

<https://doi.org/10.3799/dqkx.2023.186>



融合处理速度和加速度记录的地震检测模型及其在新丰江水库的应用

蒋策^{1,2}, 吕作勇², 房立华^{1,3*}

1. 中国地震局地球物理研究所, 北京 100081

2. 广东省地震局, 广东广州 510070

3. 中国地震局震源物理重点实验室, 北京 100081

摘要: 随着国家地震烈度速报与预警工程的建设, 加速度记录在地震科学中将得到越来越多的应用. 但目前的地震检测模型多使用速度记录训练, 对加速度记录的检测效果较差. 利用广东地震台网数据, 训练得到了可检测速度记录的 PhaseNet_GD 模型和检测加速度记录的 PhaseNet_ITS 模型. 在此基础上, 结合 GaMMA 震相关联和 HYPOSAT 地震定位方法, 发展了一套新的地震数据智能处理流程, 并处理了 2023 年新丰江水库 M_L 4.8 地震序列, 检测出的事件数量是人工目录的 3.8 倍, 匹配率为 93.2%, 误检测率为 0.38%. 这一系统可快速产出完备性高、高精度的地震目录, 为水库地震监测和区域地震台网的数据实时处理提供技术支撑.

关键词: 区域台网; 深度学习; 地震检测; PhaseNet; 新丰江水库; 水库地震.

中图分类号: P315.61

文章编号: 1000-2383(2024)02-469-11

收稿日期: 2023-01-22

Earthquake Detection Model Trained on Velocity and Acceleration Records and Its Application in Xinfengjiang Reservoir

Jiang Ce^{1,2}, Lü Zuoyong², Fang Lihua^{1,3*}

1. Institute of Geophysics, China Earthquake Administration, Beijing 100081, China

2. Guangdong Earthquake Agency, Guangzhou 510070, China

3. Key Laboratory of Earthquake Source Physics, CEA 100081, China

Abstract: With the construction of the "National Seismic Intensity Rapid Reporting and Early Warning" project, acceleration records data will be increasingly applied in earthquake science research. However, most current earthquake detection models use velocity records for training, which results in poor detection performance for acceleration records. This study utilized seismic records from the Guangdong Earthquake Network to train the PhaseNet_GD model for detecting velocity records and the PhaseNet_ITS model for detecting acceleration records. Based on this, a new intelligent earthquake data processing system was developed by combining the GaMMA, phase association method, and the HYPOSAT, earthquake location method. This system was used to process the 2023 M_L 4.8 earthquake sequence in Xinfengjiang Reservoir, Heyuan, and detected events 3.8 times more

基金项目: 国家重点研发专项 (No. 2021YFC3000702); 国家自然科学基金项目 (No. U2139205); 广东省地震局青年地震科研基金 (No. GDDZY202301).

作者简介: 蒋策 (1991—), 男, 工程师, 主要从事地震观测技术与方法研究. ORCID: 0000-0003-1521-4549. E-mail: cehasone@outlook.com

通讯作者: 房立华, ORCID: 0000-0003-2156-4406. E-mail: flh@cea-igp.ac.cn.

引用格式: 蒋策, 吕作勇, 房立华, 2024. 融合处理速度和加速度记录的地震检测模型及其在新丰江水库的应用. 地球科学, 49(2): 469-479.

Citation: Jiang Ce, Lü Zuoyong, Fang Lihua, 2024. Earthquake Detection Model Trained on Velocity and Acceleration Records and Its Application in Xinfengjiang Reservoir. *Earth Science*, 49(2): 469-479.

than the manual catalog, with a matching rate of 93.2% and a false detection rate of 0.38%. This system can provide technical support for reservoir seismic monitoring and real-time data processing of regional earthquake networks.

Key words: regional seismic network; deep learning; seismic detection; PhaseNet; Xinfengjiang Reservoir; reservoir induced earthquake.

0 引言

我国共有 31 个省级地震台网,其主要功能是监测本省的地震活动.省级地震台网监测范围广、观测周期长,其台站分布在全省各地,监测范围一般大于流动台网/台阵,且需要连续不断地监测全省的地震活动.随着国家地震烈度速报与预警工程(简称“预警工程”)的建设,省级地震台网的台站数量快速增加,检测到的地震活动,尤其是微小地震数量也成倍增长.依靠传统的人工分析方式已难以满足地震行业发展需求,亟需发展一套适用于区域地震台网的地震数据智能处理系统.

目前,地震台网多数采用 STA/LTA 算法(Allen, 1978)检测地震,该算法通过振幅变化检测地震,在大震活动中表现优异,但在微震检测中效果较差,抗干扰能力和后续震相拾取效果较差,难以用于微震处理中.近年来,以模板匹配和深度学习为主的算法在地震科学研究中得到了广泛应用(王德涛和陈国雄, 2022, 2023).模板匹配算法检测出的地震目录完备性高,抗干扰性强,但计算效率低,且需要事先准备模板事件(Peng and Zhao, 2009; Kato *et al.*, 2012; Meng *et al.*, 2018).最近,基于人工智能的震相拾取算法得到了快速发展和应用,包括利用卷积神经网络识别震相的 GPD 方法(Ross *et al.*, 2018),结合多尺度多层次信息的震相拾取算法 PickNet(Wang *et al.*, 2019),结合卷积神经网络与循环神经网络的 DetNet&PpkNet 方法(Zhou *et al.*, 2019),基于 Unet 结构改进的 PhaseNet 方法(Zhu and Beroza, 2019),轻量化震相拾取网络 LPPN(Yu and Wang, 2022),基于 Unet++ 改进的 CSESnet 方法(Li *et al.*, 2023),以及结合卷积神经网络、循环神经网络和自注意力机制的 EQTransformer 方法(Mousavi *et al.*, 2020, Xiao *et al.*, 2021).上述人工智能算法在测试集中均取得了不错的效果,但在实际应用时,尤其是用于连续波形数据处理时,其拾取精度和匹配率等指标均出现不同程度的下降(Jiang *et al.*, 2021; Lapins *et al.*, 2021).

现阶段的深度学习模型(简称“AI 模型”)一般

都存在泛化性的问题,主要原因有两个:一是,不同地区的地壳速度结构和台站噪声干扰存在差异,会导致地震波形出现较强的区域差异,已有的深度学习模型不一定能准确检测出新地区发生的地震;二是,区域台网每年的地震活动分布会有变化,如果在地震活动比较弱的地方突然发生中强地震,如汶川地震、芦山地震、马尔康地震、九寨沟地震、鲁甸地震等,对于发生在新区域的地震,其波形可能与已有波形差别较大,此时无论使用模板匹配还是 AI 模型算法,都可能出现地震漏检测的情况.此外,由于各个省份的地震和台站数量不同,采用全国或更大空间范围的数据训练模型时, AI 模型可能重点学习训练数据占比较大区域的波形特征,对于地震较少的省份未必能取得好的检测效果.

地震检测是地震数据前端处理中的一个重要环节.在数据量足够的情况下,采用本省台网积累的地震和噪声数据构建训练集,并不断增加新处理的地震数据进行迭代训练是提升地震检测效果、降低误检测的一个重要途径.此外,目前的地震检测模型多数采用宽频带速度型地震记录训练,公开的数据集也以宽频带速度型地震记录为主.通过预警工程建设,我国已建成约 2 000 个配置宽频带地震仪和加速度仪的基准站、3 000 个配置加速度仪的基准站和 10 000 个配置简易烈度计的台站,其中加速度计占比超过 85%.这些 AI 模型是否能直接用于检测预警工程的加速度波形数据?其检测效果如何?目前还缺少相关的研究和测试.

本文以广东省地震台网为例,选取以 Unet 骨干模型发展起来的 PhaseNet 进行模型训练(Zhu and Beroza, 2019).训练数据集包含广东省地震台网 2008—2021 年观测到的 340 804 条宽频带地震事件波形和 2022—2023 年观测到的 694 条加速度事件波形,分别重新训练了两个震相拾取模型 PhaseNet_GD 和 PhaseNet_ITS.根据仪器类型选定对应模型完成震相拾取,采用 GaMMA(Zhu *et al.*, 2021)算法和 HYPOSAT(Schweitzer, 2001)算法分别完成震相关联和地震定位,产出地震目录,形成新的地震数据自动处理流程.以广东河源市东源县 M_L 4.8 级地震序列为例,探讨了利用深度学习方法,

在区域地震台网同时实现宽频带速度型地震波形和加速度地震波形的自动处理. 本文对测试结果进行统计分析和评估, 为完善相关算法, 发展实时处理系统提供了有益的参考.

1 数据

1.1 广东地震台网简介

广东地处东南沿海地震带, 是华南地区地震相对活跃的省份. 有历史记录以来, 广东及其近海共发生5级以上地震47次, 其中6.0~6.9级地震14次, 7.0级以上地震2次, 最大的为1918年汕头南澳7.3级地震(潘华等, 2009).

广东省地震台网目前有地震台站70个(图1a), 其中宽频带台站有66个, 短周期台站有4个. 地震计主要为TDV-60B、BBVS-60和BBVS-120三类, 采样率统一为100 Hz. 为了对邻省地震进行监测, 广东省地震台网还实时接入了周边福建省12个、江西省8个、广西省11个、海南省5个和湖南省4个台站, 日常人工地震编目使用的台站总数为110个(含邻省40个台站). 广东省地震台站分布并不均匀, 河源新丰江水库地区最为密集(图1b), 平均台间距为9.6 km; 阳江地区平均台间距为77.2 km, 汕头南澳地区平均台间距为82.4 km, 全省平均台间距为49.8 km. 随着预警工程项目建设的完成, 广东省现已建成1 174个烈度计台站(图1a), 其仪器类型为101MEMS和JS-A2.

1.2 宽频带速度波形训练集

本文采用的训练集来自广东省地震台网2008

—2021年的宽频带地震事件波形(图1a). 训练集共包含48 534个地震事件, 340 804条地震波形, 其中最小震级为 M_L 0.0, 最大震级为 M_L 5.2. 该时段内广东地震主要分布在河源新丰江地区、阳江电白地区和汕头南澳地区. 其中, 河源新丰江地区的事件占比52.3%, 阳江电白地区占比16.07%, 汕头南澳地区占比18.51%, 通过该训练集训练形成针对速度波形的震相拾取模型PhaseNet_GD.

1.3 加速度波形训练集

加速度训练波形截取自广东省地震台网2022年2月至2023年4月基本站和一般站记录到的地震事件. 该时段共包含50个地震事件, 其中最大震级为 M_L 4.8, 最小震级为 M_L 1.9, 从中选取694条Pg与Sg成对拾取波形作为加速度训练波形. 由于可供训练的加速度波形远少于速度波形, 单靠这些加速度波形不满足模型收敛的需求, 因而采取速度波形与加速度波形混合训练的方式, 构建两套训练集, 一套训练集由694条加速度事件波形和2 000条速度事件波形组合, 另一套训练集由694条加速度波形和34 776条速度事件波形组合. 采用前者训练的加速度波形震相拾取模型称为PhaseNet_ITS1, 后者训练的模型称为PhaseNet_ITS2.

1.4 测试集

本研究构建了两套测试集. 第一套测试集将测试重训练对模型的震相拾取性能和对地震数据处理系统的影响. 采用广东省地震台网2023年3月7日至3月14日共计一周的三分量宽频带连续波形, 该时段包含广东河源市东源县 M_L 4.8级和地震序

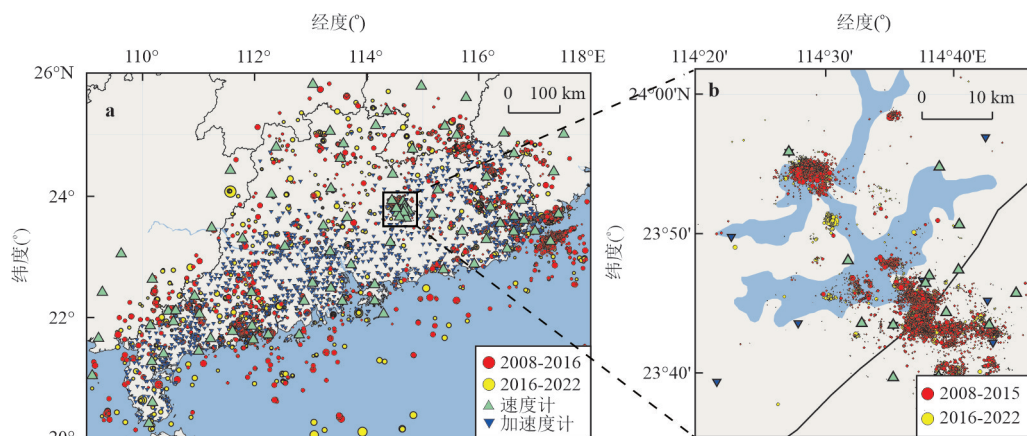


图1 广东及新丰江水库地震分布图

Fig.1 Earthquake distribution in Guangdong and Xinfengjiang Reservoir

a. 广东及周边地区台站与地震分布图(2008—2021年, $M_L \geq 0$); b. 新丰江水库周边的地震分布(2008—2021年); 图中红点为2008—2015年地震, 黄点为2016—2021年地震, 圆点大小与地震大小正相关, 绿色正三角为速度计, 蓝色倒三角为加速度计

列,共 207 次地震.第二套测试集将测试 PhaseNet 原始模型、宽频带速度波形重训练的模型 PhaseNet_GD、加速度波形与少量速度波形混合训练的模型 PhaseNet_ITS1、加速度波形与大量速度波形混合训练的模型 PhaseNet_ITS2 在加速度波形中的震相拾取效果.测试集取自 2022 年 2 月至 2023 年 4 月广东省地震台网分析处理的加速度事件波形,共包含 2 986 条震相到时,其中 Pg 震相有 962 条,Sg 震相有 2 024 条.通过上述两个测试集探索基于深度学习的地震数据处理方法,在区域地震台网完成宽频带速度和加速度波形的自动处理,实现省级地震台网的地震自动编目.

2 地震检测模型训练和测试

2.1 模型训练

本研究采用由 Unet 架构改进的 PhaseNet 算法作为震相拾取算法,该模型的震相拾取性能优异,在国内外很多地区得到了较好应用(廖诗荣等, 2021).为满足广东省地震台网业务工作要求,需对整个模型重新训练,获得广东省地震数据训练的震相拾取模型.

已有学者采用 Pytorch 框架将 PhaseNet 与其他深度学习震相拾取网络整合至同一平台(Woollam *et al.*, 2022).但是为了保证拾取结果能够与 Zhu 等(Zhu and Beroza, 2019)在美国北加州的结果进行比较,本研究采用前人公布在 GITHUB 上的源代码为基础,进行模型训练.

本研究在训练前进行了地震波形和到时数据筛选.地震波形随震中距会出现较大变化,超过一定距离,Pn 震相与 Sn 震相会出现.若把 Pn 与 Pg 设为标签 P,Sn 与 Sg 设为标签 S,会对模型提取特征产生较大的干扰,因而本研究中仅采用 Pg 与 Sg 作为训练数据.此外,震中距较大时,Pg 出现在 Pn 之后,易被 Pn 的波形干扰,致使人工标注的 Pg 到时精度不足,若采用这些震相到时可能会干扰模型的拾取性能.因而本研究根据广东地区莫霍面的大致深度,选取震中距 120 km 作为筛选距离,剔除震中距大于 120 km 的地震波形.

数据筛选后,首先对波形数据去均值、去倾斜和归一化处理;其次,在人工拾取的震相到时附近附近设置到时标签,形状为高斯窗,宽度为 0.2 s;再次,对波形窗口加入随机滑动,滑动最大值使得 S 波震相在训练波形的开头,滑动最小值使得 P 波震相在训练波

形的结尾,从而使模型在检测窗口地震波形不完整时同样具备震相拾取能力;最后,将波形长度截取为 30 s,完成 30 轮次的训练后产出新模型.

2.2 宽频带数据测试

本文的基准地震目录为人工地震编目结果,所有震相均由分析师进行处理,并由两名分析师校核,因而震相到时拾取精准.本文震相匹配条件为自动目录与人工目录的 Pg 到时差在 ± 0.2 s 范围内,Sg 到时差在 ± 0.5 s 范围内.地震事件匹配条件为震中位置偏差在 15 km 范围内,发震时刻小于 5 s,且两个事件至少有 3 条震相能互相匹配.震相与地震事件采用的评测指标为匹配率(Recall)与精确率(Precision).精确率是预测结果正确的数据集与预测为真的数据集的比值,匹配率是预测结果正确的数据集与标签为真的数据集的比值.其计算公式如下:

$$\text{Precision} = \frac{TP}{TP + FP},$$

$$\text{Recall} = \frac{TP}{TP + FN},$$

其中:TP 为预测为真,实际为真的数据集,即预测正确的数据集;FP 为预测为真,实际为假的数据集,即预测错误的数据集;FN 为预测为假,实际为真的数据集,即漏报部分;TN 为预测为假,实际也为假的数据集(Zhu and Beroza, 2019).

2023 年 3 月 7 日至 3 月 13 日,人工目录共包含震相 5 489 条,其中 Pg 震相 2 663 条,Sg 震相 2 826 条.采用原 PhaseNet 模型进行震相检测(图 2a),获得 Pg 到时 23 198 条,Sg 到时 21 432 条.Pg 震相平均到时差异为 (0.019 ± 0.042) s,Pg 震相拾取偏差小于 ± 0.1 s 占 91.8%,小于 ± 0.2 s 占 96.5%.Sg 震相平均到时差异为 (0.061 ± 0.084) s,Sg 震相拾取偏差小于 ± 0.1 s 占 82.3%,Sg 震相拾取偏差小于 ± 0.2 s 占 94.2%.采用重训练后的 PhaseNet_GD 模型进行震相检测(图 2b),获得 Pg 到时 42 733 条,Sg 到时 20 495 条.Pg 震相平均到时差异为 (0.012 ± 0.037) s,Pg 震相拾取偏差小于 ± 0.1 s 占 95.6%,小于 ± 0.2 s 占 98.4%.Sg 震相平均到时差异为 (0.036 ± 0.085) s,Sg 震相拾取偏差小于 ± 0.1 s 占 87.1%,Sg 震相拾取偏差小于 ± 0.2 s 占 95.0%.

为与原 PhaseNet 模型相比较,本研究采用与原文相同的标准,即人工与自动拾取震相到时偏差小于 0.1 s 记为 TP,并将所有波形截取为地震事件波形(表 1).相比北加州,PhaseNet 模型在广东省的震相拾取指标出现了一定程度的下降,Pg 匹配率下降

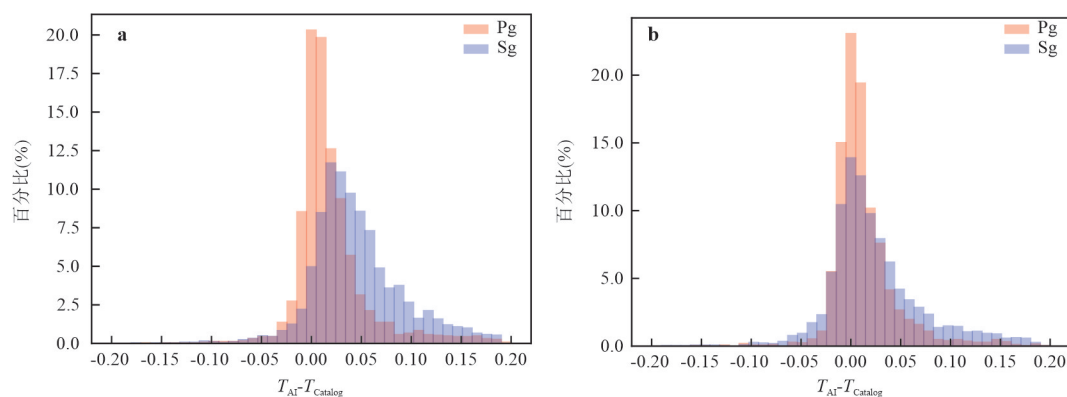


图 2 宽频带波形震相拾取偏差

Fig.2 Picking difference of broad-band waveform seismic phases

a. PhaseNet 模型;b. PhaseNet_GD 模型

表 1 宽频带速度计震相拾取结果统计

Table 1 Statistics of phase pickup results of broadband velocimetry

场景	Pg			Sg		
	Recall(%)	Precision(%)	F1	Recall(%)	Precision(%)	F1
PhaseNet_GD 在广东	87.5	83.5	0.855	80.8	79.5	0.801
PhaseNet 在广东	76.2	72.7	0.744	68.3	67.2	0.677
PhaseNet 在北加州	85.7	93.9	0.896	75.5	85.3	0.801

注:残差小于 0.1 s 记为 TP, Precision 用事件波形作为基础.

了 9.5%, 精确率下降了 21.2%, Sg 匹配率下降了 7.2%, 精确率下降了 18.1%. 然而重新训练后的 PhaseNet_GD 模型在广东省的震相拾取指标有了较大提升, Pg 匹配率提升了 11.3%, 精确率提升了 10.8%, Sg 匹配率提升了 12.5%, 精确率提升了 12.3%. PhaseNet_GD 模型在广东地区表现与 PhaseNet 在北加州的表现在同一水准.

2.3 加速度数据测试

本研究训练了两套加速度震相拾取模型 PhaseNet_ITS1 和 PhaseNet_ITS2, 在 2022 年 2 月至 2023 年 4 月期间的 2 986 条加速度波形中拾取的震相到时进行测试, 其中 Pg 震相有 962 条, Sg 震相有 2 024 条. 为进一步分析对比, 本研究还采用广东省宽频带数据重训练的 PhaseNet_GD 模型和 PhaseNet 原始模型处理加速度测试集.

统计表明(表 2), 在加速度波形测试集中, PhaseNet_ITS1 的匹配率和精确率最高, 分别为 72.0% 和 60.5%, 在 4 个模型中表现最好. PhaseNet_ITS2 匹配率 64.8%, 精确率 59.0%, 表现仅次于 PhaseNet_ITS1. 原 PhaseNet 匹配率 64.2%, 精确率 38.4%. PhaseNet_GD 匹配率 57.4%, 精确率

58.0%.

图 3 展示 PhaseNet_GD、Phase_ITS1、PhaseNet_ITS2 和 PhaseNet 模型拾取震相与人工拾取震相到时到时的统计情况. PhaseNet_GD 模型的 Pg 震相平均到时差异 (0.046 ± 0.080) s, Pg 震相拾取偏差小于 ± 0.1 s 占 84.1%, 小于 ± 0.2 s 占 94.6%. Sg 震相平均到时差异 (0.035 ± 0.131) s, Sg 震相拾取偏差小于 ± 0.1 s 占 78.2%, Sg 震相拾取偏差小于 ± 0.2 s 占 91.6%. PhaseNet_ITS1 模型的 Pg 震相平均到时差异为 (0.002 ± 0.075) s, Pg 震相拾取偏差小于 ± 0.1 s 占 92.5%, 小于 ± 0.2 s 占 96.9%. Sg 震相平均到时差异为 (0.029 ± 0.114) s, Sg 震相拾取偏差小于 ± 0.1 s 占 83.1%, Sg 震相拾取偏差小于 ± 0.2 s 占 93.4%. PhaseNet_ITS2 模型的 Pg 震相平均到时差异为 (0.010 ± 0.084) s, Pg 震相拾取偏差小于 ± 0.1 s 占 91.5%, 小于 ± 0.2 s 占 96.9%. Sg 震相平均到时差异为 (0.028 ± 0.114) s, Sg 震相拾取偏差小于 ± 0.1 s 占 83.1%, Sg 震相拾取偏差小于 ± 0.2 s 占 93.6%. 原 PhaseNet 的 Pg 震相平均到时差异为 (0.011 ± 0.085) s, Pg 震相拾取偏差小于 ± 0.1 s 占 92.9%, 小于 ± 0.2 s 占 97.8%. Sg 震相平均到时差异为 (0.063 ± 0.109)

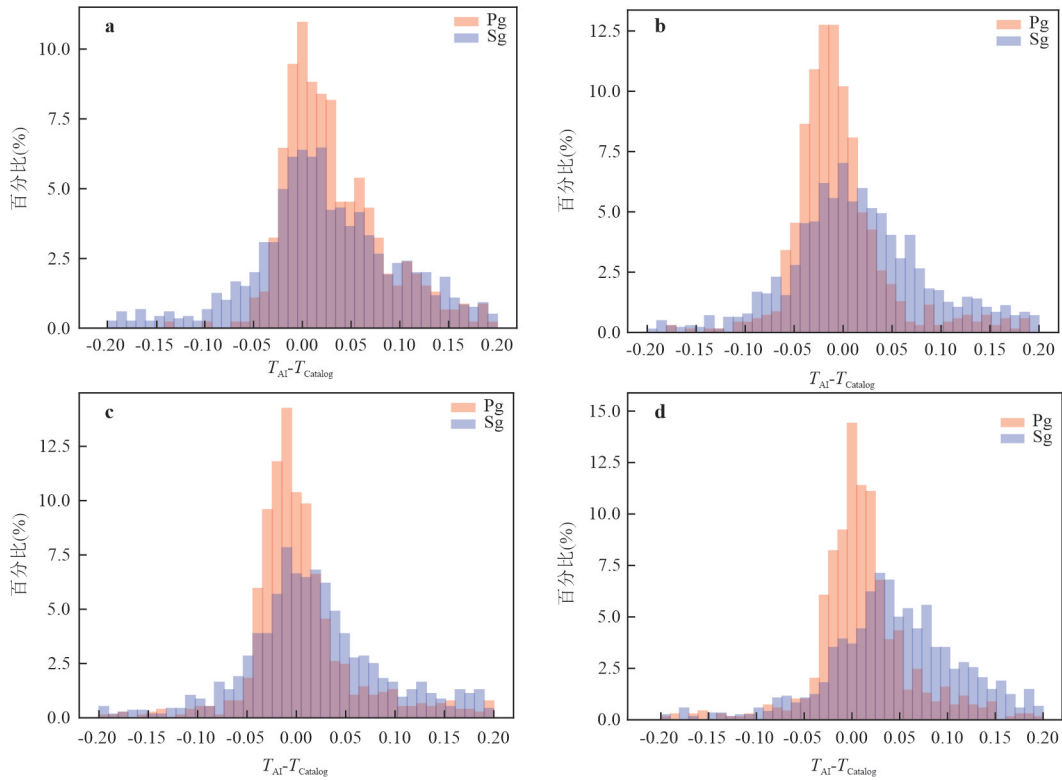


图3 加速度波形震相拾取偏差

Fig.3 Picking difference of of acceleration waveform seismic phases

a. PhaseNet_GD; b. PhaseNet_ITS1; c. PhaseNet_ITS2; d. PhaseNet

表2 加速度计震相拾取结果统计

Table 2 Statistics of accelerometer phase pickup results

训练集	Pg		Sg		Total	
	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)
PhaseNet_GD	48.4	56.3	59.8	56.8	56.1	56.7
PhaseNet_ITS1	73.5	60.4	71.2	60.5	72.0	60.5
PhaseNet_ITS2	80.2	58.4	57.4	59.5	64.8	59.0
PhaseNet	72.1	46.1	60.4	34.8	64.2	38.2

注:残差小于0.5 s记为TP.

s, Sg震相拾取偏差小于 ± 0.1 s占74.8%, Sg震相拾取偏差小于 ± 0.2 s占91.9%.

3 地震数据自动化处理系统

本研究构建的地震数据自动化处理系统主要包含波形预处理、AI震相拾取、震相关联、地震定位、震级测定5个处理模块(图3).

数据预处理模块:该模块读入宽频带数据与加速度数据,对数据进行去均值(demean)、去倾斜(detrend)和归一化处理(normalization),将波形数据写

入震相拾取模块的输入文件.

AI震相拾取模块:采用PhaseNet_GD模型在速度波形数据上进行震相拾取.采用四套训练方案中性能最好的PhaseNet_ITS1模型作为系统中正式采用的PhaseNet_ITS模型,在加速度波形上数据进行震相拾取.震相拾取窗长为30 s,重叠15 s,滑动时窗15 s.

关联模块:该模块获得自动拾取震相到时数据后,采用GaMMA方法(Zhu *et al.* 2021)完成震相关联,该算法通过dbscan算法(Ester *et al.*, 1996)完成数据时间段的划分,采用贝叶斯高斯混合模型(Bayesian Gaussian mixture model)完成震相到时的关联和震中位置的初步运算,关联要求P与S震相数之和不少于6个.

定位模块:该模块采用HYPOSAT定位程序(Schweitzer, 2001)进行绝对定位,该方法除了使用Pg和Sg震相到时,还可以使用Sg-Pg到时差,可进一步排除震相类型和到时的误检测.定位时采用的速度模型为华南模型(范玉兰等, 1990).

震级测定模块:该模块在获得震相到时和地震定位结果后,根据国内 M_L 震级计算标准来计算震

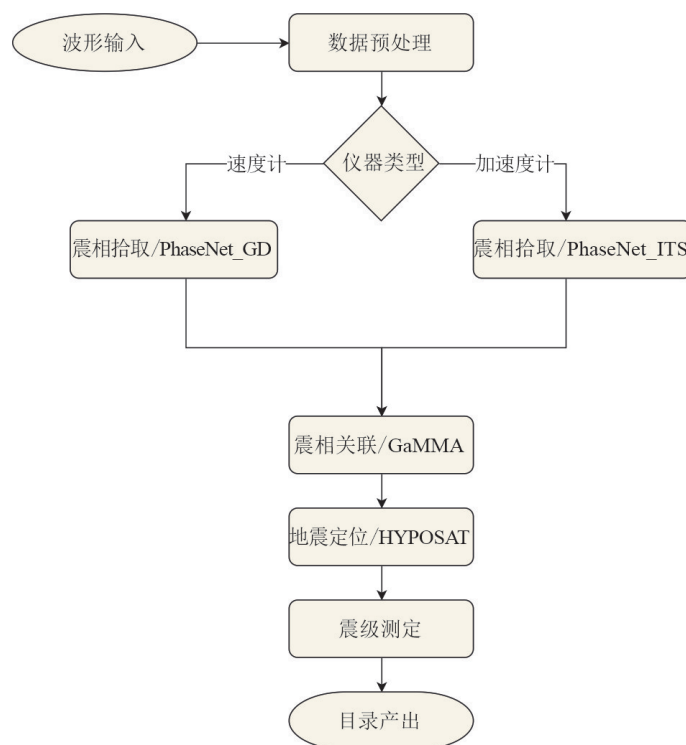


图 4 地震数据智能处理流程图

Fig.4 Flowchart of intelligent seismic data processing

级. 首先将原始波形仿真为 DD-1 仪器波形, 根据 Pg 和 Sg 震相到时, 计算各台站水平分量的最大震幅; 其次, 根据国标给出的 M_L 计算公式, 计算获得各台站的 M_L 震级; 最后, 对各台站震级舍弃最大最小值后取均值来获得该事件的 M_L 震级.

4 2023 年河源 M_L 4.8 地震序列检测

本文采用广东省地震台网 2023 年 3 月 7 日至 3 月 14 日河源 M_L 4.8 地震序列测试地震数据自动化处理系统的实际效果. 先前已用原 PhaseNet 和 PhaseNet_GD 在完成该时段内震相到时的拾取, 接着采用 PhaseNet_ITS 拾取加速度波形的震相, 随后通过关联模块、定位模块与震级计算模块完成后续地震处理, 产出基于原 PhaseNet 和 PhaseNet_GD 模型的两套地震目录.

统计表明, 原 PhaseNet 模型对应的自动目录中包含的震相总数为 7 370, 基于 PhaseNet_GD 模型的自动目录包含的震相总数为 11 191, 两套目录包含的震相数量均少于关联前的数量, 大量误拾取的干扰和噪声被剔除. 原 PhaseNet 的自动目录中, 已匹配事件的 Pg 震相的匹配率为 90.4%, 精确率为 48.9%; Sg 震相的匹配率为 90.3%, 精确率为

49.9%; PhaseNet_GD 的自动目录中, 已匹配事件的 Pg 震相匹配率为 95.9%, 精确率为 33.3%; Sg 震相匹配率为 94.3%, 精确率为 36.7%.

本文对比了加速度拾取模型 PhaseNet_ITS 在 3 月 8 日 M_L 4.8 级主震事件中的表现. 该事件共包含加速度震相 329 条, 其中 Pg141 条, Sg188 条. PhaseNet_ITS 拾取到 Pg135 条, 匹配率 73.7%, 精确率 77.0%, Sg187 条, 匹配率 77.66%, 精确率 78.7%.

基于 PhaseNet 模型的自动目录共包含地震 528 条, 基于 PhaseNet_GD 模型的自动目录共包含地震 788 条, 两套自动目录均包含比人工目录更多的地震事件. 经人工校验, 确定原 PhaseNet 模型对应的目录中共有 467 条为天然事件, 58 条为非天然事件, 3 条为误触发, 误触发率为 0.57%; PhaseNet_GD 对应目录中共有 743 条为天然事件, 42 条为非天然事件, 3 条为误触发, 误触发率为 0.38%. 原 PhaseNet 自动目录的匹配率为 90.8%, PhaseNet_GD 自动目录的匹配率为 93.2%. PhaseNet_GD 的自动目录相比原 PhaseNet 的自动目录匹配率更高, 检测出了更多人工遗漏事件, 误触发率更低.

虽然 PhaseNet_GD 产出了较为完备的自动目录, 但目录质量仍需进一步分析, 图 5 统计了

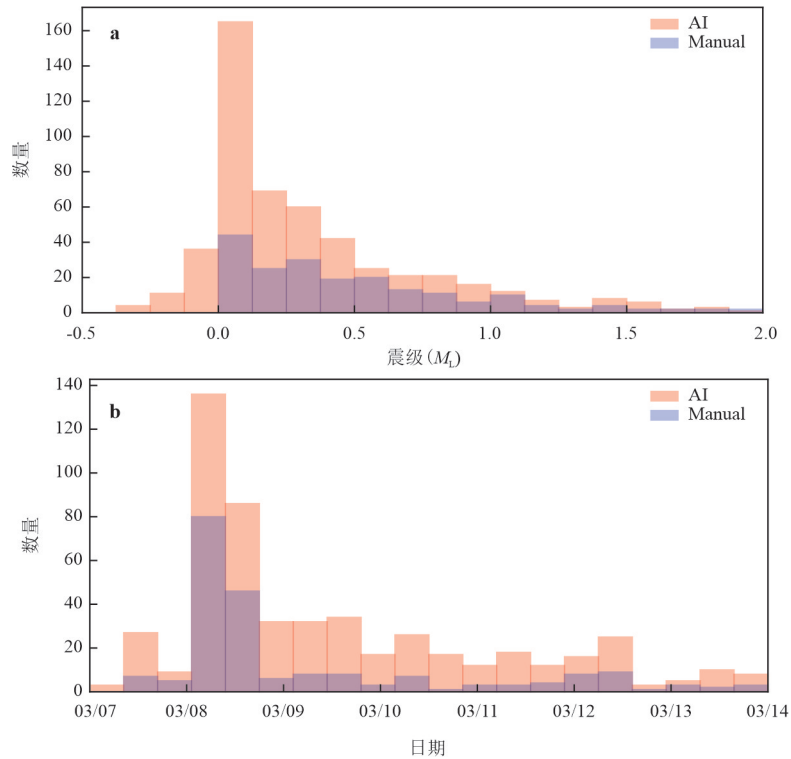


图5 自动目录与人工目录事件数量对比

Fig.5 Comparison of the number of automatic and manual events

a. 不同震级段的地震数量对比; b. 每天地震数量对比

PhaseNet_GD 目录事件数量随时间和随震级变化关系. 自动处理系统每天检测出的地震事件数量明显多于人工分析结果, 产出的事件数量是人工分析事件的 3~4 倍, 且事件数量与震级变化总体趋势与人工分析结果一致. 对于 M_L 小于 1.0 地震自动处理系统检测优势更为明显, 使得自动目录的完整性震级下限可达 M_L 0.0 级左右 (图 5a). 图 6 对比了自动目录与人工目录在河源主震区的地震分布, 自动目录的地震分布在新回龙镇和河源市区附近, 与人工分布基本一致. 经核验, 发现与人工目录偏差较大的事件主要有两类: 一类是震级较小的非天然地震事件, 占比约 5.33%; 另一类是到时误拾取和震相误关联的天然地震事件, 占比约 0.38%.

图 7 展示了基于重训练模型 PhaseNet_GD 处理的自动目录与人工目录差异统计, 发震时刻偏差 (0.060 ± 0.133) s, 偏差小于 ± 1.0 s 的事件占 99.5%; 震中位置偏差为 (0.426 ± 0.633) km, 偏差小于 5 km 的事件占 99.0%; 震源深度偏差 0.236 ± 1.553 km, 偏差小于 10 km 的事件占 98.4%; 震级偏差为 0.058 ± 0.343 , 偏差小于 ± 0.3 级的事件占 82.2%.

5 讨论

本研究对比了 PhaseNet 原始模型和利用广东地震台网数据重新训练的 PhaseNet_GD 模型在宽频带速度波形数据中的检测效果. 从表 1 可以看出, PhaseNet 在美国北加州地区表现优异, 但在广东地区匹配率和准确率均有了一定程度的下降. 主要原因在于原 PhaseNet 模型缺乏对广地区地震波形区域性特征的识别能力. 采用广东本地数据重训练的 PhaseNet_GD 大幅提升了模型拾取的匹配率, 增加了拾取的震相数量, 与原 PhaseNet 在北加州地区表现基本一致. 此外, PhaseNet_GD 拾取的震相到时与人工偏差更小, 离散程度更低.

除测试了 PhaseNet_GD 在宽频带连续波形测试集中的拾取效果外, 本研究还测试了原 PhaseNet、PhaseNet_GD、PhaseNet_ITS1、PhaseNet_ITS2 在加速度测试集中的拾取效果. 由表 1 可以看出, 单纯依靠宽频带速度波形训练的 PhaseNet_GD 和原 PhaseNet 在加速度计波形中的表现效果较差, 这是由于宽频带速度波形与加速度波形差异较大所导致. 在加速度波形的基础上少量添加宽频带波形, 训练出的模型 PhaseNet_ITS1 效果最好. 因此, 在原训练集不足以使

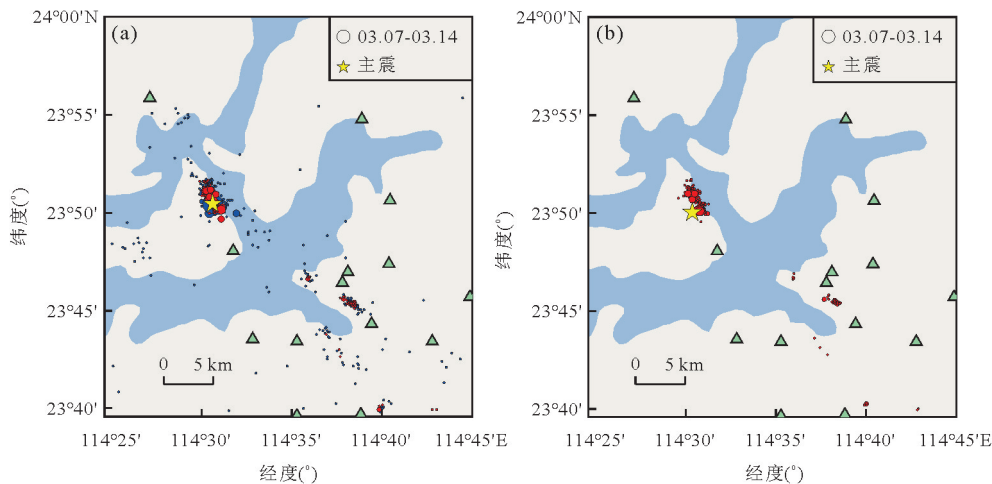


图 6 PhaseNet_GD 自动目录和人工目录震中位置对比

Fig.6 Comparison of the epicenter locations between the PhaseNet_GD automatic catalog and the manual catalog

a. 基于 PhaseNet_GD 的自动目录; b. 人工目录; 红色圆点为匹配事件, 蓝色圆点为自动系统多检测出的地震事件

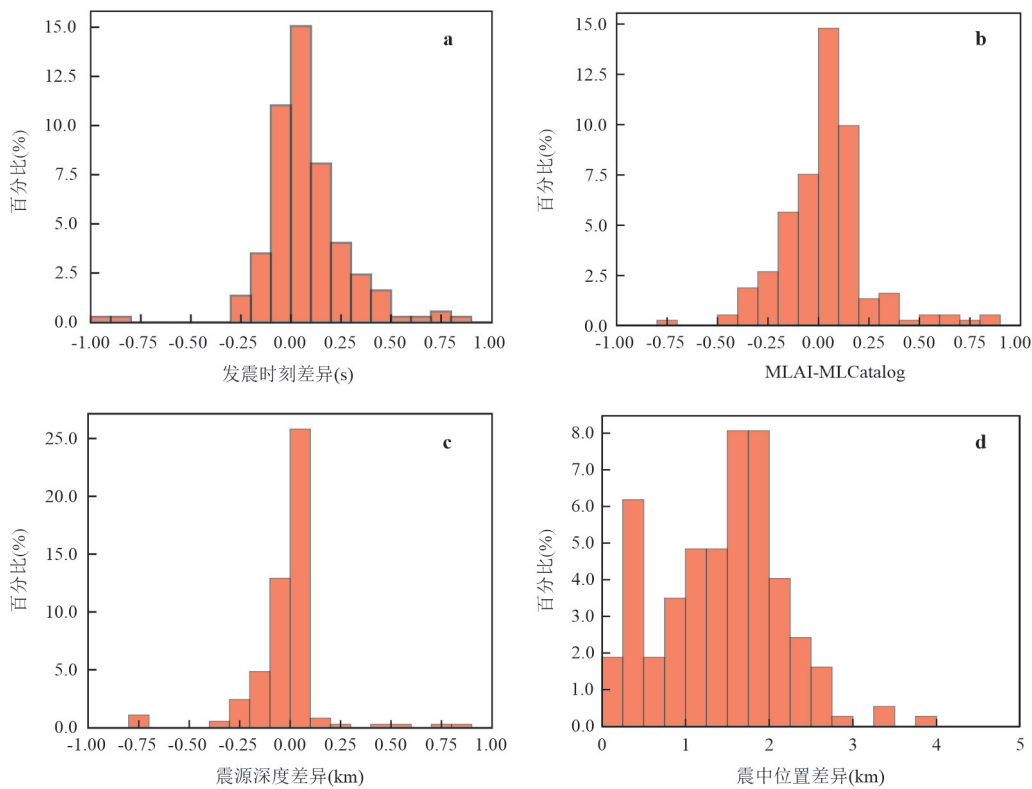


图 7 自动目录与人工目录差异统计图

Fig.7 Statistical analysis of differences between automatic and manual catalogs

a、b、c、d 分别为发震时刻、震级、震中位置、震源深度差异

得损失函数收敛的情况下,可适当增加其他地区或其他特征近似的训练集辅助训练,但添加的辅助训练集数量若远大于原训练集,会导致网络模型学习大量辅助训练集的特征,从而使得模型在目标区域的性能下降.本研究采取的地震数据处理系统将采用 PhaseNet_ITS1 作为加速度波形的拾取模型,随加速

度事件波形的不断积累,后续将迭代式训练该模型,进一步提升模型性能.

获得震相数据后,本研究采用能满足大尺度台网计算效率需求的 GaMMA 算法进行震相关联.关联后,基于原 PhaseNet 模型的自动目录共包含地震 528 条,基于 PhaseNet_GD 的自动目录共包含地震 788 条,两

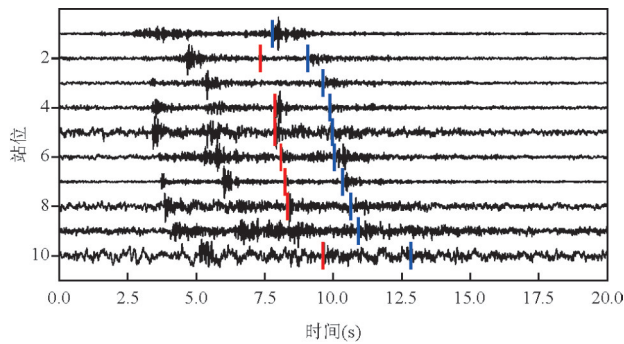


图8 漏检地震事件波形示例

Fig.8 Example waveforms of a missed event

红线为遗漏的Pg震相,蓝线为遗漏的Sg震相,纵轴为台站编号,台站波形按震中距由小到大的顺序排列

套目录都比人工目录多检测出大量的事件.经人工校验,确定两套模型对应的目录中多拾取的事件基本为地震事件,PhaseNet_GD自动目录相比原PhaseNet自动目录检测的天然地震事件数更多. PhaseNet_GD自动目录误触发概率也比原PhaseNet自动目录更低,经分析,确定误触发事件均由震相拾取模型将Pn震相识别为Pg,Pg震相识别为Sg所导致,后续工作中增加模型对Pn和Sn等远台震相的拾取能力,并考虑不同震相的走时规律,这可有效降低误触发.此外,根据已匹配事件的震相匹配率可知,对于同一地震事件而言,PhaseNet_GD的自动目录具有更完备的震相到时.由此可以得出以下结论,基于本地数据重训练模型产生的自动目录相比基于原始模型产生的目录更为完备,包含的震相到时更为齐全,震相到时误差更小.图7表明,本研究采用的地震数据自动处理系统产生的目录发震时刻、震中位置、震源深度、震级均与人工目录偏差非常小,与人工目录质量接近.

虽然自动目录的数量远多于人工目录,但是自动目录仍然遗漏了14个地震,漏检率为6.8%.漏检事件的最大震级为 $M_L 1.0$,最小震级为 $M_L 0.1$.为分析漏检原因,本研究对每一个漏检事件的波形进行分析,发现有4个地震在人工目录中触发台站和震相数量也不满足关联要求,其余10个地震均存在发震时刻接近且波形互相叠加的邻震(图8),关联算法误将两个地震合并一个事件进行关联,导致地震遗漏.上述事件的遗漏均由关联算法性能所导致,而非事件的震相未被拾取.

6 结论

本研究采用在加速度波形训练集中添加宽频带

速度波形样本的方式开展辅助训练,获得了可处理加速度波形的震相拾取模型PhaseNet_ITS,并为地震数量较少的区域台网提供本地化AI模型训练方案.利用本地数据重新训练的速度波形检测模型PhaseNet_GD和加速度波形检测模型PhaseNet_ITS在震相匹配率和精确率等方面较原始PhaseNet均有较大提升.

本文构建了一套适用于区域地震台网的数据自动处理流程,该流程包含数据预处理、震相拾取、震相关联、地震定位和 M_L 震级测定等关键模块.该流程使用广东本地数据重训练后的PhaseNet_GD和PhaseNet_ITS模型进行震相拾取,使用GaMMA算法进行震相关联,采用可考虑S-P到时差的HYPOSAT方法进行定位,可在震相关联的基础上进一步排除震相误检测,提高定位精度.

采用2023年河源 $M_L 4.8$ 级地震序列,对本文新发展的地震数据处理流程进行评估,表明基于重训练模型产生的自动目录相比基于原始模型产生的目录更为完备,包含的震相到时更为齐全,拾取误差更小.此外,基于重训练模型PhaseNet_GD产生的自动目录在发震时刻、震源位置、震级等方面与人工目录差异较小,且目录完备性高于人工目录.

本研究虽然使用了加速度波形数据进行训练,但目前搜集到的样本较少,后续将继续搜集四川、云南等多震省份的加速度波形数据,扩大训练样本集,进一步提升模型对加速度波形数据的处理效果.本文的数据处理流程可以快速产出高完备性和高可靠性的地震目录,但是仍存在少量的误检测和漏检测,后续将改进关联算法,提升关联算法的抗干扰能力和对密集地震序列的处理能力.

致谢:本研究使用的数据由广东省地震局提供.感谢中国地质大学(武汉)王墩教授和中国海洋大学邹志辉教授对本文修改过程中提出的意见和建议.

References

- Allen, R. V., 1978. Automatic Earthquake Recognition and Timing from Single Traces. *Bulletin of the Seismological Society of America*, 68(5): 1521–1532. <https://doi.org/10.1785/bssa0680051521>
- Ester, M., Kriegel, H.P., Sander, J., et al., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press, Portland, Oregon, 226–231.
- Fan, Y., Lin, J., Hu, R., et al., 1990. The Development of

- Traveltime table for near Earthquake in South China. *South China Seismological Journal*, 10(2): 1—16(in Chinese).
- Hu, J., Ding, Y., Zhang, H., Jin, C., et al., 2023. A Real-Time Seismic Intensity Prediction Model Based on Long Short-Term Memory Neural Network. *Earth Science*, 48(5): 1853—1864(in Chinese with English abstract).
- Jiang, C., Fang, L. H., Fan, L. P., et al., 2021. Comparison of the Earthquake Detection Abilities of PhaseNet and EQTransformer with the Yangbi and Maduo Earthquakes. *Earthquake Science*, 34(5): 425—435. <https://doi.org/10.29382/eqs-2021-0038>
- Kato, A., Obara, K., Igarashi, T., et al., 2012. Propagation of Slow Slip Leading up to the 2011 Mw9.0 Tohoku-Oki Earthquake. *Science*, 335(6069): 705—708. <https://doi.org/10.1126/science.1215141>
- Lapins, S., Goitom, B., Kendall, J. M., et al., 2021. A Little Data Goes a Long Way: Automating Seismic Phase Arrival Picking at Nabro Volcano with Transfer Learning. *Journal of Geophysical Research: Solid Earth*, 126(7). <https://doi.org/10.1029/2021jb021910>
- Li, B. R., Fan, L. P., Jiang, C., et al., 2023. CSESnet: A Deep Learning P-Wave Detection Model Based on UNet++ Designed for China Seismic Experimental Site. *Frontiers in Earth Science*, 10. <https://doi.org/10.3389/feart.2022.1032839>
- Liao, S., Zhang, H., Fan, L., et al., 2021. Development of a Real-Time Intelligent Seismic Processing System and Its Application in the 2021 Yunnan Yangbi M_s 6.4 Earthquake. *Chinese J. Geophys.*, 64(10): 3632—3645(in Chinese with English abstract).
- Meng, X., Yang, H., Peng, Z., et al., 2018. Foreshocks, b Value Map, and Aftershock Triggering for the 2011 Mw 5.7 Virginia Earthquake. *Journal of Geophysical Research: Solid Earth*, 123(6): 5082—5098. <https://doi.org/10.1029/2017jb015136>
- Mousavi, S. M., Ellsworth, W. L., Zhu, W. Q., et al., 2020. Earthquake Transformer: an Attentive Deep-Learning Model for Simultaneous Earthquake Detection and Phase Picking. *Nature Communications*, 11(1):3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Pan, H., Yan, J., Zhang, Z., et al., 2009. Review on 1918 Nan'ao M_s 7.3 Earthquake and Its Strong Aftershocks. *Technology for Earthquake Disaster Prevention*, 4(1):40—48(in Chinese with English abstract).
- Peng, Z. G., Zhao, P., 2009. Migration of Early Aftershocks Following the 2004 Parkfield Earthquake. *Nature Geoscience*, 2(12): 877—881. <https://doi.org/10.1038/ngeo697>
- Ross, Z. E., Meier, M. A., Hauksson, E., et al., 2018. Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America*, 108(5A): 2894—2901. <https://doi.org/10.1785/0120180080>
- Schweitzer, J., 2001. HYPOSAT: an Enhanced Routine to Locate Seismic Events. *Pure and Applied Geophysics*, 158(1): 277—289. <https://doi.org/10.1007/pl00001160>
- Wang, D., Chen, G. X., 2022. Seismic Wave Impedance Inversion Based on Temporal Convolutional Network. *Earth Science*, 47(4): 1492—1506(in Chinese with English abstract).
- Wang, J., Xiao, Z., Liu, C., et al., 2019. Deep Learning for Picking Seismic Arrival Times. *Journal of Geophysical Research: Solid Earth*, 124(7): 6612—6624. <https://doi.org/10.1029/2019jb017536>
- Woollam, J., Münchmeyer, J., Tilmann, F., et al., 2022. SeisBench - A Toolbox for Machine Learning in Seismology. *Seismological Research Letters*, 93(3): 1695—1709. <https://doi.org/10.1785/0220210324>
- Xiao, Z., Wang, J., Liu, C., et al., 2021. Siamese Earthquake Transformer: A Pair-Input Deep-Learning Model for Earthquake Detection and Phase Picking on a Seismic Array. *Journal of Geophysical Research: Solid Earth*, 126(5): e2020JB021444. <https://doi.org/10.1029/2020jb021444>
- Yu, Z., Wang, W. 2022. LPPN: A Lightweight Network for Fast Phase Picking. *Seismological Research Letters*, 93(5): 2834—2846. <https://doi.org/10.1785/0220210309>
- Zhou, Y. J., Yue, H., Kong, Q. K., et al., 2019. Hybrid Event Detection and Phase-Picking Algorithm Using Convolutional and Recurrent Neural Networks. *Seismological Research Letters*, 90(3): 1079—1087. <https://doi.org/10.1785/0220180319>
- Zhu, W. Q., Beroza, G. C., 2019. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. *Geophysical Journal International*, 216(1): 261—273. <https://doi.org/10.1093/gji/ggy423>
- Zhu, W. Q., McBrearty, I. W., Mousavi, S. M., et al., 2021. Earthquake Phase Association Using a Bayesian Gaussian Mixture Model. *Journal of Geophysical Research: Solid Earth*, 127(5): 1—10. <https://doi.org/10.1029/2021jb023249>

附中文参考文献

- 范玉兰, 林纪曾, 胡瑞贺, 等, 1990. 华南地区近震走时表的研制. *华南地震*, 10(2): 1—16.
- 胡进军, 丁祎天, 张辉, 等, 2023. 基于长短期记忆神经网络的实时地震烈度预测模型. *地球科学*, 48(5): 1853—1864.
- 廖诗荣, 张红才, 范莉苹, 等, 2021. 实时智能地震处理系统研发及其在2021年云南漾濞 M_s 6.4地震中的应用. *地球物理学报*, 64: 3632—3645.
- 潘华, 鄢家全, 张志中, 等, 2009. 1918年南澳7.3级地震与强余震之参数复核. *震灾防御技术*, 4(1): 40—48.
- 王德涛, 陈国雄, 2022. 基于时间卷积网络的地震波阻抗反演. *地球科学*, 47(4): 1492—1506.