Leading Edge*
- , 2008, 27(2): 240-248.
-
- [94] CHEN Y, FOMEL S. EMD-Seislet Transform[J].
- Society of Exploration Geophysicists*
- , 2015.
-
- [95] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J].
- Science*
- , 2006, 313(5786): 504-507.
-
- [96] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J].
- JMLR.org*
- , 2015.
-
- [97] MA H, YAO H, LI Y, et al. Deep Residual Encoder-Decoder Networks for Desert Seismic Noise Suppression[J].
- IEEE Geoscience and Remote Sensing Letters*
- , 2019, PP(99):1-5.
-
- [98] KINGMA D P, BA J. Adam: A method for stochastic optimization[J].
- arXiv preprint arXiv:1412.6980*
- , 2014.
-
- [99] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J].
- IEEE Transactions on Image Processing*
- , 2004, 13(4): 600-612.
-
- [100] WANG J, SUN K, CHENG T, et al. Deep high-resolution representation learning for visual recognition[J].
- IEEE Transactions on Pattern Analysis and Machine Intelligence*
- , 2020, 43(10): 3349-3364.
-
- [101] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J].
- IEEE Transactions on Pattern Analysis and Machine Intelligence*
- , 2015, 38(2): 295-307.
-
- [102] 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.
-
- [103] TSAI K C, HU W, WU X, et al. First-break automatic picking with deep semisupervised learning neural network[M].
- SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists*
- , 2018: 2181-2185.
-
- [104] ZHANG G, LIN C, CHEN Y. Convolutional neural networks for microseismic waveform

- classification and arrival picking[J]. *Geophysics*, 2020, 85(4): WA227-WA240.
- [105] YUAN P, HU W, WU X, et al. First arrival picking using U-net with Lovasz loss and nearest point picking method[M]. *SEG Technical Program Expanded Abstracts 2019*. Society of Exploration Geophysicists, 2019: 2624-2628.
- [106] ZHANG G, WANG Z, CHEN Y. Deep learning for seismic lithology prediction[J]. *Geophysical Journal International*, 2018, 215(2): 1368-1387.
- [107] LI Y, ALKHALIFAH T, ZHANG Z. Deep-learning assisted regularized elastic full waveform inversion using the velocity distribution information from wells[J]. *Geophysical Journal International*, 2021, 226(2): 1322-1335.
- [108] LI Y, ALKHALIFAH T, GUO Q. Target-oriented time-lapse waveform inversion using deep learning-assisted regularization[J]. *Geophysics*, 2021, 86(4): R485-R495.
- [109] LI Y, ALKHALIFAH T. Target-oriented time-lapse elastic full-waveform inversion constrained by deep learning-based prior model[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-12.
- [110] WANG B, ZHANG N, LU W, et al. Deep-learning-based seismic data interpolation: A preliminary result[J]. *Geophysics*, 2019, 84(1): V11-V20.
- [111] SUN J, INNANEN K A, HUANG C. Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis[J]. *Geophysics*, 2021, 86(3): R303-R317.
- [112] SUN J, NIU Z, INNANEN K A, et al. A theory-guided deep-learning formulation and optimization of seismic waveform inversion[J]. *Geophysics*, 2020, 85(2): R87-R99.
- [113] DAHLKE T, ARAYA-POLO M, ZHANG C, et al. Predicting geological features in 3D seismic data[J]. *Advances in Neural Information Processing Systems (NIPS)*, 2016, 29.
- [114] FAN L, YAN-JIE Z, XI-JUN H, et al. Full waveform inversion based on deep learning and optimal nearly analytic discrete method[J]. *Applied Geophysics*, 2021, 18(4): 483-498.
- [115] GENG Z, ZHAO Z, SHI Y, et al. Deep learning for velocity model building with common-image gather volumes[J]. *Geophysical Journal International*, 2022, 228(2): 1054-1070.
- [116] ZHAN C, DAI Z, SOLTANIAN M R, et al. Stage - wise stochastic deep learning inversion framework for subsurface sedimentary structure identification[J]. *Geophysical research letters*, 2022, 49(1): e2021GL095823.
- [117] ZHONG T, CHENG M, DONG X, et al. Seismic random noise attenuation by applying multiscale denoising convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-13.
- [118] YILMAZ Ö. Seismic data analysis: Processing, inversion, and interpretation of seismic data[M]. Society of exploration geophysicists, 2001.
- [119] COOPER H W, COOK R E. Seismic data gathering[J]. *Proceedings of the IEEE*, 1984, 72(10): 1266-1275.
- [120] SPIKES K T, TISATO N, HESS T E, et al. Comparison of geophone and surface-deployed distributed acoustic sensing seismic data[J]. *Geophysics*, 2019, 84(2): A25-A29.

- [121] SEARS T J, SINGH S C, BARTON P J. Elastic full waveform inversion of multi - component OBC seismic data[J]. Geophysical Prospecting, 2008, 56(6): 843-862.
- [122] DONG X, ZHONG T, LI Y. New suppression technology for low-frequency noise in desert region: The improved robust principal component analysis based on prediction of neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(7): 4680-4690.
- [123] ZHOU C, XIA J, PANG J, et al. Near-surface geothermal reservoir imaging based on the customized dense seismic network[J]. Surveys in Geophysics, 2021, 42: 673-697.
- [124] BINDER G, TITOV A, LIU Y, et al. Modeling the seismic response of individual hydraulic fracturing stages observed in a time-lapse distributed acoustic sensing vertical seismic profiling survey[J]. Geophysics, 2020, 85(4): T225-T235.
- [125] WANG S, LI Y, ZHAO Y. Attribute-guided target data separation network for DAS VSP data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-16.
- [126] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [127] SAAD O M, OBOUE Y A S I, BAI M, et al. Self-Attention Deep Image Prior Network for Unsupervised 3D Seismic Data Enhancement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021.
- [128] HE K, CHEN X, XIE S, et al. Masked autoencoders are scalable vision learners[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 16000-16009.
- [129] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]. European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.
- [130] CAO H, WANG Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]. European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 205-218.
- [131] CHEN H, WANG Y, GUO T, et al. Pre-trained image processing transformer[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 12299-12310.
- [132] LI Y, ZHANG K, CAO J, et al. Localvit: Bringing locality to vision transformers[J]. arXiv preprint arXiv:2104.05707, 2021.
- [133] WU H, XIAO B, CODELLA N, et al. Cvt: Introducing convolutions to vision transformers[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 22-31.
- [134] 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.
- [135] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelus)[J]. arXiv preprint

- arXiv:1606.08415, 2016.
- [136] BA J L, KIROS J R, HINTON G E. Layer normalization[J]. arXiv preprint arXiv:1607.06450, 2016.
- [137] XIONG R, YANG Y, HE D, et al. On layer normalization in the transformer architecture[C]. International Conference on Machine Learning. PMLR, 2020: 10524-10533.
- [138] SHAW P, USZKOREIT J, VASWANI A. Self-attention with relative position representations[J]. arXiv preprint arXiv:1803.02155, 2018.
- [139] ZINKEVICH M. Online convex programming and generalized infinitesimal gradient ascent[C]. Proceedings of the 20th international conference on machine learning (icml-03). 2003: 928-936.
- [140] SHAMIR O, ZHANG T. Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes[C]. International conference on machine learning. PMLR, 2013: 71-79.
- [141] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [142] LOEWENTHAL D, LEE S S, GARDNER G H F. Deterministic estimation of a wavelet using impedance type technique[J]. Geophysical Prospecting, 1985, 33(7): 956-969.
- [143] LI J, FAN W, LI Y, et al. Low-frequency noise suppression in desert seismic data based on an improved weighted nuclear norm minimization algorithm[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(11): 1993-1997.
- [144] TIAN Y, SUI J, LI Y, et al. A novel iterative PA-MRNet: Multiple noise suppression and weak signals recovery for downhole DAS Data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.
- [145] GU J, DONG C. Interpreting Super-Resolution Networks with Local Attribution Maps[C]. Computer Vision and Pattern Recognition. IEEE, 2021.
- [146] KOMATITSCH D, TROMP J. A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation[J]. Geophysical Journal International, 2003, 154(1): 146-153.
- [147] WANG Z, CUN X, BAO J, et al. Uformer: A general u-shaped transformer for image restoration[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 17683-17693.
- [148] CHARBONNIER P, BLANC-FERAUD L, AUBERT G, et al. Two deterministic half-quadratic regularization algorithms for computed imaging[C]. Proceedings of 1st international conference on image processing. IEEE, 1994, 2: 168-172.
- [149] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. IEEE, 2003, 2: 1398-1402.
- [150] BARR F J, SANDERS J I. Attenuation of water-column reverberations using pressure and

- velocity detectors in a water-bottom cable[J]. Seg Technical Program Expanded Abstracts, 1989:1375.
- [151] BALE R. Plane wave deghosting of hydrophone and geophone OBC data[J]. Seg Technical Program Expanded Abstracts, 1949, 17(1):2092.
- [152] BALL V, CORRIGAN D. Dual-sensor summation of noisy ocean-bottom data[J]. Seg Technical Program Expanded Abstracts, 1996:28.
- [153] P. J B F, JUINITI R, MORAIS M G G. Water depth production record set off Brazil[J]. Oil and Gas Journal, 1997(39):95.
- [154] KNAPP S. ABSTRACT: Using Multicomponent OBC Seismic Data for the Evaluation of Deep Gas Prospects on the Shelf in the Gulf of Mexico[J]. Gcags Transactions, 2002.
- [155] MICHAEL, V, DEANGELO, et al. Integrated 2D 4-C OBC velocity analysis of near-seafloor sediments, Green Canyon, Gulf of Mexico[J]. Geophysics, 2008, 73(6):B109-B115.
- [156] JIAO J, TRICKETT S, LINK B. Ocean-bottom-cable dual-sensor summation: A robust approach[J]. Cspg Special Publications, 1998.
- [157] WANG X, MA J. Adaptive dictionary learning for blind seismic data denoising[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(7): 1273-1277.
- [158] YARMAN C E, KUMAR R, RICKETT J. A model - based data - driven dictionary learning for seismic data representation[J]. Geophysical Prospecting, 2018, 66(1): 98-123.
- [159] TONG T, LI G, LIU X, et al. Image super-resolution using dense skip connections[C]. Proceedings of the IEEE international conference on computer vision. 2017: 4799-4807.
- [160] SYED Z, ADITYA A, SALMAN K, et al. Learning Enriched Features for Real Image Restoration and Enhancement[C]. Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [161] DAI J, QI H, XIONG Y, et al. Deformable Convolutional Networks[J]. IEEE, 2017.
- [162] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020.
- [163] LINES L R, SLAWINSKI R, BORDING R P. A recipe for stability of finite-difference wave-equation computations[J]. Geophysics, 1999, 64(3):967-969.
- [164] VIRIEUX J, OPERTO S. An overview of full-waveform inversion in exploration geophysics[J]. Geophysics, 2009, 74(6): WCC127-WCC152.
- [165] LIU X, CHEN X, LI J, et al. Nonlocal Weighted Robust Principal Component Analysis for Seismic Noise Attenuation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(2):1745-1756.
- [166] SIRGUE L, ETGEN J T, ALBERTIN U. 3D Frequency Domain Waveform Inversion Using Time Domain Finite Difference Methods[C]. 70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008.2014.
- [167] 董士琦. 基于波场局部尺度匹配的时间域全波形反演研究[D]. 吉林大学.
- [168] 董新桐. 基于深度学习的复杂陆地地震数据噪声压制方法研究[D]. 吉林大学

-
- [169] YANG L, CHEN W, WANG H, et al. Deep learning seismic random noise attenuation via improved residual convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(9): 7968-7981.
- [170] DONG X, LI Y. Denoising the optical fiber seismic data by using convolutional adversarial network based on loss balance[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(12): 10544-10554.
- [171] LIU N, WANG J, GAO J, et al. Similarity-informed self-learning and its application on seismic image denoising[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.
- [172] LIU N, WANG J, GAO J, et al. NS2NS: Self-Learning for Seismic Image Denoising, IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-11.