



地球物理学进展
Progress in Geophysics
ISSN 1004-2903, CN 11-2982/P

《地球物理学进展》网络首发论文

题目: 卷积神经网络中批量规范化层的使用对地震数据去噪的影响分析
作者: 张鑫, 杨万祥, 李小斌, 阎建国, 胡善政, 吴育林
收稿日期: 2023-02-20
网络首发日期: 2023-11-14
引用格式: 张鑫, 杨万祥, 李小斌, 阎建国, 胡善政, 吴育林. 卷积神经网络中批量规范化层的使用对地震数据去噪的影响分析[J/OL]. 地球物理学进展.
<https://link.cnki.net/urlid/11.2982.P.20231114.1052.002>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

卷积神经网络中批量规范化层的使用对地震数据去噪的影响分析

张鑫¹, 杨万祥¹, 李小斌^{2**}, 阎建国², 胡善政³, 吴育林⁴

1. 中国石油新疆油田分公司勘探事业部, 新疆克拉玛依 834000

2. 成都理工大学地球物理学院, 成都 610059

3. 中国石油东方地球物理勘探有限责任公司西南物探分公司, 成都 610213

4. 中国石油东方地球物理勘探有限责任公司西南物探研究院, 成都 610213

*通信作者 李小斌, 男, 1996年生, 硕士, 主要从事油气地震勘探方面的数据处理研究工作. E-mail: lixb1996@foxmail.com

摘要 深度学习算法已广泛应用于地震数据处理分析领域, 并在地震数据去噪等方面取得了较好的应用效果。目前业界关注重点在于各种不同深度学习算法和相关的网络结构形式, 以及不同标签数据对算法效果的影响, 较少关注数据集本身的差异对深度学习算法的应用效果影响。本文以卷积神经网络(convolutional neural network, CNN)算法中批量规范化层(Batch Normalization)对地震数据去噪的影响分析为例, 通过理论公式和应用效果的对比分析, 提出了基于地震数据特征分析下的批量规范化层的使用建议。批量规范化层的使用依赖于数据集的统计分布特征, 只有当训练集的归一化能量分布集中在能量较强的区域, 批量规范化层的使用才会提升网络的效果。但通常情况下, 在地震数据去噪的应用中, 不建议使用批量规范化层。这些特征为深度学习算法在地震数据去噪应用中的网络结构设计提供了有价值的参考。

关键词 深度学习; 批量规范化; 结构优化; 去噪; 数据特征分析

Analysis of the Effect of Using Batch Normalization Layers in Convolutional Neural Networks on Seismic Data Denoising

ZHANG Xin¹, YANG Wan-Xiang¹, LI Xiao-Bin^{2**}, YAN Jian-Guo², HU Shan-Zheng³, WU Yu-Lin⁴

1. Exploration Division of Xinjiang Oilfield, CNPC, Karamay 834000, China

2. School of Geophysics, Chengdu University of Technology, Chengdu 610059, China

3. Southwest Geophysical Company, CNPC BGP Inc., Chengdu 610213, China

4. Southwest Geophysical Exploration Research Institute, CNPC BGP Inc., Chengdu 610213, China

Abstract Deep learning algorithms have been widely applied in seismic data processing, and have achieved many good applications in seismic data denoising and other related domains. Current research primarily focuses on selecting and applying different deep learning algorithms, network structures, and labeling methods. However, less attention is paid to the impact of inherent dataset variations on the application of deep learning algorithms. This paper analyzes the impact of Batch Normalization in convolutional neural networks (CNN) on seismic data denoising. By employing theoretical formulas and conducting comparative numerical calculations, this study proposes recommendations for utilizing batch normalization layers based on the analysis of seismic data features. The suitability of incorporating batch normalization layers relies on the statistical distribution characteristics of the dataset. Effective improvements in network performance can be achieved only when the normalized energy distribution of the training set is concentrated within a strong amplitude region. Nevertheless, in seismic data denoising, it is generally advised to refrain from using batch normalization layers. These findings offer valuable insights for the improved application of deep learning algorithms in seismic data denoising.

Key words Deep learning; Batch normalization; Structure optimization; Denoising; Data characteristics analysis

收稿日期: 2023-02-20

基金项目: 国家“973”重点基础研究发展计划(项目编号: 2014CB239002); 中国石油集团科技重大专项(2019D-07)资助。

作者简介: 张鑫, 男, 1972年生, 高级工程师, 主要从事物探采集处理解释技术管理工作. E-mail: Zhx209@petrochina.com.cn

网络首发时间: 2023-11-14 14:34:59 网络首发地址: <https://link.cnki.net/urlid/11.2982.P.20231114.1052.002>

0 引言

近年来,深度学习算法在地震数据处理领域已得到较广泛的应用,特别是在初至拾取,去噪,插值等方面已有许多成功的工业化案例(郑浩等, 2020; 崔家豪等, 2022; 诸峰等, 2022)。在最初应用机器学习算法解决地震数据处理问题时,人们主要关注深度学习算法及网络结构的选择;随着应用的不断深入,人们开始关注深度学习算法的一些常用网络结构对不同应用场景下的应用效果,并基于所应用的地震数据特征,进行这些参数的选择和优化,从而取得更好的应用效果(王昊等, 2020)。

实际地震资料中不可避免的包含噪声,如随机噪声,异常噪声,面波,多次波等。这些噪声会干扰地震资料的成像反演效果,因此必须在地震数据的处理中去除这些噪声。利用深度学习去除地震资料中的噪声,拥有去噪效果好,操作简单,无需过多的先验假设等特点。Yu et al. (2019) 使用 DnCNN 网络,较细致地研究了随机噪声,线性干扰和多次波的去噪,展示了深度学习在地震去噪中的应用潜力。深度学习算法中有监督学习去噪方法(Liu et al., 2020; Dong et al., 2022)的去噪结果通常信噪比更高,主要难点在于标签制作方面。目前有监督学习的标签制作方法,主要为(1)采用正演来获得干净的地震资料,(2)使用传统方法去噪后的数据作为标签。方法(1)无法与实际资料进行良好的匹配,而方法(2)结果也并不可靠;深度学习算法中无监督深度学习去噪(Zhang et al., 2019; Liu et al., 2022; Yang et al., 2022b)通常伴随着过拟合,需要提前中断训练,且去噪的类型有限;深度学习算法中的自监督学习方法,如 Noise2Noise network(Lehtinen et al., 2018),通过对噪声的特征进行假设,训练过程只需要含噪声的数据。自监督学习具有无需无噪声的标签数据和较好的去噪效果等优点,是目前基于深度学习算法的地震数据去噪研究方向之一(Sun et al., 2020; Oliveira et al., 2021; Birnie et al., 2022; Meng et al., 2022; Wang X et al., 2022)。

但不管选用哪种深度学习算法,都涉及到网络结构的选择,包括网络层数,连接方式等;训练数据集的选择,包括标签数据制作等。以上这些选择可以随算法的不同而不同,从而带来更好的应用效果。在一些常用网络结构中,批量规范化层(Batch Normalization)对于地震数据去噪的最后效果影响较大。目前较多的去噪网络会选择使用批量规范化层,这主要是因为(1)目前地震勘探领域使用的网络,多为图像处理领域迁移过来,一些内在结构多为默认情况。但在地震数据处理和图像处理的不同应用场景中,最大的差异在于数据的分布形式或特征不同,这些差异会带来同一网络结构下的不同应用效果;(2)在地震去噪应用方面,由于噪声类型多样,振幅变化大,信号和噪音之间具有较大的特征差异,而所应用的深度学习算法大部分均为点对点的学习,训练和预测,处理前后数据的维度是不变的。输入数据的变化范围及分布特征对最后的去噪效果具有较大的影响。如在图像处理领域中,色标值通常较少,灰度图为 255 个级别,而彩色图为 255^3 个级别。对于地震数据而言,无论是数值的范围,还是级别数量,都远高于传统的图像处理领域。因此在基于地震数据特征分析的基础上,进行网络结构的选择和优化,是提高深度学习的地震去噪效果的重要途径之一。

基于以上思路,我们着重探讨了两种常见的卷积神经网络结构,解码编码网络结构(Unet 形式的 FCDNet)(Yang et al., 2022a)和常规等维度的网络结构(DnCNN)(Yu et al., 2019)下的 Batchnorm 层的使用效果,提出了基于地震数据特征分析下的 Batchnorm 层使用建议。结果表明,Batchnorm 的使用,取决于网络的输入资料的归一化能量统计分布特征。对于地震资料中的点对点的预测问题,训练数据集各个切片能量相近,且归一化能量集中在较强区域时,如经过一些列处理后的叠后资料或经过较强的振幅补偿后的叠前资料,Batchnorm 的使用才可以加速收敛。但数据分布不均匀时,如叠前未做能量均衡的资料,振幅分布不均匀,特征差异较大,Batchnorm 层的使用会起反作用。相关结论为后续深度学习网络参数选择和优化以及改善深度学习算法在地震数据去噪中的应用效果提供了重要的参考。

1 方法原理

1.1 深度学习的卷积网络结构

对于地震数据的去噪处理，各种卷积深度网络的数学表达，可以用以下一个简单的公式来描述，

$$cleandata = f(noisydata; \theta) \quad (1)$$

式中， $cleandata$ 表示网络输出的预测结果， $noisydata$ 表示输入的含噪声记录， $f()$ 代表卷积网络模型， θ 表示网络模型的参数，常指网络模型所要求的各个结构需要学习的一系列参数，包括卷积核权重，偏执，规范化层参数等。

本文重点阐述在利用卷积深度网络进行地震数据去噪时，批量规范化算法（Batch Nomination）对具有不同数据特征的地震数据去噪效果的影响。在卷积神经网络中，常见结构有常规等维度的网络结构，和解码编码类型的网络结构。为此，我们选取了两个参考网络，DnCNN(Yu et al., 2019)和FCDNet(Unet 形式)(Yang et al., 2022a)，作为对比研究的具体网络模型。这两个网络模型结构如图 1 和图 2 所示。这两个网络主要为卷积结构，均使用了 Batchnorm 层。Yu et al. (2019) 使用 DnCNN 网络，探讨了深度学习在去噪应用中的潜力，对随机噪声，线性干扰以及多次波等，进行了相关研究。而 Yang et al. (2022a) 使用 FCDNet 主要针对的是 DAS 资料的环境噪声。在本文实际应用中，除 FCDNet 没有使用原文中避免过拟合的 Droupout 层外，两个网络结构均与原文保持一致。

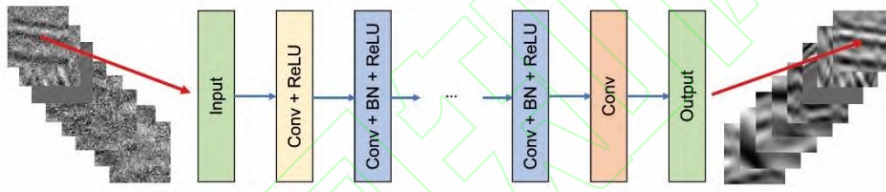


图 1. DnCNN 网络结构(Yu et al., 2019)
Fig. 1 DnCNN network structure (Yu et al., 2019)

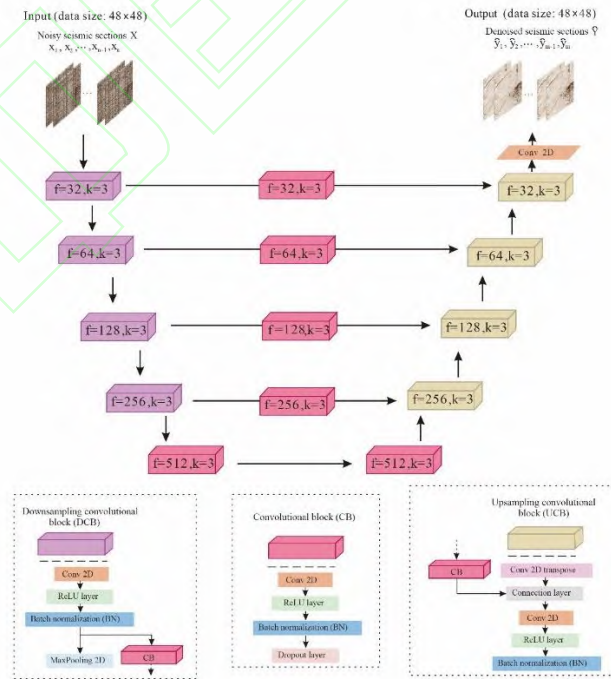


图 2. FCDNet 网络结构(Yang et al., 2022a)
Fig. 2 FCDNet network structure (Yang et al., 2022a)

1.2 Batchnorm 层

根据前人研究(Ioffe et al., 2015), 网络层数加深的同时, 收敛会更加棘手。批量规范化(Batch Normalization)可以加速深层网络的收敛速度。因此在卷积深度网络中, 一般都应用了批量规范化算法, 由 Batchnorm 层来实现。现在大多数网络采用的批量规范化的算法及参数可由式(2)表示:

$$\text{BN}(x) = \gamma \odot \frac{x - \hat{\mu}_B}{\hat{\sigma}_B} + \beta \quad (2)$$

式中, X 表示输入的小批量数据 (mini batch), $\hat{\mu}_B$ 和 $\hat{\sigma}_B$ 为小批量数据的均值和标准差。 γ 和 β 分别为拉伸参数和偏移参数。

Batchnorm 操作, 让每一层的输入数据分布变得稳定, 但降低了数据的表达能力, 也改变了原有的数据信息表达, 可能造成信息丢失。为了解决这些问题。 γ 和 β 这两个参数的目的就是保证每一次数据经过归一化后还保留原有学习来的特征, 恢复数据的表达能力, 同时又能完成归一化操作, 这两个参数需要和其他模型参数一起学习, 维度和 X 一致。

图 3 给出了一个 Batchnorm 操作前后地震数据中噪音和信号的分布特征对比的例子, 从中可以看到, 经过 Batchnorm 后 (图 3-b), 这个切片的信号和标签结果 (图 3-c) 差异变得更大, 将地震信号能量变强, 影响了其保幅性。因此训练网络参数时, 还需要额外弥补相关振幅的损失, 这对于深度学习去噪而言是不利的。

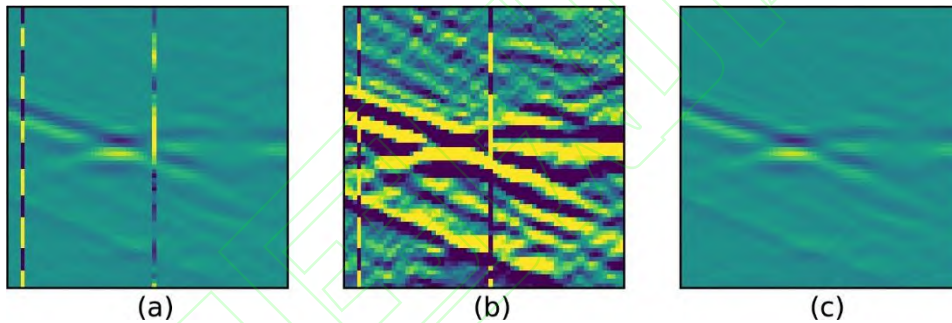


图 3 单个 Batchnorm 层的结果展示, (a) 为输入数据, (b) 为 Batchnorm 操作后的结果, (c) 为地震数据标签
Fig. 3 Results of a single Batchnorm layer. (a) is the input data, (b) is the result after Batchnorm, and (c) is the clean label.

地震数据去噪在深度学习中, 其基本原理是将每个点的值分为两个部分, 一部分为噪声, 一部分为有效信号, 并将地震资料切片后, 小批量的放入网络中进行训练; 而且不同于一些分类或识别的任务, 地震数据去噪属于点对点的预测, 会将切片中每个采样点的值分为两个互补的部分 (含噪信号 = 干净地震信号 + 噪音信号)。这种点对点的预测问题, 十分依赖与原始数据的分布信息。例如, 在随机噪音去除中, 一般情况下, 在含噪地震信号中振幅较强的采样点, 预测后的无噪音地震信号振幅也是较强的。去噪前后, 地震剖面的各采样点的整体相对大小关系变化不大。也就是说, 原有的数据分布是一个非常重要的特征。若在一个小批量样本中, 切片之间的地震数据振幅差异较大时, 经过 Batchnorm 层后, 会改变其原来的振幅值。这种规范化的操作, 在一定程度上会影响地震数据的保幅性。虽然引入了拉伸参数和偏移参数来弥补对数据分布的改变, 但在实际应用时, 这些参数的弥补效果并不明显, Batchnorm 的使用在一定程度上仍然可能受到数据分布差异的影响。本文将通过具体的数值实验, 来探讨数据分布的差异对去噪结果会产生怎样的影响。

2 深度学习去噪效果分析

2.1 数据准备

本次数据资料, 采用 sigsbee 模型正演记录, 得到干净的地震数据标签。通过添加高斯噪声和采集的环境记录噪声, 来形成相关的训练和预测数据集。我们制作了两组数据集。数据集一共 350 炮,

选取其中的 326 炮作为训练集，24 炮数据作为测试或验证。对数据进行切片，所得到的数据数量和大小如表 1 所示。

表1 数据集参数表
Table 1. The parameters of the datasets

类型	切片数量	切片大小
随机噪声数据集	训练集	41076 64×64
	测试或验证集	24 1248×352
异常环境噪音数据集	训练集	61614 64×64
	测试或验证集	24 1248×352

测试或验证集为同一批数据，不参与到训练之中。并且在计算测试集相关精度时，我们直接将整炮的数据资料放入到网络中进行预测。对于 U-net 网络，上下采样的操作限制了数据的维度大小，需要用零值来扩展数据的维度，预测结束后，再裁去边缘填充的值。

2.2 去除随机噪声的效果分析

数值模拟中，每次训练网络的超参数是相同的，batch size 为 64，损失函数为 L_2 范数，优化器算法为 Adam，采用混合精度技术加速训练。图 4 为随机噪声数据集的训练结果，展示了验证集精度随训练次数的变化。

图 4 分别展示两种不同网络模型下，保留批量规范化（Validation SNR with Batchnorm）和去除批量规范化（Validation SNR without Batchnorm）时，两个学习率（ $lr=0.001$ 和 0.0001 ）的训练结果。其中，资料信噪比的计算公式为(李勇等, 2019; Wang et al., 2020):

$$S/N = 10 \log_{10} \frac{\|seis_{pre}\|_2^2}{\|seis_{label} - seis_{pre}\|_2^2} \quad (3)$$

式中， $seis_{pre}$ 表示网络输出的预测结果， $seis_{label}$ 表示无噪声的标签数据。

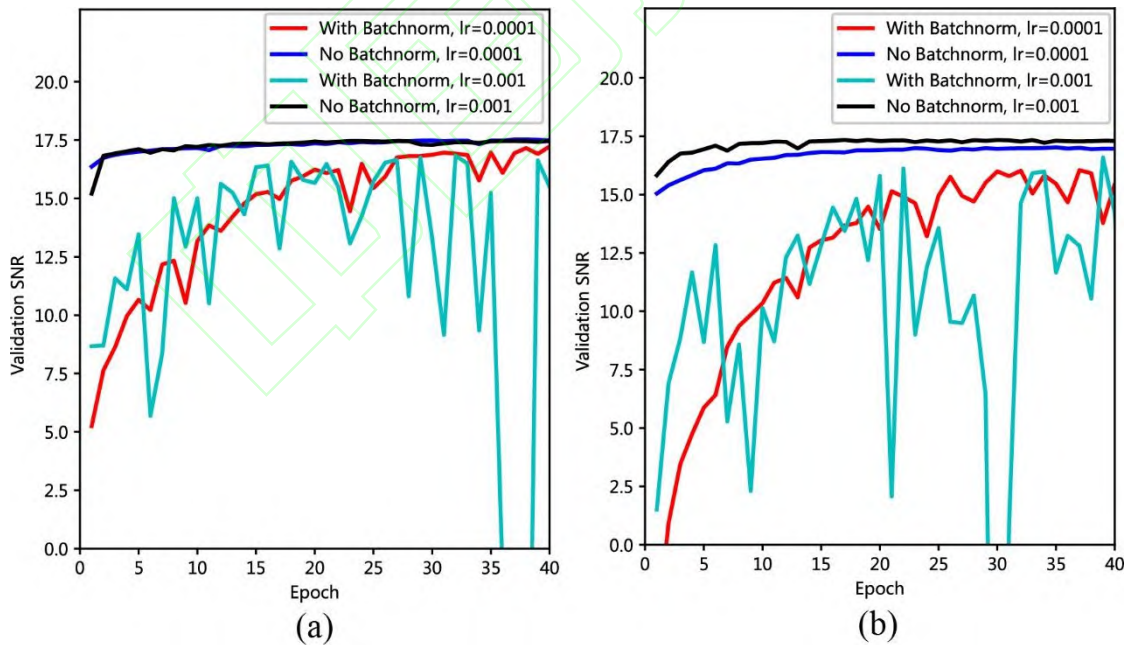


图 4 随机噪音训练结果。(a)为 DnCNN，(b)为 FCDNet
Fig. 4 Training results on the random noisy data. (a) is DnCNN, (b) is FCDNet

从图 4 中，可以明显观察到，当去除网络结构中的 Batchnorm 层之后，信噪比明显提高，特别是前期训练，可以快速达到一个较高的精度。这就是因为 Batchnorm 操作影响地震数据的保幅性，因此需要额外的训练来恢复相关振幅信息。此外，去除 Batchnorm 层之后，整体的训练稳定性也是有很大

提升。图 5 和图 6 展示了地震资料的去噪效果。整体而言，在相同参数下，FCDNet 的去噪效果，在弱能量方面的恢复，有较大优势。两个网络的去噪结果中，均是不使用 Batchnorm 层的效果更好。

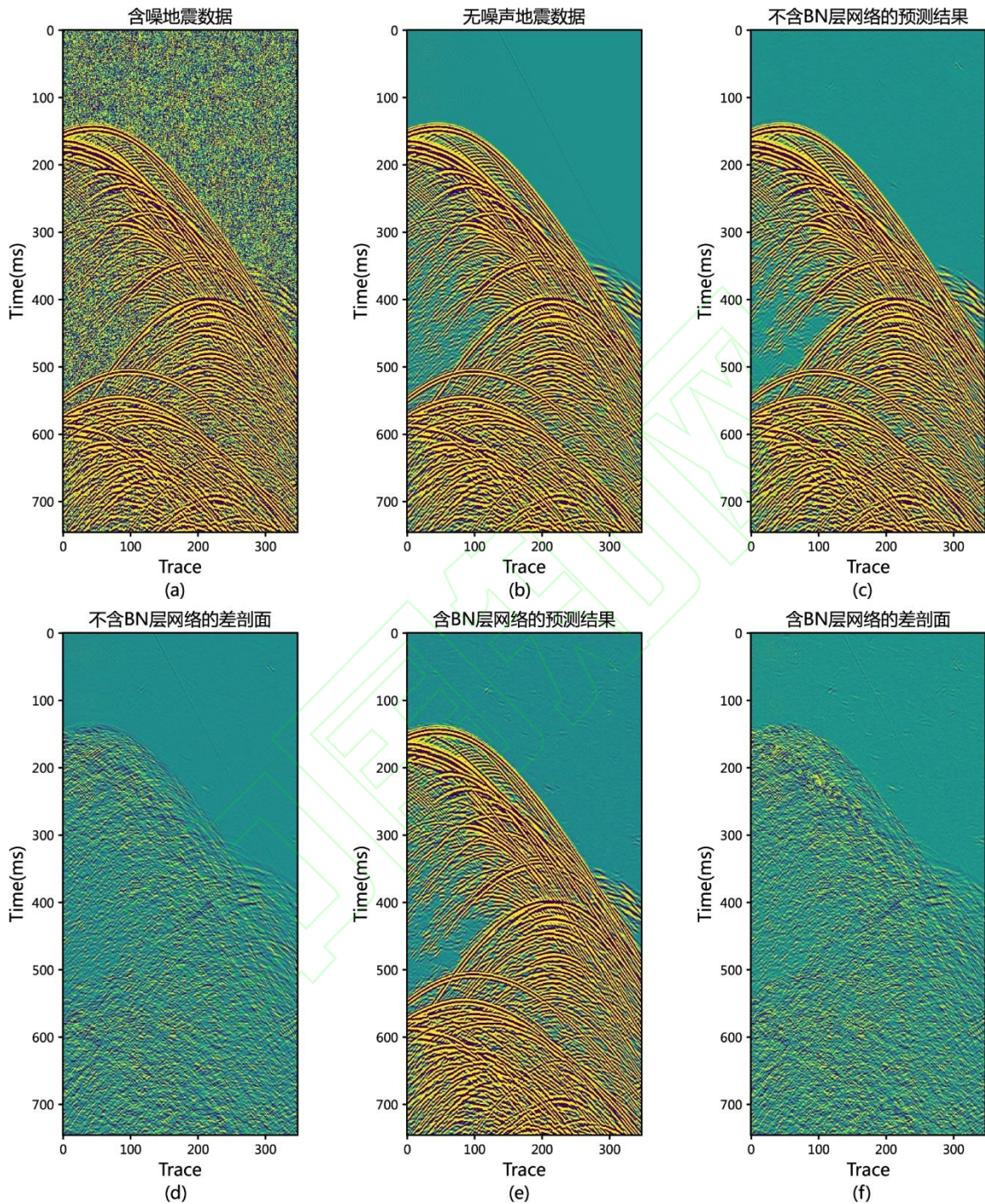


图 5 DnCNN 网络去噪结果展示。(a) 为含噪声的数据，(b) 为干净的地震数据，(c) 为不使用 BN 层的预测结果，(d) = (c) - (b) 的结果，(e) 为使用 BN 层的预测结果，(f) 为 (e) 减去 (b) 的结果。
Fig. 5 Denoising results of the DnCNN network. (a) Noisy seismic data, (b) Clean seismic data, (c) Denoising results without BN layer, (d)=(c)-(b), (e) Denoising results with BN layer, (f)=(e)-(b)

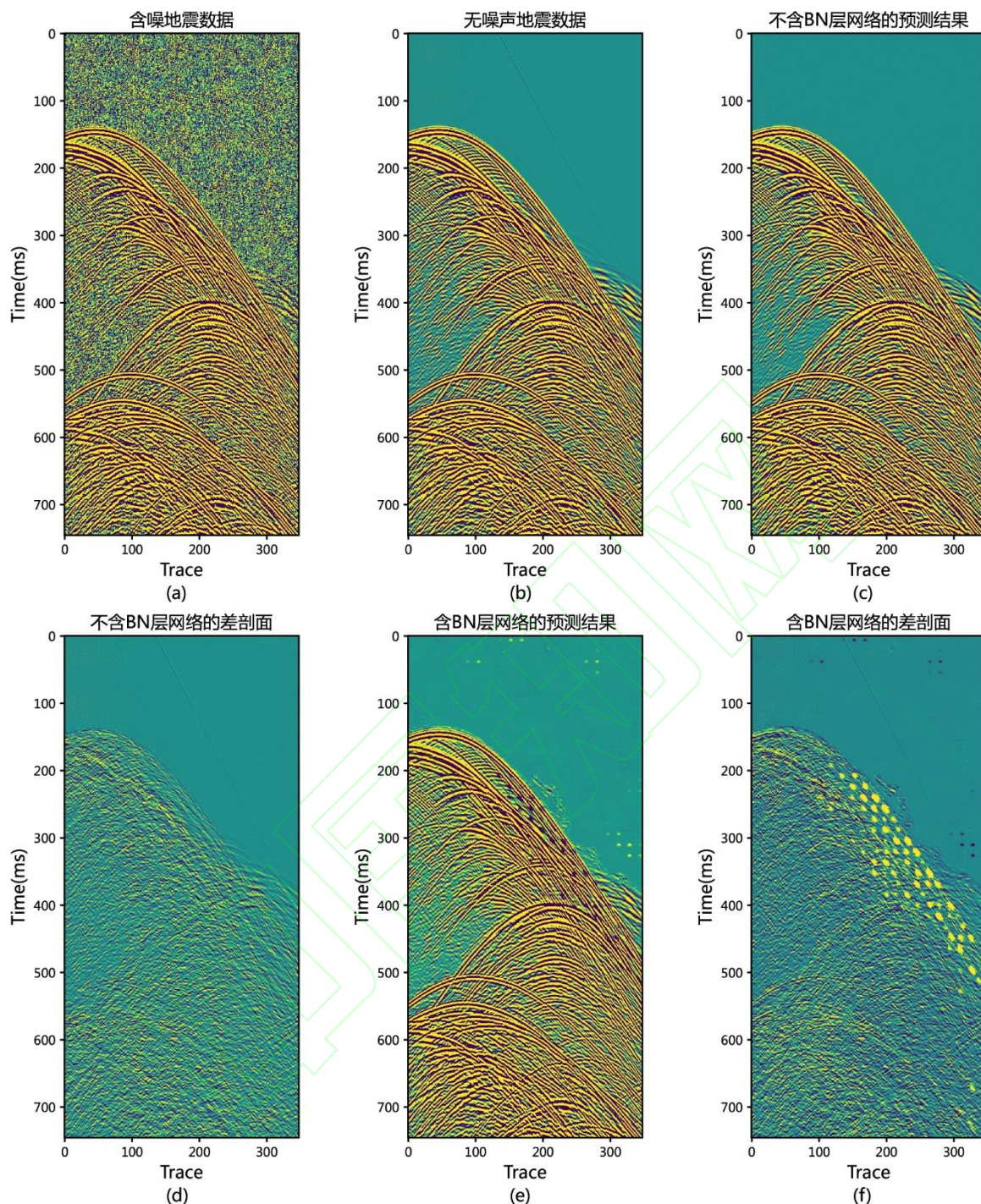


图6 FCDNet 网络去噪结果展示。(a) 为含噪声的数据, (b) 为干净的地震数据, (c) 为不使用 BN 层的预测结果, (d) 为 (c) 减去 (b) 的结果, (e) 为使用 BN 层的预测结果, (f) 为 (e) 减去 (b) 的结果。
 Fig. 6 Denoising results of the FCDNet network. (a) Noisy seismic data, (b) Clean seismic data, (c) Denoising results without BN layer, (d)=(c)-(b), (e) Denoising results with BN layer, (f)=(e)-(b)

2.3 去除异常噪声的效果分析

采用和随机噪声相同的超参数和试验方式, 对异常噪声进行训练, 结果如图 7 所示, 图示方式与图 4 类似。所谓异常噪音是指由环境因素引起的强振幅, 强干扰噪音, 如人员走动, 打钻, 工厂, 大型机械等活动, 这类型的噪音频带范围宽, 能量强, 而且目前去除方法主要依靠道切除或是直接能量压制。从上面的分析可知, 异常噪音的存在, 不同切片的能量分布差异势必较大。因此 Batchnorm 层

的存在会造成负面影响。从图 7 中也能得出同样的规律，不使用 Batchnorm 层时，收敛速度和精度都得到较大提升。图 8 和图 9 展示了地震资料的相关去噪效果。从这些图中可以观察到，使用 Batchnorm 层后，异常环境噪音的残留加重，这就是因为批量规范化，会将异常部分的噪音振幅值改变，使得网络难以学习到真振幅的恢复。

规范化方法除了 Batch Normalization，还有 Instance Normalization, Layer Normalization 等。对于不同的规范化方法，都有近似的结论。一个小批量的输入数据，包含四个维度（数量，通道，高，宽），这几种方法的主要差异在于在不同的维度上做规范化。地震数据应用在深度网络中，由于单个剖面数据量大，需要进行切片。因此一个小批次数据中，切片可能来自于同一剖面，也可能来自于不同剖面，且地震数据最初的输入一般属于单通道。因此无论是何种规范化方法，或多或少都会影响数据的原始分布。

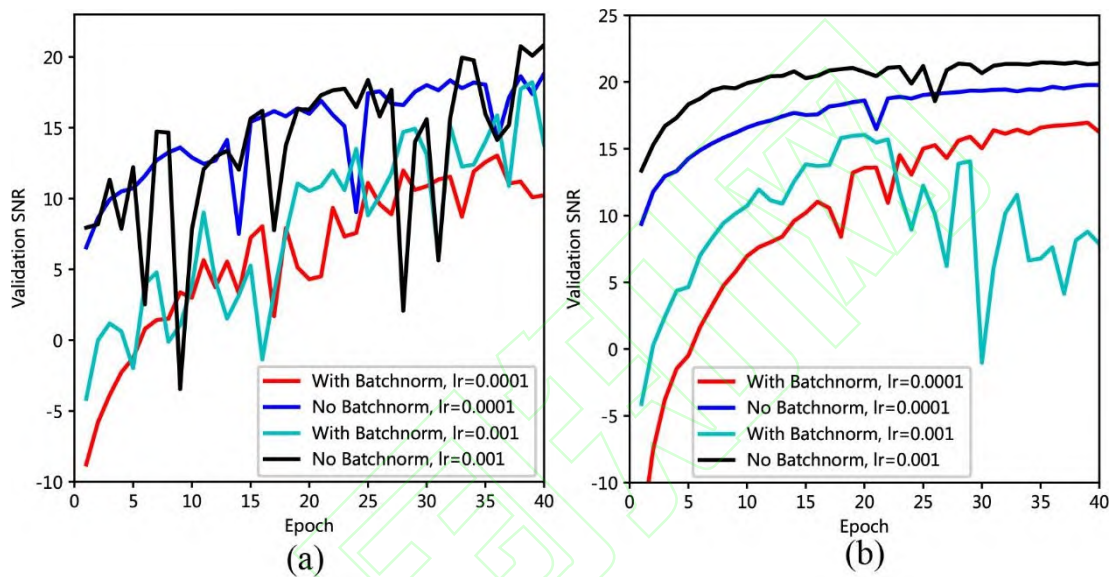


图 7 异常环境噪音训练结果。(a)为 DnCNN，(b)为 FCDNet
Fig .7 Training results on noisy data with abnormal environmental conditions. (a) is DnCNN, (b) is FCDNet

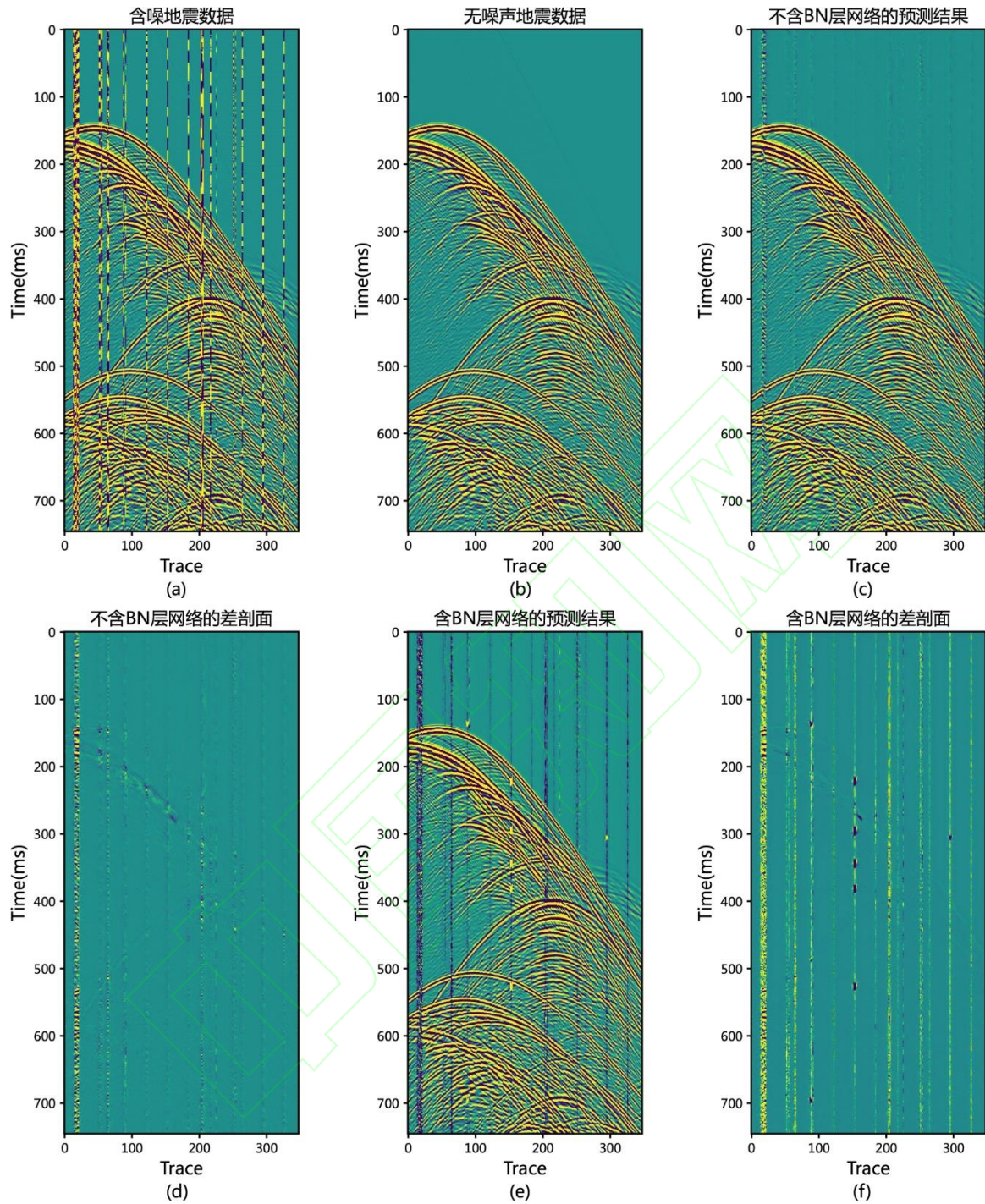


图 8 DnCNN 网络去噪结果展示。(a) 为含噪声的数据, (b) 为干净的地震数据, (c) 为不使用 BN 层的预测结果, (d) 为 (c) 减去 (b) 的结果, (e) 为使用 BN 层的预测结果, (f) 为 (e) 减去 (b) 的结果。
 Fig .8 Denoising results of the DnCNN network. (a) Noisy seismic data, (b) Clean seismic data, (c) Denoising results without BN layer, (d)=(c)-(b), (e) Denoising results with BN layer, (f)=(e)-(b)

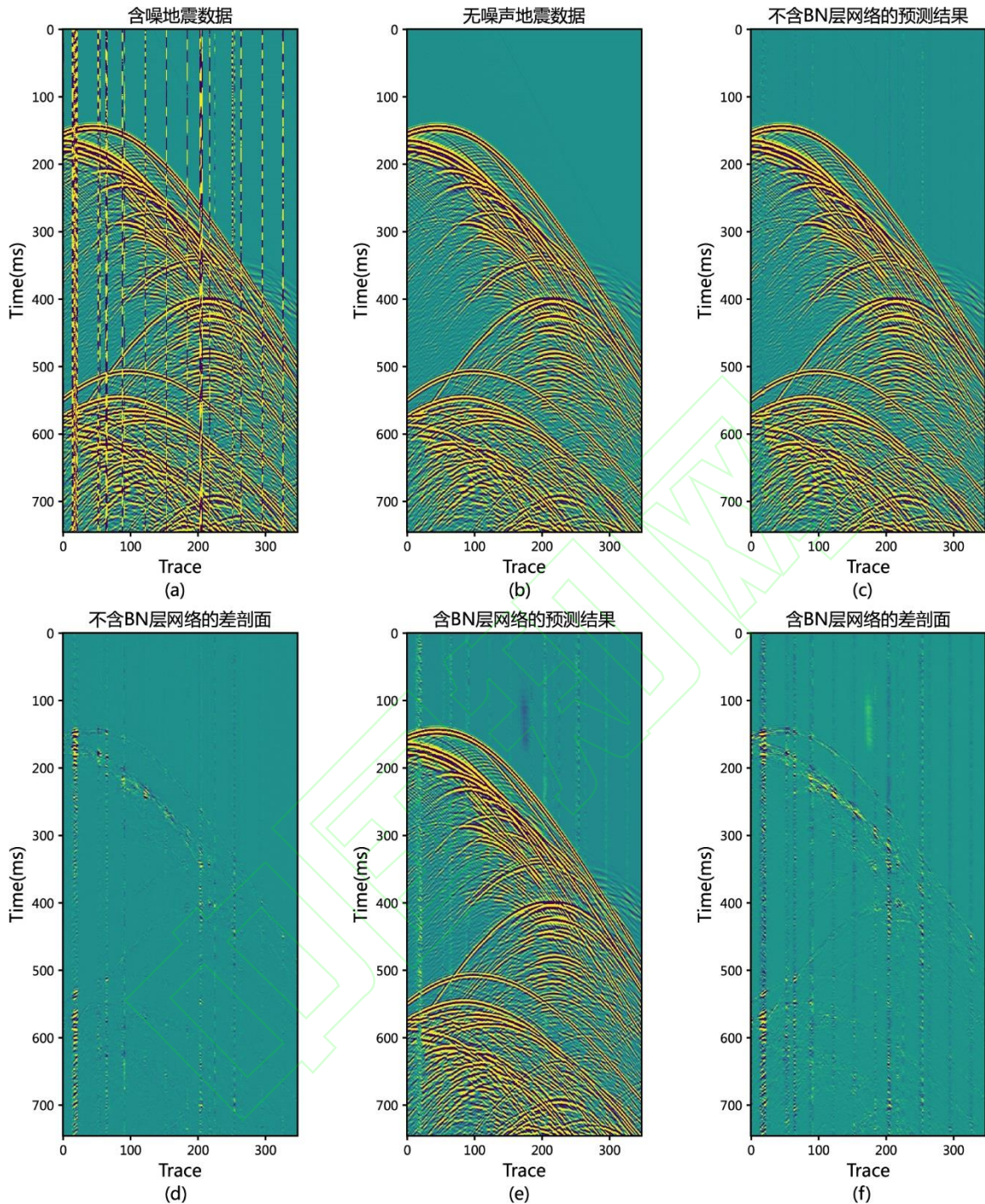


图9 FCDNet 网络去噪结果展示。(a) 为含噪声的数据, (b) 为干净的地震数据, (c) 为不使用 BN 层的预测结果, (d) 为 (c) 减去 (b) 的结果, (e) 为使用 BN 层的预测结果, (f) 为 (e) 减去 (b) 的结果。
 Fig. 9 Denoising results of the FCDNet network. (a) Noisy seismic data, (b) Clean seismic data, (c) Denoising results without BN layer, (d)=(c)-(b), (e) Denoising results with BN layer, (f)=(e)-(b)

3 数据特征与批量规范化的使用

3.1 数据分布分析

从上面这些结果可以发现, Batchnorm 层对于最终的结果影响是负作用。但是在前人的研究中, 用于去噪处理的深度网络较多采用了 Batchnorm 结构。如 Yu et al. (2019) 使用了 DnCNN 网络结构研究了随机噪音的效果。通过其开源的代码重现可以发现, 由于 Yu 等人对地震数据使用了振幅均衡

(AGC)，地震数据值的整体分布较均匀，且能量较强。而本文采用的数据为原始炮集数据，不同的切片能量分布差异较大，更接近实际情况。

为了展示 Batchnorm 层的在不同数据分布上的效果差异，我们将直接使用 Yu et al. (2019) 开源的代码和开源的数据进行分析。这一部分我们只分析了随机噪音的情况，因为异常环境噪音去除属于地震数据最先处理的几个步骤，一般是在原始地震记录上进行处理，而随机噪音的去除情况要多一些。我们创建了 3 个数据集，分别使用不同的能量均衡窗，并添加随机噪音，初始信噪比均为 2.03。

表2 数据集参数表

Table 2. The parameters of the datasets

数据集类型	时窗大小	编号	切片大小及数量	初始信噪比
原始炮集	无	A	64×64×41186	2.03
大时窗能量均衡	1 秒	B	64×64×41186	2.03
小时窗能量均衡	0.1 秒	C	64×64×41186	2.03

在图 10 中，我们统计了相关切片的能量归一化分布。对于未处理的原始炮集资料而言（图 10- (a)），大部分的切片能量较低。这是因为波在传播过程中出现衰减。在使用较大的时窗，对数据进行能量均衡后（图 10- (b)），可以看见振幅值明显增加。振幅均衡过程使用的窗口越小，整个剖面中的振幅增加越明显。图 10- (c) 的数据振幅最强，采用较小的时窗均衡后，不同切片的振幅差异也较小。

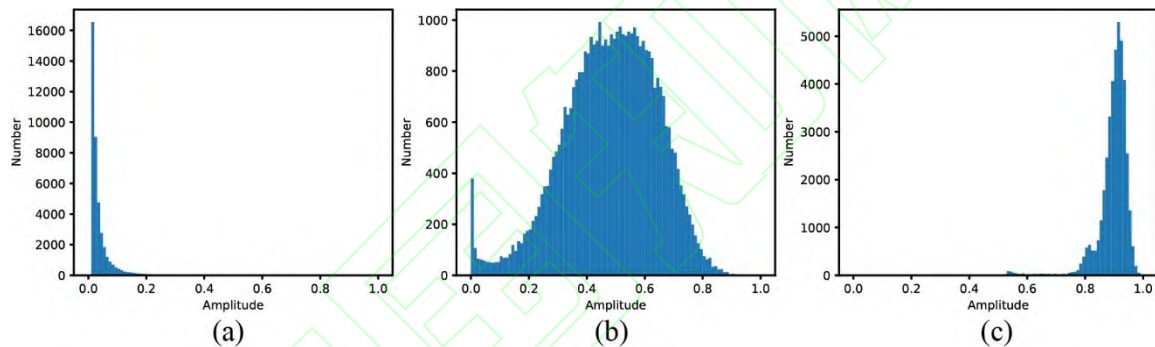


图 10 不同数据集的训练集的切片振幅统计，(a)、(b)、(c) 分别对应表 2 中的数据编号 A、B、C

Fig. 10 Patches amplitude statistics of training sets, where (a), (b), and (c) correspond to data numbers A, B, and C in Table 2, respectively

为了探讨 Batchnorm 层的效果。我们使用 Batchnorm 公式（公式 (2)）对表 2 中的数据进行处理，其中 γ 取 1，和 β 取 0，采取 64 的批量大小进行计算，结果如图 11 所示。对比图 11 和图 10，可以观察到其中的图 11- (a) 数据分布变化最大。而图 11- (c) 的数据分布变化最小。因此对于数据 A，Batchnorm 层的计算会严重破坏原始数据的分布。对于去噪而言，原始资料的数据振幅和去噪后的数据振幅差异较小，网络的输入数据切片和输出数据切片的振幅值相差也不大。对于数据 A 而言，这种破坏分布的现象，会严重影响网络的训练。

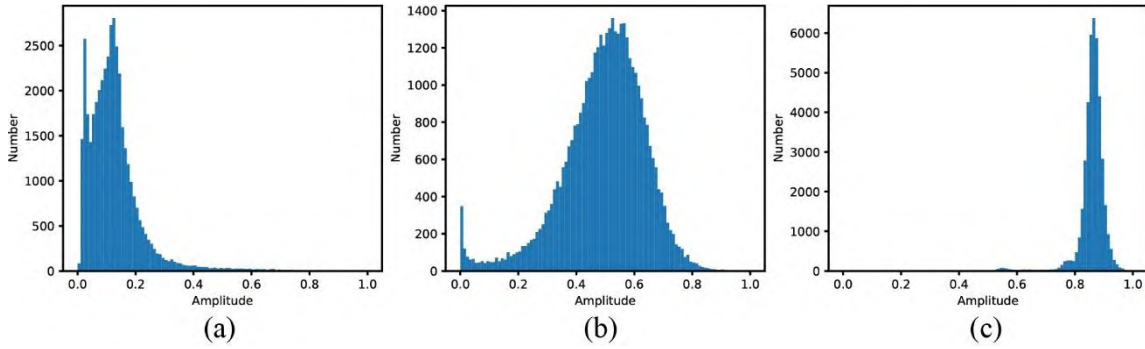


图 11 单次 Batchnorm 计算后的切片振幅统计，(a)、(b)、(c) 分别对应表 2 中的数据编号 A、B、C
 Fig .11 Patches amplitude statistics of training sets after Batchnorm operation, where (a), (b), and (c) correspond to data numbers A, B, and C in Table 2, respectively

除了观察图 10 或 11 的方法，我们也构建一个 Batchnorm 判别结构，如图 12 所示。整个网络结构只有 Batchnorm 结构，无其他任何计算。网络的输入数据和标签数据一样。该网络结构的目的是为了判断数据经过 Batchnorm 层后，是否可以恢复，从而判断数据是否适用于具有 Batchnorm 层的卷积网络结构。

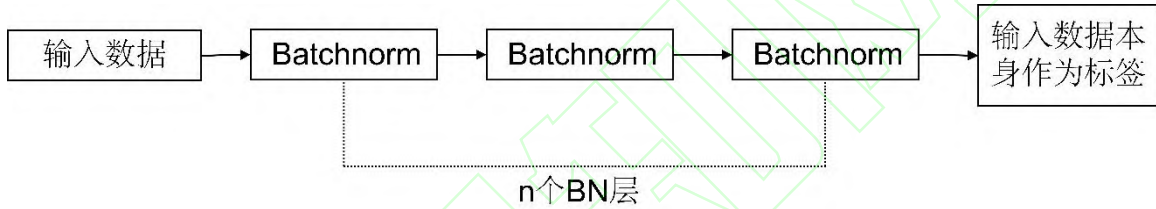


图 12 Batchnorm 判别网络
 Fig .12 Batchnorm discriminant network

我们采用 64 的批量大小进行训练，和 DnCNN 结构一样，一共采用 15 个 Batchnorm 层进行计算，结果如图 13 所示。图 13 也明确展示了，数据 A 的信噪比极低，完全无法恢复数据，而数据 B 和数据 C 的信噪比相对较高。图 13 中的信噪比高低，也预示着在使用 Batchnorm 层的网络中，数据 C 的效果相对数据 B 和数据 C 会更好，而数据 A 在具有 Batchnorm 层的网络中，效果最差。

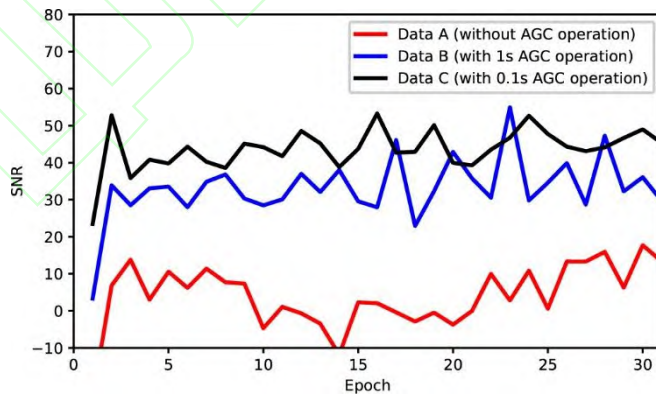


图 13 不同数据的 Batchnorm 判别网络训练效果
 Fig .13 Training results of Batchnorm discriminant network with different data

3.2 训练结果对比

为了判断我们上述的分析是否正确，我们使用 DnCNN 网络和 FCDNet 网络训练三个数据集去除随机噪音。图 14 展示三个数据集的 DnCNN 网络训练结果。图中虚线为采用 Batchnorm 层的网络，实线结果采用的是无 Batchnorm 层的网络。数据集 A 的结果（图 14- (a)）和之前的分析一致，具有

Batchnorm 结构的信噪比明显低于不采用 Batchnorm 层的网络。当数据集做过能量均衡后, 图 14-(b) 和图 14-(c) 的效果中, 具有 Batchnorm 层的效果略好于没采用 Batchnorm 层的结果, 且使用 Batchnorm 层的训练结果趋势更稳定。但总体而言, 使用 Batchnorm 和不使用 Batchnorm 层的结果差异不大。

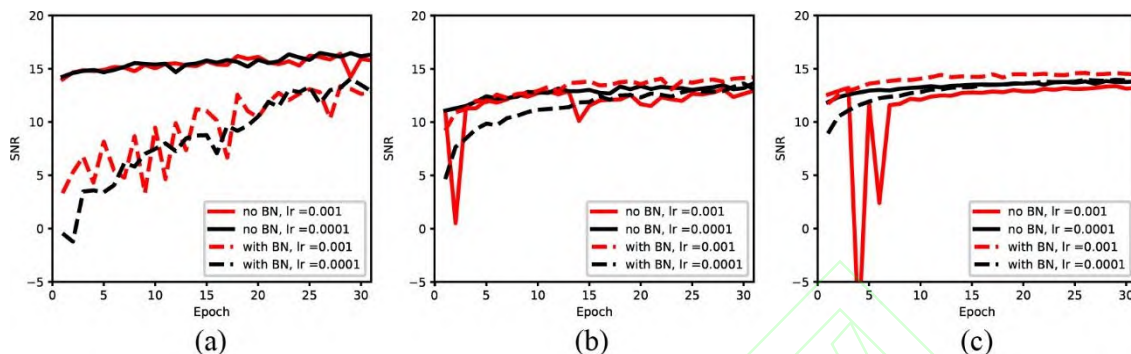


图 14 不同数据集的 DnCNN 网络训练结果, (a)、(b)、(c) 分别对应表 2 中的数据编号 A、B、C
Fig. 14 Training results of DnCNN networks with Different Datasets, where (a), (b), and (c) correspond to data numbers A, B, and C in Table 2, respectively

对于 FCDNet 而言 (图 15), 在对数据做能量均衡后 (图 15-(b) 和图 15-(c)), 仍然是不使用 Batchnorm 层的网络结果稍微优于采用 Batchnorm 层的网络。但不同数据集的整体变化趋势和图 14 保持一致, 做能量均衡之后, 采用 Batchnorm 层的网络和不采用 Batchnorm 层的网络的差异减小。

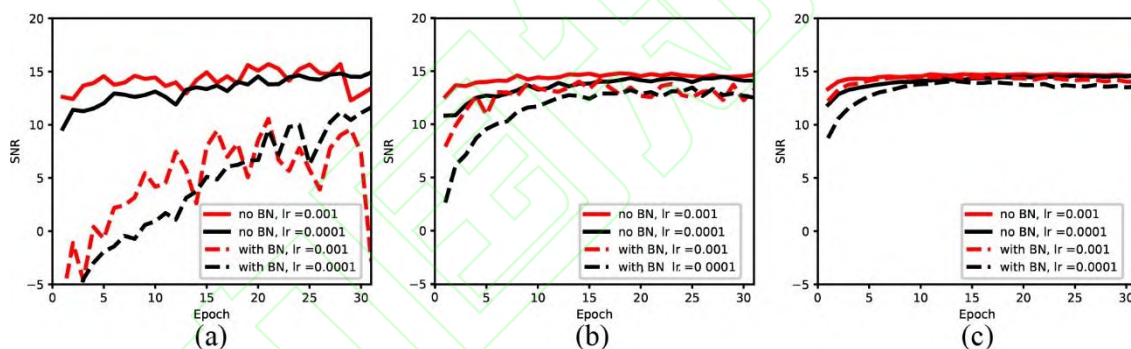


图 15 不同数据集的 FCDNet 网络训练结果, (a)、(b)、(c) 分别对应表 2 中的数据编号 A、B、C
Fig. 15 Training results of FCDNet networks with Different Datasets, where (a), (b), and (c) correspond to data numbers A, B, and C in Table 2, respectively

通过对不同训练集的对比分析, 可以明显观察到不同的数据分布, 对于 Batchnorm 层的使用有着明显的差异。虽然这种差异在不同的卷积网络结构中, 稍有不同, 但整体的变化趋势是一致的。当数据比较均衡, 整体能量较强时, 使用 Batchnorm 层才可能有优势。

4 结论

本次研究以两个卷积深度网络结构为例, 通过对 Batchnorm 的原理以及实验分析, 基于数据分布特征的分析来判断使用 Batchnorm 层的优劣。可以在设计网络结构时, 提供有价值的参考。在实际应用中, 可以通过分析训练集的特征, 来判断 Batchnorm 层是否应该使用:

(1) 总体而言, 如果对数据的情况不了解, 不建议采用 Batchnorm 层的网络来处理地震数据去噪。本文的测试结果展示, 使用了 Batchnorm 层后, 即使效果有提升, 通常提升也是微弱的。而当数据不合适时, 使用 Batchnorm 层后的效果会明显降低。

(2) 对于原始资料, 特别是未做能量均衡时的资料, 不建议采用 Batchnorm 层的网络。例如, 在处理流程中, 通常最先进行异常环境噪音的去除, 而这类噪音通常存在于原始炮集中, 所以对于异常

环境噪声的去除, 不建议使用 **Batchnorm** 层。

(3) 如果对数据做过较强的能量均衡, 或着处理叠后资料。可以分析训练集切片的振幅统计情况。当整体能量较强, 且切片的能量比较统一时(参照图 10-(c)), 可以考虑使用 **Batchnorm** 层。

(4) 通过 **Batchnorm** 判别网络(图 12), 对输入资料进行训练, 标签和输入为同一数据。通过 **Batchnorm** 判别网络, 来观察数据是否可以恢复原有分布情况。当信噪比很差时, 不建议使用 **Batchnorm** 层, 信噪比足够高时, 可以考虑使用 **Batchnorm** 层结构。

致谢 感谢审稿专家提出的修改意见和编辑部的大力支持!

参考文献

- [1] Birnie C, Alkhalifah T. 2022. Boosting self-supervised blind-spot networks via transfer learning[M]//Second International Meeting for Applied Geoscience & Energy. Society of Exploration Geophysicists and American Association of Petroleum Geologists, doi:10.1190/image2022-3748691.1.
- [2] Cui J H, Yang P, Wang H Q, et al. 2022. Research on automatic picking of seismic velocity spectrum based on deep learning[J]. Chinese Journal of Geophysics (in Chinese), 65(12): 4832-4845, doi: 10.6038/cjg2022P0769
- [3] Dong X, Lin J, Lu S, et al. 2022. Seismic Shot Gather Denoising by Using a Supervised-Deep-Learning Method with Weak Dependence on Real Noise Data: A Solution to the Lack of Real Noise Data[J]. Surveys in Geophysics, 43(5): 1363-1394, doi:10.1007/s10712-022-09702-7.
- [4] Ioffe S, Szegedy C. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[M]. arXiv[2023-01-07], doi: 10.48550/arXiv.1502.03167.
- [5] Lehtinen J, Munkberg J, Hasselgren J, et al. 2018. Noise2Noise: Learning Image Restoration without Clean Data[M]. arXiv[2022-11-22], doi:10.48550/arXiv.1803.04189.
- [6] Li Y, Zhang Y M, Lei Q, et al. 2019. Online dictionary learning seismic weak signal denoising method under model constraints[J]. Chinese Journal of Geophysics (in Chinese), 62(1): 411-420, doi: 10.6038/cjg2019M0373
- [7] Liu D, Deng Z, Wang C, et al. 2022. An Unsupervised Deep Learning Method for Denoising Prestack Random Noise[J]. IEEE Geoscience and Remote Sensing Letters, 19: 1-5, doi:10.1109/LGRS.2020.3019400.
- [8] Liu D, Wang W, Wang X, et al. 2020. Poststack Seismic Data Denoising Based on 3-D Convolutional Neural Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 58(3): 1598-1629, doi:10.1109/TGRS.2019.2947149.
- [9] Meng F, Fan Q, Li Y. 2022. Self-Supervised Learning for Seismic Data Reconstruction and Denoising[J]. IEEE Geoscience and Remote Sensing Letters, 19: 1-5, doi:10.1109/LGRS.2021.3068132.
- [10] Oliveira D A B, Semin D G, Zaytsev S. 2021. Self-Supervised Ground-Roll Noise Attenuation Using Self-Labeling and Paired Data Synthesis[J]. IEEE Transactions on Geoscience and Remote Sensing, 59(8): 7147-7159, doi:10.1109/TGRS.2020.3029914.
- [11] Sun Y, Gurevich B, Glubokovskikh S, et al. 2020. A solid/fluid substitution scheme constrained by pore-scale numerical simulations[J]. Geophysical Journal International, 220(3): 1804-1812, doi:10.1093/gji/ggz556.
- [12] Wang H, Yan J Y, Fu G M, et al. 2020. Current status and application prospect of deep learning in geophysics[J]. Progress in Geophysics, 35(2): 642-655. doi: 10.6038/pg2020CC0476
- [13] Wang H, Zhang Q, Zhang G, et al. 2020. Self-training and learning the waveform features of microseismic data using an adaptive dictionary[J]. GEOPHYSICS, 85(3): KS51-KS61, doi:10.1190/geo2019-0213.1.
- [14] Wang X, Sui Y, Wang W, et al. 2022. Random Noise Attenuation by Self-supervised Learning from Single Seismic Data[J]. Mathematical Geosciences, doi:10.1007/s11004-022-10032-y.
- [15] Yang L, Fomel S, Wang S, et al. 2022a. Denoising of distributed acoustic sensing (DAS) data using supervised deep learning[J]. GEOPHYSICS: 1-93. doi:10.1190/geo2022-0138.1.
- [16] Yang L, Wang S, Chen X, et al. 2022b. Unsupervised 3-D Random Noise Attenuation Using Deep Skip Autoencoder[J]. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-16, doi:10.1109/TGRS.2021.3100455.
- [17] Yu S, Ma J, Wang W. 2019. Deep learning for denoising[J]. GEOPHYSICS, 84(6): V333-V350, doi:10.1190/geo2018-0668.1.
- [18] Zhang M, Liu Y, Chen Y. 2019. Unsupervised Seismic Random Noise Attenuation Based on Deep Convolutional Neural Network[J]. IEEE Access, 7: 179810-179822, doi:10.1109/ACCESS.2019.2959238.
- [19] Zheng H, Zhang B. 2020. Intelligent seismic data interpolation via convolutional neural network[J]. Progress in Geophysics, 35(2): 721-727. doi: 10.6038/pg2020DD0013
- [20] Zhu F, Cheng Q, Li S S, et al. 2022. Research on noise suppression of seismic migration based on deep learning[J]. Progress in Geophysics, 37(2): 746-755. doi: 10.6038/pg2022FF0121

附中文参考文献

- [21] 崔家豪, 杨平, 王洪强等. 2022. 基于深度学习的地震速度谱自动拾取研究[J]. 地球物理学报, 65(12): 4832-4845, doi: 10.6038/cjg2022P0769.
- [22] 李勇, 张益明, 雷钦等. 2019. 模型约束下的在线字典学习地震弱信号去噪方法[J]. 地球物理学报, 62(1): 411-420, doi: 10.6038/cjg2019M0373.
- [23] 王昊, 严加永, 付光明等. 2020. 深度学习在地球物理中的应用现状与前景[J]. 地球物理学进展, 35(2): 642-655, doi: 10.6038/pg2020CC0476.
- [24] 诸峰, 程前, 李帅帅等. 2022. 基于深度学习的地震偏移噪声压制研究[J]. 地球物理学进展, 37(2): 746-755, doi: 10.6038/pg2022FF0121.
- [25] 郑浩, 张兵. 2020. 基于卷积神经网络的智能化地震数据插值技术[J]. 地球物理学进展, 35(2): 721-727, doi: 10.6038/pg2020DD0013.