doi:10.3799/dqkx.2025.287

云南预警台网异构波形数据集构建与震相拾取模型性能评估

吕帅[1,2],房立华[3*],彭钰翔[1],曹颖[1],夏登科[4],范莉苹[4],朱杰[1],郭亚茹[1]

1. 云南省地震局,云南昆明 650224
2. 中国地质大学(北京),北京 100083
3. 中国地震局地震预测研究所,北京 100036
4. 中国地震局地球物理研究所,北京 100081

摘　要：近年来,深度学习方法在地震检测和震相拾取中得到广泛的应用.然而,现有模型主要基于高信噪比的速度型波形数据进行训练,缺乏对加速度计与烈度计数据的泛化性评估.为探究现有模型对加速度数据的处理效果及在云南地区的泛化能力,本文基于云南预警台网的最新观测数据,构建了包括速度计、加速度计和烈度计的多源异构高质量波形数据集,且所有震相到时均由人工标注.结合 PhaseNet、USTC-Pickers 等五种专业模型,以及 SeisMoLLM 和 SeisT 等四种大模型,系统评估了不同模型在云南数据集上的震相拾取性能.结果表明：本地迁移优化的 USTC-Pickers 综合性能最优,其 Pg 和 Sg 震相拾取的平均 F1 值达 0.779(到时拾取差异△t≤0.1 s),显著优于其他模型,且在检测加速度计与烈度计数据时,较好解决了震相拾取滞后问题;大模型在 Sg 拾取等复杂环境中展现出更强的泛化能力.研究还揭示了主流地震检测模型在不同波形长度、震级、震中距条件下的性能变化,强调了本地化训练与模型选取在实际应用中的重要性.研究结果为地震预警系统中的地震检测和震相识别,以及中国地震科学实验场地震观测数据的实时自动处理提供参考.

关键词：震相拾取;异构波形数据集;大模型;性能评估;地震预警.

中图分类号：P315.61　　**收稿日期：**2025-08-17

Development of a Heterogeneous Waveform Dataset and Evaluation of Phase Picking Models for the Yunnan Earthquake Early Warning Network

Lv Shuai[1,2], Fang Lihua[3*], Peng Yuxiang[1], Cao Ying[1], Xia Dengke[4],Fan Liping[4], Zhu Jie[1], Guo Yaru[1]

1. Yunnan Earthquake Agency, Kunming 650224, Yunnan, China

基金项目：地震科技星火计划项目(No.XH25033YB)和国家自然科学基金(No.42374081).

作者简介：吕　帅(1991-),男,高级工程师,主要从事地震监测和地震信息化工作.E-mail：lv_303494@163.com.

通信作者：房立华(1981-),男,博士,研究员,主要从事人工智能地震学研究.E-mail：fanglihua@ief.ac.cn.

2. School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China
3. Institute of Earthquake Forecasting, China Earthquake Administration, Beijing 100036, China
4. Institute of Geophysics, China Earthquake Administration, Beijing 100081, China

Abstract: In recent years, deep learning methods have been widely applied to seismic detection and phase picking. However, existing models are mainly trained on high signal-to-noise ratio (SNR) velocity-type waveform data, with limited evaluation of their generalization to accelerometer and intensity meter data. To investigate the performance of existing models on accelerometer data and their generalization capability in Yunnan, this study constructed a high-quality, multi-source heterogeneous waveform dataset based on the latest observations from the Yunnan Earthquake Early Warning (EEW) network, including velocity meters, accelerometers, and intensity meters, with all phase arrival times manually annotated. We systematically evaluated the phase-picking performance of nine models—five domain-specific models (e.g., PhaseNet, USTC-Pickers) and four large models (e.g., SeisMoLLM, SeisT)—on the Yunnan dataset. The locally fine-tuned USTC-Pickers achieved the best overall performance, with mean F1 scores of 0.779 for Pg and Sg phase picking ($\Delta t \leqslant 0.1$ s), significantly outperforming other models, and effectively mitigating phase-picking delays for accelerometer and intensity meter data. Large models demonstrated stronger generalization in Sg picking and low-SNR conditions. The study also revealed performance variations of mainstream seismic detection models under different waveform lengths, magnitudes, and epicentral distances, underscoring the importance of localized training and model architecture selection in practical applications. The research findings provide references for seismic detection and phase picking in earthquake early warning systems, as well as for the real-time automatic processing of seismic data at the China Earthquake Science Experiment Site.

Key words: seismic phase picking; dataset; large language model; performance evaluation; earthquake early warning.

0 引言

近年来,人工智能(Artificial Intelligence, AI)技术在地震监测领域取得了显著进展.基于深度神经网络的多尺度特征自动学习能力,AI 方法能有效从复杂背景噪声中识别地震震相,提升自动检测与定位效率.国内外学者利用不同的数据集、不同的网络模型训练得到了GPD(Ross et al., 2018)、PhaseNet(Zhu and Beroza, 2019)、EQTransformer(Mousavi et al., 2020)、LPPN(Yu and Wang, 2022)、RNN(Yu et al., 2023)、USTC-Pickers(Zhu et al., 2023)、SeisT(Li et

al., 2024)、SeisLM(Liu et al., 2024)、PRIME-DP(Yu et al., 2024)和 SeisMoLLM(Wang et al., 2025)等多个地震检测模型,推动了地震数据处理由基于规则的算法向数据驱动的特征学习转变,显著提高了地震数据处理的效率以及微弱地震信号的自动识别能力(Chai et al., 2020; Mousavi and Beroza, 2022; Münchmeyer et al., 2022; Bornstein et al., 2023; Feng et al., 2023; Si et al., 2024). 然而,现有模型多基于信噪比较高的速度型波形数据训练,数据源以速度型地震计记录为主. 例如,PhaseNet 训练时使用的北加州数据中 72%为速度型数据,STEAD 数据集(Mousavi et al., 2019)、DiTing 数据集(Zhao et al., 2023)和 CSNCD 数据集(An, 2024)等均是速度型波形数据. 这类数据通常来自安装环境良好、灵敏度高且频带较宽的速度型地震计,具有信噪比高、噪声干扰少的特点.

2024 年 7 月,国家地震烈度速报与预警工程(以下简称“预警工程”)通过竣工验收,我国地震台站数量超过 20000 个,标志着我国建成了全球规模最大的地震监测台网(Peng et al., 2022). 预警工程在全国范围内部署了宽频带速度型地震计(以下简称速度计,用 HH 表示)、力平衡加速度型地震计(以下简称加速度计,用 HN 表示)及 MEMS(micro-electro-mechanical system)简易烈度计(以下简称烈度计,用 EI 表示)三种类型传感器.以云南为例,全省共部署速度计 202 套、加速度计 430 套、烈度计 1230 套(朱杰和钟玉盛, 2025) ,其中加速度型传感器占比超过 85%.加速度计和烈度计在频带响应、灵敏度及安装环境等方面与速度计不同,导致加速度波形与速度型波形特征存在显著差异.将基于速度型数据训练的震相识别模型直接应用于加速度和烈度计数据时,其检测精度和识别准确率会下降(图 1).

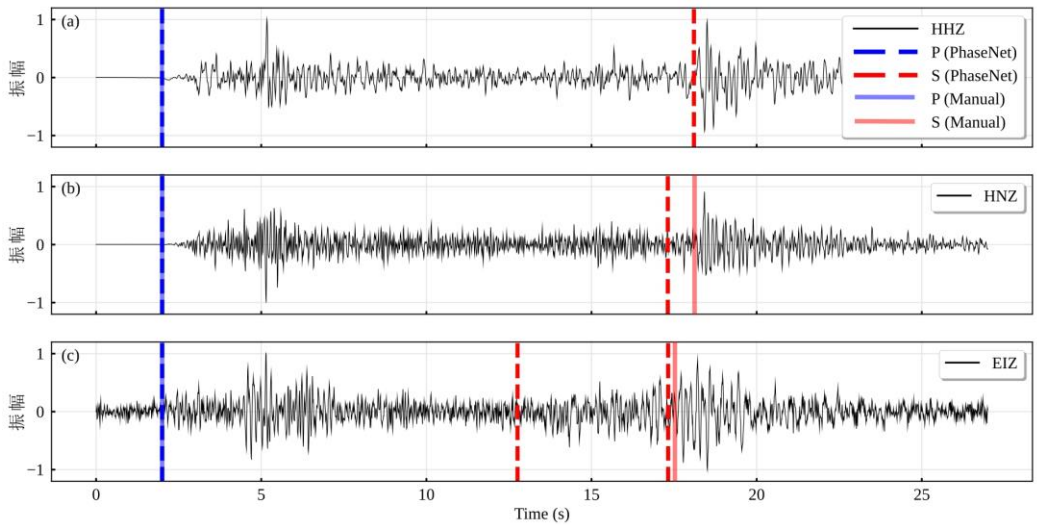

图 1 不同类型仪器记录的地震波形和 PhaseNet 检测的到时对比

Fig.1 Comparison of seismic waveforms recorded by different instruments and PhaseNet-detected arrival times

云南预警台网 SYX02 台(同址安装速度计和加速度计)和距离其最近的 J2811 台(安装烈度计)记录到的 2023 年 9 月 3 日云南耿马 M_L 3.4 地震波形.黑色实线为归一化后的地震波形,蓝色实线为人工标记的 Pg 到时、红色实线为人工标记的 Sg

到时,蓝色虚线为 PhaseNet 拾取的 Pg 到时,红色虚线为 PhaseNet 拾取的 Sg 到时;(a)为速度计,(b)为加速度计,(c)为烈度计;从图中可以看出,PhaseNet 在速度计数据上检测的震相到时与人工标注的震相到时较为一致,在加速度计或者烈度计数据上检测的震相到时与人工标注的震相到时存在一定差异.

本文基于云南预警台网 2023 - 2025 年的观测数据,制作了高质量的地震事件波形数据集和震相到时数据集,对 PhaseNet、EQTransformer 和 SeisMoLLM 等 9 种主流地震检测模型进行评估,分析了不同模型在速度计数据、加速度计数据和烈度计数据中的检测效果,探讨了迁移学习和大模型的泛化能力及适用边界,为发展基于多源异构数据的地震检测方法提供参考.

1 数据

1.1 云南地震台网简介

云南地处青藏高原东南缘,构造运动剧烈,地震活动频繁(皇甫岗, 2009).预警工程建设前,云南地区共有 68 个固定测震台站,平均台间距约为 76 km(图 2 a).

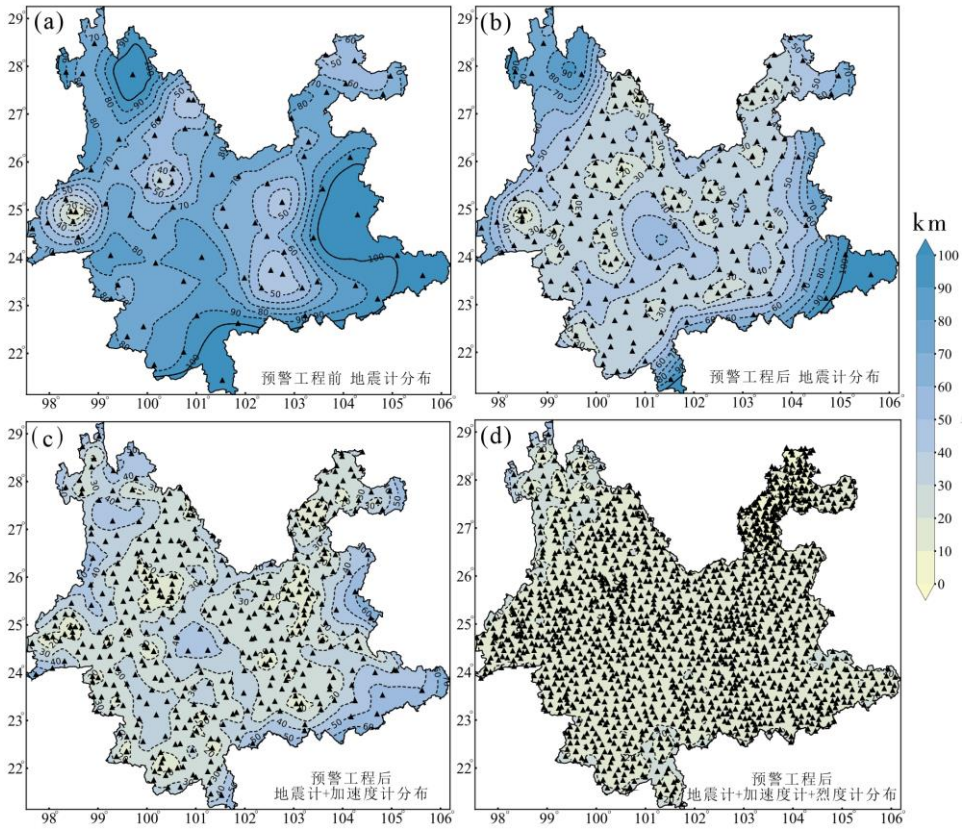

图 2 预警工程建设前后云南地震台网密度及台站分布图

Fig.2 Station density and distribution of the Yunnan Seismic Network before and after the earthquake early warning project construction

预警工程建成后,云南地震台网除原测震台站外,新增地震台站 1660 个,其中基准站 202 个、基本站 228 个、一般站 1230 个,云南地区部署的地震计类型见表 1.基准站同时安装速度计和加速度计,速度计(通道代码为 HHZ、HHN、HHE),加速度计(通道代码为 HNZ、HNN、

HNE).基本站配备加速度计(通道代码为 HNZ、HNN、HNE).一般站配备烈度计(通道代码为 EIZ、EIN、EIE).所有台站实时回传数据的采样率均为 100 Hz.预警工程显著提升了云南地区的地震台站分布密度和监测能力：速度计平均台间距缩减至约 43 km,全台网平均间距缩小至约 15 km(图 2).

表 1. 云南地区部署的地震计类型汇总表

Table 1 Summary of Seismometer Types Deployed in Yunnan Region

台站	背景噪声(m/s^2)	类型	设备型号	频带	灵敏度	通道	数量	制造商
基准站	$1.0\times10^{-9}\sim10^{-8}$	速度计	GL_CS60	60s~50Hz	2000[a]	HH*	78	港震
			GL_CS120	120s~50Hz	2000[a]	HH*	33	港震
			ITC-60A	60s~50Hz	2000[a]	HH*	43	天元
			ITC-120A	120s~50Hz	2000[a]	HH*	31	天元
			BBVS-60	60s~50Hz	2000[a]	HH*	10	港震
			JS-60	60s~50Hz	2000[a]	HH*	4	深研院
			JS-120	120s~50Hz	2000[a]	HH*	3	深研院
		加速度计	JS-A2	DC~80Hz	2.5[b]	HN*	101	深研院
			TDA-33M	DC~80Hz	2.5[b]	HN*	101	泰德
基本站	$1.0\times10^{-6}\sim10^{-5}$	加速度计	TDA-33M	DC~80Hz	2.5[b]	HN*	113	泰德
			JS-A2	DC~80Hz	2.5[b]	HN*	115	深研院
一般站	$1.0\times10^{-4}\sim10^{-3}$	烈度计	GL-P2B	DC~80Hz	1.0×10^{6}[c]	EI*	20	港震
			Palert Advance	DC~80Hz	-	EI*	100	勤联
			TMA-33	DC~80Hz	-	EI*	705	泰德
			VH-GL-LDY	DC~80Hz	1.0×10^{6}[c]	EI*	405	瑞琪

注：[a]单位为 V/(m/s); [b]单位为 V/(m/s^2); [c]单位为 count/(m/s^2)

1.2 多源异构数据集

本文收集了 2023 年 5 月至 2025 年 4 月期间云南预警台网观测的 1194 次地震的波形数据.经人工挑选与复核,共标注 Pg 震相到时 24,411 条、Sg 震相到时 18,452 条(部分波形仅包含 Pg 震相),仪器包括 152 台速度计、337 台加速度计和 897 台烈度计(图 3).

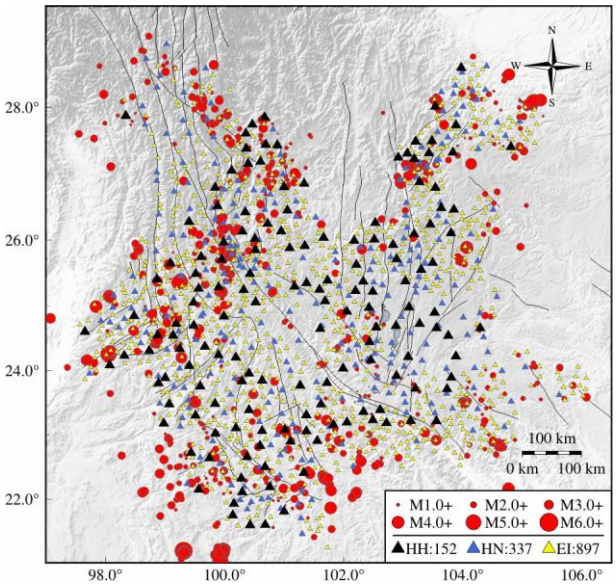

图 3 数据集中地震事件震中分布和台站分布

Fig.3 Epicenter and station distribution of earthquakes in the dataset

图中红色圆圈代表不同震级地震事件,黑色三角形代表速度计,蓝色三角形代表加速度计,黄色三角形代表烈度计

从连续波形中截取震前 10 s 和震后 50 s 三分量数据,构建长度为 60 s(6,000 个采样点)的波形数据集.由于同一地震事件不同台站的震中距存在差异,震相到时在截取窗口内具有随机性.数据集中包含速度计波形 9,098 条,加速度计波形 9,355 条,烈度计波形 5,958 条,分别占比 37.3%、38.3%和 24.4%.1194 次地震的震级范围为 M 0.0～5.2(图 4 a),震中距为 0～120 km(图 4 b),采用 SeisBench(Woollam et al., 2022)将数据集封装为 hdf5 格式.参照 PhaseNet 的信噪比计算方法,以 Pg 到时后 5 s 波形与前 5 s 背景噪声标准差比值的对数作为信噪比指标,对于未标注 Pg 震相的记录,则改用 Sg 到时后 5 s 波形进行计算.图 4(c)展示了数据集中 HH、HN、EI 三类数据归一化后的信噪比分布,图 4(d)、(e)、(f)分别为三类数据的震相走时曲线图.人工标注震相时以能清晰判断震相到时和类型为准则,从图 4(a)和图 4(b)可以看出,速度计样本相较于加速度计和烈度计,震级较小且震中距较远.从图 4(c)可以观察到,烈度计数据信噪比峰值主要位于 0.5-1.5 之间,超过 85%的数据信噪比低于 1.5;加速度计数据的信噪比分布特征与速度计形态相似，但其在高信噪比区间(>2.0)的占比略低于速度计;速度计数据信噪比分布更加均衡，在高信噪比区间保持较高占比.从图 4(d)、(e)、(f)可以看出,数据集中 HH、HN、EI 三类数据 Pg 和 Sg 震相走时较为一致,相同震中距下震级逐渐增大,EI 的 Pg 震相数量随震中距增加逐渐减少.

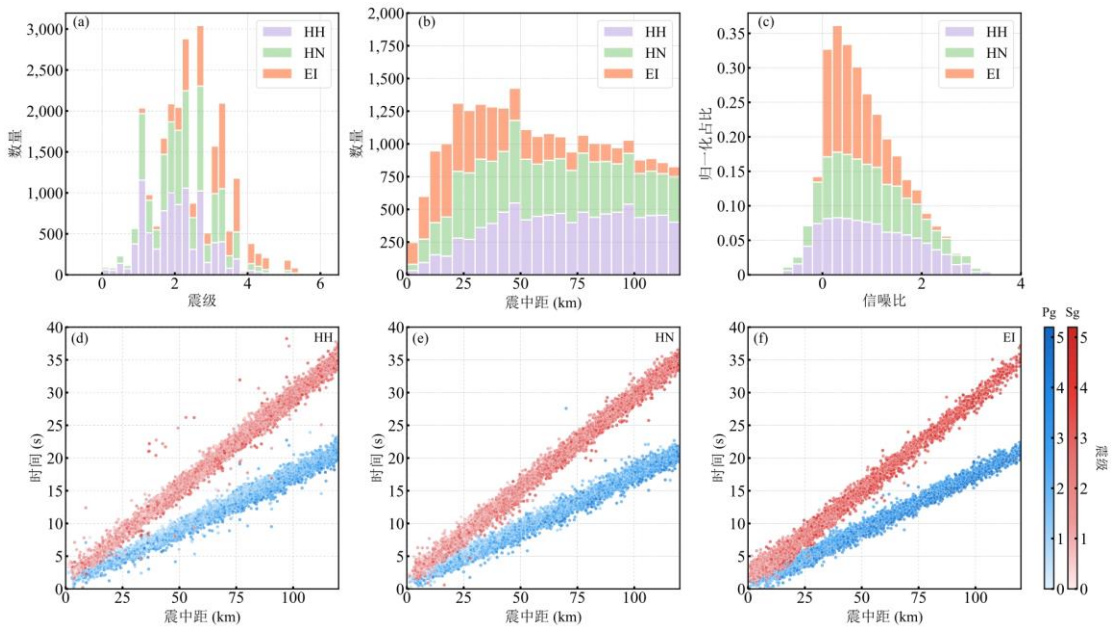

图 4 数据集特征统计图

Fig.4 Statistical characteristics of the dataset

(a)震级统计; (b)震中距; (c)信噪比; (d)速度计数据震相走时曲线; (e)加速度计数据震相走时曲线; (f) 烈度计数据震相走时曲线. (d)、(e)、(f)图中蓝色为 Pg, 红色为 Sg, 散点颜色深浅代表不同震级, 颜色越深震级越大.

2 方法

2.1 地震检测模型测评

为系统评估地震检测模型性能,本文选取了 PhaseNet、EQTransformer、RNN、LPPN 和 USTC-Pickers 5 个代表性的专业模型与 SeisMoLLM、SeisT、PRIME-DP 和 SeisLM 4 个较新的大模型进行对比分析.这些模型覆盖了从传统卷积网络、循环网络到 Transformer、大语言模型微调等主流架构范式,涉及不同的数据规模与训练策略(如区域迁移训练、全球预训练、本地微调等),所有模型均未使用本文所构建的数据集进行训练,避免数据泄露问题.为保障评测客观性,筛选数据集中震中距 60 km 范围内 M_L2.0 以上地震进行评测,其中速度计数据 1330 条、加速度计数据 2106 条,烈度计数据 3378 条.

专业模型中,PhaseNet 是基于 U-Net 的卷积神经网络结构,由四层卷积下采样与四层反卷积上采样模块组成,使用北加州地震数据中心的 77 万条三分量地震波形进行训练.EQTransformer 融合 CNN 的局部特征提取能力、LSTM 的时序建模能力与 Transformer 的全局注意力机制,采用共享编码器和三个独立解码器结构,基于 STEAD 数据集训练.RNN 基于移除了 Transformer 仅保留 CNN 与 LSTM 的 EQTransformer 架构,采用 CSNCD 数据集进行训练.LPPN 是一种轻量化神经网络模型,结合残差连接与深度可分离卷积,基于 STEAD 数

据集进行训练.USTC-Pickers 采用 U-Net 架构,通过迁移学习策略实现模型区域适应性,初始模型在 DiTing 数据集上进行全国范围预训练,后续引入各省区域数据进行微调,本文测试所用模型为云南区域子模型.

大模型中,SeisT 采用多尺度混合卷积和多路径 Transformer 架构,支持震相拾取和地震定位等多种任务联合推理,基于 STEAD 和 DiTing 数据集训练.PRIME-DP 采用 CNN-Transformer 混合编码器以及多任务解码器架构,使用 CSNCD 数据集训练.SeisMoLLM 采用跨模态方式将波形分块编码为 Token 序列后微调 GPT-2 架构,实现多任务端到端推理,模型采用 STEAD 和 DiTing 数据集训练.SeisLM 采用 ConvNet + Transformer 结构,采用掩码建模和对比学习进行自监督预训练,通过 STEAD 数据集训练.表 2 列出了本文评测的模型、训练数据集和训练策略.

表 2　地震检测模型信息统计表

Table 2　Summary of information on earthquake detection models

模型类型	模型名称	训练数据集	训练策略	参考文献
专业模型	PhaseNet	NCEDC	全监督训练	Zhu and Beroza, 2019
	EQTransformer	STEAD	全监督训练	Mousavi et al., 2020
	RNN	CSNCD	全监督训练	Yu et al., 2023
	LPPN	STEAD	全监督训练	Yu and Wang, 2022
	USTC-Pickers	DiTing	预训练+增量微调	Zhu et al., 2023
大模型	SeisT	STEAD/DiTing	联合训练,多任务优化	Li et al., 2024
	PRIME_DP	CSNCD	预训练+解码器微调	Yu et al., 2024
	SeisMoLLM	STEAD/DiTing	跨模态微调	Wang et al., 2025
	SeisLM	STEAD	自监督预训练+微调	Liu et al., 2024

2.2 性能评估指标

本文采用多维度指标评估不同模型的性能表现,分别根据公式(1)、(2)、(3)计算模型精确率(Precision)、召回率(Recall)和 F1 值.

$$\mathrm{Precision}=\frac{TP}{TP+FP} \tag{1}$$

$$\mathrm{Recall}=\frac{TP}{TP+FN} \tag{2}$$

$$\mathrm{F1}=2\times\frac{\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{3}$$

式中,人工标记的震相到时与模型拾取的震相到时差的绝对值在△t 内为正确拾取,记为 *TP*;超过△t 为误拾取,记为 *FP*;模型未检测到人工标记的震相到时为漏拾取,记为 *FN*.

在常规评估指标基础上,通过计算模型预测到时与人工标注到时的标准差(σ)和平均绝对误差(Mean Absolute Error,MAE),进一步对震相到时误差进行统计分析. σ 表征拾取结果的

离散程度, σ 值越小,表明模型的一致性越好.MAE 用于评估模型平均误差幅度,值越小,表明模型预测到时越接近真实标签.为消除仪器故障等异常值干扰,参考 PhaseNet 和 LPPN 的做法,仅统计误差绝对值在 0.5 s 范围内的有效样本,确保评估结果的代表性和可比性. σ 、MAE 分别通过公式(5)、(6)计算.

$$e_i = T_{true,i} - T_{pre,i} \tag{4}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - \mu)^2}, \quad \mu = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{5}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{6}$$

式中, $T_{true,i}$ 为样本 i 人工标注的震相到时, $T_{pre,i}$ 为样本 i 模型预测的震相到时, e_i 为样本 i 的到时误差, μ 为误差均值, n 为样本数量.

3 结果

3.1 模型性能评测结果

德国地球科学研究中心在评估地震定位结果可靠性时,将 0.1 s 视为震相到时的典型误差阈值,并以此作为定位算法的输入参数,分析其对震源位置反演精度的影响.实验结果表明,该误差量级满足高精度定位标准(Bormann et al., 2013).在震相拾取模型的性能评价中,PhaseNet、EQTransformer 等模型也普遍采用“拾取误差 0.1 s 内”作为正确拾取的判定依据(Zhu and Beroza, 2019; Mousavi et al., 2020).本文沿用这一标准,对各模型的 Pg、Sg 拾取性能进行了系统评估,计算了精确率、召回率和 F1 值,结果如表 3 所示.不同模型在相同误差阈值下的表现差异显著,反映了模型对台站响应特性与区域地壳结构的适应性.

震相拾取测试结果表明,不同模型在不同类型数据和震相上的性能存在显著差异：9 个模型在速度计数据上的平均 F1 值为 0.712,高于加速度计的 0.684 和烈度计的 0.672.各模型对 Pg 的拾取性能普遍优于 Sg,其中 SeisMoLLM、SeisT 和 EQTransformer 在两类震相间的性能差异相对较小,而 PhaseNet、RNN、LPPN、PRIME_DP 和 SeisLM 在 Pg 上的优势更为明显.USTC-Pickers 在三类数据上的平均 F1 值最高,为 0.779;在速度计和加速度计上的 F1 值分别为 0.768 和 0.798,均为各模型中最高值;在烈度计上的 F1 值为 0.771,仅次于 SeisT 的 0.793.SeisT 的综合表现仅次于 USTC-Pickers,平均 F1 值为 0.775. SeisLM 和 SeisMoLLM 紧随其后,平均 F1 值分别为 0.726 和 0.723.图 5 展示了不同模型在三类数据上的精确率、召回

率和 F1 值对比雷达图.从图中可以看出,不同模型在速度计数据,尤其是 Pg 上的性能差异相对较小,在加速度计和烈度计上的性能差异相对较大.

表 3　不同模型在本文数据集上的拾取结果统计(误差在 0.1 s 为 *TP*)

Table 3　Summary of picking results of different models on the dataset (error within 0.1 s counted as *TP*)

测试数据	测试模型	Pg			Sg		
		Precision	Recall	F1	Precision	Recall	F1
速度计	PhaseNet	0.827	0.867	0.847	0.528	0.577	0.551
	EQTransformer	0.859	0.870	0.864	**0.702**	0.631	0.665
	RNN	0.842	0.822	0.832	0.434	0.454	0.444
	LPPN	0.801	0.765	0.782	0.576	0.413	0.481
	USTC-Pickers	**0.870**	0.879	**0.874**	0.637	**0.688**	0.662
	SeisT	0.858	0.856	0.857	0.663	0.634	0.648
	PRIME_DP	0.806	0.803	0.805	0.479	0.539	0.507
	SeisMoLLM	0.804	0.865	0.833	0.653	0.681	**0.666**
	SeisLM	0.863	**0.889**	0.876	0.579	0.658	0.616
	平均值	0.837	0.846	0.841	0.584	0.586	0.582
加速度计	PhaseNet	0.802	0.846	0.823	0.422	0.565	0.484
	EQTransformer	0.876	0.832	0.853	0.579	0.622	0.600
	RNN	0.720	0.691	0.705	0.399	0.494	0.441
	LPPN	0.676	0.595	0.633	0.501	0.440	0.468
	USTC-Pickers	**0.936**	**0.935**	**0.936**	0.588	**0.753**	**0.660**
	SeisT	0.918	0.904	0.911	**0.605**	0.701	0.650
	PRIME_DP	0.760	0.735	0.747	0.393	0.525	0.449
	SeisMoLLM	0.880	0.858	0.869	0.568	0.711	0.632
	SeisLM	0.859	0.903	0.881	0.475	0.667	0.555
	平均值	0.825	0.811	0.818	0.503	0.609	0.549
烈度计	PhaseNet	0.810	**0.831**	0.820	0.479	0.611	0.537
	EQTransformer	0.900	0.617	0.732	0.665	0.529	0.589
	RNN	0.779	0.724	0.750	0.499	0.590	0.541
	LPPN	0.663	0.474	0.553	0.554	0.464	0.505
	USTC-Pickers	0.827	0.819	0.823	0.648	**0.803**	0.718
	SeisT	0.883	0.828	**0.855**	**0.682**	0.785	**0.730**
	PRIME_DP	0.766	0.602	0.674	0.471	0.533	0.500
	SeisMoLLM	**0.951**	0.535	0.684	0.617	0.698	0.655
	SeisLM	0.865	0.776	0.818	0.541	0.701	0.611
	平均值	0.827	0.689	0.746	0.573	0.635	0.598

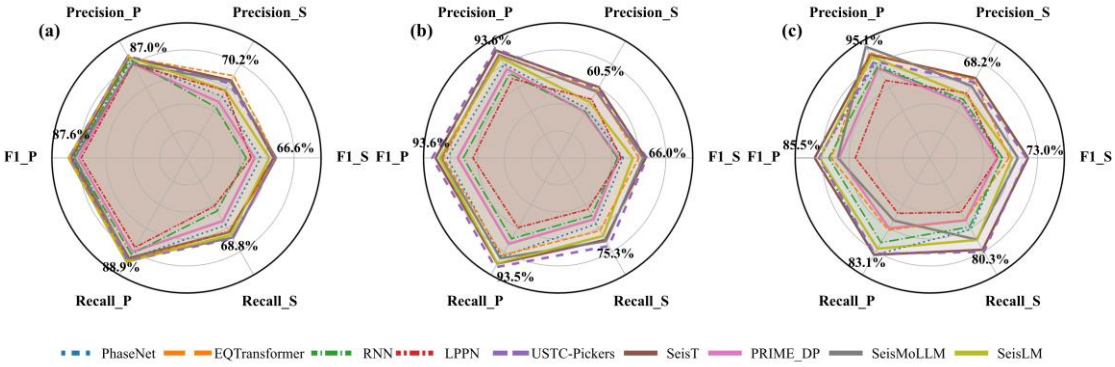

图 5 不同模型在三类数据上的精确率、召回率和 F1 雷达图

Fig.5 Precision, recall, and F1-score radar chart of different models on three data types

(a)速度计;(b)加速度计;(c)烈度计

为进一步了解模型性能,本文计算了±0.5 s 误差范围内的各模型拾取结果的标准差 σ 和平均绝对误差 MAE,图 6 展示了 9 个模型在三类数据上的震相到时误差标准差及平均绝对误差对比.SeisMoLLM、USTC-Pickers 和 SeisT 的误差标准差和平均绝对值误差最低,尤其是 SeisMoLLM,在烈度计数据上一致性最好.图 7 和图 8 分别展示了各模型 Pg 和 Sg 震相到时误差的分布直方图.结果显示,所有模型的预测误差主要集中在±0.2 s 区间内,USTC-Pickers 和 SeisT 等大模型的误差分布相对均衡,而 PhaseNet 和 EQTransformer 等专业模型误差分布呈明显右偏态分布,表明专业模型拾取的震相到时相对滞后,而经过本区域数据迁移训练的模型能有效抑制滞后趋势.

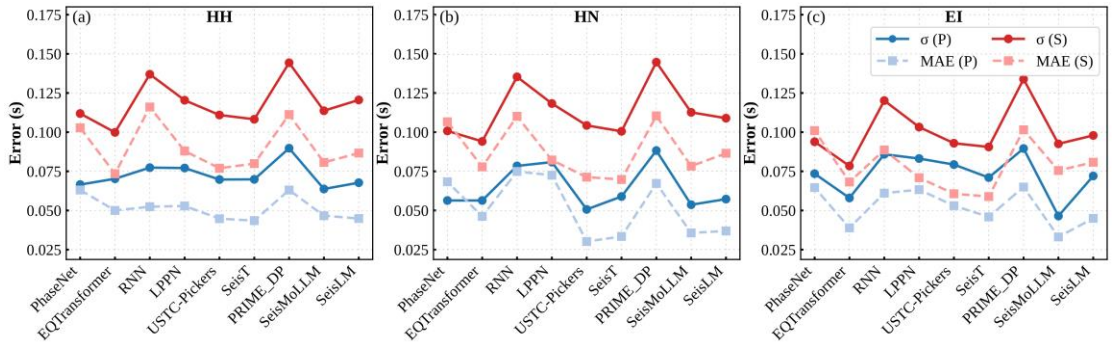

图 6 不同模型在三类数据上的震相到时误差标准差及平均绝对误差对比

Fig.6 Comparison of standard deviation and mean absolute error of phase arrival times for different models on three data types

(a)速度计;(b)加速度计;(c)烈度计

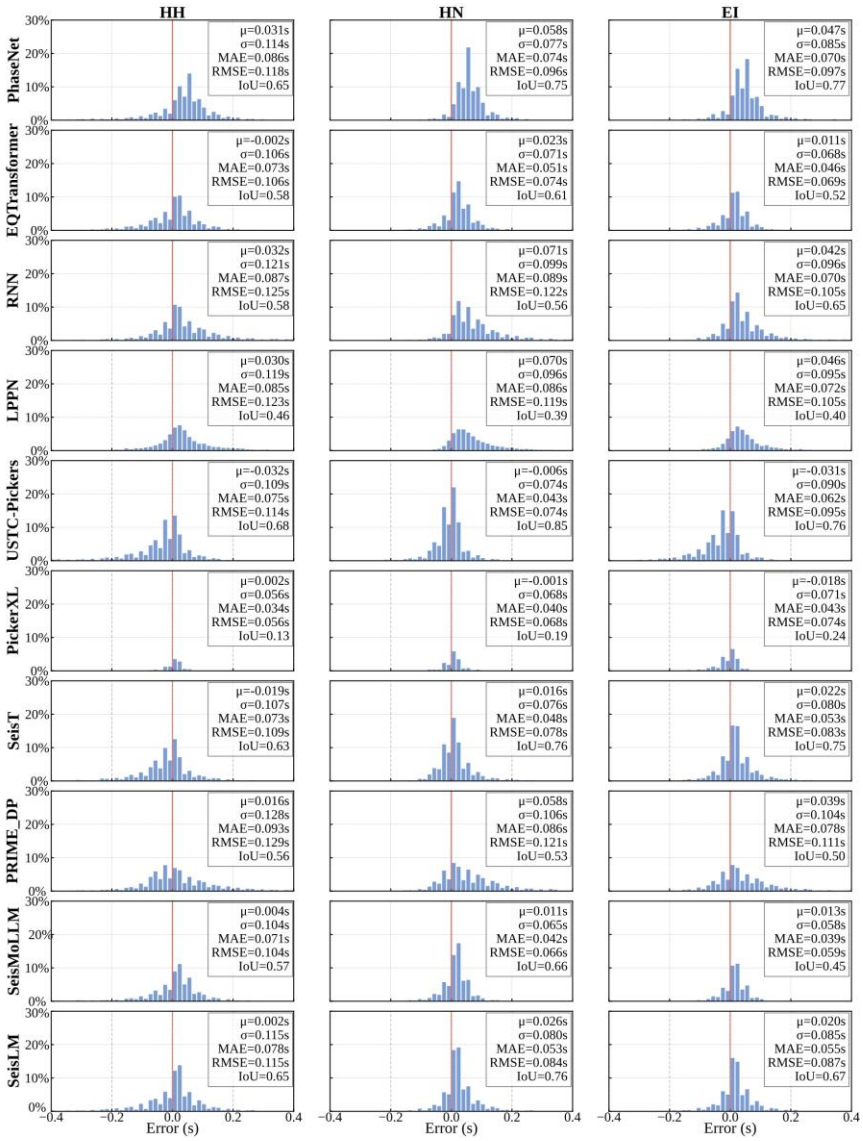

图 7 不同模型在三类数据上的 Pg 震相到时误差($T_{pred} - T_{true}$)分布直方图

Fig.7 Histogram of Pg arrival time errors ($T_{pred} - T_{true}$) for different models on three data types

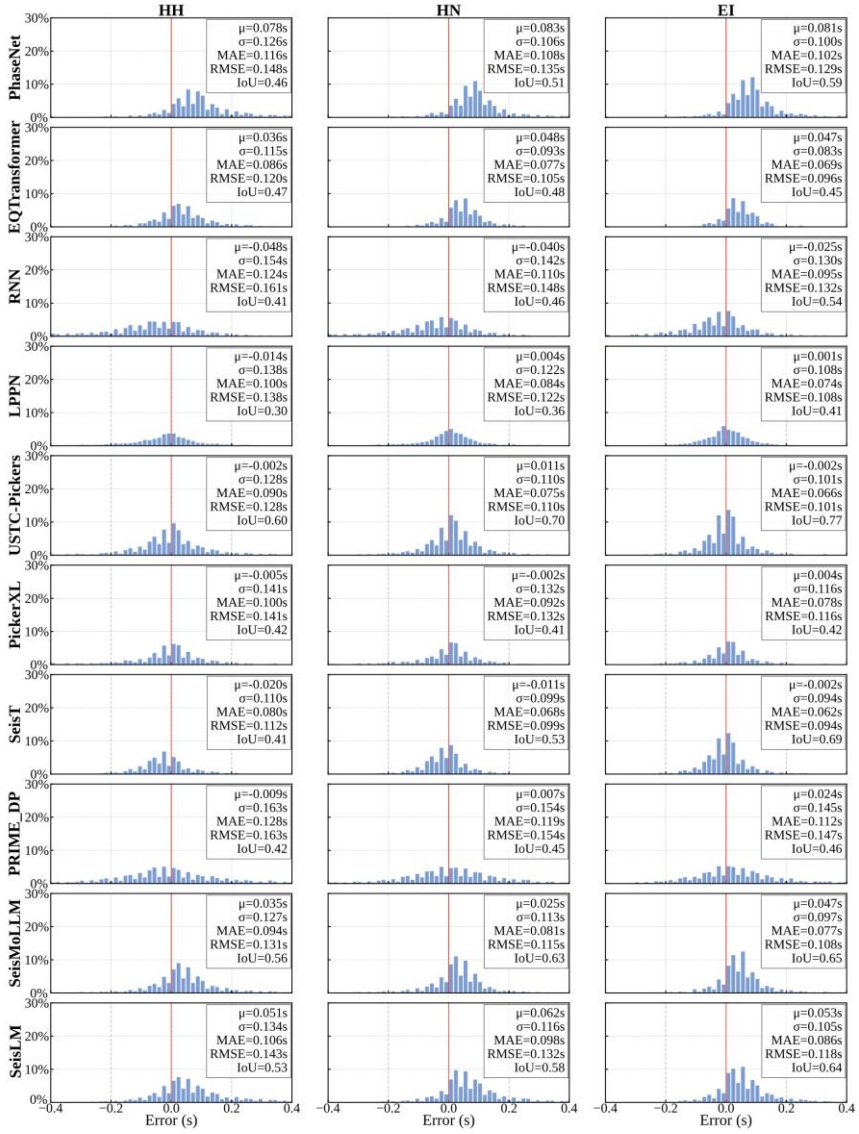

图 8 不同模型在三类数据上的 Sg 震相到时误差($T_{pred} - T_{true}$)分布直方图

Fig.8 Histogram of Sg arrival time errors ($T_{pred} - T_{true}$) for different models on three data types

3.2 不同时窗长度拾取效果对比

从数据集中筛选出信噪比较高的加速度计波形样本,对真实震相概率以及 PhaseNet、USTC-Pickers 和 SeisT 三个具有代表性模型的输出概率曲线进行对比分析.如图 9 所示,在时长为 30 s 的波形数据上(图 9 a),三个模型均能准确识别 Pg 和 Sg 震相,在窗长为 12 s 的数据上(图 9 f),PhaseNet 拾取的 Pg 相较于真实到时表现出明显的滞后特征,而经过云南地区数据迁移训练的 USTC-Pickers 有效抑制了这种滞后现象.此外,USTC-Pickers 在不同时长数据上的表现具有较高一致性,这种现象可能源于专业模型学习了更多的波形细节特征.相比之下,SeisT 在 30 s 和 12 s 不同时长数据上呈现出显著差异的拾取结果,这种特性可能与其融合

Transformer 结构后同时关注全局信息有关.因此,当数据长度不同时,不同模型的拾取结果会有差异.

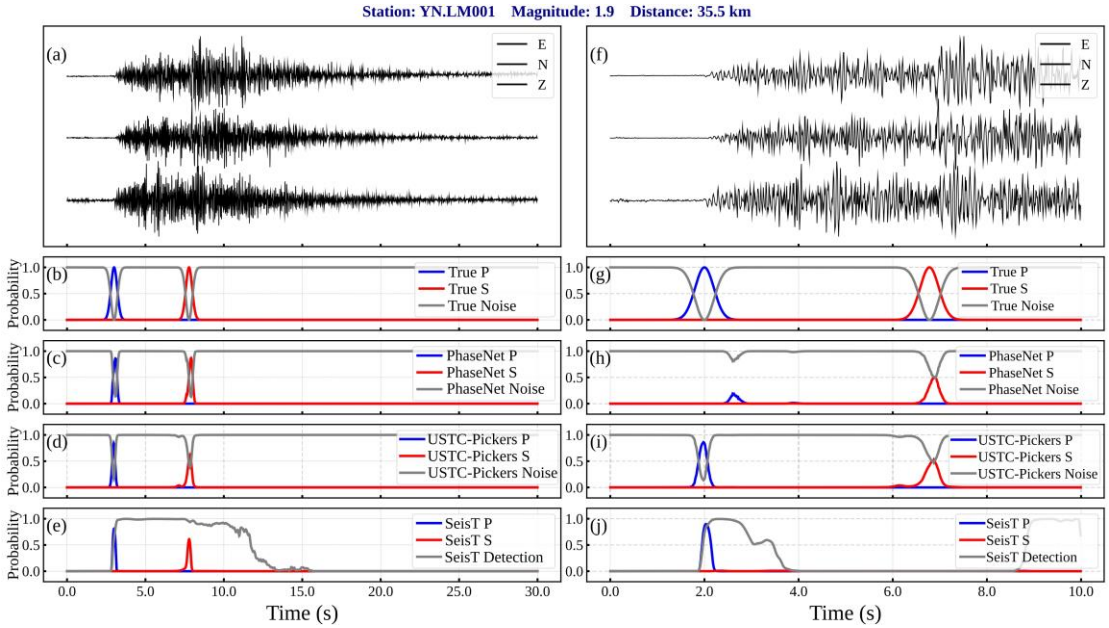

图 9 模型在同一波形不同时窗长度数据上的拾取效果对比

Fig.9 Comparison of picking performance of different models on the same waveform with different time window lengths

左图为 30 s 窗长;右图为 12 s 窗长;(a)和(f)为三分量原始波形;(b)和(g)为真实标签到时扩展为高斯窗的概率曲线;(c)和(h)、(d)和(i)、(e)和(j)分别为 PhaseNet、USTC-Pickers 和 SeisT 输出的概率曲线

3.3 不同数据条件下模型性能变化

为进一步分析不同数据条件下模型的拾取性能变化,基于 0.1 s 的判定阈值统计各模型的正确拾取数量,绘制了模型在不同信噪比、震级和震中距条件下对三类数据的拾取效果对比图.对速度计的拾取结果(图 10)显示,模型整体性能趋势与数据集的原始分布特征保持一致,模型在 Pg 上的性能相较于 Sg 更为接近,表明模型在 Pg 上的拾取差异较小,而在 Sg 上的拾取差异较大,尤其是震中距大于 40 km 后,差异进一步增大.

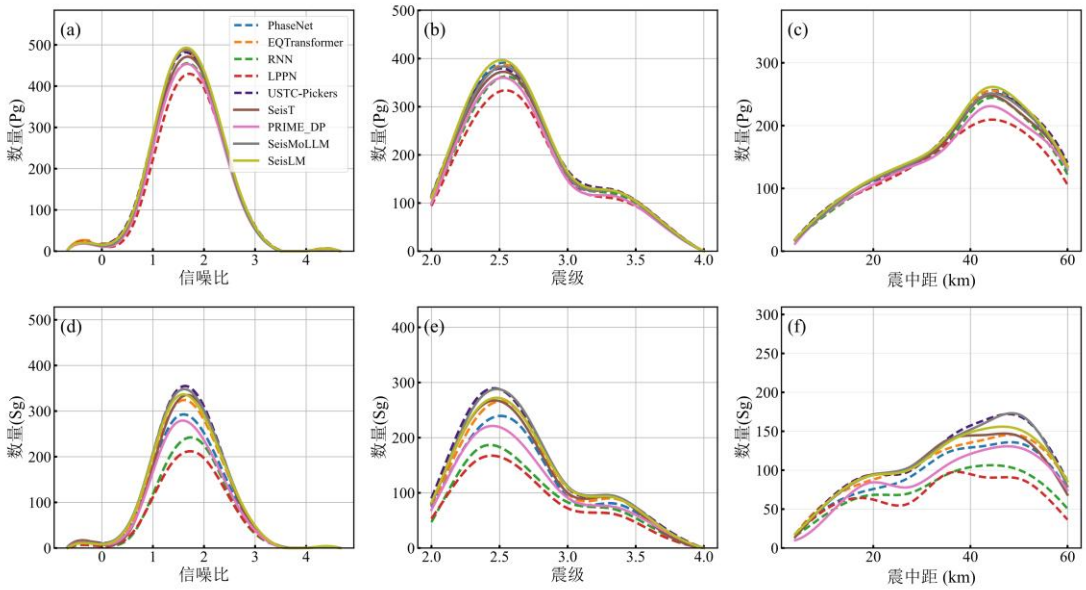

图 10 模型在不同信噪比、震级和震中距离的速度波形上的拾取效果对比

Fig.10 Phase picking performance on velocity data under different SNR, magnitude, and epicentral distance

在加速度计数据上,模型间的拾取差异相较于速度计进一步增大,USTC-Pickers 模型展现出最好的性能,在其他所有测试条件下均保持最优表现.此外,SeisT、SeisMoLLM 和 SeisLM 等大模型在 Sg 拾取任务中表现突出,其性能优于除 USTC-Pickers 外的其他专业模型,这一现象可能源于大模型具有提取更丰富波形特征的能力(图 11).

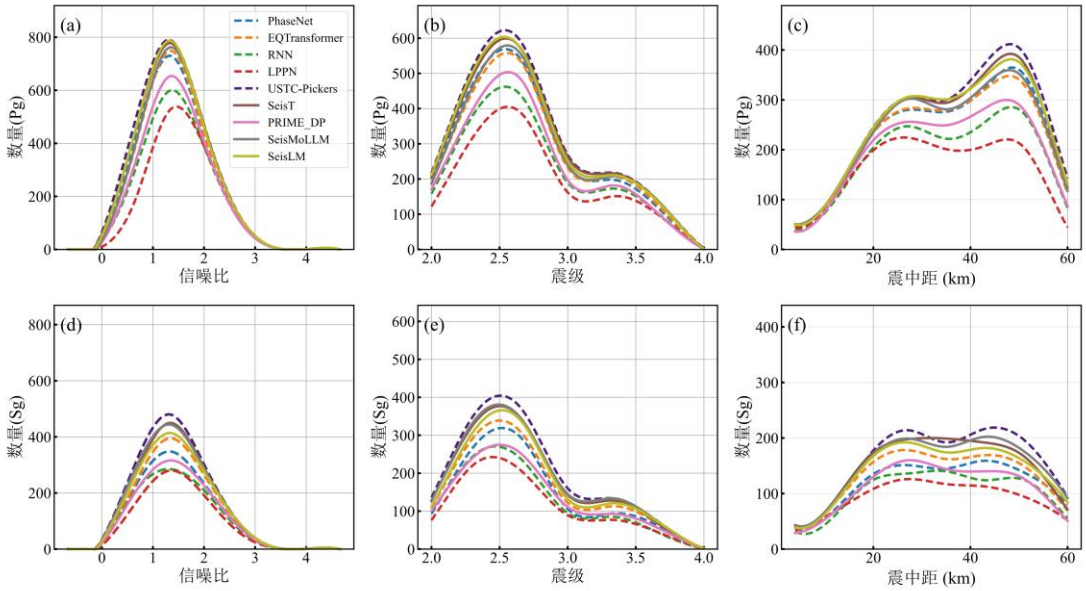

图 11 模型在不同信噪比、震级和震中距的加速度计数据的拾取效果对比

Fig.11 Phase picking performance on accelerogram under different SNR, magnitude, and epicentral distance

与加速度计类似,模型在烈度计数据的拾取差异相较于速度计同样较大.USTC-Pickers、

SeisT、SeisMoLLM 和 SeisLM 依然拥有较好表现(图 12).

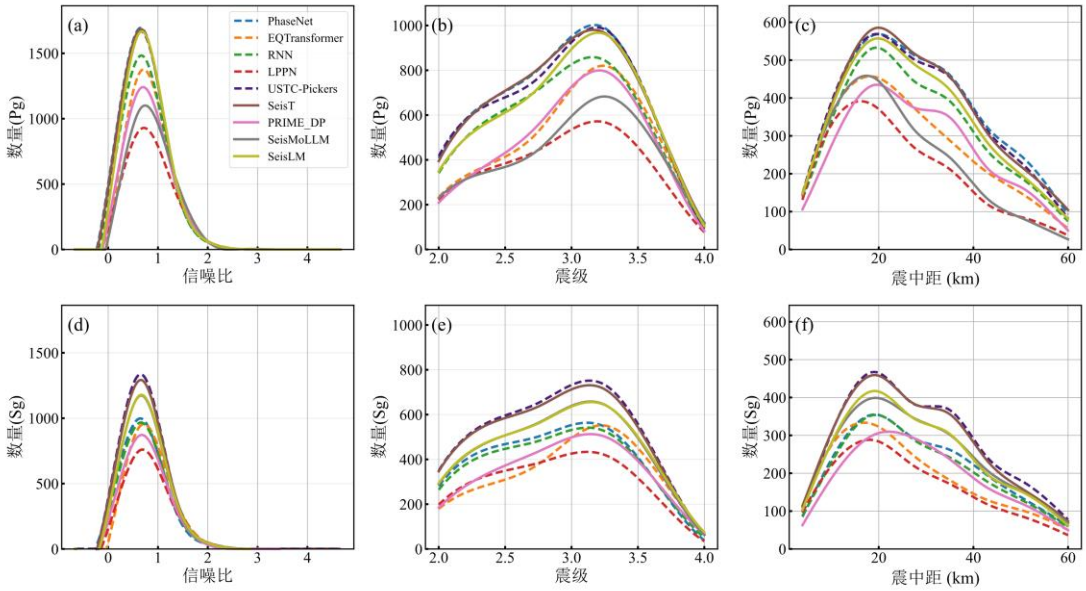

图 12 模型在不同信噪比、震级和震中距的烈度计数据的拾取效果对比

Fig.12 Phase picking performance on MEMS data under different SNR, magnitude, and epicentral distance

3.4 同址速度计和加速度计数据拾取性能比较

依托预警基准站同步部署的观测优势,选取 2023 年 5 月 2 日云南保山隆阳 M_S 5.2 地震 100 km 范围内 14 个基准站的速度计和同址加速度计数据(震相检测时段为 2023 年 5 月 2 日 0 时至 5 月 5 日 0 时)进行速度计和加速度计震相拾取同址对比.采用 PhaseNet 进行震相拾取,结果表明:加速度计的震相拾取数量仅为速度计的 50%左右;速度计数据上 Pg 拾取数量多于 Sg,而加速度计则呈现相反趋势(图 13 a),这种现象可能源于宽频带速度计更宽的频带响应范围和更高的灵敏度使其能记录更丰富的信号特征.统计显示,同台站速度计与加速度计的 Pg(图 13 b)和 Sg(图 13 c)到时误差均值分别为-0.031 s 和-0.039 s,表明加速度计震相识别存在约 0.03 - 0.04 s 的系统性延迟;二者的标准差分别为 0.103 和 0.121,反映出较高的一致性,其中 P 波误差波动更小,这与 P 波初动清晰、更易识别特性相符.P 和 S 的偏度系数分别为-1.91 和-1.37,表明某些条件下的加速度计的震相到时会出现显著延迟.

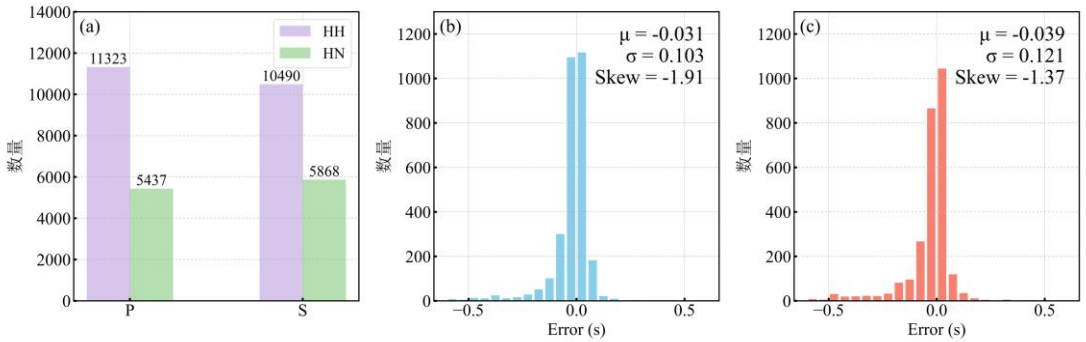

图 13 PhaseNet 对同台址速度计和加速度计震相拾取对比

Fig.13 Phase picking comparison on co-located velocity and acceleration data

(a)为 P 和 S 的震相数量;(b)为速度计与加速度计 P 到时误差 *T*(HH) - *T*(HN)分布;(c)为速度计与加速度计 S 到时误差 *T*(HH) - *T*(HN)分布

4 讨论

4.1 不同模型在多元异构数据上的适用性分析

从各模型在数据集的整体表现来看,USTC-Pickers 综合表现最优,平均 F1 值为 0.779,SeisT、SeisLM 和 SeisMoLLM 紧随其后,平均 F1 值均超过 0.7.但是,所有模型在 Sg 震相拾取上的表现均相对较差,后续模型开发应着重提升对 Sg 震相的识别能力.误差分布结果表明,PhaseNet、EQTransformer 等专业模型在三类数据尤其是加速度型数据上均存在明显的震相到时滞后现象.这种现象可能源于加速度计和烈度计的灵敏度较低,当地震信号较小时其响应能力不足导致初至震相拾取困难,需要依赖后续较大振幅才能准确识别震相.相比之下速度计凭借高灵敏度和低噪声环境优势能够清晰记录震相初至.经过本地数据训练的 USTC-Pickers 即便在仪器类型和观测环境发生变化的情况下,仍然可以有效抑制这种滞后趋势.相关研究表明,迁移学习可以显著提高模型的适应能力(Chai et al., 2020; Niksejel and Zhang, 2024; Saad et al., 2024),利用历史数据进行迁移训练,模型能够更好地掌握本地地质构造特征,进而提高在不同类型数据上的拾取效果.从测试结果来看,相较于更换模型或重新设计模型架构,迁移训练带来的性能提升更为显著.同时,由于加速度型地震计存在灵敏度低、安装环境受人类活动干扰大、震相到时滞后等问题,在实际地震定位中应优先采用速度计数据以提高定位精度.

4.2 专业化模型与通用大模型性能比较

专业模型通常被认为在特定场景中更高效(Pecher et al., 2025),而大模型更强调通用泛化能力(Du et al., 2024).用于地震检测的专业模型自 2017 年以来得到了快速发展(Fang et al., 2017),而地震通用大模型最近 2 年才刚刚起步,目前全球已公布的用于地震检测的大模型只有 SeisT、SeisLM、PRIME-DP、SeisMoLLM 和 DiTing(http://www.esdc.ac.cn/).震相拾取作为一种密集型时序语义分割任务,要求模型对每个采样点给出精确的分类概率,因此无论是专业模型还是大模型,都需要具备对震相到时的敏感性.SeisMoLLM 的结构消融实验显示：即便是具备时序特征理解能力的大模型,在缺乏卷积模块的情况下,震相拾取性能下降显著.这一现象表明模型依赖卷积的细节特征提取能力(Wang et al., 2025),因此,即便是通用模型,也必须具备局部细节感知能力,才能胜任高精度震相拾取任务.

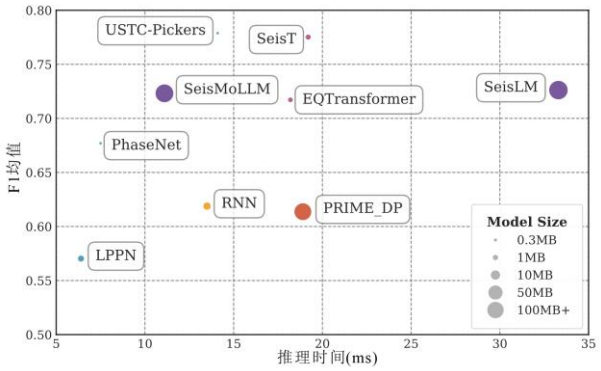

图 14 各模型在异构数据集上的 F1 均值与推理速度对比

Fig.14 Comparison of average F1-score and inference speed of different models on heterogeneous datasets

从同一波形不同时窗长度的测试结果来看,包含 Transformer 结构的 SeisT 在处理不同长度波形时,输出结果会因上下文长度变化而显著波动,但其全局感知特性使其在复杂、低信噪比波形中的鲁棒性优于专业模型,特别是在连续波形、Sg 识别等场景中,往往能获得更优的拾取表现.当波形长度增加时,大模型可更好地整合长距离依赖关系,从而提供更贴近真实标签的结果;相对而言,专业模型对波形长度变化的敏感性较小,更关注于短时窗内的局部特征.此外,大规模预训练模型在低信噪比和跨域场景下展现了更强泛化性(Li et al., 2024; Liu et al., 2024).以 SeisMoLLM 为例,虽然其在 STEAD 数据集上训练,但在本研究云南地区的异构数据集中仍保持了较优表现,表明其具有一定程度的区域泛化能力,这对缺乏高质量标注数据的新区域极具价值.随着硬件与资源的持续进步,构建和部署具备一定规模的大模型变得越来越可行,其在地震预警系统中的应用潜力正日益凸显.

尽管大模型具有非常大的潜力,但目前专业模型在特定场景中依然具有不可替代的优势.首先,专业模型结构更轻量,训练成本低,便于部署在资源受限的前端设备或小型台网.其次,基于本地数据的迁移学习策略可在较低成本下显著提升模型在特定区域或仪器类型上的性能.例如,USTC-Pickers 尽管使用的是相对简单的 PhaseNet 架构,但通过在云南地区多源数据上的迁移训练,表现出良好的适应性与拾取精度,能够有效缓解由于地质构造差异或仪器类型不同所带来的系统性偏差.此外,专业模型由于参数量较小,在训练和推理中的效率往往较高.本文的测试硬件环境为一台搭载双路 CPU 的服务器,具体配置为：两颗 Intel Xeon Gold 6230 处理器（每颗 CPU 基频为 2.10GHz,共 48 个物理核心）,以及 512GB DDR4 内存。所有模型均在统一数据集下开展测试,且采用相同长度的输入数据进行评估。测试结果显示,专业模型推理时间一般小于 15 ms,大模型约为 20 – 30 ms(图 14).大模型多基于 Transformer 架构,模型参数量大,训练需要使用的数据多,计算资源大,训练成本较高(Li et al., 2023).

综上所述,通用大模型与专业模型各具优势：前者适合构建具备跨区域泛化能力的统一框架,后者则在面向特定区域或任务场景时具备更高性价比.在当前尚未形成统一高质量大规模地震数据集的背景下,利用本地数据进行迁移学习仍是提升模型性能的有效路径;而随着数据规模与模型能力的不断增长,探索兼具泛化性与局部适应性的融合模型将是未来的重要研究方向.

5 结论

本文构建了覆盖云南预警台网速度计、加速度计和烈度计的异构波形数据集,通过对PhaseNet、EQTransformer 以及 SeisT 等 9 种主流地震检测模型进行评估,主要研究结论如下：

(1)USTC-Pickers 作为迁移优化的专业模型,在云南三类仪器数据中综合性能最优,尤其在速度计和加速度计数据中, Pg 和 Sg 震相拾取的平均 F1 值达 0.779(△t≤0.1 s),且有效缓解了震相拾取滞后问题.

(2)SeisT、SeisMoLLM 和 SeisLM 等大模型在 Sg 和低信噪比条件下展现出优越的泛化能力,在烈度计等新型仪器数据中具备更大潜力.

(3)专业模型更关注波形细节、输出一致性高;大模型则在上下文理解和复杂信号中更具优势.在当前缺乏统一大规模训练数据的背景下,结合迁移学习优化专业模型仍是提升区域泛化性能的重要手段.未来可进一步探索融合本地训练与大模型预训练的混合策略,推动震相识别模型向更高精度与广泛适用性发展.

致谢：衷心感谢审稿专家付出的宝贵时间以及提出的建设性意见，对进一步完善论文质量起到了至关重要的作用。本文使用的地震数据由云南地震台提供，数据处理和计算使用了Obspy、SeisBench 工具箱以及 9 个震相拾取模型，在此一并表示感谢.

References

An, Y. R., 2024. Introduction to a Recently Released Dataset Entitled CSNCD: A Comprehensive Dataset of Chinese Seismic Network. *Earthquake Research Advances*, 4(1): 100255. https://doi.org/10.1016/j.eqrea.2023.100255

Bormann, P., Wendt, S., Klinge, K., 2013. Data Analysis and Seismogram Interpretation. *New Manual of Seismological Observatory Practice 2* (NMSOP-2), 11: 1-34. https://doi.org/10.2312/GFZ.NMSOP-2_CH11

Bornstein, T., Lange, D., Münchmeyer, J., et al., 2023. PickBlue: Seismic Phase Picking for Ocean Bottom Seismometers with Deep Learning. *Earth and Space Science*, 11(1): e2023EA003332. https://doi.org/10.1029/2023EA003332

Chai, C. P., Maceira, M., Santos-Villalobos, H. J., et al., 2020. Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking. *Geophysical Research Letters*, 47(16): e2020GL088651. https://doi.org/10.1029/2020GL088651

Du, M. N., He, F. X., Zou, N., et al., 2024. Shortcut Learning of Large Language Models in Natural Language Understanding. *Communications of the ACM*, 67(1): 110-120. https://doi.org/10.1145/3596490

Fang, L. H., Wu, Z. L., Song, K., 2017. SeismOlympics. *Seismological Research Letters*, 88(6): 1429-1430. https://doi.org/10.1785/0220170134

Feng, C. Y., Yang, Y., Hu, X., et al., 2023. MSNet: A Seismic Phase Picking Network Applicable to Microseismic Monitoring. *IEEE Geoscience and Remote Sensing Letters*, 20:1-5. https://doi.org/10.1109/LGRS.2023.3314029

Huangfu, G., 2009. Research on the Seismicity in Yunnan, China (Dissertation). University of Science and Technology of China, Hefei: 2-3(in Chinese with English abstract).

Li, J., Chen, J. Y., Tang, Y. C., et al., 2023. Transforming Medical Imaging with Transformers? A Comparative Review of Key Properties, Current Progresses, and Future Perspectives. *Medical Image Analysis*, 85: 102762. https://doi.org/10.1016/j.media.2023.102762

Li, S., Yang, X., Cao, A. Y., et al., 2024. SeisT: A Foundational Deep-Learning Model for Earthquake Monitoring Tasks. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-15. https://doi.org/10.1109/TGRS.2024.3371503

Liu, T. L., Münchmeyer, J., Laurenti, L., et al., 2024. SeisLM: a Foundation Model for Seismic Waveforms. *arXiv*. https://doi.org/10.48550/arXiv.2410.15765

Mousavi, S. M., Beroza, G. C., 2022. Deep-learning Seismology. *Science*, 377(6607): 4470. https://doi.org/10.1126/science.abm4470

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

Mousavi, S. M., Sheng, Y. X., Zhu, W. Q., et al., 2019. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access*, 7: 179464-179476. https://doi.org/10.1109/ACCESS.2019.2947848

Münchmeyer, J., Woollam, J., Rietbrock, A., et al., 2022. Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers. *Journal of Geophysical Research: Solid Earth*, 127(1): e2021JB023499. https://doi.org/10.1029/2021JB023499

Niksejel, A., Zhang, M., 2024. OBSTransformer: a Deep-learning Seismic Phase Picker for OBS Data Using Automated Labelling and Transfer Learning. *Geophysical Journal International*, 237(1): 485-505. https://doi.org/10.1093/gji/ggae049

Pecher, B., Srba, I., Bielikova, M., 2025. Comparing Specialised Small and General Large Language Models on Text Classification: 100 Labelled Samples to Achieve Break-Even Performance. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Suzhou, 165-184. https://10.18653/v1/2025.emnlp-main.9

Peng, C. Y., Jiang, P., Ma, Q., et al., 2022. Chinese Nationwide Earthquake Early Warning System and Its Performance in the 2022 Lushan M6.1 Earthquake. *Remote Sensing*, 14(17): 4269. https://doi.org/10.3390/rs14174269

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

Saad, O. M., Savvaidis, A., Chen, Y. K., 2024. Transfer Learning for Seismic Phase Picking With Significantly Higher Precision in Faraway Seismic Stations. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-12. https://doi.org/10.1109/TGRS.2024.3422612

Si, X., Wu, X. M., Li, Z. F., et al., 2024. An All-in-one Seismic Phase Picking, Location, and Association Network for Multi-task Multi-station Earthquake Monitoring. *Communications Earth & Environment*, 5(1): 22.

https://doi.org/10.1038/s43247-023-01188-4

Wang, X. H., Liu, F., Su, R., et al., 2025. SeisMoLLM: Advancing Seismic Monitoring via Cross-modal Transfer with Pre-trained Large Language Model. *arXiv*. https://doi.org/10.48550/arXiv.2502.19960

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

Yu, Z. Y., Cai, Y. Q., Wang, W. T., et al., 2024. PRIME-DP: Pre-trained Integrated Model for Earthquake Data Processing. *arXiv*. https://doi.org/10.48550/arXiv.2408.01919

Yu, Z. Y., Wang, W. T., 2022. LPPN: A Lightweight Network for Fast Phase Picking. *Seismological Research Letters*, 93(5): 2834-2846. https://doi.org/10.1785/0220210309

Yu, Z. Y., Wang, W. T., Chen, Y. N., 2023. Benchmark on the Accuracy and Efficiency of Several Neural Network Based Phase Pickers Using Datasets from China Seismic Network. *Earthquake Science*, 36(2): 113-131. https://doi.org/10.1016/j.eqs.2022.10.001

Zhao, M., Xiao, Z. W., Chen, S., et al., 2023. DiTing: A large-scale Chinese Seismic Benchmark Dataset for Artificial Intelligence in Seismology. *Earthquake Science*, 36(2): 84-94. https://doi.org/10.1016/j.eqs.2022.01.022

Zhu, J., Li, Z. F., Fang, L. H., 2023. USTC-Pickers: a Unified Set of Seismic Phase Pickers Transfer Learned for China. *Earthquake Science*, 36(2): 95-112. https://doi.org/10.1016/j.eqs.2023.03.001

Zhu, J., Zhong, Y. S., 2025. Evaluation of the Monitoring Capability of the Yunnan Earthquake Early Warning Network. *Journal of Seismological Research*, 48(3):422-431(in Chinese with English abstract).

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

附中文参考文献

皇甫岗, 2009. 云南地震活动性研究(博士学位论文). 合肥: 中国科学技术大学, 2-3.

朱杰, 钟玉盛, 2025. 云南地震预警站网监测效能评估. 地震研究, 48(3): 422-431.