



地球科学

Earth Science

ISSN 1000-2383, CN 42-1874/P

《地球科学》网络首发论文

题目: 地震自动编目技术的进展与反思: 在智能处理时代
作者: 周一剑, 周仕勇
收稿日期: 2025-07-15
网络首发日期: 2025-09-10
引用格式: 周一剑, 周仕勇. 地震自动编目技术的进展与反思: 在智能处理时代[J/OL]. 地球科学. <https://link.cnki.net/urlid/42.1874.P.20250910.1032.002>



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。

doi:10.3799/dqkx.2025.177

地震自动编目技术的进展与反思：在智能处理时代

周一剑^{1*}，周仕勇^{2*}

1 美国加州理工学院，地质与行星科学系

2 广州新华学院，人工智能与数据科学系

Rethinking the Advances and Challenges of Contemporary Auto-cataloging Workflows: In the AI Processing Era

Yijian ZHOU^{1*}, Shiyong ZHOU^{2*}

1 Division of Geological and Planetary Sciences, California Institute of Technology, CA, USA

2 Department of Artificial Intelligence and Data Sciences, Xinhua University, Guangzhou, China

摘要：基于 AI 算法的自动编目技术逐步成为主流的今天，预训练模型的泛化性问题也在成为共识。本文通过对一些最新成果的综述和一些简单的测试，试图指出这一技术瓶颈问题，并阐述关于未来发展的想法。一方面，AI 模型评价体系亟需更新，目前主流的基于人工标注的评价方式存在一些局限性，且对于用户的具体案例而言缺乏实用性；另一方面，关于训练数据与模型表现的关系的研究尚处萌芽状态，各家解决泛化性问题的策略也各有不同，但针对这个复杂的问题都缺少系统性讨论。本文旨在给出方向性的建议，即这些技术难题有可能通过何种方式取得进展，希望对钻研 AI 编目技术的研究者有所帮助。

关键词：AI 智能算法，地震编目，泛化性问题，评价指标，训练集

中图分类号：P315

收稿日期：2025-07-15

基金项目：国家重点研发项目(课题号：2022YFF0800602)和国家自然科学基金(课题号：42474069)联合资助。

第一作者信息：周一剑，美国加州理工学院博士后。主要研究方向为：基于人工智能的地震自动编目，地震与断层物理学，慢地震信号检测，以及诱发地震相关的物理与统计建模，Email: zhouyj@caltech.edu，OCRID: 0000-0002-7205-1769

引言

随着海量地震数据的积累与深度学习的飞速发展,地震学领域也进行了大量关于应用智能处理技术(本文简称为 AI 技术)的探索(Kong et al., 2018; Bergen et al., 2019; Yu and Ma, 2021; Arrowsmith et al., 2022; Mousavi and Beroza, 2022; 2023)。其中,在地震自动编目任务中 AI 技术的落地最为成功:以 PhaseNet (Zhu and Beroza, 2018)、EQTransformer (Mousavi et al., 2020)等为代表的、基于大区域或全球地震数据的预训练模型被直接应用在世界各地的上百个案例中(如, Jiang et al., 2021; Tan et al., 2021; Wilding et al., 2023; Kato, 2024; Yoon and Shelly, 2024; Yao et al., 2025),显示出远优于传统算法的地震检测能力和震相拾取精度,使地震目录在事件数量和定位质量上都取得了显著进步(如, Chiaraluce et al., 2022; Zhou et al., 2025),为小地震相关的统计学研究和断层精细结构的挖掘提供了新机遇(Ross et al., 2019b; Herrmann and Marzocchi, 2020; Waldhauser et al., 2021; Hsu et al., 2024; Cui et al., 2025)。这些模型层面的进展还催生了数据和流程上的研究热点,例如,以 SeisBench (Woollam et al., 2022)为代表的、整合了大量预训练模型和数据集的高层封装,可以让用户简易灵活的应用各种模型,并进行对比测试研究(如, Münchmeyer et al., 2022);以 STEAD (Mousavi et al., 2019a)、DiTing (Zhao et al., 2023)等为代表的大规模数据集,可以使大模型训练和跨区域迁移成为可能(如, Zhu et al., 2023; Chai et al., 2025)。

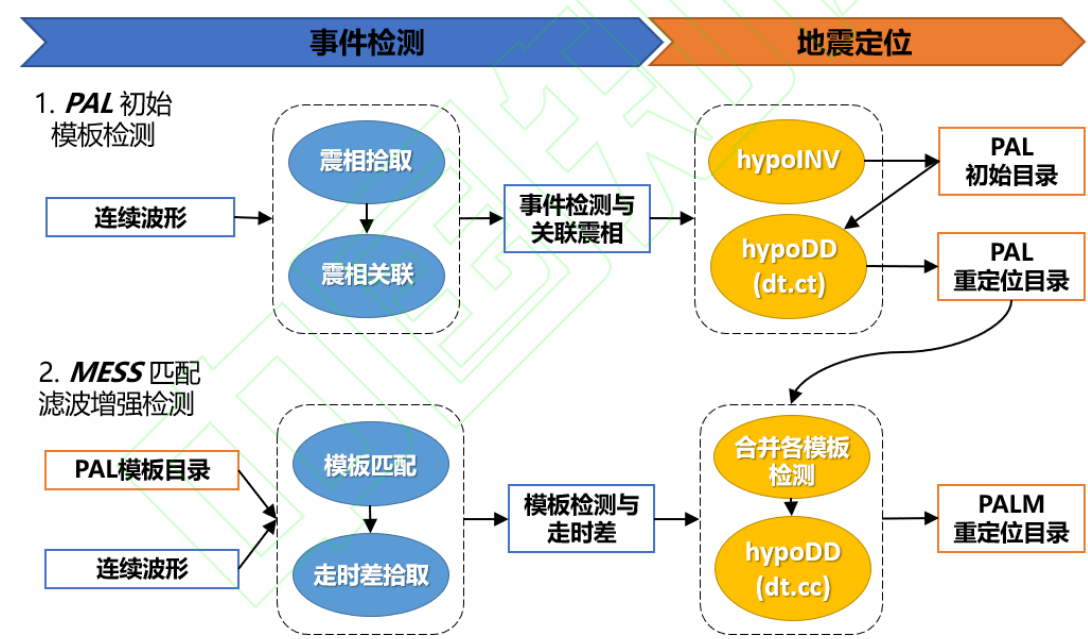


图 1. PALM 地震编目流程图。改自 Zhou et al. (2021b)。

Figure 1. PALM earthquake cataloging workflow. Modified from Zhou et al. (2021b)。

具体而言,地震自动编目一般包括三个主要步骤(如图 1, Li, 2021; Zhou et al., 2021b; Zhang et al., 2022; Zhu et al., 2022a): (1)震相拾取,即获得连续波形中记录到的大量 P 波和 S 波到时; (2)震相关联,即找到不同台站的 P & S 波到时所对应的地震事件; 以及(3)地震定位,即利用每个地震在不同台站的 P & S 波到时反演其发震时刻和震源位置。其中,震相拾取从根本上决定了地震目录的完备性,并且由于地震和噪声波形特征多样,基于规则的传统算法往往难以在保证准确性的同时拥有较高的地震检测能力;而震相关联和地震定位则遵循确定的物理过程,至今仍由传统算法主导,无需 AI 介入。需要指出,这类流程在前 AI 时代

即已存在(如, Chen and Holland, 2016), 但传统算法在震相拾取的完备性和精度上都难以和人工识别相媲美, 因此高精度目录的构建常基于人工目录加上模板匹配的半自动化手段(如图 1, Peng and Zhao, 2009; Shelly, 2020)。然而模板匹配的计算量极大, 理想情况下应当使用所有模板扫描全部研究时段的连续波形、计算互相关, 但这种策略即使在 GPU 的加持下也难以应用于长时间大尺度的目录构建(Ross et al., 2019b)。因此, AI 地震学最初就聚焦于这个最有挑战性的地震检测和震相拾取任务中, 希望在高效计算的同时获得近似于模板匹配的检测能力(如, Perol et al., 2018; Ross et al., 2018; Zhu and Beroza, 2018; Mousavi et al., 2019b; Wang et al., 2019; Zhou et al., 2019)。随着应用案例和对比测试的积累, 人们发现直接应用预训练 AI 模型就可以成倍提升传统算法或人工目录的事件检测数量和定位精度, 尽管在一些案例中 AI 仍被证明检测能力略低于模板匹配(如, Mousavi et al., 2019b; Zhou et al., 2021a; Liu et al., 2022; Naoi et al., 2025)。另一方面, 人们也逐步认识到 AI 模型仍存在某些局限性, 例如, Zhu et al. (2023)发现预训练的 PhaseNet 模型在中国各省的表现存在较大差异, 并提出用少量人工标注数据进行迁移学习; Zhong and Tan (2024)指出现有的预训练模型无法有效识别火山构造下的长周期地震事件; 而这种跨区域的泛化性问题的海底地震仪(Ocean Bottom Seismometer, OBS)数据上格外显著, 甚至会造成大量漏检测或不良定位, 导致断层结构的成像不完备、地震活动性的分布特征出现偏差(Xi et al., 2024; Liu and Tan, 2025), 因此需要重新构建大规模 OBS 训练集以完成模型迁移(Bornstein et al., 2024; Niksejel and Zhang, 2024)。

这些潜在问题应当引起我们的反思: AI 目录的可靠性应该如何评估? 如何克服 AI 模型的泛化性问题? AI 自动编目技术还存在哪些问题? 未来该如何发展? 针对以上问题, 本文将通过若干实例进行探讨, 希望在全流程目录构建的层面提供一些新的见解。需要声明, 本文旨在讨论阻碍 AI 地震编目技术发展的瓶颈问题, 而非对已有成果的综述; 另外, 由于地震编目技术本身涉及数据与方法的耦合(例如, 基于全球台网和基于小区域密集台阵的监测显然应当采用截然不同的手段), 我们将讨论范围设定在最常规的、地方至区域尺度下(约 150 km 内)的面状台网、使用 AI 算法进行单台 P & S 直达波拾取的情形。

1. AI 模型的评价是否需要依赖于人工标注?

如引言所述, 人们利用 AI 算法进行震相拾取是为了在高效的基础上获得更强的地震检测能力: 即, 相比人工目录, 我们希望 AI 算法召回尽可能多的事件检测, 并获得人工目录遗漏的小地震。然而, 当我们需要评价 AI 算法在某个案例的具体表现时, 这些增加的事件检测的准确性显然是无法通过对比人工目录实现的。但众所周知, 当前业内绝大部分对比测试研究仍基于 AI 预测与人工标注的对比, 并以 Precision/Recall/F1 等指标评价震相检测的准确性、以到时误差分布的均值与方差等评价震相拾取的精度(如, Mousavi et al., 2020; Jiang et al., 2021; Münchmeyer et al., 2022; Yu et al., 2023; Lim et al., 2024)。实践表明, 这类基于人工标注的评价策略无法真正反映 AI 算法在目录构建中的表现。一个典型的例子就是 Xi et al. (2024)在汤加俯冲带的工作, 他们在对 PhaseNet-TF 模型进行第一轮优化后的结果相比二到三轮优化后, 在相似定位精度下少了约 30%的事件, 但是三个模型相比人工拾取的 Precision/Recall/F1 指标几乎完全相同, 前者甚至在拾取精度的指标上更高(即与人工标注更一致)。因此, AI 模型的评价不应完全基于人工标注, 并且要更多的关注地震目录本身的质量。

出于以上目的, Zhou et al. (2025)提出可以通过震相关联率来评价 AI 震相拾取的准确性。其定义为:

$$R = \frac{N_{\text{pick}}}{N_{\text{assoc}}} \quad (1)$$

其中 N_{pick} 为震相拾取总数, N_{assoc} 为被关联震相数, 因此震相关联率 R 衡量被关联震相占总震相拾取量的比例。背后的逻辑是, 现代的震相关联算法(如, Zhang et al., 2019; Zhou et al., 2021b)要求较为严格的震相与台站的时空关联, 误识别的震相几乎不会被成功关联, 因此可以近似认为被关联震相为正确拾取的震相。这个假设可以被实际应用中较低的震相关联率侧面验证, 如 Zhou et al. (2025)中测试的不同震相拾取和关联算法, 在相似的定位精度下、在美国 Ridgecrest-Coso 地区和土耳其 East Anatolian 断裂带上的震相关联率在大约 4%至 23% 不等: 这意味着即使全部震相拾取都为误识别, 震相关联也成功剔除了其中的 96%。这里我们给出一个实际例子, 如图 2a 所示为土耳其 East Anatolian 断裂带上各台站上的未关联震相数, 我们对比了 PAL 算法和两个 AI 算法: 预训练 PhaseNet 模型和 Self-Attention RNN (SAR) 模型。由于 PAL 算法的拾取由长短时窗算法完成, 准确性相对较差, 来自人类活动等的脉冲状噪声都会造成误识别。因此, 若 AI 算法在某台站造成了远多于 PAL 算法的未关联震相, 则说明 AI 算法在该台站的地震识别准确率较差。例如, 在 East Anatolian 断裂带这个例子里, 根据以上分析就可以判断 PhaseNet 在 TU.MAYA 台站产生了大量误识别(图 1a)。经波形检查, 我们发现此台站出现了一些仪器故障, 噪声水平较高, 但 PhaseNet 仍在如图 2b 所示的这种没有信号的波形上进行了很多错误的 P & S 波识别。这些误识别显然是难以预料的, 并且由于 AI 模型的黑箱特性, 无法解释成因。具体到 TU.MAYA 台站, 这种误识别是危险的, 因为该台站处于一个余震丛集正上方, 而很多误识别的 S-P 到时差又刚好符合震相传播规律, 故没有被震相关联算法去除。这种包含大量近台误识别的震相文件, 显然又会产生地震定位的偏差。因此, 这种结合震相关联率和传统长短时窗算法的对比验证, 可以帮助我们发现 AI 算法可能的异常表现, 规避目录中的假象。

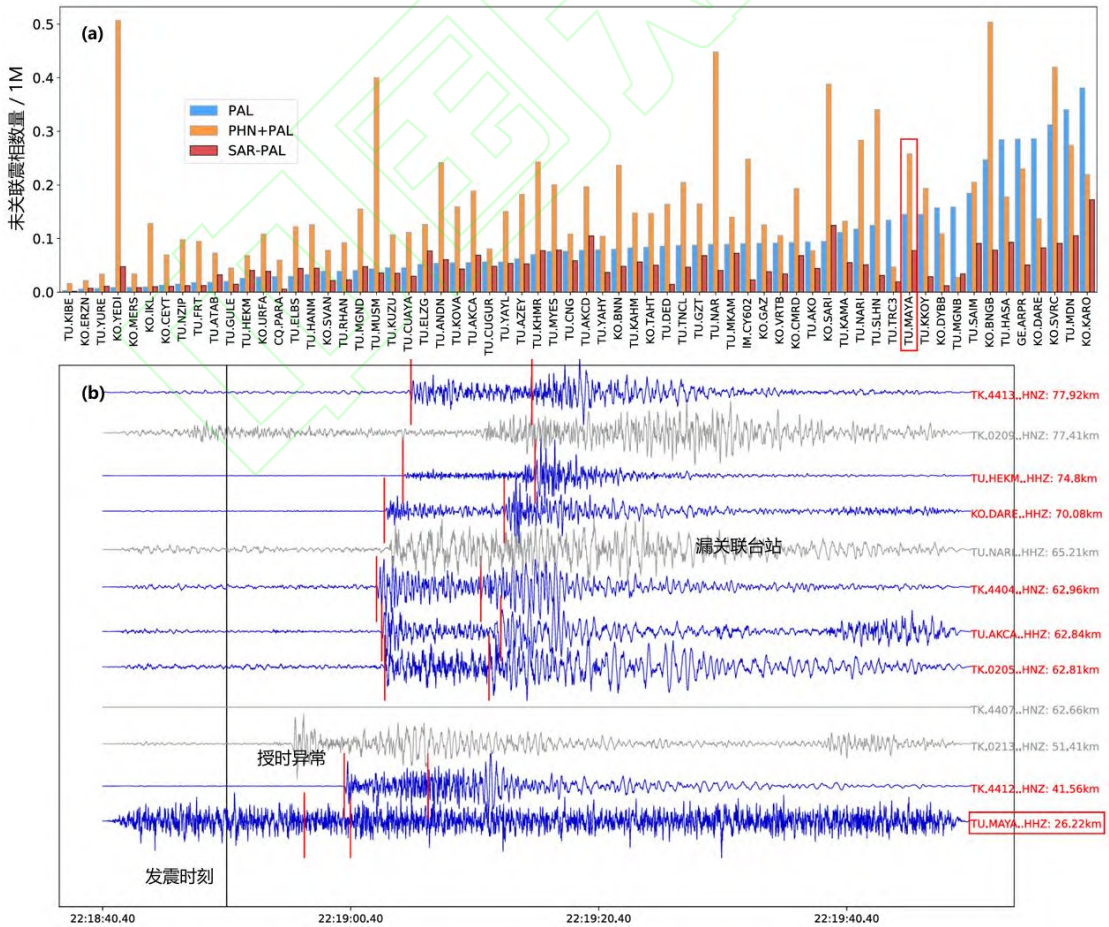


图 2. (a) PAL, PhaseNet, 以及 AI-PAL 的未关联震相数在各台站的分布(改自 Zhou et al., 2025); (b) PhaseNet 误识别波形示例, 其中蓝色和灰色波形分别为被关联台站和漏关联台站, 红色竖线为 P & S 波到时拾取。

Figure 2. (a) Number of unassociated picks by PAL, PhaseNet, and AI-PAL across different stations (modified from Zhou et al., 2025); (b) Example false detection by PhaseNet. The blue and gray waveform denote associated and miss associated stations. The vertical red lines mark the P & S picks.

另一方面, 地震目录本身的质量也强烈依赖于震相文件的准确性, 因此异常的震级频度分布(Gulia et al., 2012; Tan et al., 2024)、较大的定位误差(Husen and Hardebeck, 2010; Liu and Tan, 2025)、重定位时丢弃事件较多(Waldhauser, 2001; Ross et al., 2019a)等, 都可以是震相拾取或/和震相关联效果不佳的体现。这种定性的质量控制在实践上简洁有效, 因为一个各方面表现正常的目录不仅需要高质量的震相文件, 还有赖于可靠的速度模型和定位参数等。我们在这里仍以 East Anatolian 断裂带为例对此进行简单说明。图 3 对比了 2023 年土耳其主震前后两个时段、用 PAL 和 GaMMA (Zhu et al., 2022b)两种算法对 PhaseNet 拾取进行关联的结果, 这两个目录均由 Hypoinverse 算法(Klein, 2002)在相同定位参数下完成定位, 并评估反演质量(即是否为良定位事件)。可以看出, 虽然 GaMMA 关联的结果多出 30%的事件数, 但 PAL 关联的准确性远胜, 因为 PAL 目录仅包含不到 3%的不良定位事件, 远低于 GaMMA 目录中 44%的不良事件比例。这个定位质量的差异也直接反映在事件分布上: GaMMA 目录的定位分布存在大量断层外事件, 特别是在余震时段的分布, 显然不符合地震分布的物理规律(图 3d)。当然, 对于目录质量的分析是复杂的, 并非所有假象都如本例这样显著, 但我们仍建议用户对地震目录不同方面的信息进行可视化、并对照其他独立观测(如震源机制、地表破裂等)进行交叉验证, 这样至少可以避免一些较为明显的错误。同时, 我们也呼吁更多研究利用统计学手段发现更隐蔽的观测假象, 如通过 Schuster 测试分析微震活动的周期性(Rydelek and Hass, 1994; Ader and Avouac, 2013), 从而识别地震目录中错误收录的爆破事件; 又如通过最近邻分析(Nearest-neighbor analysis)找出误识别造成的假异地同发事件(Hsu et al., 2024)。若能够最终形成一些有效的评价指标, 则有助于进一步找出和理解现有算法的问题并加以解决。

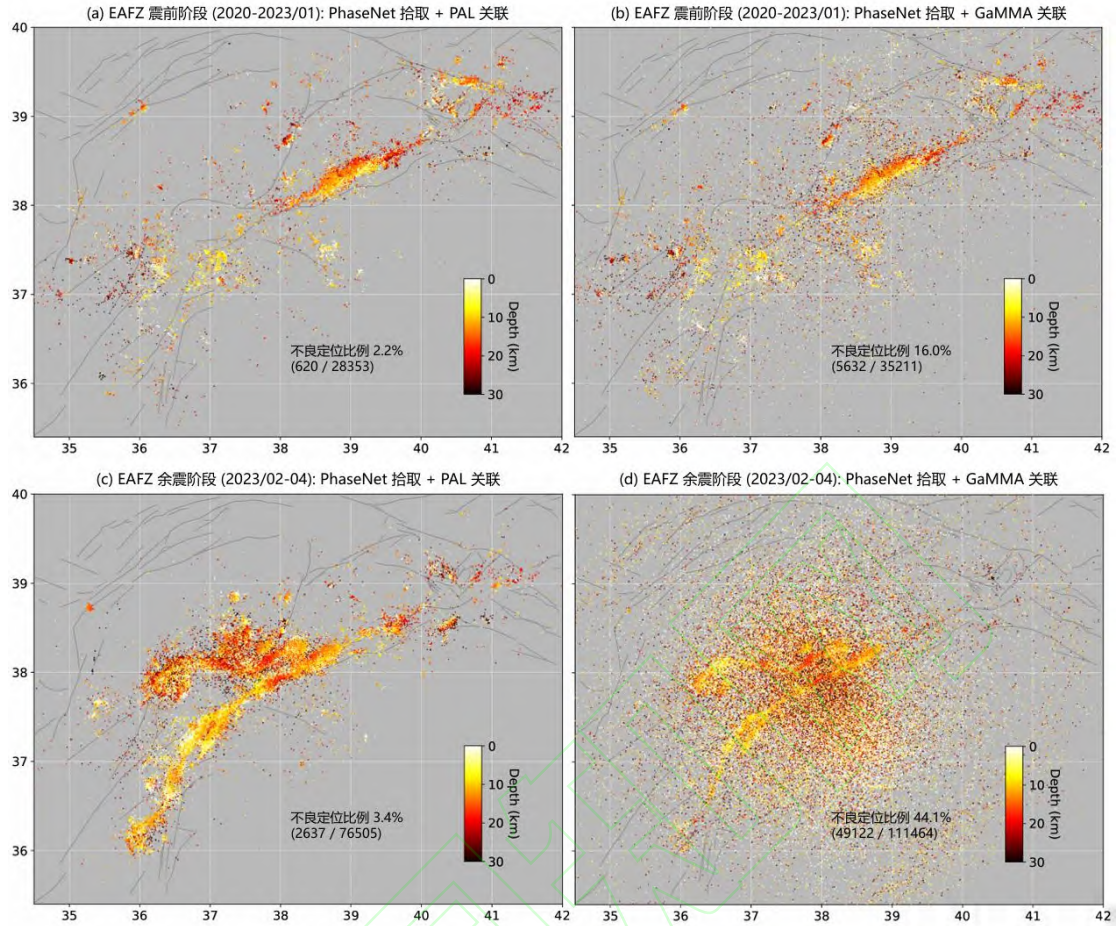


图 3. PAL 与 GaMMA 关联的定位效果对比：在 EAFZ 断裂带的震前阶段(a-b)与余震阶段(c-d)。改自 Zhou et al. (2025)。

Figure 3. Comparison of location distribution along the EAFZ using PAL and GaMMA for phase association. (a-b) plot the preseismic period and (c-d) for aftershock period. Modified from Zhou et al. (2025).

2. AI 模型的泛化性问题：以 2020 Mount Cristo 震群为例

如前所述，尽管预训练 AI 模型的泛化性问题已被不少研究指出，但基于人工标注的评价指标存在一些局限性，我们仍难以全面的评估训练数据对 AI 算法的影响。另一方面，大部分这类研究并未控制变量，例如，测试了不同预训练模型在不同区域的表现，但模型结构与训练数据都有差异，因此无法得知性能的差异来自何种因素。为此，我们以美国内华达州 Mount Cristo 在 2020 年发生的震群(Bormann et al., 2021; Ruhl et al., 2021)为例，设计了一个简单的测试，讨论同一个模型结构在不同训练集的优化后会产生多大的差异。

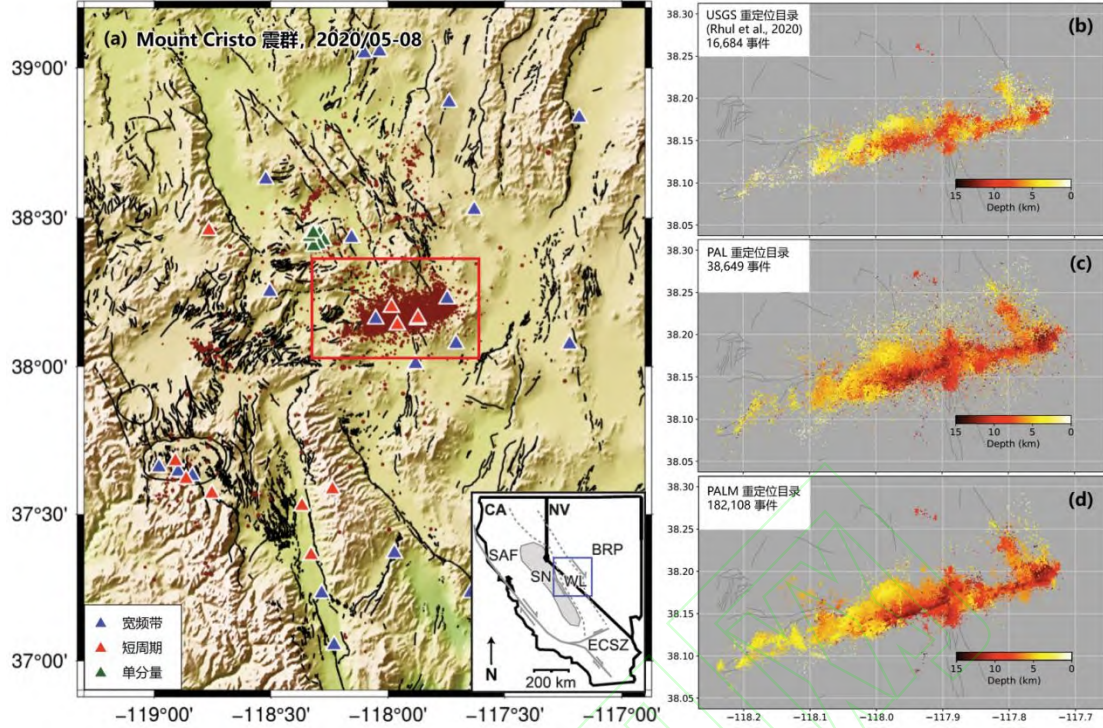


图 4. 美国内华达州 2020 Mount Cristo 震群及目录对比。(a)台站分布和构造背景；(b-d)分别为 USGS 重定位目录、PAL 重定位目录，以及 PALM 重定位目录的事件水平分布。

Figure 4. Catalog comparison for the 2020 Mount Cristo swarm in Nevada, USA. (a) Station distribution and tectonic background; (b-d) plot the map view distribution of the relocated USGS catalog, the PAL relocated catalog, and the PALM relocated, respectively.

我们下载了震源区约 100 km 以内的、2020 年 5-8 月的连续波形数据(图 4a)，并使用 PAL，PALM (Zhou et al., 2021b)，以及 SeisBench (Woollam et al., 2022)提供的不同版本的 PhaseNet 预训练模型构建了地震目录。在运行这些算法时，我们均保持了程序的默认参数，并确保其适合的台站分布与本例一致，故相关细节可以直接参考原文，此处不再赘述。对比 USGS 提供的公开目录(图 4b-d)，PAL 重定位目录实现了超过两倍的事件检测数量(38,649 vs. 16,684 事件)，并在经过模板匹配后进一步获得了近 5 倍的检测数量提升(PALM 目录 182,108 事件)。值得注意的是，由于 PALM 目录的重定位来自互相关测量的走时差数据，其定位精度也远高于另外两个目录，显示出了更紧凑的微震分布特征，与我们大量来自其他案例的经验一致。这些基于规则的算法提供了可靠的参考，我们可以据此评价 AI 算法的完备性和准确性。

我们对比了三个预训练 PhaseNet 模型，分别是基于北加州数据的原版模型(PHN_org)、基于 STEAD 全球数据的模型(PHN_stead)，以及基于南加州数据的模型(PHN_scedc)。三套震相拾取结果均由 PAL 算法在同一套参数下进行震相关联。在评价指标上，我们采用上一节所介绍的震相关联率、被关联震相数，以及良定位地震数，对震相拾取的完备性和准确性进行评价。如图 5 所示，原版 PhaseNet 取得了近 3 倍于 PAL 的被关联震相数和良定位事件，而另外两版 PhaseNet 的检测完备性甚至低于 PAL。这个结果给我们两方面的启示：

- (1) 与若干前人研究一致，我们发现 AI 算法在检测完备性上与模板匹配仍有差距：尽管基于波形相似性的算法在 2020 Mount Cristo 这种密集序列上天然具有优势，但 PhaseNet 仅获得约 60%于 PALM 的事件检测数仍是非常显著的差异。后续研究可以在不同构造环境与数据情况下进行这类对比，从而给出各种场景下的最优策略。

(2) 即使模型结构完全相同，不同训练集也会对 AI 性能产生相当大的影响：在我们的结果中，事件和震相检测的数量最大可达 8 倍左右。值得注意的是，这种严重的泛化性问题甚至并不来自构造环境和数据质量，因为表现最差的一组结果就来自南加州数据，与原版 PhaseNet 所基于的北加州数据构造环境相似、台站规格相当。后续研究应当关注这个问题的成因，是训练样本的数量、采样方式，训练参数(如学习率、优化轮数等)，还是其他什么因素？

在预训练模型存在泛化性问题逐步形成共识的今天，以上问题值得持续关注。一方面，直接使用预训练模型存在风险，更加需要注意自动编目的质量控制；另一方面，如何构建合适的训练数据集去微调模型、甚至重新训练，从而获得适合于用户具体目标数据的 AI 模型，需要更多探索。

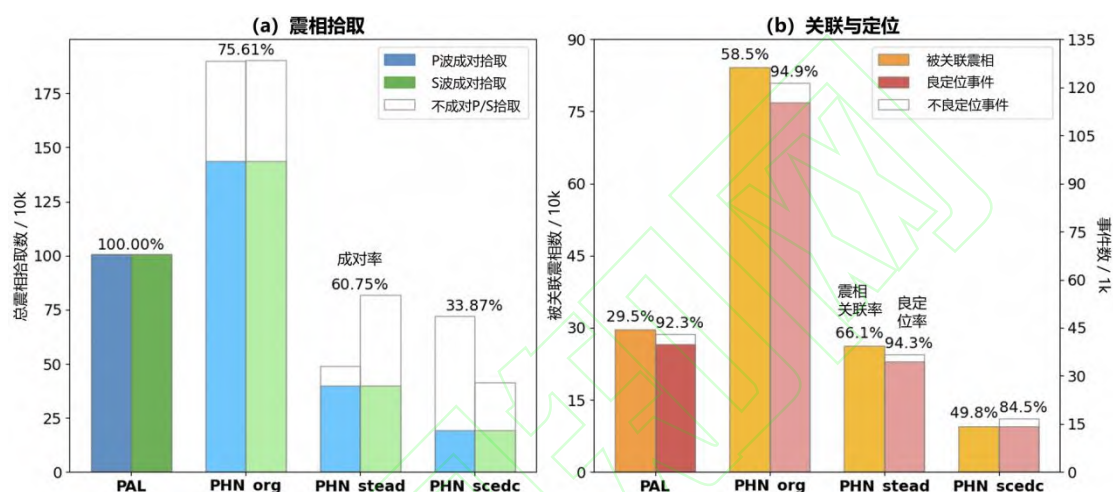


图 5. PAL 与不同版本的 PhaseNet (图中缩写为 PHN) 在 Mount Cristo 震群的震相拾取与地震检测效果对比。(a)柱状图每组以蓝色与绿色分别代表 P 波和 S 波的震相拾取数，空白为不成对的 P/S 拾取。(b)柱状图每组以橙色与红色分别代表被关联震相数和事件检测数，空白为不良定位事件。各版本 PhaseNet 分别为：PHN_org – 原始版本，PHN_stead – STEAD 数据集训练版本，以及 PHN_scedc – 南加州台网中心数据训练版本。

Figure 5. Comparison of phase picking, association, and event detection performance between PAL and various versions of PhaseNet (abbreviated as PHN) in the 2020 Mount Cristo swarm. (a) The blue and green bars denote the number of P and S picks, respectively. Blank portion mark the unpaired P/S picks. (b) Orange and red bars denote the number of associated picks and event detections, respectively. Blank portion mark the poorly-located events. The PhaseNet versions include: PHN_org – original version, PHN_stead – trained by STEAD dataset, and PHN_scedc – trained by Southern California Earthquake Data Center

3. 训练数据中的负样本(噪声)对 AI 模型的影响

从上一节我们知道，AI 算法的表现不仅依赖于模型结构，还取决于训练数据，而後者的影响甚至可能更大。但如何构建训练集去解决泛化性问题，学界目前尚未形成共识(如，Chai et al., 2020; Zhu et al., 2023; Bornstein et al., 2024; Niksejel and Zhang, 2024; Xi et al., 2024; Zhong and Tan, 2024; Zhou et al., 2025)。这类讨论要求程序开发者不仅汇报成功的策略，也应该探索某策略中不同参数的适用范围，从而为用户提供更有指导性的建议。这里，我们对

AI-PAL 流程(Zhou et al., 2025)中提出的训练样本构建策略进行了这样的测试，探索负样本(噪声)数据对 AI 模型的影响。

在 AI-PAL 中，Zhou et al. (2025)提出了一套根据震相文件从连续波形中采样训练样本的策略(图 6)。这套策略中不仅包括对正样本(地震)的随机切取和数据增强，还包括负样本(噪声)的采样策略。具体而言，我们在某台某天随机切出的噪声样本量由如下公式给出：

$$N_{neg}^{ij} = N_{unassoc}^{ij} \cdot \frac{N_{aug} \cdot N_{assoc}}{N_{unassoc}} \cdot \frac{0.5 - \frac{N_{assoc}^{ij}}{N_{pick}^{ij}}}{0.5}, \quad (2)$$

其中，上标 i & j 分别是日期和台站的编号，指具体某台某天的数值；不带上标的 N_{assoc} 和 $N_{unassoc}$ 则是在所有台站所有时段上的总关联震相数和未关联震相数； N_{aug} 为用户设定的数据增强倍数。在此规则下，某台某天的负样本量首先由未关联震相数决定，因为噪声水平越高，就需要越多的负样本来提升模型准确性；公式(2)的第二项则是为了平衡正负样本的数量；第三项则是为了矫正未关联震相数量大但震相关联率高的情形(如余震序列)，若震相关联率大于 50% 则不在此台此天截取负样本。在美国 Ridgecrest-Coso 地区，我们从 2008-2019 年震前约 11 年的数据中，通过 PAL 获得了 61,053 个事件检测和 451,694 个被关联震相。进一步的，利用这些被关联震相和各台站各天的未关联震相数，Zhou et al. (2025)对正样本进行了一次数据增强，获得了 812,771 个正样本，以及根据图 6 的策略自适应采样的 451,689 个负样本。这套训练样本采样策略，结合 AI-PAL 流程中的其他步骤，被证明可以将 PAL 的事件检测数提升约 2.5 倍以上(Zhou et al., 2025)。那么负样本数在模型训练中具体起到什么作用？训练结果是否对其敏感？这类问题在此前研究中尚未被系统测试，尽管绝大部分预训练模型的构建都使用了噪声样本。

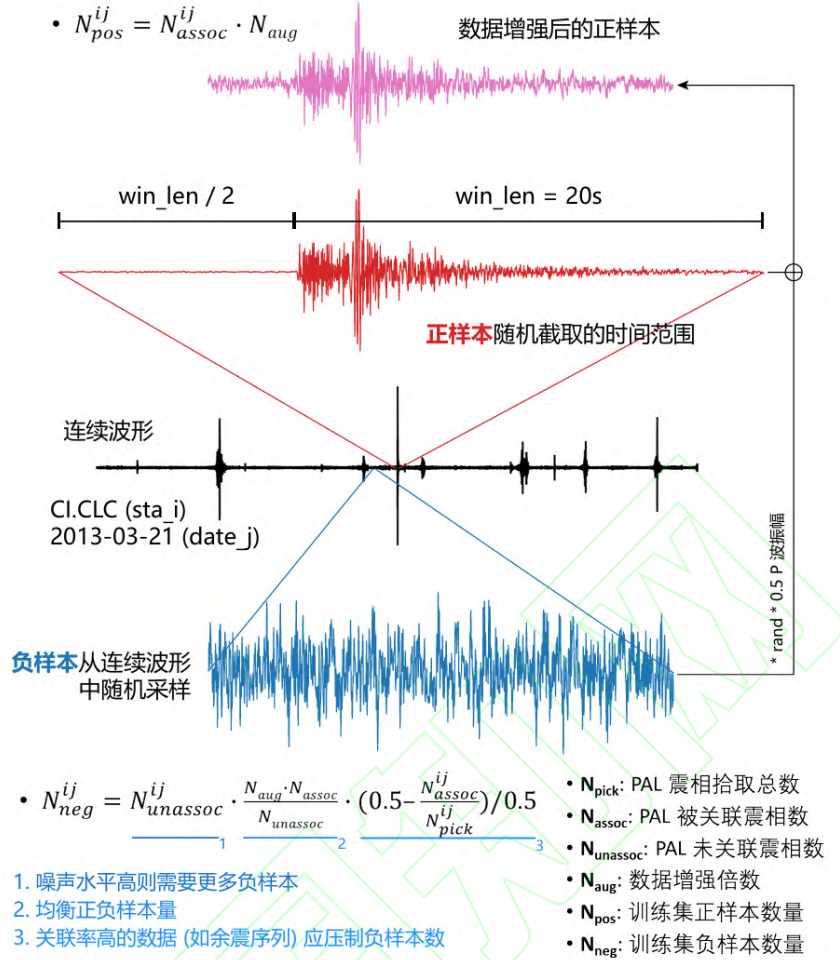


图 6. AI-PAL 流程中训练集正负样本的采样策略。改自 Zhou et al. (2025)。

Figure 6. Sampling strategy of noise data in the AI-PAL workflow. Modified from Zhou et al. (2025).

我们以 AI-PAL 的策略为基准，在美国 Ridgecrest-Coso 区域尝试用不同数量的负样本对 SAR 模型和 PhaseNet 模型进行训练，评估其性能差异(图 7)。注意，此测试中我们仅改变了负样本总量，但保持了各台各天中的相对数量，因此训练集的特征分布与 Zhou et al. (2025) 中一致。结果显示，训练集负样本量与 AI 模型的检测完备性呈负相关(图 7a & c)，但对关联震相数仅会造成约 10% 以内的差异；与检测准确率呈正相关(图 7b)，在 0.25 倍 AI-PAL 初始策略的负样本量时，SAR 和 PhaseNet 模型的震相关联率分别为 7.5% 和 5.5%，远低于原策略的 14.5% 和 9.1%，仅略高于 PAL 的 4.4%。综合来说，由于震相关联算法的有效性，我们倾向于使 AI 算法更加敏感，可以在一定程度上容忍假阳性的情况。因此，AI-PAL 基于未关联震相数和震相关联率的负样本采样策略，结合 PAL 震相关联算法，实际上在相当大的范围内(从 0.25 倍到 1.75 倍原始策略)可以给出相似数量的被关联震相和事件检测。

我们同时也要指出，在这组测试中，SAR 模型所给出的真阳性预测数量基本比 PhaseNet 模型高出 20% 左右(图 7c)，而震相关联率所显示出的预测准确率并没有明显差异，SAR 在某几个负样本量时甚至更准确(图 7b)。因此，即使略微降低触发阈值，SAR 模型也预期会优于 PhaseNet 的检测能力(即，SAR 模型拥有更好的 ROC 曲线)。但我们不应简单的给出“SAR 模型优于 U-net 结构”这样的结论，因为不同模型很可能对应不同的最优训练集构建策略。未来研究可进一步探索“模型结构-训练数据-触发阈值”的优化组合，以提升流程整体性能。

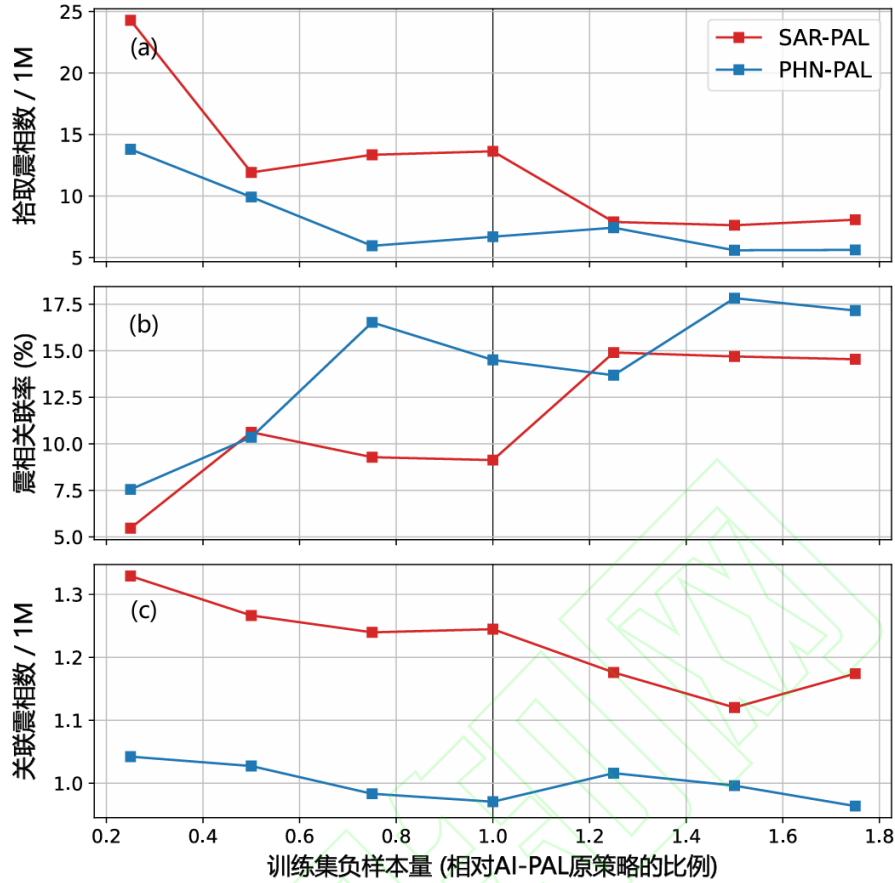


图 7. 训练集负样本量对 AI 模型震相拾取效果的影响。红色与蓝色分别表示 SAR 拾取模型和 PhaseNet 模型结合 PAL 关联算法的结果。

Figure 7. Effects of negative training samples on the AI picker performance. The red and blue markers denote the SAR model and PhaseNet model combined with PAL associator, respectively.

4. 结论与展望

AI 自动编目技术发展至今，对于大陆环境下常规台网的处理已经相当成熟，甚至可以部署成为实时自动处理系统(如，廖诗荣 et al., 2021; Li et al., 2022)。但在应对较为复杂的构造环境、大时空尺度的批处理，和特殊的数据类型(如 OBS)时，预训练模型的泛化性问题仍无法回避。本文就通过一些简单的测试，直观的说明了训练数据对 AI 模型表现的影响之大，亦即这种泛化性问题的严重性。

关于如何处理这种问题，近年来涌现出不少策略。例如，Niksejel and Zhang (2024)利用多种已有 AI 算法对大量 OBS 数据进行自动标注，通过算法间的一致性进行质量控制。但 Liu and Tan (2025)的测试说明，这种自动标注策略并不如直接收集大量人工标注并进行迁移学习效果好(Bornstein et al., 2024)，所以利用多个预训练模型制作训练集并不能凭空生成泛化能力。类似的收集人工标注并进行迁移学习的策略还被成功应用在火山长周期地震的识别模型(Zhong and Tan, 2024)和全国各省的本地化模型(Zhu et al., 2023)构建中。与上述迁移学习的方式略有差异，Xi et al. (2024)在汤加俯冲带使用了一种“利用本轮模型的预测作为下一轮优化的标注”的迭代式优化策略，但仍需要从人工目录训练出初始模型。Zhou et al. (2025)提出的自监督 AI-PAL 流程则彻底摆脱了对人工标注的依赖(图 8)，利用基于规则的算法 PAL (Zhou et al., 2021b)获取大量 P & S 波到时的标注，直接训练较为轻量化的 SAR 模型，从而

保持了跨区域的性能一致性。除了这类构建本地化训练集的思路，最近还有一些基于全球数据构建大模型的尝试(如，Chai et al., 2025)，也有希望通过直接提升预训练模型的复杂度解决泛化性问题。

进一步的，对不同 AI 编目流程的全面评价依赖于跨区域数据集的构建与标准化的评测框架。如前文所述，尽管学界已涌现了大量聚焦于方法的研究，但基于人工标注的评价指标无法全面反映算法在连续波形上实际应用的效果。除了应用和寻找类似震相关联率这样的指标，更新共享数据集的构建策略也是值得探索的方向。如引言中所述，近年来的确涌现出不少大规模的公开数据集，也为一批新的 AI 模型的训练和测试提供了基础，但除了增加震相数量和空间范围，还应当注意数据特征层面的丰富性，包括构造环境、仪器类型，甚至地震序列与波形的特征等。另一方面，将小规模数据做精细是值得尝试的方向：例如，选取一个典型台网，针对一个小规模的地震序列，将所有人眼可识别的 P & S 波到时全部检出，并人工进行跨台站的关联，构建集连续波形、震相到时和事件检测为一体的评测数据集。这类工作将填补人工震相受限于专家精力而造成的缺漏，以及仅包含事件波形和震相到时的局限性。最后，从发现有前景的泛化性识别策略，到用户能够有方向性的完成目录构建，中间需要经过一个理解算法原理和处理流程的过程，这对于数据驱动的具有黑箱特性的 AI 算法来说格外困难。这需要我们不仅去比较不同策略的优劣，还要理解其中的控制因素和参数的容错范围。本文对训练集负样本影响的探索仅是这个方向的一小步，更多的关于训练集特征分布及其影响的研究应当在日后开展，避免盲目应用 AI 带来难以预料的隐藏假象。

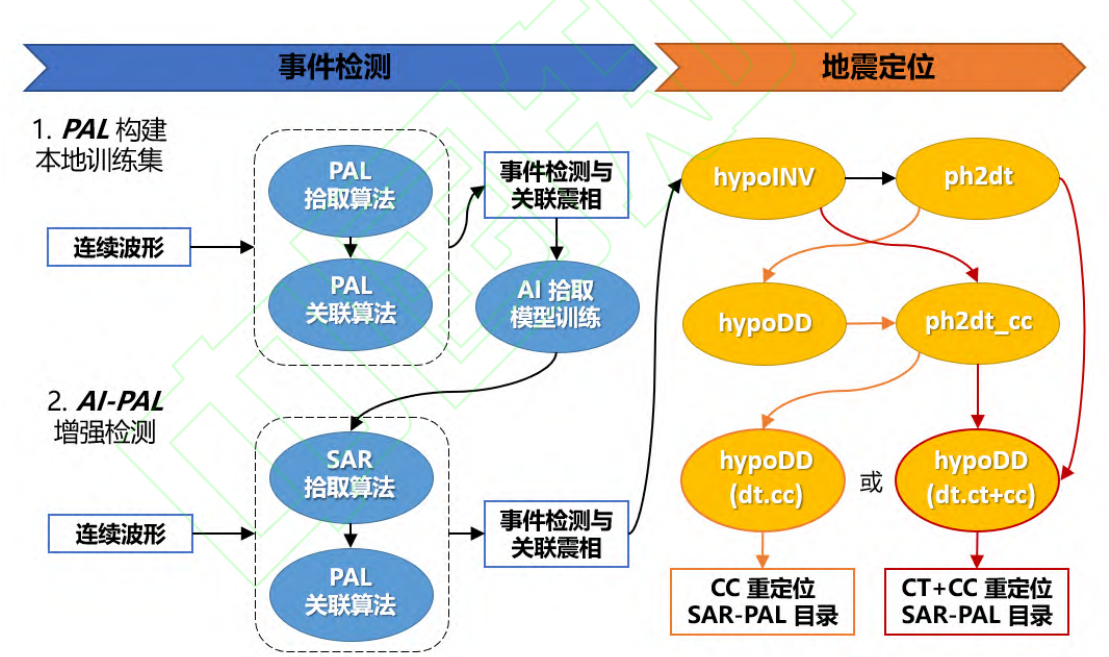


图 8. AI-PAL 地震编目流程图。改自 Zhou et al. (2025)。

Figure 8. AI-PAL earthquake cataloging workflow. Modified from Zhou et al. (2025).

致谢

本研究得到国家重点研发项目(课题号: 2022YFF0800602)和国家自然科学基金(课题号: 42474069)联合资助。

参考文献

- Ader, T. J., and J.-P. Avouac (2013). Detecting periodicities and declustering in earthquake catalogs using the Schuster spectrum, application to Himalayan seismicity, *Earth and Planetary Science Letters*, 377-378, 97-105, doi:<https://doi.org/10.1016/j.epsl.2013.06.032>.
- Arrowsmith, S. J., D. T. Trugman, J. MacCarthy, K. J. Bergen, D. Lumley, and M. B. Magnani (2022). Big Data Seismology, *Reviews of Geophysics*, 60(2), e2021RG000769, doi:<https://doi.org/10.1029/2021RG000769>.
- Bergen, K. J., P. A. Johnson, M. V. de Hoop, and G. C. Beroza (2019). Machine learning for data-driven discovery in solid Earth geoscience, *Science*, 363(6433), eaau0323, doi:10.1126/science.aau0323.
- Bormann, J. M., E. A. Morton, K. D. Smith, G. M. Kent, W. S. Honjas, G. L. Plank, and M. C. Williams (2021). Nevada Seismological Laboratory Rapid Seismic Monitoring Deployment and Data Availability for the 2020 Mw 6.5 Monte Cristo Range, Nevada, Earthquake Sequence, *Seismological Research Letters*, 92(2A), 810-822, doi:10.1785/0220200344.
- Bornstein, T., D. Lange, J. Münchmeyer, J. Woollam, A. Rietbrock, G. Barcheck, I. Grevenmeyer, and F. Tilmann (2024). PickBlue: Seismic Phase Picking for Ocean Bottom Seismometers With Deep Learning, *Earth and Space Science*, 11(1), e2023EA003332, doi:<https://doi.org/10.1029/2023EA003332>.
- Chai, C., M. Maceira, H. J. Santos-Villalobos, S. V. Venkatakrishnan, M. Schoenball, W. Zhu, G. C. Beroza, C. Thurber, and E. C. Team (2020). Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking, *Geophysical Research Letters*, 47(16), e2020GL088651, doi:<https://doi.org/10.1029/2020GL088651>.
- Chai, C., D. Rose, S. Stewart, N. Martindale, M. Adams, L. Linville, C. Stanley, A. Torres Polanco, and P. Bingham (2025). PickerXL, A Large Deep Learning Model to Measure Arrival Times from Noisy Seismic Signals, *Seismological Research Letters*, 96(4), 2394-2404, doi:10.1785/0220240353.
- Chen, C., and A. A. Holland (2016). PhasePapy: A Robust Pure Python Package for Automatic Identification of Seismic Phases, *Seismological Research Letters*, 87(6), 1384-1396, doi:10.1785/0220160019.
- Chiaraluce, L., et al. (2022). A comprehensive suite of earthquake catalogues for the 2016-2017 Central Italy seismic sequence, *Scientific Data*, 9(1), 710, doi:10.1038/s41597-022-01827-z.
- Cui, X., Z. Li, J.-P. Ampuero, and L. De Barros (2025). Does Foreshock Identification Depend on Seismic Monitoring Capability?, *Geophysical Research Letters*, 52(11), e2025GL115394, doi:<https://doi.org/10.1029/2025GL115394>.
- Gulia, L., S. Wiemer, and M. Wyss (2012). Catalog artifacts and quality control, *Community Online Resource for Statistical Seismicity Analysis*, doi:<http://dx.doi.org/10.5078/corssa-93722864>.
- Herrmann, M., and W. Marzocchi (2020). Inconsistencies and Lurking Pitfalls in the Magnitude - Frequency Distribution of High - Resolution Earthquake Catalogs, *Seismological Research Letters*, 92(2A), 909-922, doi:10.1785/0220200337.
- Hsu, Y.-F., I. Zaliapin, and Y. Ben-Zion (2024). Informative Modes of Seismicity in Nearest-Neighbor Earthquake Proximities, *Journal of Geophysical Research: Solid Earth*, 129(3), e2023JB027826, doi:<https://doi.org/10.1029/2023JB027826>.

- Husen, S., and J. Hardebeck (2010). Earthquake location accuracy, *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa-55815573.
- Jiang, C., L. Fang, L. Fan, and B. Li (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes, *Earthquake Science*, 34(5), 425-435, doi:10.29382/eqs-2021-0038.
- Kato, A. (2024). Implications of Fault-Valve Behavior From Immediate Aftershocks Following the 2023 Mj6.5 Earthquake Beneath the Noto Peninsula, Central Japan, *Geophysical Research Letters*, 51(1), e2023GL106444, doi:https://doi.org/10.1029/2023GL106444.
- Klein, F. W. (2002). User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes *Rep. 2331-1258*, US Geological Survey.
- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2018). Machine Learning in Seismology: Turning Data into Insights, *Seismological Research Letters*, 90(1), 3-14, doi:10.1785/0220180259.
- Li, J., et al. (2022). A real-time AI-assisted seismic monitoring system based on new nodal stations with 4G telemetry and its application in the Yangbi MS 6.4 aftershock monitoring in southwest China, *Earthquake Research Advances*, 2(2), 100033, doi:https://doi.org/10.1016/j.eqrea.2021.100033.
- Li, Z. (2021). Recent advances in earthquake monitoring II: Emergence of next-generation intelligent systems, *Earthquake Science*, 34(6), 531-540, doi:https://doi.org/10.29382/eqs-2021-0054.
- Lim, C. S. Y., S. Lapins, M. Segou, and M. J. Werner (2024). Deep learning phase pickers: how well can existing models detect hydraulic-fracturing induced microseismicity from a borehole array?, *Geophysical Journal International*, 240(1), 535-549, doi:10.1093/gji/ggae386.
- Liu, M., H. Li, L. Li, M. Zhang, and W. Wang (2022). Multistage Nucleation of the 2021 Yangbi MS 6.4 Earthquake, Yunnan, China and Its Foreshocks, *Journal of Geophysical Research: Solid Earth*, 127(5), e2022JB024091, doi:https://doi.org/10.1029/2022JB024091.
- Liu, M., and Y. J. Tan (2025). Evaluating the performance of machine-learning-based phase pickers when applied to ocean bottom seismic data: Blanco oceanic transform fault as a case study, *Geophysical Journal International*, doi:10.1093/gji/ggaf256.
- Mousavi, S. M., Y. Sheng, W. Zhu, and G. C. Beroza (2019a). STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI, *IEEE Access*, 7, 179464-179476, doi:10.1109/ACCESS.2019.2947848.
- Mousavi, S. M., W. Zhu, Y. Sheng, and G. C. Beroza (2019b). CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection, *Scientific Reports*, 9(1), 10267, doi:10.1038/s41598-019-45748-1.
- Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nature Communications*, 11(1), 3952, doi:10.1038/s41467-020-17591-w.
- Mousavi, S. M., and G. C. Beroza (2022). Deep-learning seismology, *Science*, 377(6607), eabm4470, doi:doi:10.1126/science.abm4470.
- Mousavi, S. M., and G. C. Beroza (2023). Machine Learning in Earthquake Seismology, *Annual Review of Earth and Planetary Sciences*, 51(Volume 51, 2023), 105-129, doi:https://doi.org/10.1146/annurev-earth-071822-100323.

- Münchmeyer, J., 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, doi:<https://doi.org/10.1029/2021JB023499>.
- Naoi, M., S. Hirano, and Y. Chen (2025). High-resolution monitoring of hydraulically induced acoustic emission activities using neural phase picking and matched filter analysis, *Progress in Earth and Planetary Science*, 12(1), 24, doi:10.1186/s40645-025-00696-5.
- Niksejel, A., and M. Zhang (2024). OBSTransformer: a deep-learning seismic phase picker for OBS data using automated labelling and transfer learning, *Geophysical Journal International*, 237(1), 485-505, doi:10.1093/gji/ggae049.
- Peng, Z., and P. Zhao (2009). Migration of early aftershocks following the 2004 Parkfield earthquake, *Nature Geoscience*, 2(12), 877-881, doi:10.1038/ngeo697.
- Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Science Advances*, 4(2), doi:10.1126/sciadv.1700578.
- Ross, Z. E., M. A. Meier, E. Hauksson, and T. H. Heaton (2018). Generalized Seismic Phase Detection with Deep Learning, *Bulletin of the Seismological Society of America*, 108(5A), 2894-2901, doi:10.1785/0120180080.
- Ross, Z. E., et al. (2019a). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence, *Science*, 366(6463), 346-351, doi:10.1126/science.aaz0109.
- Ross, Z. E., D. T. Trugman, E. Hauksson, and P. M. Shearer (2019b). Searching for hidden earthquakes in Southern California, *Science*, 364(6442), 767-771, doi:10.1126/science.aaw6888.
- Ruhl, C. J., E. A. Morton, J. M. Bormann, R. Hatch - Ibarra, G. Ichinose, and K. D. Smith (2021). Complex Fault Geometry of the 2020 Mww 6.5 Monte Cristo Range, Nevada, Earthquake Sequence, *Seismological Research Letters*, 92(3), 1876-1890, doi:10.1785/0220200345.
- Rydelek, P. A., and L. Hass (1994). On estimating the amount of blasts in seismic catalogs with Schuster's method, *Bulletin of the Seismological Society of America*, 84(4), 1256-1259, doi:10.1785/bssa0840041256.
- Shelly, D. R. (2020). A High - Resolution Seismic Catalog for the Initial 2019 Ridgecrest Earthquake Sequence: Foreshocks, Aftershocks, and Faulting Complexity, *Seismological Research Letters*, 91(4), 1971-1978, doi:10.1785/0220190309.
- Tan, F., H. Kao, K. M. Yi, E. Nissen, C. Goerzen, J. Hutchinson, D. Gao, and A. M. Farahbod (2024). Next Generation Seismic Source Detection by Computer Vision: Untangling the Complexity of the 2016 Kaikōura Earthquake Sequence, *Journal of Geophysical Research: Solid Earth*, 129(5), e2024JB028735, doi:<https://doi.org/10.1029/2024JB028735>.
- Tan, Y. J., F. Waldhauser, W. L. Ellsworth, M. Zhang, W. Zhu, M. Michele, L. Chiaraluce, G. C. Beroza, and M. Segou (2021). Machine - Learning - Based High - Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence, *The Seismic Record*, 1(1), 11-19, doi:10.1785/0320210001.
- Waldhauser, F. (2001). hypoDD--A program to compute double-difference hypocenter locations, *U.S. Geological Survey Open-File Report 01-113*, 25 pp.
- Waldhauser, F., M. Michele, L. Chiaraluce, R. Di Stefano, and D. P. Schaff (2021). Fault Planes, Fault Zone Structure and Detachment Fragmentation Resolved With High-Precision Aftershock Locations of the 2016–2017 Central Italy Sequence, *Geophysical Research Letters*, 48(16), e2021GL092918, doi:<https://doi.org/10.1029/2021GL092918>.

- Wang, J., Z. Xiao, C. Liu, D. Zhao, and Z. Yao (2019). Deep Learning for Picking Seismic Arrival Times, *Journal of Geophysical Research: Solid Earth*, 124(7), 6612-6624, doi:<https://doi.org/10.1029/2019JB017536>.
- Wilding, J. D., W. Zhu, Z. E. Ross, and J. M. Jackson (2023). The magmatic web beneath Hawai'i, *Science*, 379(6631), 462-468, doi:[doi:10.1126/science.ade5755](https://doi.org/10.1126/science.ade5755).
- Woollam, J., et al. (2022). SeisBench—A Toolbox for Machine Learning in Seismology, *Seismological Research Letters*, 93(3), 1695-1709, doi:[10.1785/0220210324](https://doi.org/10.1785/0220210324).
- Xi, Z., S. S. Wei, W. Zhu, G. C. Beroza, Y. Jie, and N. Saloor (2024). Deep learning for deep earthquakes: insights from OBS observations of the Tonga subduction zone, *Geophysical Journal International*, 238(2), 1073-1088, doi:[10.1093/gji/ggae200](https://doi.org/10.1093/gji/ggae200).
- Yao, J., D. Yao, F. Chen, M. Zhi, L. Sun, and D. Wang (2025). A Preliminary Catalog of Early Aftershocks Following the 7 January 2025 MS6.8 Dingri, Xizang Earthquake, *Journal of Earth Science*, 36(2), 856-860, doi:[10.1007/s12583-025-0210-9](https://doi.org/10.1007/s12583-025-0210-9).
- Yoon, C. E., and D. R. Shelly (2024). Distinct Yet Adjacent Earthquake Sequences near the Mendocino Triple Junction: 20 December 2021 Mw 6.1 and 6.0 Petrolia, and 20 December 2022 Mw 6.4 Ferndale, *The Seismic Record*, 4(1), 81-92, doi:[10.1785/0320230053](https://doi.org/10.1785/0320230053).
- Yu, S., and J. Ma (2021). Deep Learning for Geophysics: Current and Future Trends, *Reviews of Geophysics*, 59(3), e2021RG000742, doi:<https://doi.org/10.1029/2021RG000742>.
- Yu, Z., W. Wang, and Y. Chen (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, doi:<https://doi.org/10.1016/j.eqs.2022.10.001>.
- Zhang, M., W. L. Ellsworth, and G. C. Beroza (2019). Rapid Earthquake Association and Location, *Seismological Research Letters*, 90(6), 2276-2284, doi:[10.1785/0220190052](https://doi.org/10.1785/0220190052).
- Zhang, M., M. Liu, T. Feng, R. Wang, and W. Zhu (2022). LOC - FLOW: An End - to - End Machine Learning - Based High - Precision Earthquake Location Workflow, *Seismological Research Letters*, 93(5), 2426-2438, doi:[10.1785/0220220019](https://doi.org/10.1785/0220220019).
- Zhao, M., Z. Xiao, S. Chen, and L. Fang (2023). DiTing: A large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology, *Earthquake Science*, 36(2), 84-94, doi:<https://doi.org/10.1016/j.eqs.2022.01.022>.
- Zhong, Y., and Y. J. Tan (2024). Deep-Learning-Based Phase Picking for Volcano-Tectonic and Long-Period Earthquakes, *Geophysical Research Letters*, 51(12), e2024GL108438, doi:<https://doi.org/10.1029/2024GL108438>.
- Zhou, Y., H. Yue, Q. Kong, and S. Zhou (2019). Hybrid Event Detection and Phase - Picking Algorithm Using Convolutional and Recurrent Neural Networks, *Seismological Research Letters*, 90(3), 1079-1087, doi:[10.1785/0220180319](https://doi.org/10.1785/0220180319).
- Zhou, Y., A. Ghosh, L. Fang, H. Yue, S. Zhou, and Y. Su (2021a). A high-resolution seismic catalog for the 2021 MS6.4/MW6.1 Yangbi earthquake sequence, Yunnan, China: Application of AI picker and matched filter, *Earthquake Science*, 34(5), 390-398, doi:<https://doi.org/10.29382/eqs-2021-0031>.
- Zhou, Y., H. Yue, L. Fang, S. Zhou, L. Zhao, and A. Ghosh (2021b). An Earthquake Detection and Location Architecture for Continuous Seismograms: Phase Picking, Association, Location, and Matched Filter (PALM), *Seismological Research Letters*, 93(1), 413-425, doi:[10.1785/0220210111](https://doi.org/10.1785/0220210111).

- Zhou, Y., H. Ding, A. Ghosh, and Z. Ge (2025). AI-PAL: Self-Supervised AI Phase Picking via Rule-Based Algorithm for Generalized Earthquake Detection, *Journal of Geophysical Research: Solid Earth*, 130(4), e2025JB031294, doi:<https://doi.org/10.1029/2025JB031294>.
- Zhu, J., Z. Li, and L. Fang (2023). USTC-Pickers: a Unified Set of seismic phase pickers Transfer learned for China, *Earthquake Science*, 36(2), 95-112, doi:<https://doi.org/10.1016/j.eqs.2023.03.001>.
- Zhu, W., and G. C. Beroza (2018). PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophysical Journal International*, 216(1), 261-273, doi:10.1093/gji/ggy423.
- Zhu, W., A. B. Hou, R. Yang, A. Datta, S. M. Mousavi, W. L. Ellsworth, and G. C. Beroza (2022a). QuakeFlow: a scalable machine-learning-based earthquake monitoring workflow with cloud computing, *Geophysical Journal International*, 232(1), 684-693, doi:10.1093/gji/ggac355.
- Zhu, W., I. W. McBrearty, S. M. Mousavi, W. L. Ellsworth, and G. C. Beroza (2022b). Earthquake Phase Association Using a Bayesian Gaussian Mixture Model, *Journal of Geophysical Research: Solid Earth*, 127(5), e2021JB023249, doi:<https://doi.org/10.1029/2021JB023249>.
- 廖诗荣, 张红才, 范莉苹, 李珀任, 黄玲珠, 房立华, and 秦敏 (2021). 实时智能地震处理系统研发及其在 2021 年云南漾濞 $M_{\text{S}}6.4$ 地震中的应用, *地球物理学报*, 64(10), 3632-3645, doi:10.6038/cjg202100532.