

8-25-2024

A non-uniform interpolation method for seismic data based on a diffusion probabilistic model

CHEN Yao

School of Mathematics, Harbin Institute of Technology, Harbin 150001, China, edisonchenyao@163.com

YU Siwei

School of Mathematics, Harbin Institute of Technology, Harbin 150001, China, siweiyu@hit.edu.cn

LIN Rongzhi

School of Mathematics, Harbin Institute of Technology, Harbin 150001, China

Follow this and additional works at: <https://cge.researchcommons.org/journal>



Part of the [Earth Sciences Commons](#), [Mining Engineering Commons](#), [Oil, Gas, and Energy Commons](#), and the [Sustainability Commons](#)

Recommended Citation

CHEN Yao, YU Siwei, LIN Rongzhi, et al. (2024) "A non-uniform interpolation method for seismic data based on a diffusion probabilistic model," *Coal Geology & Exploration*: Vol. 52: Iss. 8, Article 17.

DOI: 10.12363/issn.1001-1986.24.03.0160

Available at: <https://cge.researchcommons.org/journal/vol52/iss8/17>

This Intelligent Geophysical Exploration is brought to you for free and open access by Coal Geology & Exploration. It has been accepted for inclusion in Coal Geology & Exploration by an authorized editor of Coal Geology & Exploration. For more information, please contact 380940179@qq.com.



移动阅读

陈尧, 于四伟, 林荣智. 基于扩散概率模型的非均一地震数据插值方法[J]. 煤田地质与勘探, 2024, 52(8): 177-186. doi: 10.12363/issn.1001-1986.24.03.0160

CHEN Yao, YU Siwei, LIN Rongzhi. A non-uniform interpolation method for seismic data based on a diffusion probabilistic model[J]. Coal Geology & Exploration, 2024, 52(8): 177-186. doi: 10.12363/issn.1001-1986.24.03.0160

基于扩散概率模型的非均一地震数据插值方法

陈尧, 于四伟*, 林荣智

(哈尔滨工业大学数学学院, 黑龙江哈尔滨 150001)

摘要: 【目的】能源勘探领域长期面临非均一地震数据插值的难题。由于检波器无法精确放置在理论网格点, 现有均一插值技术常导致结果存在偏差和细节失真。【方法】为解决此问题, 提出一种基于扩散概率模型的非均一插值新方法。扩散概率模型是一种新兴的生成式深度学习模型, 分为扩散和生成两个过程。在扩散过程中, 对完整地震数据迭代施加噪声, 训练神经网络的去噪能力; 在生成过程中, 使用神经网络对噪声数据迭代去噪, 得到重构数据。利用插值算子计算迭代数据与采集数据的偏差, 作为神经网络的额外输入, 提高扩散概率模型的非均一插值能力。在数值实验部分, 使用二维合成数据集与实际数据集对非均一采样进行了测试, 并使用均一插值模型与所提模型进行比较。【结果和结论】结果表明: 该方法显著提升了扩散概率模型对非均一采样的处理能力, 在合成数据与实际数据测试中, 信噪比大约提高了 7 dB, 有效提高了深度学习对非均一插值的精度, 为非均一地震数据插值算法提供了新的思路。

关键词: 地震数据插值; 非均一采样; 深度学习; 生成式模型; 扩散概率模型

中图分类号: P631 **文献标志码:** A **文章编号:** 1001-1986(2024)08-0177-10

A non-uniform interpolation method for seismic data based on a diffusion probabilistic model

CHEN Yao, YU Siwei*, LIN Rongzhi

(School of Mathematics, Harbin Institute of Technology, Harbin 150001, China)

Abstract: [Objective] The non-uniform interpolation of seismic data is identified as a prolonged challenge in energy exploration. Since geophones cannot be precisely placed at positions corresponding to theoretical grid points, current uniform interpolation techniques frequently suffer deviations and detail distortion. [Methods] This study proposed a novel non-uniform interpolation method based on a diffusion probabilistic model, which is an emerging generative model in deep learning that involves the diffusion and generation processes. In the diffusion process, noise is added to the complete seismic data iteratively to train the denoising capability of the neural network. In the generation process, the neural network is employed for iterative denoising of data containing noise to obtain the reconstructed data. In this study, interpolation operators were employed to calculate the deviations between iterative and sampled data. These deviations were then used as the additional inputs of the neural network to improve the non-uniform interpolation capability of the diffusion probabilistic model. In the numerical experiments, the non-uniform sampling was tested using 2D synthetic and actual datasets, and the uniform interpolation model was compared with the model in this proposed study. [Results and Conclusions] The results indicate that the proposed method significantly enhanced the processing capability of the diffusion probabilistic model for non-uniform sampling. The tests of synthetic and actual data revealed an increase of approximately 7 dB in the signal-to-noise ratio. Therefore, the proposed method can effectively improve the precision of deep learning for non-uniform interpolation, providing a new approach for non-uniform interpolation algorithms of seismic data.

收稿日期: 2024-03-06; 接收日期: 2024-06-02

基金项目: 国家自然科学基金项目(42074156)

第一作者: 陈尧, 2000 年生, 男, 河北保定人, 博士研究生, 研究方向为优化算法, 深度学习, 地震数据处理. E-mail: edisonchenyao@163.com

*通信作者: 于四伟, 1987 年生, 男, 天津人, 博士, 教授, 博士生导师, 研究方向为优化算法, 深度学习, 地震数据处理.

E-mail: siweiyu@hit.edu.cn

Keywords: interpolation of seismic data; non-uniform sampling; deep learning; generative model; diffusion probabilistic model

受自然环境、经济条件等影响,采集的地震数据通常含有缺失现象,需使用插值技术对数据进行重建。高分辨率地震数据是获取高质量地下成像的前提。因此,地震数据插值是勘探地震数据处理过程的重要一环。现阶段插值技术可分为模型驱动与数据驱动两类。

模型驱动方法根据完整地震数据的先验信息重构数据,主要包括 4 类方法:(1) 预测误差滤波方法,如基于线性同相轴的 f-x 域抗假频插值技术等^[1-3];(2) 波动方程方法,利用地下模型先验信息模拟地震波传播过程,需要大量计算资源^[4];(3) 稀疏变换类方法,对完整地震数据在变换域下的稀疏性进行假设,结合不同的稀疏变换重构地震数据,如傅里叶变换^[5]、拉东变换^[6]等;(4) 降秩类方法,对完整地震数据在特定域下的低秩性进行假设,缺失道及噪声导致了地震数据秩的增加^[7-8]。

数据驱动方法直接从大量数据中学习插值模型,如在稀疏变换类方法框架下直接学习一种稀疏变换^[9-10]。深度学习作为该类型方法的主流技术,被广泛应用于地球物理领域^[11-13]。目前,应用较广的方法是利用现有的网络架构,使用有标签数据集对网络进行训练,如 U-net^[14-15]、Transformer^[16-18]等。生成式模型无须直接学习从缺失数据到完整数据的映射,凭借其可处理各种复杂的采样情况成为新兴技术。扩散概率模型^[19]是一种主流的生成式模型,它将复杂的映射模型转换为迭代去噪,降低了模型训练的难度。目前,它已广泛应用于视觉领域,如图像超分辨率重建^[20]、图像分割^[21]、图像修复^[22]等。2024 年, Liu Qi 和 Ma Jianwei^[23] 基于扩散概率模型提出了均一插值模型,该模型对不同缺失情况具备较强的适应能力。

现阶段插值技术大都认定检波器坐标与理论坐标一致,即已知地震道均位于理论网格点上,统称为“均一”插值。而在实际勘探过程中,检波器受障碍物的干扰,无法精确地放置在理论网格上。因此,除重构缺失地震道外,需对采集的地震数据进行“校正”,此类插值技术称为“非均一”插值。目前,已有部分学者将常规的均一插值技术拓展为非均一插值技术,如将凸集投影算法中的采样算子与重心拉格朗日插值算子结合,实现了同时校正与插值的功能^[24];将基于非均一傅里叶变换的插值算子插入压缩感知框架^[25];将非均一插值算子插入至多通道奇异谱分解框架^[26]。部分学者针对非均一采样情况提出了非均一变换,如非均一傅里叶变换^[27]、非均一曲波变换^[28]等。

笔者拟将扩散概率生成式模型与插值算子相结合,以期扩展扩散概率模型的应用领域至非均一插值问题,

旨在为地震数据处理等领域提供更为精准和高效的深度学习新方法。

1 基本原理

本节首先对均一插值模型与非均一插值模型进行建模。对扩散概率模型进行概括介绍,最后提出基于扩散概率模型的非均一插值模型。

1.1 均一插值模型与非均一插值模型

以二维地震数据为例,均一采样可使用下式描述:

$$\mathbf{X}_{\text{miss}} = \mathbf{M} \circ \mathbf{X} \quad (1)$$

\mathbf{X}_{miss} 对应的采样坐标网格记为 (y_1, y_2, \dots, y_n) , 坐标点采样自 \mathbf{X} 对应的均一网格,例如 $y_1 = x_3, y_2 = x_7, y_3 = x_{10}, \dots, y_n = x_{n-2}$; 均一插值模型通过 \mathbf{X}_{miss} 得到 \mathbf{X} , 仅需对 \mathbf{X}_{miss} 中的缺失数据进行重构。

非均一采样过程可描述为:

$$\mathbf{X}_{\text{ir}} = \mathbf{A}(\mathbf{X}) \quad (2)$$

\mathbf{X}_{ir} 对应的采样坐标网格记为 $(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$, 受障碍物的影响, \tilde{y}_1 不一定等于 y_1 。插值算子 $\mathbf{A}(\cdot)$ 的功能如图 1 所示,表示由均一网格 $(x_1, x_2, \dots, x_{n_x})$ 到非均一采样网格 $(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$ 的插值算子,非均一采样网格相较于均一网格,采样点除缺失外存在扰动现象。 $\mathbf{A}(\cdot)$ 的逆算子理论上能够完全正确地非均一采样网格插值到均一网格,但实际上该算子不可逆。需要基于均一网格到非均一网格的映射建立反问题模型,利用该非均一插值算子的伴随算子以及迭代算法求解该反问题。

1.2 扩散概率模型

扩散概率模型分为扩散过程与生成过程。将完整地震数据 \mathbf{X} 所属分布记作 $p(\mathbf{X})$ 。扩散过程将 $p(\mathbf{X})$ 转换为标准高斯分布,生成过程则将标准高斯分布中的样本逆向转换为 $p(\mathbf{X})$ 中的样本^[19]。如图 2 所示,可将扩散过程理解为逐步施加噪声的过程,生成过程为利用神经网络逐步去除噪声的过程。其中扩散过程 $q(\mathbf{X}_t | \mathbf{X}_{t-1})$ 是模型驱动的,逐步将 $p(\mathbf{X})$ 转变为标准高斯分布;而生成过程 $p_\theta(\mathbf{X}_{t-1} | \mathbf{X}_t)$ 则是数据驱动的,通过训练神经网络逐步去噪,将随机高斯噪声逐步转变为 $p(\mathbf{X})$ 中的样本。

扩散过程 $q(\mathbf{X}_t | \mathbf{X}_{t-1})$ 可使用下式描述:

$$q(\mathbf{X}_t | \mathbf{X}_{t-1}) = N(\mathbf{X}_t | \sqrt{1 - \beta_t} \mathbf{X}_{t-1}, \beta_t \mathbf{I}) \quad (3)$$

式(3)还可写为:

$$\mathbf{X}_t = \sqrt{1 - \beta_t} \mathbf{X}_{t-1} + \sqrt{\beta_t} \mathbf{I} \quad (4)$$

通过逐步施加噪声,当 T 足够大时, \mathbf{X}_T 可近似为高斯随机噪声。由于扩散过程是显式的,且 \mathbf{X}_0 已知,可由

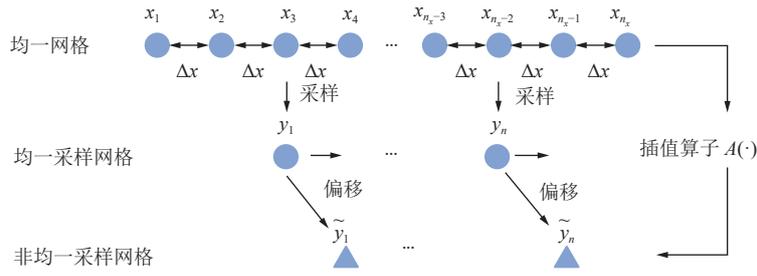


图 1 插值算子 $A(\cdot)$ 功能及非均一采样过程

Fig.1 Functions of interpolation operator $A(\cdot)$ and the non-uniform sampling process

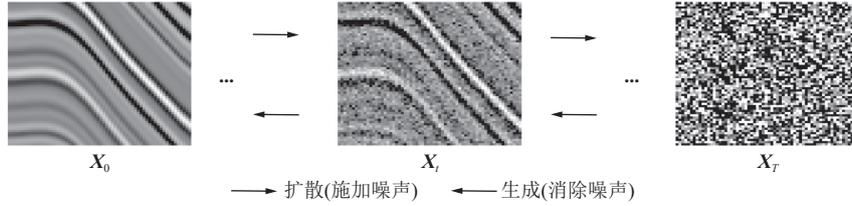


图 2 扩散概率模型流程

Fig.2 Workflow of a diffusion probabilistic model

X_0 直接得到 X_t :

$$X_t = \sqrt{\alpha_t} X_0 + \sqrt{1 - \alpha_t} \varepsilon_t \quad (5)$$

其中, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$, ε_t 服从标准高斯分布。

由式(5), X_t 可由 X_0 直接计算得到, 提高了扩散过程的计算效率。

生成过程即为扩散过程的逆过程, 可表示为 $q(X_{t-1}|X_t)$, 加入 X_0 作为额外条件, 由贝叶斯公式及式(4)、式(5), 可求解 $q(X_{t-1}|X_t, X_0)$, 并得到生成过程公式:

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(X_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \varepsilon_t \right) + \sqrt{\beta_t} \varepsilon \quad (6)$$

其中, ε 服从标准高斯分布, $\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t$, 当 T 足够大时, X_T 可看作标准高斯分布中的样本。在生成过程中, ε_t 未知, 但与 X_t 相关, 因此, 可使用神经网络近似预测:

$$\varepsilon_t \approx \varepsilon_\theta(X_t, t) \quad (7)$$

X_t 作为输入, 并对迭代步数 t 进行位置编码, 具体编码方式可参考模型原论文^[19]。

因此, 生成过程变为了神经网络预测噪声的过程。在网络训练阶段, 对训练集中的样本进行随机扩散, 训练神经网络预测噪声的能力。

1.3 基于扩散概率模型的均一插值模型

Liu Qi 等^[23] 基于 1.2 节所述扩散概率模型提出均一插值模型, 将式(7)中的神经网络加入额外输入条件 X_{miss} , 即:

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(X_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \varepsilon_\theta(X_t, X_{\text{miss}}, t) \right) + \sqrt{\beta_t} \varepsilon \triangleq f_\theta(X_t, X_{\text{miss}}, t) \quad (8)$$

具体而言, 将 X_t 与 X_{miss} 作为二通道数据输入, 为便于在图中展示, 使用 $f_\theta(\cdot, \cdot, \cdot)$ 表示式(8)的生成过程。图 3a 展示了该模型的生成过程, 扩散过程保持不变。

1.4 基于扩散概率模型的非均一插值模型

本文借鉴式(8)提出基于扩散概率模型的非均一插值模型, 具体生成过程如下:

$$X_t^{\text{dis}} = B(A(X_t) - X_{\text{ir}}) \quad (9)$$

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(X_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \varepsilon_\theta(X_t, X_t^{\text{dis}}, t) \right) + \sqrt{\beta_t} \varepsilon = f_\theta(X_t, X_t^{\text{dis}}, t) \quad (10)$$

式(9)中 $A(\cdot)$ 与公式(2)意义一致, 表示由均一网格到非均一网格的插值算子, $B(\cdot)$ 为由非均一网格到均一网格的插值算子, $A(\cdot)$ 与 $B(\cdot)$ 的具体计算形式参考文后附录 A; X_t^{dis} 借助插值算子表示了规则网格下的 X_t 与非规则网格下的 X_{ir} 之间的差异性。式(10)与式(8)类似, 网络以 X_t 与 X_t^{dis} 作为二通道数据输入, 图 3b 展示了该模型的生成过程。

2 数值实验

本节采用信噪比(Signal-to-Noise Ratio, SNR)对数据的重构质量进行度量, 信噪比 $R_{S/N}$ 的计算公式如下:

$$R_{S/N} = 10 \lg \frac{\|X\|_F^2}{\|X - X_{\text{rec}}\|_F^2} \quad (11)$$

实验中采用 1.3 节所述均一插值模型与本文所提模型进行对比。均一插值模型在非均一插值问题中的重构实现方法细节如下: 将 X_{ir} 中的地震道分至最近的均一采样点, 如地震道对应的采样坐标为 \tilde{y}_2, \tilde{y}_2 与 x_4 最近, 则将该地震道放置于矩阵中的第 4 列。构成与 X_{miss}

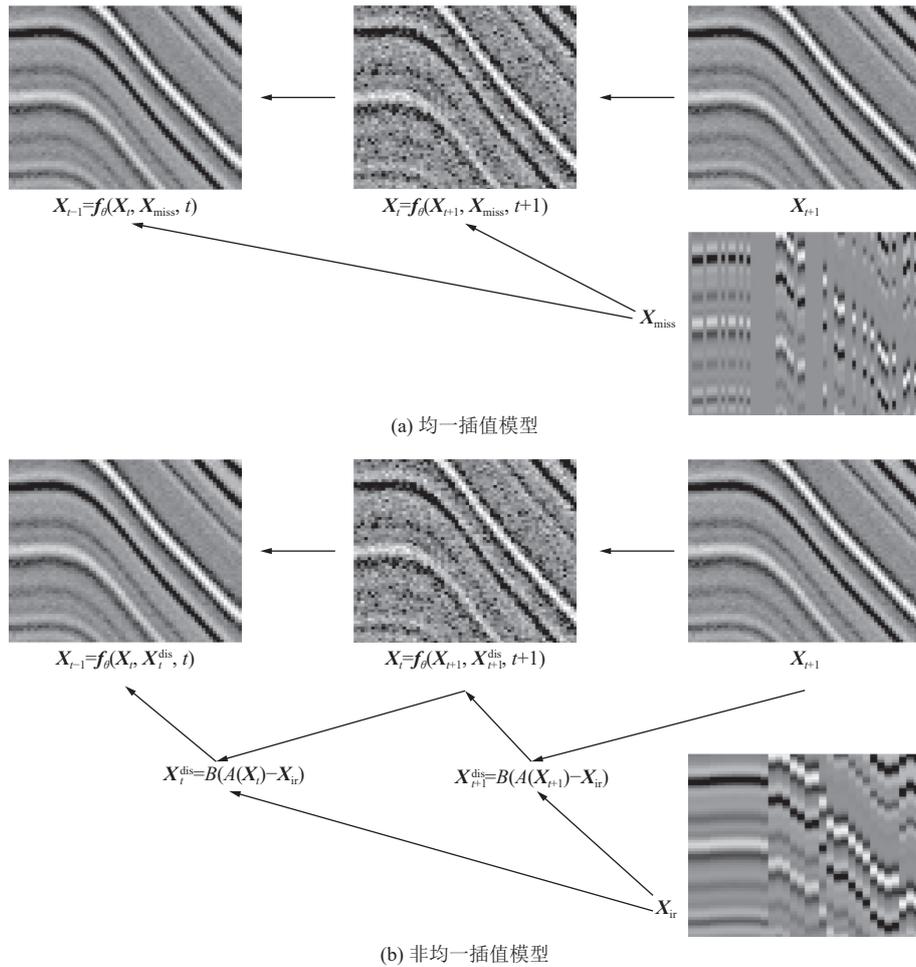


图 3 基于扩散概率模型的均一与非均一插值模型

Fig.3 Uniform and non-uniform interpolation models based on a diffusion probabilistic model

近似的含有缺失地震道的地震数据(缺失处为 0), 并使用该数据近似替代 X_{miss} 。

2.1 二维合成数据测试

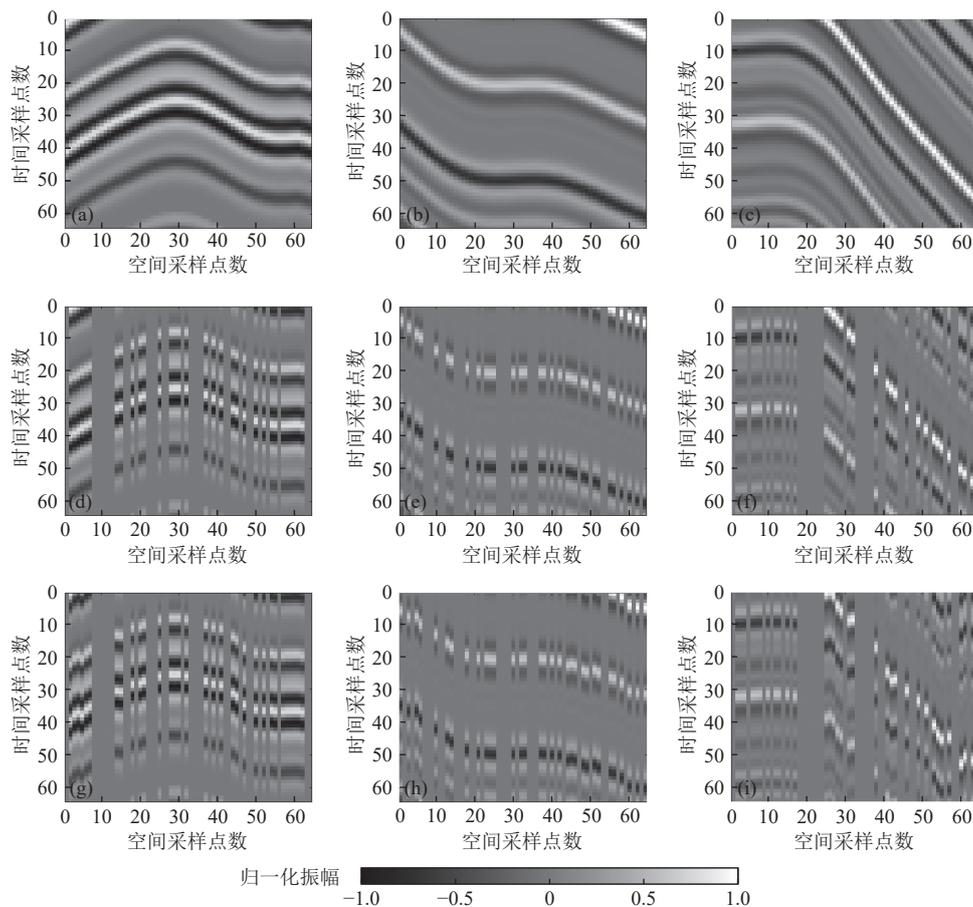
本节测试数据集裁剪自一组二维合成数据。该组数据含有若干弯曲同相轴, 共计 12 800 个 64×64 大小的数据块。随机选取 10 000 个作为训练数据集, 2 800 个作为验证数据集。图 4 展示了验证数据集中的 3 个样本, 其中图 4a—图 4c 为原始完整地震数据; 对图 4a—图 4c 所示的完整数据进行 50% 随机下采样, 得到如图 4d—图 4f 所示的缺失数据, 采样的地震道坐标均位于均一网格上, 为均一采样情况。对图 4d—图 4f 的采样坐标进行随机抖动, 并使用重心拉格朗日插值算子将对应的数据进行偏移, 用于模拟非均一采样情况, 偏移后的数据如图 4g—图 4i 所示, 其中图 4i 的同相轴倾斜角度较大, 抖动也较为明显。图 5 是对图 4a 的进行坐标抖动, 默认理论坐标间距归一化为 1, 坐标的随机抖动范围为 $[-2, 2]$ 。

使用 1.3 节所述的均一插值模型及 1.4 节本文提出的非均一插值模型进行非均一插值训练。采用的神经网络为含有残差模块的 U-net 架构^[29], 训练轮次为 300,

批次大小为 8, 优化器为 Adam^[30], 学习率为 10^{-5} , 生成扩散的迭代步数为 1 000, 采用均方误差作为损失函数。训练过程中损失函数值的变化如图 6 所示, 在 100~300 轮次, 非均一插值模型的损失函数值明显低于均一插值模型。

使用两模型分别对图 4d—图 4f 进行重构, 插值结果及残差如图 7 与图 8 所示, 观察残差图及信噪比, 以图 7f 与图 8f 为例, 对于较大斜率的同相轴, 非均一插值模型的重构效果远好于均一插值模型; 对于前 2 个数据, 非均一插值模型在同相轴倾斜位置较大的区域重构结果明显好于均一插值模型。上述结果及分析证明了本文所提出的非均一插值模型能够较好地实现非均一插值功能, 较大地提高了扩散概率模型对非均一插值问题的处理能力。

为凸显扩散概率模型处理非均一插值问题的优势, 采用传统的有监督 U-net 模型^[31]与上述 2 个模型(图 3)进行对比。在下采样过程中, U-net 的卷积输出通道数依次为 64、128、256、512; 在上采样过程中, 采用最近邻插值算子。有监督训练方式为: 将缺失抖动数



(a)–(c) 完整数据; (d)–(f) 缺失数据; (g)–(i) 缺失数据, 对应坐标进行了随机抖动

图 4 二维合成数据集部分样本

Fig.4 Part samples from the 2D synthetic dataset

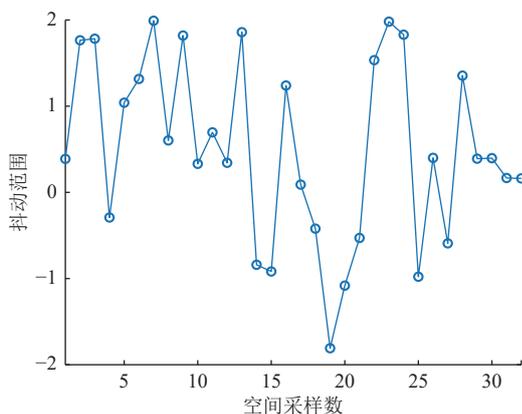


图 5 图 4a 的坐标抖动

Fig.5 Coordinate jitter applied in Fig.4a

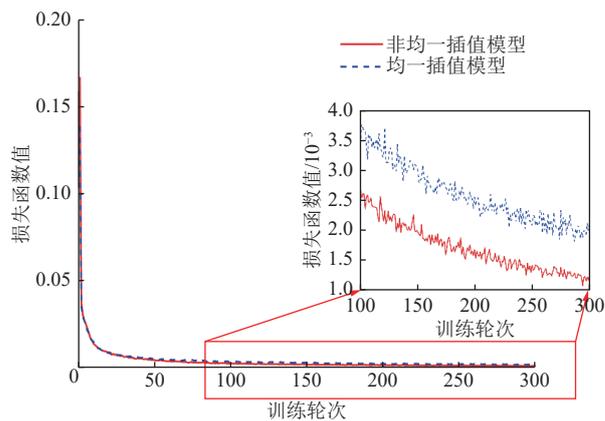


图 6 损失函数值变化

Fig.6 Changes in values of the loss function

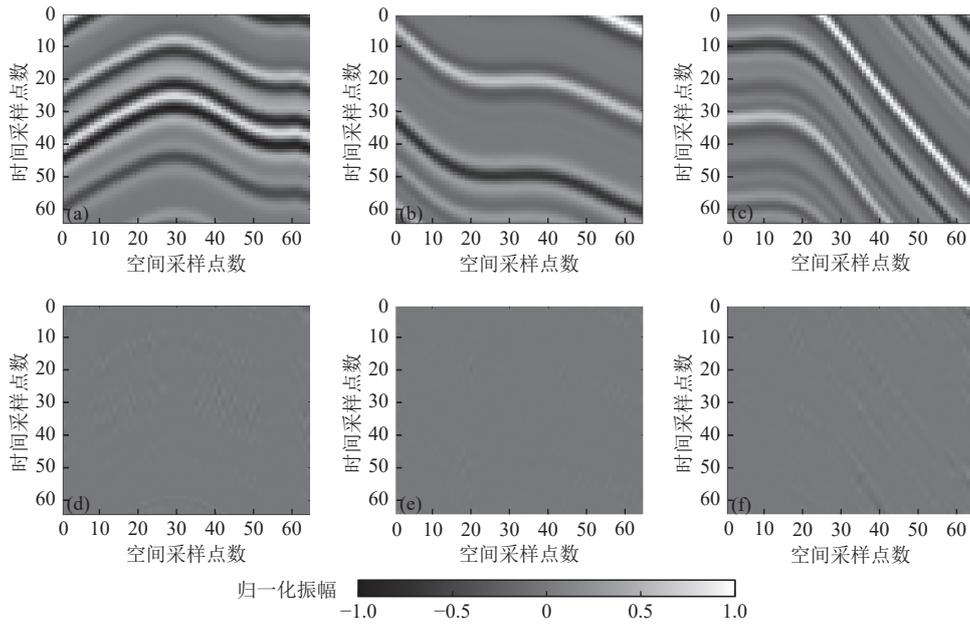
据(图 3d—图 3f) 作为特征, 完整数据(图 4g—图 4i) 作为标签, 采用输出与标签的均方误差作为损失函数。训练轮次为 500, 其余参数与上述 2 模型(图 3) 一致。插值结果如图 9 所示, 观察可知, 扩散概率模型的重构同相轴更接近原始数据, 残差图能量更弱, 信噪比更高。

2.2 二维合成叠前数据测试

本节测试数据集裁剪自一个三维合成叠前数据, 时

间采样数为 178, 炮点数为 178, 检波点数为 128。从该数据体中裁剪得到 40 个 128×128 的二维合成叠前数据, 对其进行 50% 随机下采样, 并对采样数据使用重心拉格朗日插值算子进行随机抖动, 仿真非均一采样情况。选取 30 个数据作为训练数据集, 10 个数据作为验证数据集。图 10 展示了验证数据集中的—一个数据。

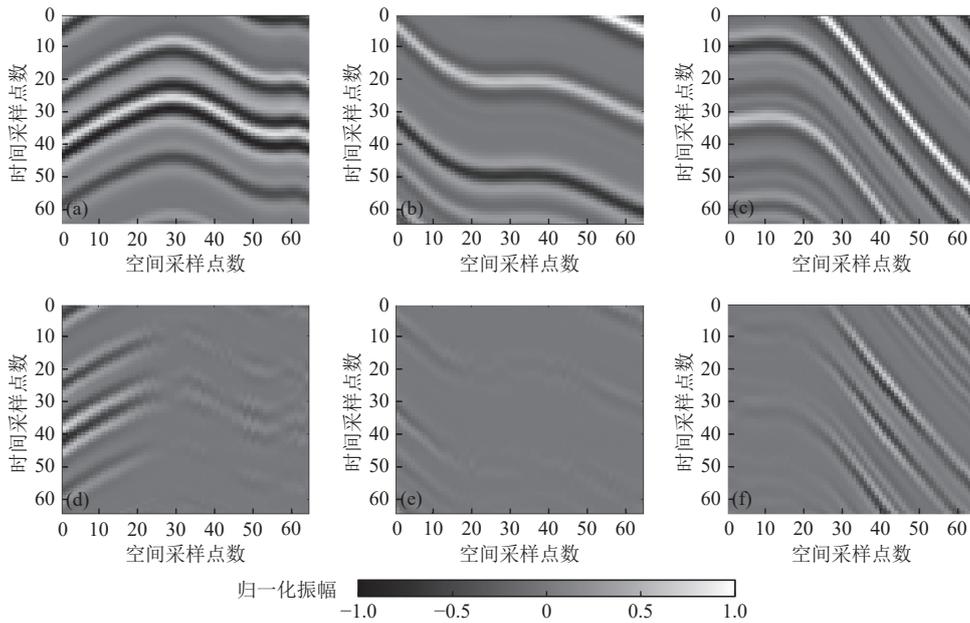
对均—插值模型与非均—插值模型进行训练, 训练



(a)–(c) 重构数据, R_{SN} 分别为 28.73、31.02、24.06 dB; (d)–(f) 残差

图 7 非均一插值模型重建结果及残差

Fig.7 Reconstructed and residual data of the non-uniform interpolation model



(a)–(c) 重构数据, R_{SN} 分别为 9.37、18.81、6.15 dB; (d)–(f) 残差

图 8 均一插值模型重建结果及残差

Fig.8 Reconstructed and residual data of the uniform interpolation model

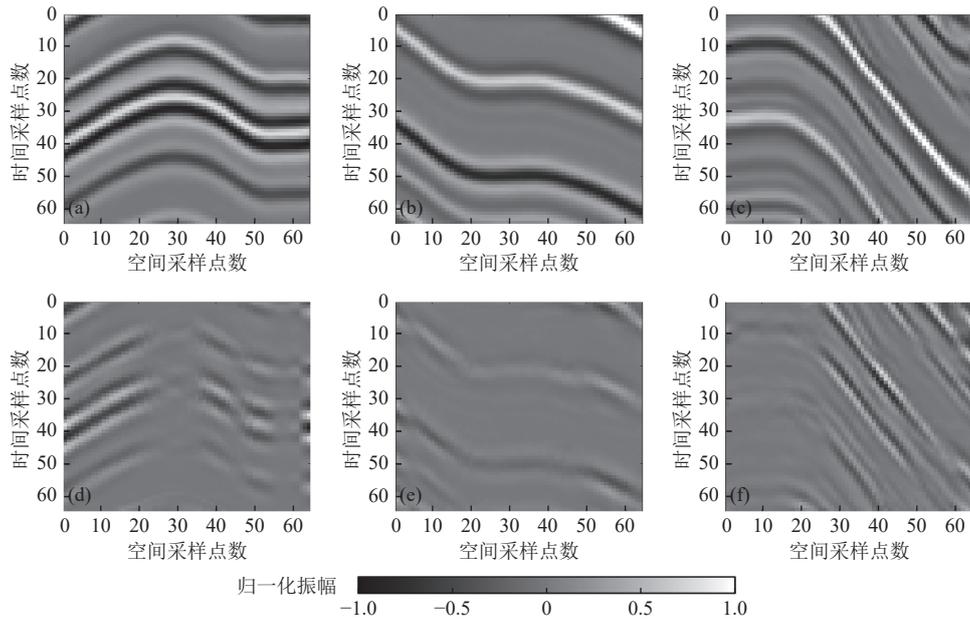
轮次为 1 000, 批次为 2, 其余参数与 2.1 节一致。使用训练后的模型对图 10b 进行测试, 重构结果及残差列于图 10c, 观察图 10d 与图 10f, 图 10f 的同相轴更为明显, 非均一插值模型取得了更高的信噪比, 重构同相轴更接近原始数据。

2.3 海上拖缆数据测试

本节测试数据来自一组处理后的海上拖缆数据, 时

间采样间隔为 0.004 s, 空间采样间隔为 50 m。裁剪为 100 个 128×128 的二维数据切片, 随机选取 80 个作为训练数据集, 20 个作为验证数据集。与 2.2 节一致, 对其进行 50% 下采样与随机抖动, 仿真非均一采样。图 11 展示了验证数据集中一个数据。

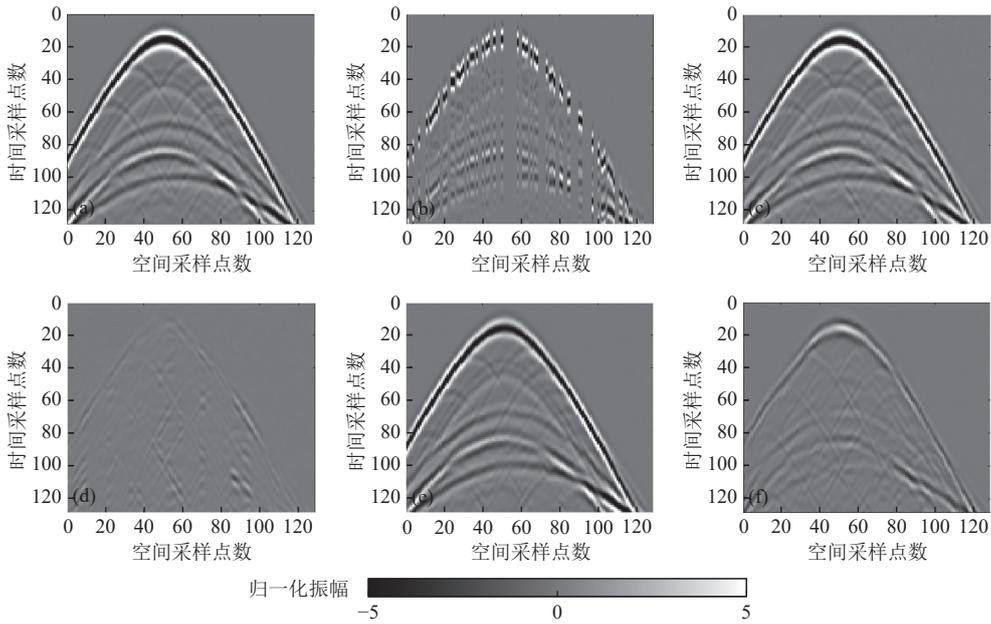
对两模型进行训练, 轮次为 1 500, 批次为 4, 其余参数与 2.1 节一致。使用训练后的模型对图 11b 进行



(a)–(c) 重构数据, $R_{S/N}$ 分别为 8.66、10.43、5.20 dB; (d)–(f) 残差

图 9 有监督 U-net 插值模型重建结果及残差

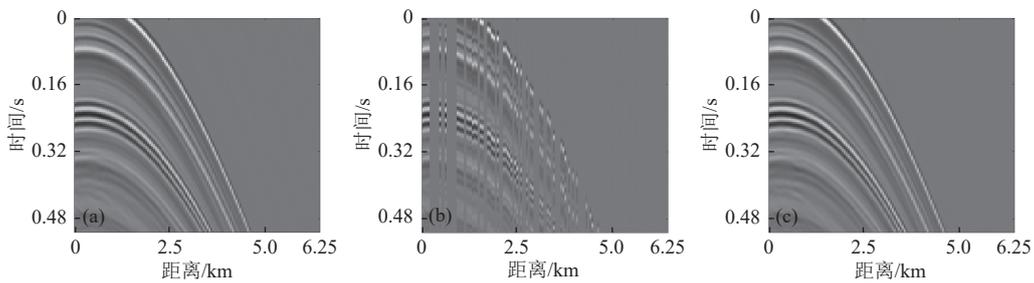
Fig.9 Reconstructed and residual data of the supervised U-net interpolation model

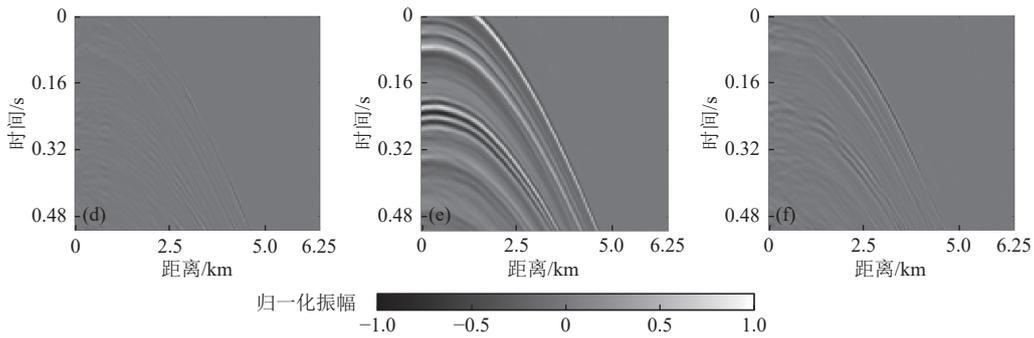


(a) 完整数据; (b) 非均一采样缺失数据; (c) 非均一插值模型重构结果, $R_{S/N} = 18.50$ dB; (d) 非均一插值模型重构残差; (e) 均一插值模型重构结果, $R_{S/N} = 10.74$ dB; (f) 均一插值模型重构残差

图 10 二维合成叠前数据测试结果

Fig.10 Test results of 2D synthetic prestack data





(a) 完整数据; (b) 非均一采样缺失数据; (c) 非均一插值模型重构结果, $R_{S/N} = 18.55$ dB; (d) 非均一插值模型重构残差; (e) 均一插值模型重构结果, $R_{S/N} = 11.79$ dB; (f) 均一插值模型重构残差

图 11 海上拖缆数据测试结果
Fig.11 Test results of marine streamer data

非均一插值, 图 11 表明非均一插值模型取得了更高的信噪比。图 12 展示了图 11 的单道对比, 观察可得, 非均一插值数据(图 11c) 相较于偏移数据(图 11b) 更接近完整数据(图 11a)。图 13 展示了图 11 的对数 $f-k$ 频谱图, 非均一插值模型的重构频谱图更接近原始数据。

在图 10 与图 11 中, 非均一插值模型相较于均一插值模型, 信噪比大约提升了 7 dB。因此, 本文所提方法显著提升了扩散概率模型的非均一插值能力。

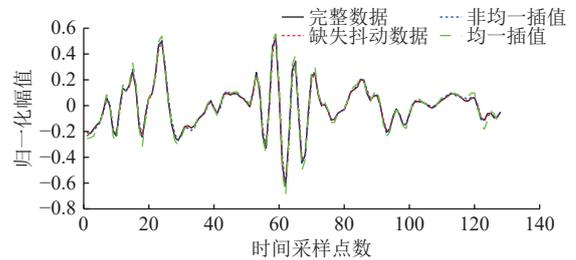
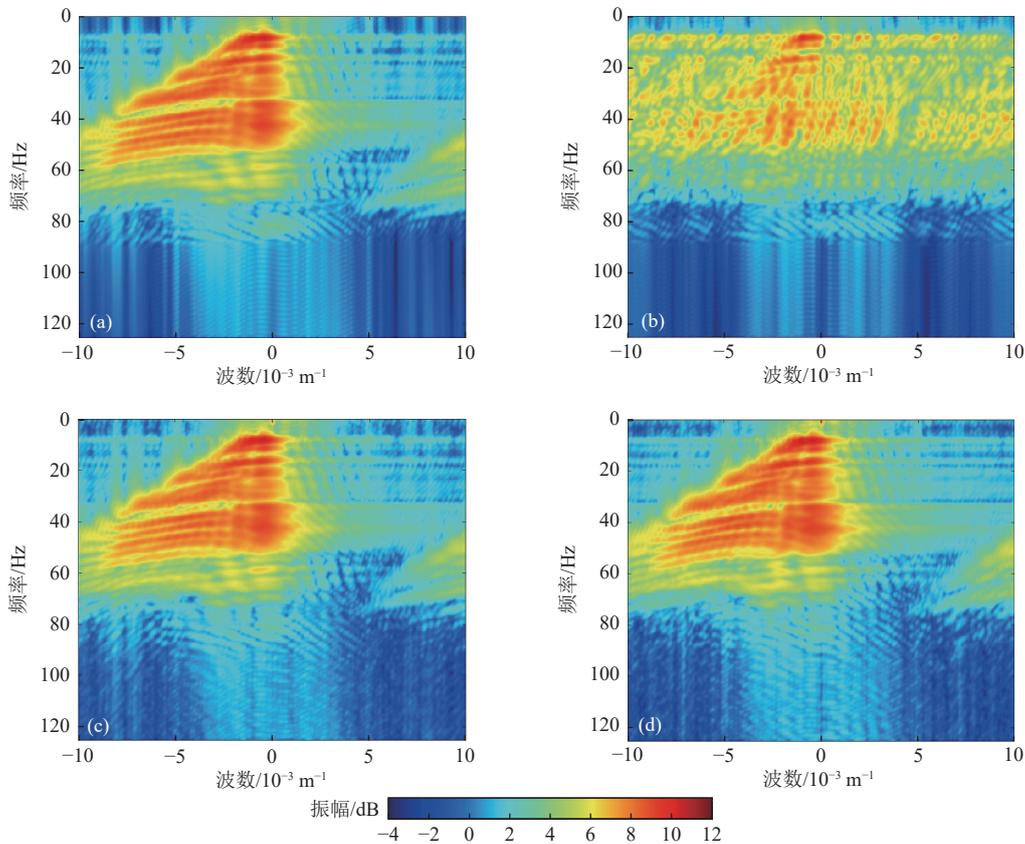


图 12 图 11 的单道对比
Fig.12 Single-trace comparison of data shown in Fig.11



(a)–(d) 分别对应图 11a, 图 11b, 图 11c, 图 11e

图 13 图 11 的对数 $f-k$ 频谱
Fig.13 Logarithmic $f-k$ spectra of data shown in Fig.11

3 结论

(1) 针对能源勘探领域的非均一插值问题, 提出了一种基于扩散概率模型的非均一插值方法。

(2) 新方法将扩散模型和插值算子相结合, 在原始扩散概率模型的基础上, 通过插值算子计算规则网格下的迭代数据与非规则网格下的采样数据的差, 并以此作为神经网络的额外输入, 迭代生成重构数据, 充分利用数据间的空间关系, 有效提高了插值精度。

(3) 二维合成数据集和实际地震数据集上的测试结果, 新方法较传统均匀插值模型的信噪比提高约 7 dB, 明显提升了插值精度, 为非均匀地震数据插值算法提供了新思路。

(4) 本文所设计模型在二维数据上进行了初步测试, 在今后工作中应考虑更为复杂的三维实际采样数据。

符号注释:

$A(\cdot)$ 为均一网格到非均一网格的插值算子; $B(\cdot)$ 为非均一网格到均一网格的插值算子; $f_{\theta}(\cdot, \cdot, \cdot)$ 为(非)均一插值模型中的生成过程; 下标 F 为 Frobenius 范数, 简称 F 范数; \mathbf{I} 为单位矩阵; \mathbf{M} 为二值采样矩阵, 元素为“1”代表采样位置, “0”代表缺失位置; n 为实际采样点数; n_t 为时间采样点数; n_x 为空间理论采样点数; $N(\cdot, \cdot)$ 为高斯分布; $p(\mathbf{X})$ 为完整地震数据所属分布; $p_{\theta}(\mathbf{X}_{t-1}|\mathbf{X}_t)$ 为生成过程; $q(\mathbf{X}_t|\mathbf{X}_{t-1})$ 为扩散过程; $R_{S/N}$ 为信噪比, dB; Δx 为采样网格的采样间隔; \mathbf{X} 为理论完整地震数据, $\mathbf{X} \in \mathbb{R}^{n_t \times n_x}$; \mathbf{X}_{ir} 为非均一采样数据, $\mathbf{X}_{ir} \in \mathbb{R}^{n_t \times n_x}$; \mathbf{X}_{miss} 为包含缺失地震道的均一采样地震数据, $\mathbf{X}_{miss} \in \mathbb{R}^{n_t \times n_x}$; \mathbf{X}_{rec} 为重构数据; \mathbf{X}_t 为迭代数据, $t = 0, \dots, T$, t 为当前迭代步数, T 为扩散过程的迭代总步数, $\mathbf{X}_0 = \mathbf{X}$; \mathbf{X}_t^{dis} 为 \mathbf{X}_{ir} 与 \mathbf{X}_t 之间的差异; $(x_1, x_2, \dots, x_{n_x})$ 为理论均一坐标网格; (y_1, y_2, \dots, y_n) 为实际均一采样坐标网格; $(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$ 为非均一采样坐标网格; $\alpha_t = 1 - \beta_t$; $\bar{\alpha}_t$ 为 $\prod_{i=0}^t \alpha_i$; β_t 为第 t 步扩散过程的扩散方差, $t = 1, 2, \dots, T$, $\beta_1 = 10^{-4}$, $\beta_T = 0.02$, 序列呈线性递增关系; $\tilde{\beta}_t$ 为 $\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t$; Δx 为理论均一坐标网格的采样间隔; ε 为标准高斯分布样本; ε_t 为扩散过程第 t 步施加的标准高斯分布噪声; $\varepsilon_{\theta}(\mathbf{X}_t, t)$ 为用于预测 ε_t 的神经网络, 其参数矩阵为 θ ; “ \circ ” 为元素乘法。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

参考文献(References)

- [1] SPITZ S. Seismic trace interpolation in the F - X domain[J]. *Geophysics*, 1991, 56(6): 785-794.
- [2] PORSANI M J. Seismic trace interpolation using half - step prediction filters[J]. *Geophysics*, 2012, 64(5): 1461-1467.
- [3] 刘保童. 一种基于傅里叶变换的去假频内插方法及应用[J]. *煤田地质与勘探*, 2009, 37(2): 63-67.
LIU Baotong. Dealiasing interpolation based on Fourier transform and its application[J]. *Coal Geology & Exploration*, 2009, 37(2): 63-67.
- [4] RONEN J. Wave-equation trace interpolation[J]. *Geophysics*, 1987, 52(7): 973-984.
- [5] ABMA R, KABIR N. 3D interpolation of irregular data with a POCS algorithm[J]. *Geophysics*, 2006, 71(6): 91-97.
- [6] SHAO Jie, WANG Yibo. Seismic data antialiasing interpolation using sparse Radon transform and dynamic mask function[J]. *Geophysics*, 2022, 87(5): 437-449.
- [7] MANENTI R, SACCHI M D. Tensor tree decomposition as a rank-reduction method for pre-stack interpolation[J]. *Geophysical Prospecting*, 2023, 71(8): 1404-1419.
- [8] NIU Xiao, FU Lihua, ZHANG Wanjuan, et al. Seismic data interpolation based on simultaneously sparse and low-rank matrix recovery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 9100613.
- [9] YU Siwei, MA Jianwei, ZHANG Xiaoqun, et al. Interpolation and denoising of high-dimensional seismic data by learning a tight frame[J]. *Geophysics*, 2015, 80(5): 119-132.
- [10] ALMADANI M, WAHEED U, MASOOD M, et al. Dictionary learning with convolutional structure for seismic data denoising and interpolation[J]. *Geophysics*, 2021, 86(5): 361-374.
- [11] WANG Benfeng, ZHANG Ning, LU Wenkai, et al. Deep-learning-based seismic data interpolation: A preliminary result[J]. *Geophysics*, 2019, 84(1): 11-20.
- [12] 李宇腾, 程建远, 鲁晶津, 等. 基于人工神经网络的矿井直流电阻率法超前预测方法[J]. *煤田地质与勘探*, 2023, 51(6): 185-193.
LI Yuteng, CHENG Jianyuan, LU Jingjin, et al. Direct current resistivity method for the advance prediction of water Hazards in coal mines based on an artificial neural network[J]. *Coal Geology & Exploration*, 2023, 51(6): 185-193.
- [13] 古瑶, 解海军, 周子鹏, 等. 基于 Attention 机制的 CNN-BiLSTM 瞬变电磁实时反演方法[J]. *煤田地质与勘探*, 2023, 51(10): 134-143.
GU Yao, XIE Haijun, ZHOU Zipeng, et al. An Attention mechanism-based CNN-BiLSTM real-time transient electromagnetic method[J]. *Coal Geology & Exploration*, 2023, 51(10): 134-143.
- [14] FANG Wenqian, FU Lihua, ZHANG Meng, et al. Seismic data interpolation based on U-net with texture loss[J]. *Geophysics*, 2021, 86(1): 41-54.
- [15] WU Geng, LIU Yang, LIU Cai, et al. Seismic data interpolation using deeply supervised U-Net++ with natural seismic training sets[J]. *Geophysical Prospecting*, 2023, 71(2): 227-244.
- [16] GUO Yuanqi, FU Lihua, LI Hongwei. Seismic data interpolation

- based on multi-scale transformer[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 7504205.
- [17] PENG Junheng, LI Yong, LIAO Zhangquan. Irregularly spatial seismic missing data reconstruction using transformer with periodic skip connection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5918613.
- [18] 郭元奇, 李志明. 基于卷积和 Transformer 联合网络的地震数据插值[J]. *工程地球物理学报*, 2023, 20(3): 393-401.
GUO Yuanqi, LI Zhiming. Seismic data interpolation based on joint convolutional and transformer networks[J]. *Chinese Journal of Engineering Geophysics*, 2023, 20(3): 393-401.
- [19] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[EB/OL]. 2020: 2006.11239. <http://arxiv.org/abs/2006.11239v2>.
- [20] XIAO Yi, YUAN Qiangqiang, JIANG Kui, et al. EDiffSR: An efficient diffusion probabilistic model for remote sensing image super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-14.
- [21] LUGMAYR A, DANELLJAN M, ROMERO A, et al. RePaint: Inpainting using denoising diffusion probabilistic models[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 11451-11461.
- [22] LI Haoying, YANG Yifan, CHANG Meng, et al. SRDiff: Single image super-resolution with diffusion probabilistic models[J]. *Neurocomputing*, 2022, 479: 47-59.
- [23] LIU Qi, MA Jianwei. Generative interpolation via a diffusion probabilistic model[J]. *Geophysics*, 2024, 89(1): 65-85.
- [24] YU Siwei, MA Jianwei. Simultaneous off-the-grid regularization and reconstruction for 3D seismic data by a new combined sampling operator[J]. *Geophysics*, 2023, 88(4): 291-302.
- [25] YU Siwei, MA Jianwei, ZHAO Bangliu. Off-the-grid vertical seismic profile data regularization by a compressive sensing method[J]. *Geophysics*, 2020, 85(2): 157-168.
- [26] CAROZZI F, SACCHI M. Interpolated multichannel singular spectrum analysis: A reconstruction method that honors true trace coordinates[J]. *Geophysics*, 2021, 86(1): 55-70.
- [27] ZWARTJES P M, SACCHI M D. Fourier reconstruction of nonuniformly sampled, aliased seismic data[J]. *Geophysics*, 2007, 72(1): 21-32.
- [28] MA J, PLONKA G. The curvelet transform[J]. *IEEE Signal Processing Magazine*, 2010, 27(2): 118-133.
- [29] DIAKOGIANNIS F I, WALDNER F, CACCETTA P, et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 94-114.
- [30] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]//Proceedings of the International Conference on Learning Representations. 2014.
- [31] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[M]//NAVAB N, HORNEGGER J, WELLS W M, et al. Lecture notes in computer science. Cham: Springer International Publishing, 2015: 234-241.

(责任编辑 聂爱兰)

附录 A

插值算子 $A(\cdot)$ 与 $B(\cdot)$ 的具体计算

已知坐标 (x_1, x_2, \dots, x_n) 与对应数据 (y_1, y_2, \dots, y_n) , 对坐标为 x_f 的数据 y_f 进行插值, 具体计算过程如下:

(1) 选取与 x_f 最近的 $k+1$ 个点 $(x'_1, x'_2, \dots, x'_k)$, 对应数据为 $(y'_1, y'_2, \dots, y'_k)$, k 为人为调整参数; 本文令 $k=1$ 。

(2) 计算求和权重 $w_i = 1 - \frac{|x'_i - x_f|}{\sum_{j=1}^k |x'_j - x_f|}$, $i = 1, 2, \dots, k$ 。

(3) 加权求和得 $y_f = \sum_{i=1}^k w_i y'_i$ 。